Original Papers

A Telehealth System Incorporating a Serious Game Intervention to Aid Occupational Therapists in Identifying and Treating Children With Difficulty Crossing the Body’s Midline: Key Informant Interviews Among Occupational Therapists (e27761)
Jonathan Jacobs, Reolyn Heymann, Jacob Greeff  ................................................................. 3

Feasibility of a Sensor-Controlled Digital Game for Heart Failure Self-management: Randomized Controlled Trial (e29044)
Kavita Radhakrishnan, Christine Julien, Tom Baranowski, Matthew O’Hair, Grace Lee, Atami Sagna De Main, Catherine Allen, Bindu Viswanathan, Edison Thomaz, Miyong Kim ................................................................. 20

Development and Validation of a Mobile Game for Culturally Sensitive Child Sexual Abuse Prevention Education in Tanzania: Mixed Methods Study (e30350)
Maria Malamsha, Elingarami Sauli, Edith Luhanga ................................................................. 42

Gamifying Sexual Education for Adolescents in a Low-Tech Setting: Quasi-Experimental Design Study (e19614)
Hussein Haruna, Kingsley Okoye, Zamzami Zainuddin, Xiao Hu, Samuel Chu, Samira Hosseini  ................................................................. 57

Evidence of Construct Validity of Computer-Based Tests for Clinical Reasoning: Instrument Validation Study (e17670)
Tianming Zuo, Baozhi Sun, Xu Guan, Bin Zheng, Bo Qu ................................................................. 72

An Interactive Physical-Cognitive Game-Based Training System Using Kinect for Older Adults: Development and Usability Study (e27848)
Teerawat Kamnardsiri, Kochaphan Phrom, Sirinun Boripuntakul, Somporn Sungkarat ................................................................. 100

Immersion Experiences in a Tablet-Based Markerless Augmented Reality Working Memory Game: Randomized Controlled Trial and User Experience Study (e27036)
Bo Zhang, Nigel Robb ................................................................. 116

Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: Methodological Approach and Exploratory Study (e18359)
Karsten Giels, Marie-Elena Vanden Abeele, Robin De Croon, Paul Dierick, Filipa Ferreira-Brito, Lies Van Assche, Katrien Verbert, Jos Tournoy, Vero Vanden Abeele ................................................................. 142
Effectiveness and Utility of Virtual Reality Simulation as an Educational Tool for Safe Performance of COVID-19 Diagnostics: Prospective, Randomized Pilot Trial (e29586)
Tanja Birrenbach, Josua Zbinden, George Papagiannakis, Aristomenis Exadaktylos, Martin Müller, Wolf Hautz, Thomas Sauter. 167

Changes in Self-Reported Web-Based Gambling Activity During the COVID-19 Pandemic: Cross-sectional Study (e30747)
Emma Claesdotter-Knutsson, Anders Håkansson. 198

Reviews

Effects of Active Video Games on Health-Related Physical Fitness and Motor Competence in Children and Adolescents With Overweight or Obesity: Systematic Review and Meta-Analysis (e29981)
Cristina Comeras-Chueca, Jorge Marin-Puyalto, Angel Matute-Llorente, German Vicente-Rodriguez, Jose Casajus, Alex Gonzalez-Aguero. 8

Software Engineering Frameworks Used for Serious Games Development in Physical Rehabilitation: Systematic Review (e25831)
Jorge Ambros-Antemate, María Beristain-Colorado, Marciano Vargas-Treviño, Jaime Gutiérrez-Gutiérrez, Pedro Hernández-Cruz, Itandehui Gallegos-Velasco, Adriana Moreno-Rodríguez. 181

Viewpoint

Serious Games Are Not Serious Enough for Myoelectric Prosthetics (e28079)
Christian Garske, Matthew Dyson, Sigrid Dupan, Graham Morgan, Kianoush Nazarpour. 130
Original Paper

A Telehealth System Incorporating a Serious Game Intervention to Aid Occupational Therapists in Identifying and Treating Children With Difficulty Crossing the Body's Midline: Key Informant Interviews Among Occupational Therapists

Jonathan Jacobs1*, BEng, MEng; Reolyn Heymann1*, BEng, MEng, DEng; Jacob Jacobus Greeff2, NDip, BTech, MTech

1Centre for Collaborative Digital Networks, Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa
2School of Computer Science and Information Systems, Faculty of Natural and Agricultural Sciences, University of North-West, North-West, South Africa
*these authors contributed equally

Corresponding Author:
Jonathan Jacobs, BEng, MEng
Centre for Collaborative Digital Networks
Department of Electrical and Electronic Engineering Science
University of Johannesburg
Engineering Building
Corner of Kingsway and University Road, Auckland Park
Johannesburg
South Africa
Phone: 27 741550219
Email: jonty.jacobs@gmail.com

Abstract

Background: The midline is an imaginary line that isolates the left and right parts of the body. Crossing the midline infers that a body part (eg, hand or foot) can spontaneously move over to the opposite side of the body to perform an action. A child who has difficulty crossing the midline can physically perform actions that cross the center of the body; however, they do not intuitively cross the midline when challenged with a task that requires this movement, as their perceptual components prevent them from engaging on the contralateral side. This requires treatment from an occupational therapist. Owing to the recent COVID-19 pandemic, access to therapeutic sessions was not possible or reduced, putting the responsibility for treatment on caretakers at home. Caretakers do not have the knowledge and skills to provide treatment, and occupational therapists do not receive adequate feedback from caretakers on the child’s progress.

Objective: The first objective is to adapt a simple serious game, or applied game, into a telehealth solution. Children will play the game at home under the supervision of a caretaker, and the results will be stored on the web. Occupational therapists can monitor progress via a web-based dashboard, receive additional valuable feedback about the child’s behavior during treatment, and easily adapt the game to target specific needs. The second objective is to evaluate whether the implemented telehealth solution is feasible as a treatment option for midline crossing difficulties and thus fit for purpose.

Methods: To meet the first objective, engineering and game development stakeholders formed a team with an occupational therapist, and through a collaborative design process combined with an agile programming approach, a telehealth solution was designed to assist remote monitoring of the serious gameplay. For the second objective, 6 different occupational therapists were introduced to the game, had the opportunity to play the game, and then provided feedback regarding the feasibility, benefits, and applicability of the system during structured interviews.

Results: A telehealth system was designed aimed to address this problem. All results are saved on the web and accessed by occupational therapists via a dashboard. In addition, observed behavioral information is also saved. During the interviews, occupational therapists indicated that the dashboard would support their treatment plan and was indeed a feasible solution.

https://games.jmir.org/2021/4/e27761
Conclusions: The feedback from the occupational therapists for this telehealth solution suggests a feasible method to treat midline crossing problems remotely. The therapists commented on the convenience of integrating both assessment and treatment into the same application, as it assists them when grading a child. The therapists collectively agreed that the quantitative aspect the serious game creates by providing measurable and standardized data proves advantageous when compared with traditional methods of assessment and treatment.

(JMIR Serious Games 2021;9(4):e27761) doi: 10.2196/27761

KEYWORDS
serious games; input device; telehealth; occupational therapy; midline crossing

Introduction

Background

Owing to the COVID-19 pandemic, numerous health care practitioners were unexpectedly required to transition their standard in-person treatment to telehealth options, often without advance preparation or training [1]. Telehealth is becoming an increasingly used service delivery model in rehabilitation services. Telehealth has the potential to alleviate provider shortages, decrease costs associated with providing therapy, and allow for treatment within a client's natural environment [2]. Telehealth can be used by occupational therapists for evaluation, intervention, education, and prevention of injury or exacerbation of conditions [3]. Telehealth facilitates collaboration and consultation with other professionals, which facilitates coordination of care [4].

The midline is an imaginary line that isolates the left and right parts of the body. Crossing the midline infers that a body part—for example, a hand or foot—can spontaneously move over to the opposite side of the body to perform an action. To clarify, a child who has difficulty crossing the midline can physically perform actions that cross the center of the body; however, they do not intuitively cross the midline when challenged with a task. The diagnosis of children with midline crossing difficulties requires a cluster of clinical observations that indicate bilateral integration dysfunction [4]. A study was performed in which 10% of the sample was considered to constitute a possible deficit range, and a further 10% was considered to be in the suspect range. The remaining 80% was considered to fall in the normal range [4].

Midline crossing is a difficulty that needs the intervention of an occupational therapist. Midline crossing difficulties may affect the child’s physical well-being as well as their future development if it is not addressed in a timely manner. If a child does not get the needed intervention due to the COVID-19 lockdown, it can affect the child’s future. Being able to cross the midline is a developmental milestone. By the age of 5 years, a child is expected to be able to cross the midline, that is, use both sides of the body simultaneously [5]. When the midline is crossed spontaneously, supporting neural networks and pathways for specific activities are developed. This is a prerequisite skill essential for the development and maintenance of motor and cognitive demands associated with specific activities. Consequently, children who have trouble crossing the body’s midline also frequently experience difficulty with reading, writing, tying their shoelaces, brushing their teeth, and participating in physical activities [6].

Current treatment approaches used by therapists are arts and crafts in which actions such as threading beads, cutting, pasting, and folding paper are used as these actions require the midline to be crossed. Finger puppets or stickers are also used by placing or sticking the puppets or stickers on one of the child's hands and then encouraging the child to remove the puppet or sticker with the opposite hand. Other common methods used by therapists include building blocks and playing Twister and marching games using arms and legs. Occupational therapists use standardized assessment tools such as the Movement Assessment Battery for Children, the Developmental Test of Visual Motor Integration, and the Draw a Person Test to measure changes in function and occupational. These standardized tests have procedures for administration and scoring [7]. However, when measuring the outcome of an intervention, there is still a lack of tools available for objective measurement, particularly among children with perceptuomotor or attention deficit disorder [7].

At present, occupational therapists observe if the child exhibits the actions given in Textbox 1.

Textbox 1. How to identify whether a child has difficulties crossing the midline.

How to identify whether a child has difficulties crossing the midline

- Swaps hands midway through a task when writing, drawing, painting, or coloring
- Uses the left hand for activities on the left side of the body and right hand for activities on the right-hand side
- Rotates their trunk to the opposite side when reaching across the body (to avoid crossing the body midline)
- Has difficulty visually tracking an object from one side of the body to the other, such as following text when reading
- Has poor pencil skills (pencil grip)
- Uses different feet to kick a ball (mixed dominance)
- Has difficulty coordinating gross motor patterns (eg, crawling, skipping, and star-jumps)
On the basis of their observations and discretion, they conclude whether the child has difficulty crossing the midline. Assessment is an integral part of the occupational therapy process and is a necessity for evidence-based practice. Without appropriate measurement, therapists cannot provide evidence for the offered interventions [8]. The assessment is subjective, and there are issues with this subjectivity, as the same level of difficulty may be rated differently by different assessors. When therapists write reports for the child, it is challenging to justify the child's improvement as there is no criterion to compare it with. In addition, numerous children up to the age of 5 years living in underdeveloped countries, including South African rural areas, face exposure to multiple risks affecting their early childhood development [8]. These children and their parents are often not aware of the functional difficulties they may have, and even if they are aware, they need to travel long distances to receive the therapy they require. Moreover, because of the recent COVID-19 pandemic and resulting lockdowns, many children who would normally receive occupational therapy could not access occupational therapists or had to decrease their sessions.

If a child has a pathology of midline crossing, frustration in the behavior of the child will be noticeable, as the child will become angry when trying to engage in fine motor activities because of less refined hand skills. The coordination of both sides of the body will be less refined, leading to difficulties experienced when playing sports or doing any physical activity. In addition, when children have difficulty with midline crossing, they may additionally have trouble visually tracking an object from left to right. Thus, when the object reaches the midline, they often blink and have to refocus, and this results in their losing their place while reading. In addition, when drawing horizontal and diagonal lines as well as writing letters such as x, they may segment these lines rather than overlap them because of midline crossing difficulties [9]. Some children may struggle to cross the body's midline easily. When a child shows hesitancy in reaching, stepping, or looking across the midline of the body, it is known as midline crossing inhibition. Sometimes, this delay can be seen when a child hesitates or is clumsy during gross motor tasks that require the arm or leg to cross over to the other side. Some children with delayed midline crossing skills may display some compensatory mechanisms in school that make writing awkward for them. Crossing the midline is a treatable affliction. If a child struggles to cross the body's midline and is treated, milestones such as developing a dominant hand will occur [6]. Pencil skills and fine motor tasks will be refined, easing the transition to an academic environment where those skills are expected to be grasped [6]. The child will be able to complete self-care tasks, for example, brushing their teeth and getting dressed. The child will be able to kick and hit balls as well as run their gross and fine motor skills will be improved. Finally, their ability to visually track across a page effectively will be better and, therefore, will result in fluent reading.

At the forefront of technological advancements in occupational therapy are serious games [10]. For the purpose of this research, serious games are described, among other things, as digital games, virtual environments, simulations, and a mixed reality that engage the player. These serious games form encounters and experiences that convey meaning [11]. Serious games are introduced as a need to meet objectives that go beyond entertainment and benefit the user in the area that needs to be mitigated. The applications created under the terminology of serious games induce motivation and engage the user [12].

**Objective**

Before proposing how serious games can be integrated, a major obstacle to treating children in the conventional sense is boredom because of the intensive and repetitive practice required. The advantage of incorporating input devices and the use of serious games into treatment methods is that it combats the boredom factor. The child is placed into a game environment that is similar to the real world in terms of the perceptual stimuli it exhibits, which then puts the child at ease [13]. The child then has the ability to manipulate and control some of the stimuli and see the outcome of their actions in real time and adjust them accordingly. This aspect, being the interactive component, creates an engagement with the environment, allowing the child to feel and be in control of their movement. This idea is explained as the perceptual illusion of nonmeditation [14]. The sensations that are familiar and present when playing the game and the ability of the child to control and manipulate the stimuli that surround them generate the psychological effects of enjoyment and, particularly, involvement [14]. The unique merging of purpose and pleasure develops intrinsic motivation in the child.

Microsoft's Kinect sensor has been highly investigated and used for the development of new complements that help improve or optimize rehabilitation processes worldwide. In 2012, Ruiz and Cantos [15] designed a therapeutic tool using the Kinect for neurorehabilitation using games to stimulate patients, cognitive functions, perceptions, and gross and visual motor skills through play. This tool verified that patients had fun while being treated in this manner; intrinsic motivation is achieved by merging purpose and pleasure.

Moreover, using the Kinect and combining purpose and pleasure, Chang et al [16] performed a study in Taiwan in 2013 that proposed the possibility of rehabilitating two 14-year-old adolescents with cerebral palsy through therapies personalized to their condition. Data showed that the 2 participants had significantly increased motivation for upper limb rehabilitation, thus improving exercise performance during the intervention phases.

Although Kinect is a useful device, it can, however, encounter the problem of misdetection when it comes to extremity angles or overlapping extremities [17]. In 2016, the Taiwan University proposed a new rehabilitation gaming system, which focuses on the upper part of the body with wireless inertial measurement units (IMUs) and a Kinect device. The Kinect was used as a base tracking system by the gaming system. Multiple sets of IMUs were integrated into the extremity of the subject to calculate the angles through algorithms. Wireless IMUs were also added to compensate for the error in the calculation of angles in the Kinect device [17]. This study exposed that the use of the Kinect by itself may not be sufficient and, therefore, additional sensors were needed to ensure correct motion capture.
Nintendo’s Wii Fit is a commercial product used for both fitness and fun and was created to encourage people to exercise as well as improve balance. The Wii Fit uses the Nintendo Wii console and a balance board. A study conducted by the University of Naples found that the Wii Fit showed better improvements regarding physical therapy in terms of balance and self-confidence than conventional treatment [18]. It was concluded that the Wii Fit is acceptable as an adjunct to virtual rehabilitation interventions and provides an exciting new therapy device [18]. The Wii console does not provide a platform or offer a game in which more specific impositions, such as difficulty crossing the midline, can be treated. If one wanted to use the Wii for treatment or training purposes, one would need to purchase not only the balance board but also the console and remote, thus making it too expensive for the communities that this study is aimed at.

Ultraleap’s Leap Motion Controller is a computer hardware sensor device analogous to a mouse [19]. Using motion capture technology, Leap Motion is able to process the input. It does not require direct hand contact with the device as input; instead, hand and finger motions are tracked. Despite the small size of Leap Motion, it is capable of capturing smaller details, such as finger movement. Leap Motion is smaller and cheaper than Kinect; however, Kinect is more precise in capturing movement. Sourial and Reichardt [20] proposed implementing a virtual therapist (VT) to help patients do their exercises at home in an engaging gamified environment. The VT artificial intelligence used a hierarchical finite-state machine architecture. Hand therapy helps the patient regain the hand’s full functionality after a certain injury or surgery. Hand therapy could be a very tedious process that implies physical exhaustion [20]. In addition, finding appointments with the therapist frequently enough for an efficient healing process is difficult and costly.

To test the efficiency of the VT, a web-based hand therapy exercise was implemented using the Unity platform to build the exercise environment. Leap Motion technology was used to detect the information of the hand movement. This exercise was tested on 19 participants. The idea of being coached by a VT was welcomed by the participants, as the exercise was fun and motivating to them. VT guidance and assessment were helpful and easy to follow. However, some modifications are needed in the pain detection part to form a more efficient exercise [20].

Another system using Leap Motion and capable of improving fine motor skills in children, was proposed by Hidalgo et al [21] through a serious game and 3D environment. The proposed system allowed the therapist to choose among different levels of serious games according to the child’s needs. The game excited the children; however, the children took time to adapt to the game because of the inaccurate readings of Leap Motion. Therefore, it can be noted that including different levels creates excitement for the children as they experience different challenges; however, when technology interferes with the game or is difficult to use, the child takes time to become familiar with the system.

Sony’s motion capture system Intel RealSense enables fine-motoric gesture recognition, and its small form factor allows for preintegration into notebooks and tablets, substituting conventional cameras [22]. This setup enables new methods of therapy in the form of serious games that are engaging and easy to set up. Chhor et al [22] developed and evaluated a serious game prototype for rehabilitation using Intel’s RealSense (called Breakout) based on a commercial game framework. Despite the fact that RealSense can easily integrate with the applications mentioned above, it is an expensive app.

Neuroplasticity refers to the ability of the brain to adapt structurally and functionally and is enhanced by training and experience. It is known that neuroplasticity is at its maximum in a critical period, which corresponds to the first 7 years of a child’s life [23]. Therefore, with the development of a serious game and given the neuroplasticity of the child’s brain at this age, the child may be able to develop the required pathways necessary to conduct movements that once seemed impossible. With the integration of technology and the innovation and creativity of this approach, a beneficial method of mitigating impositions such as midline crossing could be developed to assist occupational therapists in the treatment of children. A serious game with an input device was initially developed to assist in midline crossing therapy [24]. An improved wireless input device that accompanied the serious game was developed [25]. The aim of this study is twofold. The first objective focuses on the design of a web-based telehealth system that consists of a serious game, an input device, and a web-based dashboard that displays relevant data to occupational therapists. The second objective focuses on whether the system is fit for its purpose and can be adopted by occupational therapists to treat patients remotely.

**Methods**

**Objective One: Designing a Remote Monitoring System for a Midline Crossing Serious Game**

**Overview**

An agile software development life cycle approach was taken to develop the end-to-end solution. This methodology was chosen because of the frequent feedback needed during the design process, allowing the design team to give recommendations through collaborative design at the end of each iteration. The design team comprised game designers, electronic engineers, and an occupational therapist. Figure 1 shows the iterations of the agile software development life cycle. It should be mentioned that each component of the solution (serious game, input device, and telehealth system) was developed separately through iterations; however, the components were integrated and tested at the end of each iteration. Furthermore, the initial serious game from the previous project was used as a foundation to produce the web-based telehealth solution presented in this study [24].

A high-level solution design is shown in Figure 2. The sequence of events is illustrated in Figure 3.
Figure 1. Iterations in the agile software development life cycle. Req: requirements.

Figure 2. High-level design of solution. OT: occupational therapist.
The innovative low-cost input device shown in Figure 4 was developed previously and used so that the movement of the child could act as an input for the game [25]. The input device is designed in such a way that when the device is turned on and the game started, a connection is established. There are no complicated installations or setup steps. The role of the caretaker is to help the child set up the device with the game. Additionally, the caretaker will explain to the child how to play and then observe the child’s movements according to the guidelines designed by the occupational therapists, which will be discussed further in the Results section. The design of the data visualization component and the new serious game is presented in this study.

**Initial Serious Game**

The game genre chosen was a casual game because of its low system requirements, accessibility in most devices, and, at the same time, not demanding high levels of concentration from the player [26]. The game was not connected to the internet, and players had to install it to play it. The fact that the game does not take much concentration is an important factor in ensuring that the child does not become fatigued and frustrated. Moreover, the duration of the game was purposely designed to be short to retain the child’s concentration. The developed game comprised a sprite controlled by the child [24]. Sprite is the terminology used to describe an avatar, shape, or character that the child has control over. The device was attached to the hand of the child, and the child moved the device in an arc shape with their arm extended in front of them while standing. The aim of the game was to collect faces falling at a calculated position from the sky. The movements in the game ensured that the child crossed the midline. The game intended to prompt the child to perform particular movements that occupational therapists would invoke when using traditional methods of treatment.

**Input Devices**

The problems encountered with the initial design of the input device and the detailed new design have been previously published [25]. For the sake of completeness, a short overview of the input devices is given in this section. In the initial input device, an accelerometer and magnetometer were used to obtain the movement of the child [25]. Tilting of the device was not the only issue that affected the feasibility of the design; before every game, a calibration was required to obtain the reference of the sensor. It did not seem to be the most suitable solution for therapists to expect the child to perform the calibration each time they wanted to play the game, ensuring that they do not tilt the sensor. Therefore, a more child-friendly design was needed that would be able to track the movement of the hand from side to side (crossing the body’s midline) and transform the position into the game. In addition to accurately transforming the movement, a device that would not require calibrating the sensor was required. When discussions were held with an occupational therapist during the design phase, it was suggested that an input device that would be placed on a surface be designed. The idea behind this was to allow the child to move their hand from side to side and exhibit the same motion required when performing tasks that cross the midline, such as writing. Through extensive research into suitable sensors or electrical devices that could be used, a clean, innovative solution was reached. The design entailed the addition of a linear potentiometer; the input device is shown in Figure 4 [25].
Objective Two: Evaluate Whether the Online Game Is Fit for Purpose

After the prototype was developed, it was crucial to evaluate whether this solution was a feasible option for occupational therapists and whether it was fit for purpose. This was achieved through semistructured interviews with occupational therapists. These occupational therapists all had between 5 to 10 years of experience. In addition, as many children have difficulty crossing the midline, these occupational therapists were competent to give feedback regarding the identification and treatment of midline crossing difficulties. Note that the occupational therapist who was part of the participatory design sessions was not included in the evaluation of the system.

The researcher took each of the participants through a presentation detailing the purpose of this research. The researcher conducted a demonstration of the full solution, taking the occupational therapist through the game and explaining how each stage of the game was played. The occupational therapists then had the opportunity to play the game and evaluate each stage on their own. A semistructured interview took place, allowing the researcher to further explore the insights provided by the therapist. Each session was conducted individually to ensure unbiased and objective responses. The interview aimed to highlight the strengths and weaknesses of the solution as well as the suitability of administering the solution remotely. The interview provided the therapists with the opportunity to suggest changes and offer recommendations. The questions were purposefully left open-ended so as not to limit the therapist's responses. The main themes that needed to be established from the interviews can be seen in Textbox 2. The questions contained in each section can be seen in Textbox 3.

Textbox 2. Main themes established from interviews.

Main themes
- The feasibility of using a serious game and input device to assist occupational therapists in identifying and treating children with midline crossing difficulties
- The benefits and applicability of the dashboard
- The applicability of the solution in areas where the availability of occupational therapy resources is limited

Textbox 3. Questions asked to occupational therapists during the interviews.

Questions
- In your professional context, comment on whether a serious game and input device would assist you as a therapist to identify children with midline crossing difficulties as well assist in your treatment process. Please include in your answer the following:
  - Advantages
  - Disadvantages
  - Suggested changes
  - Recommendations
- Comment on whether the use of the dashboard is beneficial and applicable to assist you as a therapist in identifying whether a child has pathology with midline crossing. Does the dashboard provide constructive tracking of the child's progress during treatment?
  - Advantages
  - Disadvantages
  - Suggested changes
  - Recommendations
- Comment whether the solution can be reasonably implemented and used effectively in areas where occupational therapy resources are limited.
  - Advantages
  - Disadvantages
  - Suggested changes
  - Recommendations
**Results**

Objective One: Designing a Remote Monitoring System for a Midline Crossing Serious Game

**Adapted Serious Game**

The initial game, as described in the previous section, was used as a base for the new telehealth system. The new system comprised an adapted serious game (with distinct levels), an input device, and a web-based backend system that enabled occupational therapists to access dashboards and behavioral information about the intervention. During the collaborative design sessions, it was decided that the serious game would be divided into different stages. Each stage has a specific aim, and different game variables can be set according to each child's individual needs. The game comprises 4 stages: an assessment stage, 2 intervention stages, and a maintenance stage [24]. In each stage, there are 3 variables: distance, speed, and time, which are used to create a specific environment for testing. The time variable, set at 2 minutes, was kept constant for all stages. The chosen duration was advised by the occupational therapist. The duration was purposely set to be short to allow the child to concentrate for a short amount of time without getting bored or losing focus. Each stage addresses a certain intervention, which will be explained below; therefore, the input to each stage is different. The occupational therapist can set these inputs however they see fit (or advice the caretaker at home on which values to use during remote treatment).

The game environment is chosen by selecting a stage and choosing the required values for the inputs for that stage. The game commences with the hat (controlled by the child using the input device) starting either on the left or right side of the screen depending on the handedness of the child. For the purpose of this explanation, a child who is right-handed will be used; therefore, the sprite starts on the left. If the child were left-handed, the sprite would start on the right so that the child’s first movement would be crossing the midline. The balls that fall from the sky are strategically placed, depending on the stage, which will be discussed below. When the balls fall, the child must catch the ball in the hat and return the hat to the starting point (either the left or right side of the screen) to earn a point. **Textbox 4** shows the variables used in the equations. **Table 1** shows the inputs and equations for the variable set. The game flow is illustrated in **Figure 5**; depending on the stage, the equations shown in **Table 1** will be inserted.

**Textbox 4. Variables used in the equations to set the game environment.**

<table>
<thead>
<tr>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_0$: starting distance</td>
</tr>
<tr>
<td>$I$: incremental distance</td>
</tr>
<tr>
<td>$S$: the score</td>
</tr>
<tr>
<td>$DV$: the distance variance</td>
</tr>
<tr>
<td>Random $(a, b)$ is the random function that provides a random number between a and b</td>
</tr>
<tr>
<td>$S_{p0}$: the initial speed</td>
</tr>
<tr>
<td>$I_{sp}$: the incremental speed</td>
</tr>
</tbody>
</table>

**Table 1. Equations used in each stage to set the game environment.**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Input</th>
<th>Distance</th>
<th>Speed</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_0^a$, $I^b$</td>
<td>$X_0+(IX_s^c)$</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>2</td>
<td>$X_0$, $DV^d$</td>
<td>$X_0+$Random $(-DV, DV)$</td>
<td>Constant</td>
<td>Constant</td>
</tr>
<tr>
<td>3</td>
<td>$X_0$, $S_{p0}^e$, $I_{sp}^f$</td>
<td>$X_0$</td>
<td>$S_{p0}+(I_{sp}SX)$</td>
<td>Constant</td>
</tr>
<tr>
<td>4</td>
<td>N/A$^g$</td>
<td>Random $(-7.5,7.5)$</td>
<td>Constant</td>
<td>Constant</td>
</tr>
</tbody>
</table>

*a* $X_0$: starting distance.  
*b* $I$: incremental distance.  
*c* $IX_s$: incremental distance as function of score.  
*d* $DV$: distance variance.  
*e* $S_{p0}$: initial speed.  
*f* $I_{sp}$: incremental speed.  
*g* N/A: not applicable.
Stage 1
Stage 1 takes the starting distance \( (X_0) \) and incremental distance \( (I) \) as input. The starting distance is the point on the x-axis where the first ball will be spawned from the sky (top of the screen). The x-axis is depicted in Figure 6. The incremental distance is the distance that is incremented from the starting distance. The point on the x-axis where the balls are spawned is a function of the score \( (S) \); as the score increases, so does the distance of the spawned ball. The score represents the number of balls caught and returned. A ball is considered returned when the hat reaches the left side of the screen if the child is right-handed; if the child is left-handed, the hat would need to be returned to the right side of the screen. The score will update only after the hat has been returned. Consequently, the point at which the next ball falls will increase by the incremental distance (increase to the right if the child is right-handed). If a ball is missed, intuitively, the score will not increase, thus remaining the same; therefore, the ball will be spawned in the same place instead of being spawned at the incremented distance.

Figure 5. General game flow. eq: equation.

Figure 6. Position on x-axis in game.
Stage 1, which is the assessment stage, aims to determine the child’s gross ability to cross the midline. In addition to the stages incorporated into the game, an interface to input the specified variables determined by the therapist was introduced. These variables create a tailored game environment for the child. In stage 1, the speed and time variables remain constant, whereas the distance variable is incremented. The input device shown in Figure 4 controls the movement of the hat illustrated in Figure 6 from side to side, simulating the movement of crossing the midline while writing. After the child catches the ball, the hat needs to be moved all the way to the other side of the screen, ensuring that the midline is crossed.

A reference point for the child’s baseline functioning, including their ability to cross the midline, was established in stage 1. The amount of external input required to achieve midline crossing would be noted by the caretaker on the system. Although the game is designed to determine the core impairments of crossing the midline, clinical reasoning, observations, and external guidance from the occupational therapist are vital for a holistic and accurate assessment. In the event that an occupational therapist is not performing the assessment (ie, caretaker, parent, teacher, or guardian), a notes page, shown in Figure 7, will need to be completed, which is presented to the caretaker at the end of the game. A guidance script with key points of reference will be provided to the caretaker before the child plays the game to assist and ensure accurate observations. Using the results, an individualized treatment plan tailored to the child’s needs would be constructed and developed by the occupational therapist, which can be done remotely.

Figure 7. Notes page to fill in after each stage is completed.

Stage 2
Stage 2 takes the starting distance \((X_0)\) and distance variance as inputs. Here, the distance variance is the distance from the starting distance, and the distance can be varied; for example, if the starting distance is 1.5 and the distance variance is 0.5, the ball can be spawned between points 1 and 2 on the x-axis.

Stage 2 is the first intervention stage, aimed at improving the child’s gross ability to cross the midline. This stage allows for the distance variable to be adjusted, whereas the speed and time variables are constant. The caretaker sets the distance at the specific distance where the child began experiencing problems with midline crossing. This distance can then be adapted and graded during the intervention.

Stage 3
Stage 3 takes the starting distance \((X_0)\), initial speed \((Sp_0)\), and incremental speed \((Is_p)\) as inputs. The initial speed is the speed at which the ball starts to fall. The incremental speed is the increase in speed from the starting point. The ball will continue to fall at the specified starting point; however, the speed at which the ball falls will increase as each ball is caught.

Stage 3, the second intervention stage, aims to enhance the child’s gross ability to cross the midline. The distance and time variables will remain constant, whereas the speed variable will be incremented. This increases sufficient accuracy and skill.

Stage 4
Stage 4 does not require any input as the position at which each ball is spawned is random. The point at which each ball is spawned is calculated using the random function \(\text{Rand()}\). As shown in Figure 6, the x-axis spans from −7.5 to 7.5; therefore, the ball can be spawned at a random point on the axis. The speed of the ball remains constant and does not increase as a function of the score. The gameplay time is constant for stage 4.

The purpose of stage 4, the unregulated round, is to maintain the skills the child has obtained in the previous stages. In this stage, the speed and time variables will remain constant, whereas the distance variable will be random.

Scoring
The occupational therapist advised on the scoring criteria for the stages during the collaborative design sessions. These scoring criteria were based on observations that occupational therapists would typically make during traditional treatments.
The first three criteria were generated from the gameplay data received from the game, and the following five criteria were recorded by observation. To confirm that the child did not display any form of pathology with midline crossing, a child would need to achieve a total of ≥4 points on their score card. It should be noted that all 8 criteria should be used in conjunction with one another and not assessed independently. Table 2 shows the 8 criteria, the observation being assessed, and the measurement to score the criterion for stage 1. If a child meets the scoring for a criterion, a point will be awarded. The criteria for the other stages are the same as those for stage 1, except for criterion 4. In the scoring criteria for stage 2, the effect of the random displacement of balls in a targeted area is determined rather than the increasing distance. In the scoring criteria for stage 3, the effect of increasing the speed of the balls is determined rather than increasing the distance. Finally, in the scoring criteria for stage 4, the effect of the balls randomly falling is determined.

Table 2. This table shows the scoring criteria for an assessment (stage 1).

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Observation</th>
<th>Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Balls caught</td>
<td>&gt;14</td>
</tr>
<tr>
<td>2</td>
<td>Balls missed</td>
<td>&lt;4</td>
</tr>
<tr>
<td>3</td>
<td>Average time between balls</td>
<td>&lt;7.5</td>
</tr>
<tr>
<td>4</td>
<td>Distance increasing</td>
<td>No pattern</td>
</tr>
<tr>
<td>5</td>
<td>Follow instructions</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>Eye movements</td>
<td>Maintain visual focus</td>
</tr>
<tr>
<td>7</td>
<td>Body movement</td>
<td>No gross adjustments</td>
</tr>
<tr>
<td>8</td>
<td>General behavior</td>
<td>No verbal cues</td>
</tr>
</tbody>
</table>

Data Visualization

Occupational therapists can support children with difficulty crossing the midline by helping them develop the skills needed to perform activities of daily living. During therapy sessions, occupational therapists use a variety of techniques to support rehabilitation, such as functional electrical stimulation, constraint-induced motor training, facilitation, and virtual reality applications [27]. This combination of therapy, exercise, and context-specific retraining is critical for neuroplasticity, as mentioned in the Introduction. Although retrospective recall and exercise diaries can gather subjective data, the quality of these data is limited and relies on the notes and observations written by an occupational therapist; as such, occupational therapists lack objective data about the degree to which exercises have been performed. Consequently, a dashboard was designed to aid occupational therapists by providing a collection of objective data about children who have difficulty crossing the midline. The proposed monitoring tool provides a solution in which the gameplay of the child during an assessment or treatment is recorded, and the data collected in the game can be visualized by the therapist. The design of the system, and the monitoring functionality in particular, allows for the game and input device to be used by parents or guardians with the children in a home environment and to be used remotely. Therefore, the game would not be restricted to only being played during therapy sessions but rather can be played outside of these times as well. A dashboard was designed using Microsoft’s PowerBI to display the data illustrated in Figure 8.
During the evaluation process, the occupational therapists were asked to play the game themselves and get familiar with how the system works. Figure 9 shows an illustrative example of how the dashboard would look for a child without midline crossing difficulties. The dashboard is accessible remotely. Figure 10 is an illustrative example of a dashboard for a child with midline crossing difficulties. The red circle indicates that as the distance increased, the child missed more balls. In this case, there are various factors that could impact the child’s results. For example, the child may have felt bored and, therefore, did not play as they should have. A child with an average performance result needs to be further assessed to see whether their results are because of pathology with midline crossing or whether there are other factors yielding these results.
Objective Two: Evaluate Whether the Web-Based Game Is Fit for Purpose

The therapists commented on the convenience of integrating both assessment and treatment into the same application as it assists the therapists when grading a child. The stages developed in the game were also commended by the therapists, as each stage has a different focus point with a specific outcome. They collectively agreed that the quantitative aspect that the game creates by providing measurable and standardized data proves advantageous when compared with traditional methods of assessment and treatment. When using a traditional method, such as asking the child to build a puzzle or draw a line across a page, there are no quantitative measures that can be deduced; instead, the assessments are limited to mere observation. The proposed solution elicits not only quantitative data but also allows observational data to be recorded using the notes page. Furthermore, they referred to the notes page as not only descriptive but also appreciated that it was designed to be understood by nonmedical professionals, thereby making the observational recording more effective and user-friendly. The notes page is viewed as an extremely beneficial aspect of the serious game as it ensures that the assessor covers all the observational components included in the assessment and treatment. As a result of the movements triggered by the game and the quantitative and observational data recorded by the application, the occupational therapists concluded that the serious game, accompanied by the input device, would assist them in identifying whether a child has pathology with midline crossing. Moreover, the solution would also assist in treating the child’s pathology. Finally, the therapists thought that the solution could assist in exposing and identifying other gross and fine motor pathologies that a child might have, or it might even reveal underlying behavioral, cognitive, or physical impairments. It was suggested that the assessment stage should be standardized per age group.

One of the ways to identify whether a child has pathology with midline crossing is by observing whether the child has a midline jerk. This can be noticed when the child is performing actions where the midline is crossed, and their eyes move from side to side in a rapid manner. Consequently, the therapist proposed incorporating eye tracking to detect if the child’s oculomotor functions are impaired. Additional monitoring through the use of a laptop camera can also be provided to observe the eye movement, head adjustments, and posture of the child.

As part of the game, when the child catches the object falling in the hat, the child is required to return the hat to the side of the screen to receive a point. Even though an alert is sounded when the object returns to the side, a suggestion was proposed to display a reminder on the screen to return the object to the side at the point where the child has caught the object. In addition, when the child is playing the game and controlling the input device, the child may release the device. Consequently, it was recommended that if the child releases the device during the game, an alert should be shown to remind the child to grasp the device.

Specific attention was given to the dashboard design as it is an important part of enabling remote monitoring. Everyone unanimously agreed that the dashboard’s ability to track the progress of the child would benefit their reports immensely. The results represented in the dashboard would assist them in supporting and verifying their observational conclusions. Furthermore, when submitting reports to their clients’ medical aids, the data from the dashboard can be used to solidify their conclusions and can be included in their reports.

The only disadvantage that was raised by the occupational therapists was that the dashboard might not be particularly easy to interpret for a layman; however, a suggestion was made to provide training to therapists on how to interpret and analyze the dashboard. The other disadvantage was that the dashboard was limited to only showing the results for a midline crossing.
assessment and no other difficulties that could be identified by the game.

There were suggestions to provide a way of showing the standardized results for different categories by searching by age, for example, and, consequently, the results for that particular age group would be displayed. Introducing standardized norms would allow the child to be compared with other children within the specific area being examined. Another suggestion that was made was to allow therapists to add additional comments on the results of the child displayed for further reference. Finally, it was proposed that artificial intelligence could be incorporated so that when comments are made on the notes page by the therapist or guardian, the observations can be translated into quantitative data that can be used to calculate the scoring criteria instead of physically inputting the scores for the criteria that require observations.

The occupational therapists agreed that this solution could be used as a telehealth device. To further enhance these benefits, it was suggested that the game be deployed as a mobile app. This would allow for the game to be available on smart devices, which would, therefore, make it even more accessible than if it were only available on a laptop or computer. A second proposition was to record the gameplay so that it could be played back for reference. It was recommended that a video or chat capability be introduced so that the child or guardian could communicate with the therapist.

Discussion

Principal Findings

A telehealth system consisting of a serious game accompanied by an input device and a dashboard can be implemented to address children’s midline crossing difficulties. The scoring system provides a quantitative aspect that proves advantageous when compared with traditional methods of assessment and treatment, where the assessment is limited to mere observation. In addition, the notes page that is completed at the end of the serious game ensures that the observational components that are vital in assessing a child are still included in the overall assessment. Therefore, the telehealth system elicits not only quantitative data but also allows observational data to be recorded using the notes page.

Although there are many input devices that can be used for therapeutic reasons, as described in the literature section, they each have shortcomings that would need to be addressed to make them more suitable when used by a child with a pathology with midline crossing. The common drawback of all the devices is their cost. It is not feasible to expect parents to buy expensive devices for treatment at home that will only be used for a limited time. A Kinect system will cost approximately US $399 and a Wii, US $164, whereas the proposed system costs approximately US $32 to manufacture. This price may also decrease if large quantities are manufactured.

The Wii provides a suitable solution for physical training and balance in particular; however, there are no games that explicitly aim to treat or assess children with midline crossing [18]. At first, Leap Motion seemed to be a suitable device for the development of a VT; however, with inaccurate readings stemming from the device affecting the game experience, the device would not be suitable for children [21]. The RealSense is a fitting tool; however, because of the high price of the device, it is not suitable [19].

Studies with a serious game and Kinect showed that individuals are more motivated, enjoy therapy, and even enable therapy to be more accessible [12]. On the basis of the feedback from the occupational therapists, the child will also experience these benefits from the system presented in this paper. However, this will only be proven in clinical studies.

Limitations

When determining whether serious games are feasible as a treatment option for midline crossing difficulties and are fit for purpose, there appears to be a lack of clinical evidence about the benefit to children from the application of serious games [28]. Owing to ethical constraints regarding testing the solution directly on children, professional opinions of occupational therapists were gathered to validate the solution. The next step would be to obtain the needed clearance to test directly on children without difficulties crossing the midline to attain a baseline. Thereafter, tests can be performed on children who have difficulty crossing the midline. Although the dashboard was configurable, it was found to be slightly difficult to interpret. It was suggested that therapists could be trained to interpret and analyze the dashboard. The telehealth system focuses on only one intervention. Therefore, the possibility of using the serious game and input device to assist in exposing and identifying other gross and fine motor difficulties that a child might have could be investigated.

Future Work

The therapists were confident that the telehealth system presented will assist them in identifying and treating children with midline crossing difficulties. Therefore, the possibility of using this solution to expose and identify other gross and fine motor difficulties that a child might have could be investigated. Furthermore, the solution could even reveal underlying behavioral, cognitive, or physical impairments. To clarify, the therapists would be able to identify other aspects, such as hand dominance, eye tracking, postural control, range of motion, attention and focus of the child, and the child’s hand functions (grips and grasps) when playing the game. Furthermore, when identifying whether a child has midline crossing difficulties, a midline jerk can be observed. When a child is confronted with actions in which the midline is crossed, their eyes move from side to side rapidly. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.

One of the actions required in the game when a falling object is caught is to return the object to the side to receive a point. As a result, adding eye-tracking capabilities to the solution was put forward. To mitigate the concern raised by the therapists regarding the fact that an older child may achieve the game objectives quicker than a younger child, a standardized assessment stage per age group can be introduced.
child may release the device. Consequently, it was suggested that in the event that the device is released during the game, an alert should be displayed to remind the child to remain holding the device. In addition, the results presented on the dashboard are limited to displaying the results for 1 player at a time. A recommendation was proposed to provide a way of showing standardized results for different categories, such as age. This would allow the child to be compared with other children within a specific area examined.

In future iterations, artificial intelligence can be incorporated so that when comments are made on the notes page by the therapist or guardian, the observations can be translated into quantitative data that can be used to calculate the scoring criteria instead of physically inputting the scores for the criteria that require observations. Finally, the solution proposed in this dissertation extracts gameplay data that can assist occupational therapists in identifying and treating children with difficulty crossing the midline. A greater number of occupational therapists adopting the solution in their practices means that more children will play the game and, ultimately, means more data are accumulated. These data can be analyzed using machine learning algorithms to find trends and enhance the assessment and treatment processes and visualizations produced.

Conclusions
A solution was designed to determine whether a telehealth system comprising a serious game can assist occupational therapists in identifying whether a child has pathology with midline crossing and can assist in treating the child remotely. Serious games are introduced as a need to meet objectives that go beyond entertainment and benefit the user in the area that needs to be mitigated. Through collaboration with occupational therapists, the telehealth system was designed to make use of different levels in the serious game, where each level addressed a different need of the therapy process. A novel, low-cost input device accompanies the serious game to track the movement of a child’s hand from side to side and transform the position into the game. All results are saved on the web, and occupational therapists can access a dashboard that displays the results of each child. In addition, observed behavioral information will also be saved to assist occupational therapists in making decisions regarding changes to the intervention. During the interviews, occupational therapists indicated that the dashboard would support their treatment plan and that the end-to-end solution was indeed feasible.

Acknowledgments
The authors would like to thank all the occupational therapists who took part in this study.

Conflicts of Interest
None declared.

References
9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games); Sep 6-8, 2017; Athens, Greece. [doi: 10.1109/vs-games.2017.8056599]


Abbreviations

IMU: inertial measurement unit
VT: virtual therapist

Edited by N Zary; submitted 05.02.21; peer-reviewed by I Manuaba, M Stein; comments to author 19.03.21; revised version received 11.05.21; accepted 19.06.21; published 01.11.21.

Please cite as:
Jacobs J, Heymann R, Greeff JJ
A Telehealth System Incorporating a Serious Game Intervention to Aid Occupational Therapists in Identifying and Treating Children With Difficulty Crossing the Body's Midline: Key Informant Interviews Among Occupational Therapists
JMIR Serious Games 2021;9(4):e27761
URL: https://games.jmir.org/2021/4/e27761
doi: 10.2196/27761
PMID: 34723821

©Jonathan Jacobs, Reolyn Heymann, Jacob Jacobus Greeff. Originally published in JMIR Serious Games (https://games.jmir.org), 01.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Feasibility of a Sensor-Controlled Digital Game for Heart Failure Self-management: Randomized Controlled Trial

Kavita Radhakrishnan¹, PhD; Christine Julien², PhD; Tom Baranowski³, PhD; Matthew O’Hair⁴, MA; Grace Lee², MS; Atami Sagna De Main¹, MS; Catherine Allen¹, MPH; Bindu Viswanathan⁴, PhD; Edison Thomaz², PhD; Miyong Kim¹, PhD

¹School of Nursing, The University of Texas Austin, Austin, TX, United States
²Department of Electrical and Computer Engineering, Cockrell School of Engineering, The University of Texas Austin, Austin, TX, United States
³Baylor College of Medicine, Houston, TX, United States
⁴Good Life Games, Inc, Austin, TX, United States
⁵Department of Statistics and Data Sciences, The University of Texas Austin, Austin, TX, United States

Corresponding Author:
Kavita Radhakrishnan, PhD
School of Nursing
The University of Texas Austin
1710 Red River St
Austin, TX, 78701
United States
Phone: 1 512 471 7936
Email: Kradhakrishnan@mail.nur.utexas.edu

Abstract

Background: Poor self-management of heart failure (HF) contributes to devastating health consequences. Our innovative sensor-controlled digital game (SCDG) integrates data from sensors to trigger game rewards, progress, and feedback based on the real-time behaviors of individuals with HF.

Objective: The aim of this study is to compare daily weight monitoring and physical activity behavior adherence by older adults using an SCDG intervention versus a sensors-only intervention in a feasibility randomized controlled trial.

Methods: English-speaking adults with HF aged 55 years or older who owned a smartphone and could walk unassisted were recruited from Texas and Oklahoma from November 2019 to August 2020. Both groups were given activity trackers and smart weighing scales to track behaviors for 12 weeks. The feasibility outcomes of recruitment, retention, intervention engagement, and satisfaction were assessed. In addition to daily weight monitoring and physical activity adherence, the participants' knowledge, functional status, quality of life, self-reported HF behaviors, motivation to engage in behaviors, and HF-related hospitalization were also compared between the groups at baseline and at 6, 12, and 24 weeks.

Results: A total of 38 participants with HF—intervention group (IG: 19/38, 50%) and control group (CG: 19/38, 50%)—were enrolled in the study. Of the 38 participants, 18 (47%) were women, 18 (47%) were aged 65 years or older, 21 (55%) had been hospitalized with HF in the past 6 months, and 29 (76%) were White. Furthermore, of these 38 participants, 31 (82%)—IG (15/19, 79%) and CG (16/19, 84%)—had both weight monitoring and physical activity data at the end of 12 weeks, and 27 (71%)—IG (14/19, 74%) and CG (13/19, 68%)—participated in follow-up assessments at 24 weeks. For the IG participants who installed the SCDG app (15/19, 79%), the number of days each player opened the game app was strongly associated with the number of days the player engaged in weight monitoring ($r=0.72; P=.04$) and the number of days with physical activity step data ($r=0.9; P<.001$). The IG participants who completed the satisfaction survey (13/19, 68%) reported that the SCDG was easy to use. Trends of improvement in daily weight monitoring and physical activity in the IG, as well as within-group improvements in HF functional status, quality of life, self-reported HF behaviors, motivation to engage in behaviors, and HF-related hospitalization were also compared between the groups at baseline and at 6, 12, and 24 weeks.

Conclusions: Playing an SCDG on smartphones was feasible and acceptable for older adults with HF for motivating daily weight monitoring and physical activity. A larger efficacy trial of the SCDG intervention will be needed to validate trends of improvement in daily weight monitoring and physical activity behaviors.

Trial Registration: ClinicalTrials.gov NCT03947983; https://clinicaltrials.gov/ct2/show/NCT03947983
Introducing the Role of Innovative Digital Health Interventions for Motivating Heart Failure Self-management Behaviors

Introduction

Need for Heart Failure Self-management Behaviors

Despite significant advances in treatment and management, heart failure (HF) continues to be the leading cause of hospitalization among older adults in the United States [1,2], with an estimated annual cost of US $32 billion [1]. Improved self-management behaviors, defined as “behaviors that maintain physiological stability” and enable “response to symptoms when they occur,” can help reduce the adverse effects of an HF diagnosis [3,4]. Daily weight monitoring is such a behavior; weight gain is typically the first sign of volume overload in patients with HF, and if weight gain is treated promptly, clinically significant HF exacerbations can be avoided [5,6]. Similarly, physical activity improves myocardial function and functional capacity [7] and reduces depressive symptoms [8]. Engagement in these 2 critical HF self-management behaviors improves quality of life and reduces health care use [9,10]. Yet, weight monitoring and physical activity show significantly poorer adherence than other HF self-management behaviors [11,12] because of poor knowledge and lack of motivation [13].

Role of Innovative Digital Health Interventions for Motivating HF Self-management Behaviors

Recent advances in wearable technology, digital sensor devices, and mobile health (mHealth) apps have allowed application of these technologies to observe and motivate health behaviors in real-world conditions. Yet, despite the advantages of portability and scalability offered by digital devices and apps, long-term adoption rates remain low. There have been many instances of abandonment of the use of activity trackers in prior research, with more than 50% of the activity tracker users stopping use within a period of 2 weeks to 6 months [14,15]. Long-term maintenance rates of such devices by patients with HF are yet to be explored.

Digital games that serve as affordable, portable, scalable tools while being enjoyable and easy to use can help contextualize health behaviors [16] and motivate clinically significant behavior changes, thereby improving health outcomes [17]. Digital games can combine appealing stories [18] and interests and hobbies [19] with active learning [20], incentives [21], and social connections [22] to offer an accessible, engaging, and immersive habit-forming medium [23] while objectively measuring these behaviors [24,25]. Well-designed digital games have significantly improved physical (eg, balance and mobility) and cognitive (eg, processing speed) health outcomes [26] as well as behavioral outcomes for chronic diseases (eg, dietary changes, physical activity, tobacco use) [24,25].

Digital games incorporating gamification principles such as competition, leaderboards, and incentives have improved learning outcomes and have been highly acceptable among participants with HF [29,30], but these games do not involve data on real-time HF self-management behaviors. In a sensor-controlled digital game (SCDG), data on behaviors from sensors, including those in wearable devices, are synchronized with a mobile gaming app to trigger game progress, rewards, personalized and contextually relevant feedback (eg, reduce fluid intake or call physician for weight gain), and incentives based on participants’ real-time behaviors [24]. Combining digital games and sensors thus offers a powerful way to improve behavior adherence and potentially make health care participatory, personalized, predictive, and preventive as defined by the Precision Medicine Initiative [31].

Thus, the primary goals of this study are to obtain preliminary efficacy data on behavior adherence and undertake a comprehensive feasibility assessment of an SCDG intervention called Heart Health Mountain. This app synchronizes with a Bluetooth-enabled weighing scale and activity tracker to activate game rewards and feedback based on the real-time weight monitoring and exercise behaviors of older adult participants with HF.

Methods

Study Design and Scope

We conducted a prospective feasibility randomized controlled trial (1:1) with 2 parallel groups (sensors only or sensors plus SCDG app) from November 2019 to January 2021 (ClinicalTrials.gov: NCT03947983). The results presented here are from the 12-week SCDG intervention measuring objective behaviors of daily weight monitoring and physical activity and the 24-week follow-up measuring quality of life, functional status, HF self-management efficacy, and HF hospitalizations.

Study Population and Recruitment

Before the COVID-19 outbreak, we identified potential participants through chart review at a cardiac rehabilitation center as well as through referrals by clinical case managers at an inpatient cardiac floor in central Texas. Research staff members provided information on the study to adults who were aged 55 years or older, English-speaking, diagnosed with HF, and classified according to the New York Heart Association’s HF classification as class II or III [32] during their inpatient stay or outpatient visit to the cardiac center. Other eligibility criteria included smartphone ownership as a proxy measure to indicate prior familiarity with smartphone use, ability to independently walk without a walker or human assistance, and a score of 4 or higher on the Mini-Cog [33] cognitive screen. The exclusion criteria included severe visual or tactile impairments (eg, legal blindness or severe arthritis), which would prevent the use of a smartphone, or end-stage renal failure...
or terminal illness (eg, cancer), both of which adversely affect HF prognosis [34].

To continue the trial during the COVID-19 pandemic, all in-person interactions were converted to remote interactions. We used a secure, Health Insurance Portability and Accountability Act–compliant email system to continue receiving referrals from clinical case managers at the inpatient HF center. In addition, we contracted with the participant recruitment company Trialfacts [35] to recruit participants on the web from the states of Texas and Oklahoma. The eligibility criteria remained the same, but the formal screening process then required a narrative description of HF history or confirmation from the potential participant’s health care provider to confirm the individual’s HF diagnosis.

**Intervention**

The description of the design, development, and usability assessment phase of the SCDG has been detailed elsewhere [25]. Guided by the Fogg behavioral model [36] and played on a smartphone, the Heart Health Mountain SCDG presented a narrative in which the older adult player helps an avatar to climb a mountain in a forested area, with the game’s goal being to avoid hospitalization. The SCDG motivates the player’s engagement in critical behaviors related to HF self-management, helping the avatar to climb the mountain. This is done by daily weight monitoring and attaining physical activity steps based on goal steps tailored to the individual player. The step goal in the SCDG ranged from 3000 to 15,000 steps.

According to the Fogg model [36], for a person to perform a target health behavior, they must (1) be sufficiently motivated, (2) have the ability to perform the behavior, and (3) be triggered to perform the behavior. For example, based on the motivation concept of the Fogg model [36], positive attitudes toward engaging in HF self-management and empathy with the game character will be aided by positive feedback, competition (eg, leaderboards), and game rewards (eg, coins to buy low-salt food items or accessories for the game character for engaging in real-time behaviors and in-game challenges). On the basis of the ability concept of the Fogg model [36], knowledge about HF self-management within the SCDG was provided using language especially chosen for those with low literacy levels (from the Living Well With Heart Failure booklet) [37,38] in bite-sized chunks (Figure 1). Built-in quizzes tested content mastery, and problem-solving strategies provided opportunities for higher game rewards (Figure 2).

Finally, based on the trigger concept of the Fogg model [36], game alerts, the avatar’s health status, messages (Figure 3), and incentives were tailored to the participant’s real-time HF self-management behaviors based on data from the behavior-tracking sensors (eg, climbing up the mountain by 2 steps if the physical activity goal was attained that day). HF game players are expected to make meaningful connections between game events and real-time HF self-management behaviors, increasing the likelihood that they retain and apply their newly acquired knowledge, skills, and habits for HF self-management behaviors outside the game’s world [39] and improve their health outcomes.
Figure 1. Knowledge embedded in the sensor-controlled digital game.

What is heart failure?

Heart failure causes water to leak out of your blood vessels. This water can fill up your lungs and make you short of breath. It can also cause your legs to swell.
Figure 2. Built-in quiz in the sensor-controlled digital game.

John starts exercising on Monday. After exercising for few minutes, he started to feel chest pain and dizziness. John decides to stop exercising right away. Did John take the correct decision?
Figure 3. Example of game alert messages based on real-time behavior.

### Intervention Procedures

CONSORT (Consolidated Standards of Reporting Trials) guidelines [40] informed this parallel-group randomized controlled trial and the reporting of outcomes (Multimedia Appendix 1). The SCDG intervention group (IG) received sensors tracking weight monitoring and physical activity and played the SCDG app on a smartphone; the sensors-only control group (CG) received sensors tracking weight monitoring and physical activity only. Both IG and CG participants were given the Withings Go activity tracker (Withings) [41], Withings Body smart weighing scale [42], and Withings Health Mate app [43] to record, store, and transmit daily weight and physical activity data, but the CG did not receive the SCDG app. The Withings Body smart weighing scales have been found to accurately measure body mass [44]. We selected the Withings Go tracker because it has long battery life, is waterproof, is reasonably accurate in measuring activity levels of people with HF [45,46], and uses a clock schema, which is a familiar interface for most adults, to represent the user’s daily physical activity. The step goals for all participants were set based on their preferences as well as the nursing research assistant’s assessment of their physical health status.

Whereas the participants in the IG received standardized HF education [6,38] embedded within the SCDG, the participants in the CG received the same information in written format. Thus, the difference between the IG and the CG was the receipt of the SCDG gaming elements. The SCDG app transmitted the IG participants’ game-playing data, and the Health Mate app transmitted all participants’ sensor data to the research team through cellular data service or home Wi-Fi. An intervention nursing research assistant remotely tracked user engagement for participants in both groups during the first week and contacted participants by phone in the absence of any data to offer help with troubleshooting.

Before the COVID-19 outbreak, an intervention nursing research assistant installed the apps on all participants’ smartphones and trained the participants to use the sensor devices and apps at either their homes or at a location of their convenience. During the COVID-19 pandemic, in-person interactions were converted to remote interactions. Contactless delivery of study equipment, videoconferencing on smartphones, and printed pictorial and video instructions allowed remote support for installation, training, and troubleshooting of the devices and apps. The study team members’ experience with the devices and apps during in-person interactions with the older adults at their homes helped
inform the training materials that were provided to the participants during the pandemic. Installation and training times by the research team member varied widely during the remote phase, from 0 minutes (participant self-installation) to 180 minutes compared with 45 to 120 minutes during the in-person phase.

Randomization

The participants were assigned to the IG or CG with a 1:1 randomization ratio such that the 2 groups were equivalent in terms of biological sex. The randomization was carried out after informed consent was obtained and the baseline survey completed. The allocation sequence list for block randomization was generated by an independent researcher at the Sealed Envelope Ltd website [47] with random block sizes of 2 and 4 and concealed until the trial group was assigned. To ensure blinding during assessment of the outcomes, the research assistant who delivered the intervention to the 2 groups was different from the research assistant who collected the baseline and follow-up surveys from the 2 groups. The researcher who performed the data analysis was blinded to the participant groups. Indication of participation in a digital health intervention study in the informed consent allowed us to blind all participants to their specific group assignment.

Outcome Measures

Feasibility Outcomes

To inform the development of a future large-scale randomized controlled clinical trial, we determined the feasibility, acceptability, and effectiveness of our SCDG in engaging older adults with HF by assessing recruitment (how many accepted the invitation to participate in the study), retention, engagement, and satisfaction. We considered the trial feasible if the enrollment rate was at least 20% of the patients with HF who were approached to participate in the study, if at least 50% of the IG and CG participants completed both the 12-week objective behavior assessment and the 24-week follow-up assessment (retention), and if at least 50% of the IG participants used the SCDG intervention for at least 50% of the days (engagement) and were satisfied with the SCDG intervention (satisfaction).

Retention

Retention was recorded as the number (proportion) of participants in both groups who used the sensor devices to measure physical activity and weight monitoring for 12 weeks and completed follow-up assessments at 24 weeks to assess maintenance of the SCDG intervention.

Engagement and Adherence

A Google Cloud console was created to store all participants’ deidentified physical activity and weight monitoring behavioral data obtained from the Withings Health Mate app as well as data used to assess the IG participants’ engagement with the SCDG and its features. The number of days of adherence to physical activity and weight monitoring as well as engagement with the SCDG were obtained from the Google Cloud console.

Satisfaction

We assessed the IG participants’ satisfaction with the SCDG using a questionnaire at the end of 12 weeks based on the 4-item interest and enjoyment subscale of the Intrinsic Motivation Inventory (α=.8) [48], a multidimensional instrument intended to assess participants’ subjective experience related to a target activity. The interest and enjoyment subscale assesses an individual’s interest and inherent pleasure in performing a specific activity on a 4-point Likert scale (1=not satisfied, 4=very satisfied). In addition, the participants were asked open-ended questions about what they liked and disliked most about using the SCDG for HF self-management. A 4-point scale (1=strongly disagree, 4=strongly agree) was used to assess the key elements that might have helped the participants maintain their motivation to continue using the SCDG. Finally, the participants were asked whether they would recommend the SCDG to others with HF.

Health and Behavior Outcomes

Overview

All participants were asked to complete surveys at 6, 12, and 24 weeks after the baseline survey to assess the immediate effect of the intervention and maintenance of behavioral changes. Before the COVID-19 outbreak, the participants were asked to complete the surveys on the Qualtrics platform (Qualtrics XM) on a study iPad (Apple Inc). During the pandemic, the participants were emailed or texted links to the Qualtrics surveys. All participants completed a sociodemographic survey questionnaire at baseline. They also completed the 2-item Patient Health Questionnaire [49] to report on depressive symptoms at baseline and at 6, 12, and 24 weeks.

Weight Monitoring Behavior

The primary outcome of the days of weight monitoring was measured by the number of days with weight monitoring data. This measure was collected from sensor logs within the Health Mate app [43], with each day of weighing measured dichotomously (yes or no).

Physical Activity Behavior

Physical activity data were derived from the Withings Go sensor logs within the Health Mate app [43] by obtaining the cumulative steps for each day and averaging the steps for each participant over 6 and 12 weeks.

HF-Related Functional Status

For this measure, we used items 1-12 from the Kansas City Cardiomyopathy Questionnaire (KCCQ) [50]. The item scores are transformed to a range of 0 to 100 by subtracting the lowest possible scale score, dividing by the range of the scale, and multiplying by 100. The summary scores for the functional status thus range from 0 to 100, with higher scores indicating better functional status. The Cronbach α value for the KCCQ functional status is excellent at .93 [49]. The scale has demonstrated criterion validity, with high correlation with the New York Heart Association HF classification and the 6-minute Walk Test [50].
Quality of Life

Quality of life was measured with items 13-15 from the KCCQ [50]. The item scores are transformed to a range of 0 to 100 by subtracting the lowest possible scale score, dividing by the range of the scale, and multiplying by 100. The summary values for the quality-of-life domain thus range from 0 to 100, with higher scores indicating better quality of life. The Cronbach $\alpha$ value was .78 in prior studies [50]. In comparison with similar quality-of-life instruments, the KCCQ is sensitive to clinical changes in HF and has shown a significant and high correlation with the New York Heart Association HF classification [50].

HF Self-management Knowledge

The instrument used was the 30-item Atlanta Heart Failure Knowledge Test [51]. Each correct answer is scored as 1, with no additional weighting of items; the correct responses are then summed. Incorrect or skipped questions are scored as 0. The total scores range from 0 to 30, and higher scores indicate better HF knowledge. The content validity ratings for relevance and clarity ranged from 0.55 to 1.0, with 81% of the items rated 0.88 to 1.0. The Cronbach $\alpha$ value for reliability was .84 [51]. Construct validity has been demonstrated by directly correlating knowledge with clinical and self-care outcomes, including dietary sodium consumption, medication adherence, and health care use [51].

Self-reported HF Self-care Behaviors

The instrument used was the 9-item European Heart Failure Self-care Behavior Scale [52]. The items were scored on a Likert scale of 1 to 5. For calculating the standardized score, each item was reverse coded and then computed using the following formula: (sum of all even-coded items minus 9) times 2.7777. The standardized scores range from 0 to 100; every item is given equal weight, with a higher score indicating better self-care. The coefficient $\alpha$ value was .80 for this instrument [52].

HF Self-efficacy

The instrument used was the 6-item Section C (self-efficacy section) of the Self-Care of Heart Failure Index [53]. The item scores were standardized using the following formula: (sum of Section C items minus 6) times 5.56. The scores on the self-efficacy scale range from 0 to 100, with higher scores reflecting better self-efficacy. The Cronbach $\alpha$ value was .88, with good evidence for construct validity and contrasting group validity [53].

Motivation for HF Self-management Behaviors

The instrument used was the 19-item Treatment Self-Regulation Questionnaire [54], adapted for HF. Each item was scored on a scale of 1 to 7. The scale has 2 subscales: autonomous regulation and controlled regulation. The score for each subscale was the average of the items in that subscale. The average for controlled regulation was subtracted from the average for autonomous regulation to calculate the Relative Autonomy Index. The Cronbach $\alpha$ value for chronic disease–related behaviors ranged from .83 to .87 [54]. Construct and convergent validity have been confirmed using factor analysis and correlations with autonomy-related subscales of other behavior scales and health outcomes [55].

HF Hospitalization

This measure was obtained through the participants’ self-report (yes or no) on the 6-, 12-, and 24-week surveys and was confirmed with hospitalization discharge summaries or communication with health care providers.

Statistical Analysis

Power

On the basis of 1:1 randomization with 80% power ($\alpha$=5%), a sample size of 38 patients per group would have been required to detect a difference of 80% versus 50% in both groups of daily monitoring of weight monitoring and physical activity. We chose 80% as the cutoff for adequate weight monitoring at 12 weeks because patients with HF who completed at least 80% of the weight diaries (5.6/7 of the days per week) were found to have significantly reduced odds for HF-related hospitalizations in comparison with patients who completed less than 80% of the weight diaries [6]. In addition, only 50% of the participants with HF in a remote monitoring sensor group recorded their weights more than 50% of the time [56]. Allowing for 10% attrition, 49 patients per group (N=98) would have had to be recruited for a fully powered clinical trial of the SCDG intervention for HF behavior of weight monitoring. The power analysis was performed using G*Power (Heinrich Heine University) [57]. With a sample of 38, our study was underpowered. However, this sample size was sufficient to conduct a feasibility study [58], which can inform implementation of a fully powered study with fewer problems to test the SCDG’s effectiveness.

Data Analysis

Descriptive statistics for feasibility included (1) the percentage of participants recruited from the total who were approached, (2) the percentage of participants (among those recruited) who were retained in the study at the end of 24 weeks in each arm and overall, (3) the number of days the IG participants played the SCDG, and (4) average satisfaction with the SCDG. All open-ended questions were coded by 2 team members (CA and AS) and then analyzed using a general inductive thematic approach [59].

The observed effect sizes (Cohen $d$) for the primary outcome of weight monitoring days at the end of 12 weeks was calculated using the following formula: mean of days in IG minus mean of days in CG divided by average SD of days in IG and days in CG. The potential clinical meaningfulness of the results (in addition to statistical significance) was based on the magnitude of the effects: small (Cohen $d$=0.20), medium (Cohen $d$=0.50), and large (Cohen $d$=0.80) [60].

All statistical analyses were conducted using SPSS software (version 26.0; IBM Corp). Baseline characteristics of the participants in the IG and CG were compared using independent 2-tailed $t$ tests for continuous variables and chi-square tests for categorical variables. Although this study was a feasibility trial, we assessed within-group trends in the IG and CG using paired-sample $t$ tests (2-tailed) at baseline, 6 weeks, 12 weeks, and 24 weeks. Within-group changes were presented as absolute changes from baseline. All data were presented as means with...
SDs. Missing values were addressed using intent-to-treat principles with maximum likelihood estimations.

**Ethical Considerations**

The institutional review board at the University of Texas at Austin approved this study on July 30, 2018 (number 2017-12-0042). Interested participants provided their phone numbers to be contacted by a nursing graduate research assistant for formal screening through phone. If potential participants were found eligible, the research assistant scheduled visits at their homes to complete the written informed consent.

During the COVID-19 pandemic, the eligible potential participants completed the verbal informed consent with the research assistant on the phone. The participants were also emailed a copy of the informed consent form. To protect the participants’ privacy, 6-digit unique ID numbers and dummy email addresses were generated for the profile information that was required by the Health Mate app [43] and Google Cloud console so that no personal patient information was shared or processed through the app or the console. Participants who scored higher than 4 on the Patient Health Questionnaire (2 items) for depression [49] were provided resources for local mental health services, including available mental health services at their HF clinical site. The eligible participants received US $75 as incentives spread over 4 time points for participating in the study. The participants who continued in the study during the COVID-19 pandemic were also allowed to keep the weighing scale and tracker to avoid reuse of the devices.

**Results**

**Baseline Characteristics**

The characteristics of the participants are presented in Table 1. There were no marked differences between the IG and the CG except for marital status, with significantly more IG participants being married or having a partner.
Table 1. Baseline characteristics of the cohort (N=38).

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>SCDG&lt;sup&gt;a&lt;/sup&gt; (IG&lt;sup&gt;b&lt;/sup&gt;; n=19), n (%)</th>
<th>Sensors only (CG&lt;sup&gt;c&lt;/sup&gt;; n=19), n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>11 (58)</td>
<td>9 (47)</td>
<td>20 (53)</td>
</tr>
<tr>
<td>65-74</td>
<td>7 (37)</td>
<td>5 (26)</td>
<td>12 (32)</td>
</tr>
<tr>
<td>≥75</td>
<td>1 (5)</td>
<td>5 (26)</td>
<td>6 (15)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td>0.99</td>
</tr>
<tr>
<td>Male</td>
<td>10 (53)</td>
<td>10 (53)</td>
<td>20 (53)</td>
</tr>
<tr>
<td>Female</td>
<td>9 (47)</td>
<td>9 (47)</td>
<td>18 (47)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>3 (16)</td>
<td>0 (0)</td>
<td>3 (8)</td>
</tr>
<tr>
<td>Non-Hispanic or Latino</td>
<td>16 (84)</td>
<td>19 (100)</td>
<td>35 (92)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td>0.25</td>
</tr>
<tr>
<td>White</td>
<td>16 (84)</td>
<td>13 (69)</td>
<td>29 (76)</td>
</tr>
<tr>
<td>African American</td>
<td>2 (11)</td>
<td>4 (21)</td>
<td>6 (16)</td>
</tr>
<tr>
<td>Native American</td>
<td>0 (0)</td>
<td>1 (5)</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (5)</td>
<td>1 (5)</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>Highest level of education</strong></td>
<td></td>
<td></td>
<td>0.90</td>
</tr>
<tr>
<td>High school</td>
<td>4 (21)</td>
<td>3 (16)</td>
<td>7 (18)</td>
</tr>
<tr>
<td>University education</td>
<td>8 (42)</td>
<td>8 (42)</td>
<td>16 (42)</td>
</tr>
<tr>
<td>Technical diploma</td>
<td>7 (37)</td>
<td>8 (42)</td>
<td>15 (40)</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>Married or has a partner</td>
<td>13 (68)</td>
<td>7 (37)</td>
<td>20 (53)</td>
</tr>
<tr>
<td>Divorced or widowed</td>
<td>6 (32)</td>
<td>12 (63)</td>
<td>18 (47)</td>
</tr>
<tr>
<td><strong>Living arrangement</strong></td>
<td></td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Living alone</td>
<td>5 (26)</td>
<td>6 (32)</td>
<td>11 (29)</td>
</tr>
<tr>
<td>Living with others</td>
<td>12 (63)</td>
<td>12 (68)</td>
<td>24 (63)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (11)</td>
<td>0 (0)</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>Duration with HF diagnosis</strong></td>
<td></td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td>&lt;6 months</td>
<td>6 (32)</td>
<td>5 (26)</td>
<td>11 (29)</td>
</tr>
<tr>
<td>7 months to 1 year</td>
<td>2 (11)</td>
<td>3 (16)</td>
<td>5 (13)</td>
</tr>
<tr>
<td>1-5 years</td>
<td>7 (37)</td>
<td>5 (26)</td>
<td>12 (32)</td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>4 (21)</td>
<td>6 (32)</td>
<td>10 (26)</td>
</tr>
<tr>
<td><strong>Last HF hospitalization</strong></td>
<td></td>
<td></td>
<td>0.51</td>
</tr>
<tr>
<td>&lt;1 month</td>
<td>5 (26)</td>
<td>6 (32)</td>
<td>11 (29)</td>
</tr>
<tr>
<td>1-6 months</td>
<td>7 (37)</td>
<td>3 (16)</td>
<td>10 (26)</td>
</tr>
<tr>
<td>7 months to 1 year</td>
<td>1 (5)</td>
<td>2 (11)</td>
<td>3 (8)</td>
</tr>
<tr>
<td>&gt;1 year</td>
<td>6 (32)</td>
<td>8 (42)</td>
<td>14 (37)</td>
</tr>
<tr>
<td><strong>Prior digital game-playing</strong></td>
<td></td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td>Yes</td>
<td>17 (90)</td>
<td>13 (68)</td>
<td>30 (79)</td>
</tr>
<tr>
<td>No</td>
<td>2 (10)</td>
<td>6 (32)</td>
<td>8 (21)</td>
</tr>
<tr>
<td><strong>Depression (&gt;2 on PHQ&lt;sup&gt;e&lt;/sup&gt;-2)</strong></td>
<td></td>
<td></td>
<td>0.99</td>
</tr>
</tbody>
</table>
### Baseline characteristics

<table>
<thead>
<tr>
<th></th>
<th>SCDG&lt;sup&gt;a&lt;/sup&gt; (IG&lt;sup&gt;b&lt;/sup&gt;; n=19), n (%)</th>
<th>Sensors only (CG&lt;sup&gt;c&lt;/sup&gt;; n=19), n (%)</th>
<th>Total, n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yes</strong></td>
<td>5 (26)</td>
<td>5 (26)</td>
<td>10 (26)</td>
<td>.62</td>
</tr>
<tr>
<td><strong>Number of comorbid conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-2</td>
<td>12 (63)</td>
<td>13 (68)</td>
<td>25 (66)</td>
<td></td>
</tr>
<tr>
<td>&gt;2</td>
<td>8 (42)</td>
<td>6 (32)</td>
<td>14 (37)</td>
<td></td>
</tr>
<tr>
<td><strong>Top 2 comorbid conditions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertension</td>
<td>10 (53)</td>
<td>10 (53)</td>
<td>20 (53)</td>
<td>.99</td>
</tr>
<tr>
<td>Diabetes</td>
<td>8 (42)</td>
<td>5 (26)</td>
<td>13 (34)</td>
<td>.55</td>
</tr>
</tbody>
</table>

<sup>a</sup>SCDG: sensor-controlled digital game.  
<sup>b</sup>IG: intervention group.  
<sup>c</sup>CG: control group.  
<sup>d</sup>HF: heart failure.  
<sup>e</sup>PHQ: Patient Health Questionnaire.

### Feasibility Outcomes

#### Recruitment and Enrollment

Before the COVID-19 outbreak, 91 older adults with HF were approached at clinical centers (cardiac rehabilitation center, cardiac hospital unit, and senior centers), of whom 25 (27%) were found ineligible for the study. The reasons for exclusion were as follows: not owning smartphones (19/25, 76%), moderate cognitive impairment (2/25, 8%), moving to another state (pre–COVID-19; 2/25, 8%), or not fluent in English (2/25, 8%). Of the 66 eligible participants, 14 (21%) enrolled in the study.

During the COVID-19 pandemic, 50 patients with HF were approached, of whom 11 (22%) were ineligible. The reasons for exclusion were as follows: moving to a hospice (2/11, 18%), not having a confirmed diagnosis of HF (6/11, 55%), or using a wheelchair or walker to ambulate (3/11, 27%). Of the 39 participants found eligible—11 (28%) from clinical sites and 28 (72%) through web-based recruitment—24 (61%) agreed to enroll in the study: 6 (25%) from clinical sites and 18 (75%) through web-based recruitment (Figure 4). The average distance of the participants’ homes from the study team increased from 14 miles (range 2-52) during the pre–COVID-19 phase to 143 miles (range 11-465) during the COVID-19 phase.
Retention

Of the 38 participants who enrolled in the study, 6 (16%) completed the baseline survey but dropped out before installation of the devices or apps. Of the 6 participants who dropped out, 4 (67%) were lost to follow-up, and 2 dropped out because of end-stage disease unrelated to HF. We obtained the 12-week objective behavior data assessments from 84% (32/38) and 79% (30/38) of the total enrolled participants for weight monitoring and physical activity, respectively. Although 100% (32/32) of the participants who installed the apps were able to transmit data on daily weight monitoring, 6% (2/32; 1 in the CG and 1 in the IG) experienced issues with synchronizing the tracker and were unable to transmit data on physical activity. As the tracker became an outdated version during the course of the study, the device company was unable to support troubleshooting for these 2 trackers. Table 2 provides retention statistics on the survey data available to the study team at each data collection time point.

Figure 4. Diagram of participant flow. HF: heart failure; LTFU: lost to follow-up.
Table 2. Retention of participants at each time point (N=38).

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12-week weight monitoring</td>
<td>32 (84)</td>
</tr>
<tr>
<td>12-week physical activity</td>
<td>30 (79)</td>
</tr>
<tr>
<td>6-week surveys</td>
<td>30 (79)</td>
</tr>
<tr>
<td>12-week surveys</td>
<td>30 (79)</td>
</tr>
<tr>
<td>24-week surveys</td>
<td>27 (71)</td>
</tr>
</tbody>
</table>

**Engagement**

Installation and training for the devices and apps was completed with 100% (32/32) of the participants who progressed in the study. Of the 32 participants, 14 (44%) owned a smartphone with the Android platform, whereas the remaining 18 (56%) owned a smartphone with the iOS platform. Of the 15 remaining IG participants, 11 (71%) played the SCDG more than 50% of the days (range 7%-96%). Using the daily data of the 15 IG participants, the number of days each player opened the game app was strongly associated with the number of days each player engaged in weighing ($r=0.72; P=.04$) and the number of days with physical activity step data ($r=0.9; P<.001$).

**Satisfaction With the SCDG**

At the end of 12 weeks of game-playing, of the 15 IG participants, 13 (87%) completed the survey regarding their satisfaction with the SCDG (Table 3).

Table 3. Interventions group participants’ satisfaction with the sensor-controlled digital game (N=13).

<table>
<thead>
<tr>
<th>Satisfaction parameters</th>
<th>Participants, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interesting</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Easy</td>
<td>13 (100)</td>
</tr>
<tr>
<td>Enjoyable</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Satisfying to play</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Satisfaction with the avatar’s look</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Satisfaction with sound in the game</td>
<td>12 (92)</td>
</tr>
<tr>
<td>Satisfaction with graphics</td>
<td>12 (92)</td>
</tr>
<tr>
<td>Satisfaction with using the sensor devices to progress in the game</td>
<td>8 (62)</td>
</tr>
<tr>
<td>Satisfaction with the content or information</td>
<td>13 (100)</td>
</tr>
<tr>
<td>Game motivated me to weigh myself daily</td>
<td>12 (92)</td>
</tr>
<tr>
<td>Game motivated me to exercise more</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Will recommend this game to others with HFa</td>
<td>11 (85)</td>
</tr>
<tr>
<td>Prefer playing digital games over other ways to learn about managing HF</td>
<td>10 (77)</td>
</tr>
</tbody>
</table>

*aHF: heart failure.*

The themes related to playing the SCDG included its competitive nature (2/13, 15%), motivation to attain health behavior goals (5/13, 38%), ease of the interface between the sensor devices and the game (2/13, 15%), opportunity to learn about managing HF (2/13, 15%), and access to behavior data (1/13, 8%). A participant said that the duration of the game play for 12 weeks was not long enough to develop healthy habits:

```
I was disappointed that it ended so soon. It was motivational for me to continue to pay attention to my heart failure regimen until it became more of a routine. I’m at a stage in my life where I can continue to strengthen my heart by adding more to my routine. [Participant]
```

We faced challenges in syncing data from some of the trackers within the SCDG in real time because the device company discontinued support by the time the study was implemented. Therefore, a common theme among the barriers to playing the game included problems with syncing steps from the Withings Go activity tracker with the game (5/13, 38%). Other barriers included inaccuracies in tracking bicycle activity as steps and the lowest attainable step goal in the game being too high for a participant with HF:

```
Steps and physical activity at unrealistic start levels.
I started out at 280 steps, got up to 550 steps a day. [Participant]
```

Of the 13 participants who completed the survey, 2 (15%) found the game to be simplistic and said they would prefer more features. A participant stated as follows:

```
Needs more bells and whistles. More color and animations. [Participant]
```
Finally, of the 13 participants who completed the survey, 3 (23%) stated that the use of digital games to motivate HF behaviors was not appealing. Of these 3 participants, 2 (67%) preferred reading to playing games, whereas 1 (33%) stated as follows:

*If it would help people do daily weight/vital sign checks it would be worthwhile. I didn’t need motivation other than my cardiologist telling me to.*

[Participant]

### Health and Behavior Outcomes

#### Weight Monitoring

At the end of 12 weeks, in comparison with baseline self-reports of weighing behaviors, trends of increase by 40% in weighing 5 days or more in a week were observed in the IG. Trends of decrease by 6% were observed in the CG for weighing 5 days or more in a week (Table 4). An effect size of 0.53 for the SCDG intervention was obtained for the primary outcome of mean days with weight monitoring.

<table>
<thead>
<tr>
<th>Outcome and group</th>
<th>Baseline (self-report)</th>
<th>1-6 weeks (sensor data)</th>
<th>7-12 weeks (sensor data)</th>
<th>1-12 weeks (sensor data)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weighing 5 days or more in a week, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG (n=15)</td>
<td>6 (40)</td>
<td>13 (87)</td>
<td>11 (73)</td>
<td>12 (80)</td>
</tr>
<tr>
<td>CG (n=17)</td>
<td>10 (53)</td>
<td>12 (71)</td>
<td>8 (47)</td>
<td>8 (47)</td>
</tr>
<tr>
<td>All (N=32)</td>
<td>16 (50)</td>
<td>25 (78)</td>
<td>19 (5)</td>
<td>20 (63)</td>
</tr>
<tr>
<td><strong>Days with weighing (maximum: 42 in 6 weeks; 84 in 12 weeks), mean (SD); median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG (n=15)</td>
<td>N/A</td>
<td>35.8 (6.0); 37</td>
<td>32.9 (10.5); 35</td>
<td>69.7 (16.1); 76</td>
</tr>
<tr>
<td>CG (n=17)</td>
<td>N/A</td>
<td>32.2 (10.2); 34</td>
<td>28.0 (10.9); 30.4</td>
<td>60.1 (18.7); 64</td>
</tr>
<tr>
<td>All (N=32)</td>
<td>N/A</td>
<td>34 (8.4); 37</td>
<td>30 (10.9); 31</td>
<td>64 (17.8); 64</td>
</tr>
<tr>
<td><strong>Physical activity steps, mean (SD); median</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IG (n=15)</td>
<td>N/A</td>
<td>2742 (2499); 1986</td>
<td>3365 (2821); 2281</td>
<td>2887 (2821); 2133</td>
</tr>
<tr>
<td>CG (n=16)</td>
<td>N/A</td>
<td>2638 (1573); 2522</td>
<td>2444 (1757); 2244</td>
<td>2541 (1604); 2601</td>
</tr>
<tr>
<td>All (N=31)</td>
<td>N/A</td>
<td>2690 (2060); 2480</td>
<td>2737 (2596); 2262</td>
<td>2713 (2270); 2482</td>
</tr>
</tbody>
</table>

aIG: intervention group.  
bCG: control group.  
cN/A: not applicable.

Overall, across the 2 groups, the female participants had lower trends in weight monitoring than the male participants (60 vs 70 days, respectively, out of 84); however, both biological sex groups had higher trends in the IG (men, 80; women, 61) compared with the CG (men, 62; women, 57) for days with weight monitoring.

#### Physical Activity

Trends in average physical activity steps from the 6th week to the 12th week were seen with modest increases in the IG and modest decreases in the CG (Table 4).

#### Multiple Behavior Engagement

HF self-management requires simultaneous engagement in multiple behaviors over an extended duration. The IG participants showed higher correlation ($r$=0.78; $P<.001$) between engaging in weight monitoring behaviors and engagement in physical activity behaviors from the 6th week to the 12th week than the CG participants ($r$=-0.01; $P=.79$; Figure 5).
**HF Functional Status**  
At the end of 6 weeks, the IG demonstrated a clinically significant increase of 7 points on the KCCQ [50,61]. At the end of 24 weeks, both groups retained a clinically and statistically significant increase in functional status over baseline levels (Table 5).

**Quality of Life**  
Both groups demonstrated a clinically [62] and statistically significant within-group increase in quality of life at 6, 12, and 24 weeks compared with baseline on the KCCQ (Table 5).

**HF Self-management Knowledge**  
Both groups demonstrated modest but statistically significant within-group improvement in knowledge at 6, 12, and 24 weeks compared with baseline (Table 5).

**Self-reported HF Self-management Behaviors**  
The IG participants consistently demonstrated improvement in self-reported self-management behaviors compared with baseline and retained a statistically significant improvement at the end of 24 weeks. Self-reported behaviors in the CG participants changed minimally in comparison with baseline (Table 5).

**HF Self-efficacy**  
Both groups demonstrated within-group improvement in self-efficacy across the different time points, although the CG participants demonstrated higher trends of improvement (Table 5).

**Motivation for HF Self-management Behaviors**  
The changes in autonomous regulation of behaviors were nonsignificant in both the groups at 6, 12, and 24 weeks (Table 5).

**HF Hospitalization**  
Decreases in 1-month and 6-month hospitalization rates were seen in both the IG and CG, albeit tempered by the worldwide COVID-19 effect of decrease in HF hospitalizations [63,64] (Table 5).
Table 5. Survey outcomes at baseline and at 6, 12, and 24 weeks (N=38).

<table>
<thead>
<tr>
<th>Paired outcome and group</th>
<th>Baseline Value</th>
<th>From baseline to 6 weeks Value</th>
<th>From baseline to 12 weeks Value</th>
<th>From baseline to 24 weeks Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF-related functional status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A b</td>
<td>30</td>
<td>N/A</td>
<td>30</td>
</tr>
<tr>
<td>IG c, mean (SD)</td>
<td>66.6 (28.4)</td>
<td>6.9 (14.1)</td>
<td>.048 c</td>
<td>3.7 (14.1)</td>
</tr>
<tr>
<td>CG d, mean (SD)</td>
<td>69.2 (20.8) 72.9</td>
<td>1.7 (11.8)</td>
<td>.54</td>
<td>4.1 (12.6)</td>
</tr>
<tr>
<td>Quality of life</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>30</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>62.03 (24.1)</td>
<td>5.8 (13.1)</td>
<td>.07</td>
<td>8.2 (13.5)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>60.2 (20.7)</td>
<td>8.7 (15.3)</td>
<td>.02 e</td>
<td>13.0 (19.3)</td>
</tr>
<tr>
<td>HF self-management knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>30</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>24.5 (2.6)</td>
<td>1.9 (2.5)</td>
<td>.003 e</td>
<td>1.4 (1.7)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>25.0 (2.4)</td>
<td>2.1 (2.5)</td>
<td>.001 e</td>
<td>1.1 (1.5)</td>
</tr>
<tr>
<td>Self-reported HF self-management behaviors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>31</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>74.1 (21.1)</td>
<td>4.9 (20.4)</td>
<td>.31</td>
<td>9.3 (16.5)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>81.7 (15.5)</td>
<td>~2.7 (9.9)</td>
<td>.26</td>
<td>2.6 (12.4)</td>
</tr>
<tr>
<td>HF self-efficacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>31</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>69.9 (20.7)</td>
<td>2.2 (25.7)</td>
<td>.7</td>
<td>1.0 (20.5)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>62.6 (18.4)</td>
<td>6.9 (18.1)</td>
<td>.11</td>
<td>10.2 (12.3)</td>
</tr>
<tr>
<td>Motivation for HF self-management behaviors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>30</td>
<td>N/A</td>
<td>30</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>1.63 (2.2)</td>
<td>~0.1 (1.2)</td>
<td>.75</td>
<td>0.1 (2.6)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>1.64 (1.6)</td>
<td>~0.2 (2.2)</td>
<td>.8</td>
<td>~0.3 (3.0)</td>
</tr>
<tr>
<td>HF hospitalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the last month</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>31</td>
<td>N/A</td>
<td>31</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>5 (26.3)</td>
<td>1 (7)</td>
<td>N/A</td>
<td>0 (0)</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>6 (31.6)</td>
<td>1 (6)</td>
<td>N/A</td>
<td>1 (6)</td>
</tr>
<tr>
<td>In the past 6 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population, n</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>IG, mean (SD)</td>
<td>12 (63)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>CG, mean (SD)</td>
<td>9 (47)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aHF: heart failure.

bN/A: not applicable.
cIG: intervention group.
dCG: control group.
eSignificant at P<.05.

https://games.jmir.org/2021/4/e29044
Discussion

Principal Findings

We have demonstrated the feasibility of conducting a randomized controlled trial to compare an SCDG intervention and a sensors-only intervention for improving older adults’ HF self-management behaviors of daily weight monitoring and physical activity and other health outcomes. Despite the constraints imposed by the COVID-19 pandemic, we successfully attained benchmark parameters for the feasibility outcomes of recruitment, retention, intervention engagement, and satisfaction with the intervention among older adults diagnosed with HF. The advantages of recruiting participants and conducting the trial remotely enabled us to increase access to our study from 2 counties to more than 18 counties, which included rural areas. The older adult participants in our study were able to use the digital game app and sensors regardless of their education level, familiarity with digital games, or smartphone platform (ie, iOS or Android).

Regarding the behavior outcomes, the IG demonstrated better trends in the primary outcome of weight monitoring behavior, with a higher average of weight monitoring days than the CG: 46% more IG participants than CG participants attained the clinically significant level of weighing 5 days or more in a week. Similarly, the IG participants demonstrated higher engagement in the dual behaviors of weight monitoring and physical activity than the CG participants. Moreover, the IG participants demonstrated significant within-group improvement in self-reported HF self-management behaviors (weight monitoring, exercise, salt restriction, medication adherence, crisis recognition, and related follow-up), which persisted at 24 weeks. This result is promising because improving overall HF outcomes depends on engagement in multiple HF self-management behaviors over long time intervals.

Ours is one of the few studies to capture daily physical activity steps of older adults with HF over 12 weeks. Although both IG and CG participants demonstrated an increase in average daily physical steps from the 6th week to the 12th week, the increase was still lower overall than recommended for physical activity steps [65]. Given the technical issues with an outdated tracker, the SCDG may motivate higher increases in step levels with better trackers. Although typical recommended levels for cardiovascular health benefits have ranged from 10,000 [65] to 15,000 [66] steps a day, a study with 16,000 older women found that even an average of 4400 steps a day resulted in significantly lower mortality rates [67]. However, older adults with HF often suffer from frailty and fatigue; therefore, it may be harder for them to reach the optimal step levels. In a recent study that examined the walking activity of adults with HF using a step counter over a year, younger age, higher ejection fraction, and lower HF classification were found to be significantly correlated with the number of daily steps [68]. Whether modest increases in steps result in health benefits among older adults with HF remains to be seen and may signify a need for additional supportive or palliative interventions that relieve physical or mental symptom burden to ensure optimal quality of life among patients with HF [69]. Nevertheless, studies like ours demonstrate actual physical activity engagement in real-world situations and examine the longitudinal relationship between clinically significant physical activity steps and optimal HF outcomes. A recent position paper on measuring physical activity by varied activity monitors for HF in real-world conditions has provided key criteria for the selection of activity monitors based on the aims of the research and observation metrics [70]. On the basis of our participants’ experience with the SCDG, future iterations could include the ability to set realistic step goal levels for participants with HF that are lower than 3000 steps and game play duration that is longer than 12 weeks.

The within-group trends of improvement in functional status and quality of life in both the IG and CG that persisted until the end of the study demonstrate the potential of digital health interventions to improve outcomes among older adults with HF. Similarly, trends of improvement in HF self-efficacy were observed in both the IG and CG. Both groups were allowed to retain the sensor devices after 12 weeks of behavior data collection, which might have contributed to persistent improvement in perceived functional status, quality of life, and self-efficacy in comparison with baseline levels. The lack of perceptible changes in autonomous regulation of behaviors in either group could be explained by the participants’ perceiving the SCDG and the sensors-only interventions as expecting accountability of their behaviors, which may be analogous to controlled regulation of behavior. Future research can help explore the need and preferences for either kind of behavior regulation for sustaining HF self-management behaviors.

Improved HF management at the clinical level may have resulted in the baseline knowledge of HF self-management being already high in both groups. Still, both groups demonstrated trends of modest improvements in knowledge at 6 and 12 weeks that were maintained at 24 weeks.

Reduction in HF hospitalization is the desired distal outcome of improved self-management behaviors. Although both groups demonstrated reduction in hospitalization, it is unclear whether the reduction resulted from the interventions or from the worldwide reduction in cardiac-related hospitalizations due to the COVID-19 pandemic [62,64]. The reduction in hospitalizations by more than 30% in both the groups persisted until the end of the 24-week study period.

Comparison With Prior Work

Recent studies have explored the feasibility of presenting data from fitness tracker sensors or Bluetooth-enabled weighing scales to track behaviors and provide feedback to patients with HF through an mHealth app. In a single-group study with 20 individuals with HF who were followed over 6 months to track weight monitoring, physical activity, and medication adherence, 60% of the participants wore trackers and 45% used the weighing scale for more than 70% of the days [71]. In our study, among the 32 participants who installed the SCDG app, 27 (80%) wore the trackers for more than 70% of the days and used the weighing scale for more than 80% of the days at the end of 12 weeks. In addition, the IG participants in our study experienced an increase in HF functional status and similar increases in knowledge in comparison with the IG participants.
in another recent study with an mHealth app intervention for HF that also employed a Bluetooth weighing scale [72]. Although our study’s results with the SCDG app compare favorably with those of similar studies that have examined mHealth apps and sensor devices with patients with HF, given the small sample size in our feasibility trial, the results and findings from our trends analysis should be treated with caution.

Limitations
Digital literacy is often a concern when digital health interventions are implemented with older adults [73]. Our study team’s prior experience in conducting usability assessments with older adults with HF as well as in installation of, and training for, devices and apps at the participants’ homes [25] helped us to design the information and content to support the participants’ self-installation of the apps and use of the devices. Nevertheless, the study might have suffered from selection bias because participation was voluntary and participants already comfortable with using technology might have been especially motivated to participate. In addition, despite block randomization and stratification by biological sex, a higher proportion of participants in the IG were married or had a partner. Having a partner to support self-management efforts might have resulted in better trends of behavior outcomes in the IG.

Another limitation was the research team’s inability to intervene immediately in the event of missing behavioral data or unsafe weight changes throughout the 12-week game play duration. Our approach of using the Twilio app (Twilio Inc) [74] for alerting the nursing research assistant to the absence of physical activity or weight monitoring data or to sudden weight changes became unmanageable because of the high volume of text messages. In future studies, developing a dashboard that will allow easier visualization of behavioral data from all participants will provide us with the opportunity to intervene immediately in the event of absent or unsafe levels of behavioral data.

Conclusions
This pilot randomized controlled trial indicates that playing an SCDG app on smartphones is feasible and acceptable for older adults with HF for motivating daily weight monitoring and physical activity. The participants who played the SCDG demonstrated positive trends in daily weight monitoring and physical activity behaviors and reduction in HF hospitalization compared with the sensors-only group. This study justifies fully powered efficacy trials with digital gaming solutions to motivate adherence to HF self-management behaviors and improve the health outcomes of older adults with HF.

Acknowledgments
This research was supported by the National Institute of Nursing Research of the National Institutes of Health under award number R21NR018229. The content is the sole responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Editorial support with manuscript development was provided by Dr John Bellquist at the Cain Center for Nursing Research at the University of Texas at Austin School of Nursing. The authors would like to thank Cathy B Spranger, DrPH, manager at Seton Medical Center Cardiac Rehab Laboratory; Dr Clay Cauthen, cardiologist at Seton Medical Center; nurse educator Rebecca Cox; and the Continuum of Care clinicians at Seton Medical Center for helping us recruit participants to this study. The authors would also like to thank Hira Asim and Katelyn Leggio, student research volunteers and assistants at the University of Texas at Austin.

Authors’ Contributions
KR conceptualized the study together with TB, CJ, MO, MK, and ET. KR, CA, and AS recruited the participants. CA collected survey outcomes data, which was supervised by KR. AS helped with technical installation and support, which was supervised by KR. MO and GL developed the technology, which was supervised by CJ, TB, and KR. Data were curated by GL, which was supervised by KR. BV conducted the data analysis. KR wrote the manuscript, which was critically reviewed by TB, CJ, BV, and MK. All authors approved the final draft of the submitted manuscript.

Conflicts of Interest
MO is the owner of Good Life Games, Inc, a company that develops health games for hire, which provided the prototype game images for this study.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2146 KB - games_v9i4e29044_app1.pdf ]

References


35. Trialfacts homepage. Trialfacts. URL: https://trialfacts.com [accessed 2020-12-31]


38. Heart Failure Society of America homepage. Heart Failure Society of America. URL: https://hfsa.org/heart-failure-educational-modules [accessed 2020-12-23]


47. Simple randomization service. Sealed Envelope. URL: https://www.sealedenvelope.com/simple-randomiser/v1 [accessed 2021-03-09]


74. APIs for reliable SMS delivery. twilio. URL: https://www.twilio.com/sms [accessed 2021-09-24]

Abbreviations
CG: control group
CONSORT: Consolidated Standards of Reporting Trials
HF: heart failure
IG: intervention group
KCCQ: Kansas City Cardiomyopathy Questionnaire
mHealth: mobile health
SCDG: sensor-controlled digital game
Development and Validation of a Mobile Game for Culturally Sensitive Child Sexual Abuse Prevention Education in Tanzania: Mixed Methods Study

Maria Proches Malamsha¹, BSc; Elingarami Sauli², PhD, MD; Edith Talina Luhanga³, PhD

¹School of Computation and Communication Science and Engineering, Nelson Mandela African Institution of Science and Technology, Arusha, United Republic of Tanzania
²School of Life Science and Bioengineering, Nelson Mandela African Institution of Science and Technology, Arusha, United Republic of Tanzania
³Carnegie Mellon University Africa, Kigali, Rwanda

Abstract

Background: Globally, 3 out of 20 children experience sexual abuse before the age of 18 years. Educating children about sexual abuse and prevention is an evidence-based strategy that is recommended for ending child sexual abuse. Digital games are increasingly being used to influence healthy behaviors in children and could be an efficient and friendly approach to educating children about sexual abuse prevention. However, little is known on the best way to develop a culturally sensitive game that targets children in Africa—where sexual education is still taboo—that would be engaging, effective, and acceptable to parents and caretakers.

Objective: This study aimed to develop a socioculturally appropriate, mobile-based game for educating young children (<5 years) and parents and caretakers in Tanzania on sexual abuse prevention.

Methods: HappyToto children’s game was co-designed with 111 parents and caretakers (females: n=58, 52.3%; male: n=53, 47.7%) of children below 18 years of age and 24 child experts in Tanzania through surveys and focus group discussions conducted from March 2020 to April 2020. From these, we derived an overview of topics, sociocultural practices, social environment, and game interface designs that should be considered when designing child sexual abuse prevention (CSAP) education interventions. We also conducted paper prototyping and storyboarding sessions for the game’s interface, storylines, and options. To validate the application’s prototype, 32 parents (females: n=18, 56%; males: n=14, 44%) of children aged 3-5 years and 5 children (females: n=2, 40%; males: n=3, 60%) of the same age group played the game for half an hour on average. The parents undertook a pre-post intervention assessment on confidence and ability to engage in CSAP education conversations, as well as exit surveys on the usability and sociocultural acceptability of the game, while children were quizzed on the topics covered and their enjoyment of the game.

Results: Parents and caregivers showed interest in the developed game during the conducted surveys, and each parent on average navigated through all the parts of the game. The confidence level of parents in talking about CSAP increased from an average of 3.56 (neutral) before using the game to 4.9 (confident) after using the game. The ability scores, calculated based on a range of topics included in CSAP education talks with children, also increased from 5.67 (out of 10) to 8.8 (out of 10) after the game was played. Both confidence level and ability scores were statistically significant ($P<.001$). All 5 children were interested in the game and enjoyed the game-provided activities.

Conclusions: The HappyToto game can thus be an effective technology-based intervention for improving the knowledge and skills of parents and children in CSAP education.
Introduction

Background
Child sexual abuse (CSA) is increasingly becoming a serious public health issue affecting both boys and girls. It has been associated with physical, mental, and behavioral health problems that may be long lasting throughout the course of a child’s life [1]. According to global evaluations and meta-analyses, the rates of CSA are estimated to be between 7.6% and 7.9% for boys and 18.0% to 19.7% for girls [2,3]. According to studies conducted in Uganda, India, Israel, and the United States, the COVID-19 pandemic has resulted in an upsurge in CSA [4-6]. In Tanzania, almost 3 out of every 10 females and 1 out of every 7 males experience some form of sexual harassment before the age of 18 [7]. Records from Tanzania’s Ministry of Health, Community Development, Gender, Elderly and Children (MoHCDGEC) show that the numbers have fluctuated regularly but have consistently increased: from 4423 cases in 2016 to 4307 in 2017 and from 9408 cases in 2018 to 10,750 cases in 2019. Such a trend is worrying and demands a whole-society approach in solving CSA. In most cases globally, the perpetrator has a close relationship with the abused child (eg, they are a parent, friend, or dating partner) [8]. Likewise, in Tanzania, the perpetrator is usually someone familiar to the child, such as a neighbor, teacher, dating partner, or relative (uncle or cousin) [7,9].

When children are sexually abused, it affects their physical, mental, and psychological health. There have been reports of both short-term effects, such as pain, bleeding, and harm to private parts, including tearing and even discharge; and long-term effects, such as psychological abuse, emotional torture, low academic performance, becoming sexual offenders, depression, anxiety, infertility in girls, and the acquisition of chronic diseases such as HIV/AIDS [1]. In recognition of these serious effects of sexual violence and exploitation of children, the United Nations sustainable development goals include a target (Target 16.2) to “end abuse, exploitation, trafficking and all forms of violence against and torture of children.” The World Health Organization, in partnership with various partners, also developed INSPIRE, a set of 7 evidence-based strategies that can be used by communities and countries to eliminate various forms of violence against children [10]. Among the strategies recommended is educating children about sexual abuse and its prevention.

CSAP education (CSAPE) is not a new concept. School-based programs have been adapted into the education systems of many countries, and they have proven to be very effective in transmitting information to children [11,12]. In Ireland for example, children aged between 6 and 12 years old are educated on topics such as safe and unsafe touches, not keeping secrets from parents, and the danger of strangers [13]. In North America, the United Kingdom, Australia, and New Zealand, school children are taught sexual abuse recognition, appropriate and inappropriate touches, the difference between good and bad secrets, how to say no and avoid unwanted advances, how to report abuse to an adult, and how abuse is never the fault of a child [12,14]. In Tanzania, no specific programs exist in the school curriculum to teach children about sexual abuse prevention [1]. However, schools may implement their own programs if needed. Unfortunately, organizing professional visits for students specifically for CSAPE is costly in terms of time and money and may therefore be out of reach for the children enrolled in primary schools, where scarcity of resources sometimes means even essential items, such as classrooms and desks, are unavailable [11,12,15].

Apart from school-based programs, CSAPE can also be conducted through places of worship and media campaigns. In Tanzania, a the 2019-2020 report by the MoHCDGEC highlighted that churches, mosques, books, and campaigns from various ministries and government agencies, among others, had offered CSAPE programs in the 2019-2020 period to 72,832 children. Similarly, in the United States, a community preventative technique that includes home visits and technological enhancement tools (eg, mobile phones, websites, phoning, SMS text messaging) to educate parents on positive parenting has shown encouraging results [16-18].

A key limitation of school and other public-based programs is that younger children are not yet in school. Younger children who are not yet in school (usually under 6 years old) are not considered in the development of the CSAPE programs. Yet, sexual abuse among this age group is present and can be perpetuated by myths, such as the belief that HIV infection can be cured through sexual contact with a child [9]. The lesser-developed communication skills of this age group may also mean that they are unable to report incidents of abuse. Home-based CSAPE is thus important to ensure their safety. From the age of 1 to 6 years, children grow sexually, recognize their biological differences, exhibit sexual gestures, and become explorative [19]. Children between 3 and 6 years look to their parents and caretakers for answers to questions such as the difference between girls and boys. Studies have proposed that parents should create home environments where they are engaged and ready to respond to and openly discuss these kinds of topics [20-23]. Child psychologists recommend that talks on sexual abuse with this age group should focus on the same topics offered to older children, including those regarding private parts, safe and unsafe touches, the importance of not keeping secrets from parents and caregivers, requesting permission before going anywhere, and reporting situations where they did not feel safe, and that these talks be offered in an easier language [24]. For many parents, this can be a daunting task. In Tanzania, there are additional sociocultural barriers, such as the discussion of sex being generally taboo, especially that occurring between parents and caretakers and their children [25].

Digital games have increasingly been used to offer health interventions to children and their parents. Mobile learning
games are games created with the defined goal of deploying on a mobile phone to motivate and engage society, especially children, on a predetermined topic [26]. Mobile learning games provide learning anywhere and anytime, and allow children to take more risks [27] and present the opportunity to retry following failure [26]. Mobile learning games geared toward health include eBug and modified Mario Brothers games. Mobile learning games specifically targeting sexual abuse prevention include Cool and Safe, Orbit, and SAP MobAPP. With the rising mobile internet use penetration in Tanzania (29 million citizens in June 2020), mobile learning games can be a useful resource for providing CSAPE to parents and their younger children. However, none of these games have been developed with the African sociocultural context in mind. For instance, body parts like the buttocks are referred to using their proper names, which is widely unacceptable in Africa, and extended families are not featured. Their general acceptability and effectiveness may therefore be hindered.

Objectives
This study aimed to develop a CSAPE mobile learning game to target the 3 to 5-year-old group, tailored for Tanzania as a case study. The acceptance of CSAPE game is associated with cultural relevance [28] and co-designed with users of digital apps. Therefore, the game was designed with a mixed methods approach to identify parents’ and caretakers’ requirements, as well as with child experts’ recommendations for the game content and design [29]. The following research questions were thus examined: What are the parents’, caretakers’, and child experts’ topics and content related to CSAPE that should be contained in the game? What are the challenges related to parents’, caretakers’, and child experts’ sociocultural and environmental condition that should be considered in the CSAPE in Tanzania? What are the parents’, caretakers, and child experts’ requirements for the game interface, features, and interaction designs? What would the acceptability and usability of a CSAPE mobile app be to society?

Related Work
Examples of computer-aided sexual abuse prevention games for children include Cool and Safe [30], which is an effective web-based prevention training program containing film clips, stories, tasks, and games developed for French and German elementary school children. Safety for Kids 3 is another mobile animation app to teach children about safety in their environment, including preventing CSA [31]. Other educational games include Orbit, which is Australian web-based software for children aged 8 to 10 years that offers information on togetherness, listening, understanding, belief, and courage [27]. There is also SAP MobAPP, a mobile app for primary school children in Korea [28]; and the Child Sexual Abuse Prevention Education Using Hybrid Application, which includes a combination of a website and a mobile app as a behavioral self-protection tool for fifth graders in South Korea [32]. Stewards of Children is also another combination of website and mobile app used for raising awareness on the prevention, recognition, and reaction to CSA using real people’s stories [33].

In designing game-based learning, cognitive, affective, motivational, and sociocultural theoretical foundations play a significant role [26]. The cognitive foundation suggests that all the tasks and activities in the game should reflect the lessons to be learned. The motivation foundation emphasizes the need for the game to have activities that will interest and engage a player in the tasks. Moreover, the affective foundation explains how features of a game make a player feel, thus affecting their learning (eg, warm colors that may attract a child). Lastly, social and cultural foundations have a role to play in game design by connecting it to the player for easy learning. Understanding a community’s sociocultural practices can significantly contribute toward a reliable and appropriate prevention program. Certain traits exist in Tanzanian communities that should be considered when developing a prevention program, such as family structures, sex and gender taboos (modesty, gender roles), sexual norms, separation (male and female), the culture of silence, foreign influence, religious teachings, unquestionable obedience by children, the value of children, poverty, and lack of explicitness in teachings [34]. As a result, any preventive measures must consider the child’s social environment (working together with the entire environment to which a child is exposed), and the appropriateness of any such approach must include parents [35]. The design must also consider children’s engagement in the game (flow theory: balancing between boredom and anxiety) [36]. This flow includes the magic circle of game-based learning, which involves response, feedback, and challenge [26].

Methods
A preliminary survey was conducted in 3 regions of Tanzania: Dar es Salaam in Ubugo district, Arusha in Meru district, and Morogoro in Mahenge district, along with other regions with a very small percentage of participation. The regions were selected to provide the sociocultural and social setting requirements for game development. Dar es Salaam was selected because of the presence of multicultural interactions in an urban setting, Morogoro was chosen because of multicultural interactions in a rural setting, and validation was done in Arusha where there is also multicultural interactions in an urban setting. The first phase of the survey, completed in Dar es Salaam, Morogoro, and other areas, involved questionnaire and focus group discussion from March 2020 to April 2020. The second phase of the survey was completed for validation and was conducted in Arusha from March 2020 to April 2020.

Sample Size
A total of 172 people participated in the study. In the first phase, 111 parents from various districts (Ubungo, Ulanga, Meru, and others) were involved in the study survey, including 58 females (52.3%) and 53 males (47.7%). Twenty four child experts forming four focus groups from two districts (Ubungo and Ulanga) took part in discussion from March 2020 to April 2020 to design the game content, features, interface, and interactions. In the second phase, 32 parents and 5 children from Meru district participated in the validation. Participants were given information about the facts and benefits of the research before they agreed to participate.
Sampling and Recruitment

To obtain a representative sample of the population in the first phase survey, a snowball sampling technique was used [37]. Parents were selected from strategic locations: outside of their children’s schools, hospitals, market places, home visits, and online Google forms via email contact lists or WhatsApp groups for answering questionnaires. Participants were asked if they would be willing to recommend another person who fulfilled the criteria, was willing to take part, and lived nearby. Similarly to the online questionnaire, participants were requested to forward it to others who met the criteria. We recognized that participation in the study may have unintended implications, such as the likelihood of emotional distress for persons who have experienced sexual abuse. To minimize this, both the introduction and the form of consent stated that the nature of questions was focused solely on app design and not on experiences. The survey questions were reviewed and approved by Kibong’oto Infectious Diseases Hospital-Nelson Mandela African Institution of Science and Technology-Centre for Educational Development in Health, Arusha Health Research Ethics Committee. The ability to exit the survey at any time was emphasized. Child experts were selected from the local government gender desks to district social service offices. One or two hosts were identified and informed about the study, and they recommended which experts should be included in the focus group discussion. During validation in the second-phase survey, parents were sampled using the snowballing technique during home and office visits.

Inclusion Criteria

To be included in the study, participants had to be 18 years of age or older, own a smartphone, and have at least 1 child aged 1 to 18 years. Both male and female parents and caretakers were eligible. To be regarded as a child expert, the individual had to possess the necessary abilities and knowledge to protect children’s safety and well-being. These participants completed a first-stage survey for the requirement selection.

To be included in the second-phase survey, participants had to be 18 years of age or older, own a smartphone, and have at least 1 child aged 3 to 5 years. Both male and female parents and caretakers were eligible. Before participating in the survey, participants were required to sign a consent form after they were informed about the study.

Parents and Caretakers Requirements

For data collection in the first survey, which was conducted face to face and online, a paper questionnaire with 33 questions separated into 4 sections was prepared (Multimedia Appendix 1). There were different types of questions, including dichotomous-response (yes or no), multiple-choice, and open-ended questions. Section 1 collected demographics, such as age, gender, marital status, number of children, and educational background. No personal information was collected, but parents and caretakers who agreed to take part in additional research phases were asked to provide their names and contact information separately after completing the survey. Parents and caretakers were free to provide or not provide such information. Section 2 included questions aimed to determine current CSAPE practices in Tanzania, such as parents’ skills, knowledge, the child’s age at which they are comfortable for their children to begin learning CSA prevention education, and challenges and enablers of parents talking to their children about CSAPE. Game specifications, mobile games, general games that children enjoy playing, and whether or not parents allow their children to use their phones were all discussed for Tanzanian children in Section 3. This section also featured elements that encourage a child to play a specific game, as well as play mode preferences for parents or caretakers and their children. The draw-write-tell method was used in this section to allow parents and caregivers to contribute in the design of the game user interface (UI) [38].

Section 4 inquired about the social environment factors that contribute to CSA in Tanzanian families, such as family structures and how often parents examine their children for symptoms of abuse.

Child Expert Requirements

The purpose of the focus groups discussion was to try to understand the challenges for parents and caretakers in talking to their children about CSAP, topics to include in CSAPE, and how the game should be designed (Multimedia Appendix 2). In each session, the facilitator introduced the purpose, and then participants could discuss them while the facilitator made sure everyone was heard. Interviews with child care experts were then conducted. The interviews were semistructured with open-ended questions and split into 4 sections. Section 1 consisted of demographic information. Section 2 comprised questions on the existing policies and methods used to protect children, topics and sociocultural requirements for prevention education against child sexual exploitation, and the obstacles that come with it in Tanzania. Section 3 asked about the social environment that contributes to child sexual abuse. Section 4 was about game UI design and using the draw-write-tell method to explain their design ideas [38]. To promote co-design with the mobile users, we also conducted paper prototyping and storyboarding sessions for the game’s interface, storylines, and options in focus groups discussions with child experts.

Game Development Approach

Transforming the user’s requirements into a software platform is referred to as development process. In this study, the game development life cycle (GDLC) model [39] was used as the development process. The GDLC involves preproduction, production, and postproduction phases. The benefit of GDLC is that it can handle the multidisciplinary nature of game development, which includes a mix of art, sound, control systems, human factors, and artificial intelligence to form a creative concept in order to entertain and teach, as opposed to conventional software development which aims at solving problems [40,41]. Due to time and budget constraints, the scope of the app was limited to include content for 3 to 5-year-old children and parents acting as a guiding presence.

The preproduction phase involved the requirements data collection. After data analysis, game design and game prototype were used to put together a comprehensive design document detailing the game goals, storyline, fun factors, level designs, gameplay mechanics, and overall blueprint of the game [40]. The production phase consisted of the development of different...
game components including storyboard production, assets creation, development platform, game engine, programming, implementation of HappyToto, and gameplay. The evolutionary prototyping approach was used, where the first version is created after obtaining input from the users, and then subsequent versions are made with additional functionality. The tools and technologies used in development of the game included a flutter platform that provided the design for the game by combining artwork, graphics, sound, code, and engine. It is a single code-based platform app that can be used to build software for both Android and iOS devices [42]. Dart, a Google-developed object programming language, was also used to develop the game. Dart focuses on front-end creation for mobile and web applications. Additionally, Adobe Illustrator software was used to create and draw 2D vector graphic objects, while Adobe After Effect was used to develop the motion graphics and visual effects of the game. Sound was recorded by sound recorder software. The postproduction phase consisted of game beta testing and validation. The researcher and 2 parents tested HappyToto, after which development and errors were rectified.

Game Validation
Validation for this HappyToto game took place between February 2021 and March 2021. The application prototype was validated using a survey with 32 parents of children aged between 3 and 5 years, including 18 females (56%) and 14 males (44%); and with 5 children, including 2 girls (40%) and 3 boys (60%) of the same age group. A pre- and postintervention assessment was used to determine the comfort of parents in allowing their children to use HappyToto as an education toolkit for CSAPE, parents’ usability of the game, and children’s’ perception of the game (Multimedia Appendix 3). Questionnaires, interviews, and observations were used for both intervention assessments. The questionnaire was used to ask parents about their comfort in talking to their children about CSA before using the game. The parents were then allowed to engage with the game while being led by the researcher. Noninterruptive data were collected based on observation as the user interacted with the system, and all the comments made by users while using the game were noted. Parents were asked to score their confidence in talking to their children about CSA after playing the game, and usability questions were asked. Open-ended questions were also asked, such as their favorite feature, the least-liked feature, and what could be added or improved to make the game more appealing and appropriate.

Data Analysis
Questionnaires, interviews, and observations were used to collect both qualitative and quantitative data. For qualitative data, the inductive method for qualitative analysis was used. For quantitative data, descriptive statistics (frequency) were used to investigate the data while t tests were conducted using R software (The Foundation for Statistical Computing) [43] to detect any significance difference between the ranks of CSAPE description from parents before and after using the game.

Ethical Considerations
Before taking part in the study, the qualified participants were informed of their rights, free will to participate, the type of research, what data were needed, how their data would be published in aggregate, and that there would be no adverse impact whether they chose to participate or not. Participation was limited to the parents and child experts whose informed consent was signed, while for the children, this involved a trusted adult. The HappyToto game development study received ethical clearance for each stage of its development and validation from Kibong’oto Infectious Diseases Hospital-Nelson Mandela African Institution of Science and Technology-Centre for Educational Development in Health, Arusha Health Research Ethics Committee.

Results
Parent, Caretaker, and Child Expert Requirements
A total of 111 parents and caretakers (females: 58, 52.3%; male: 53, 47.7%) were distributed across the regions of Arusha (n=11, 9.9%), Dar es Salaam (n=54, 48.7%), Morogoro (n=20, 18%), and others (Dodoma, Iringa, Kagera, Kigoma, Kilimanjaro, Manyara, Mbeya, Mtwara, Mwanza, Pwani, Shinyanga, Songwe, and Zanzibar: 26, 23.4%) during phase 1 of the survey. Interviews involved 24 participants from Dar es Salaam (n=19, 79%) and Morogoro (n=5, 21%) and included 18 (75%) females and 6 males (25%) as indicated in Table 1.
### Table 1. Socioeconomic and demographic characteristics of participant parents, caretakers, and experts.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Phase 1 (requirement gathering)</th>
<th>Phase 2 interview (validation), n (%) (N=32)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Survey, n (%) (N=111)</td>
<td>Interview, n (%) (N=24)</td>
</tr>
<tr>
<td><strong>Regions reached</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dar-es Salaam</td>
<td>54 (48.7)</td>
<td>19 (79)</td>
</tr>
<tr>
<td>Morogoro</td>
<td>20 (18)</td>
<td>5 (21)</td>
</tr>
<tr>
<td>Arusha</td>
<td>11 (9.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Others&lt;sup&gt;a&lt;/sup&gt;</td>
<td>26 (23.4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Gender identity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>58 (52.3)</td>
<td>18 (75)</td>
</tr>
<tr>
<td>Men</td>
<td>53 (47.7)</td>
<td>6 (25)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20 -30</td>
<td>43 (38.7)</td>
<td>4 (17)</td>
</tr>
<tr>
<td>31-40</td>
<td>40 (36)</td>
<td>15 (63)</td>
</tr>
<tr>
<td>41 and above</td>
<td>28 (25.2)</td>
<td>6 (21)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>9 (8.1)</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>12 (10.8)</td>
<td>5 (21)</td>
</tr>
<tr>
<td>Tertiary level</td>
<td>81.1 (90)</td>
<td>16 (67)</td>
</tr>
<tr>
<td><strong>Employment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed</td>
<td>70 (63.1)</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Social officer</td>
<td>N/A</td>
<td>19 (79)</td>
</tr>
<tr>
<td>Teacher</td>
<td>N/A</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Police officer</td>
<td>N/A</td>
<td>1 (4)</td>
</tr>
<tr>
<td>House maid</td>
<td>N/A</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Self-employed</td>
<td>35 (31.5)</td>
<td>N/A</td>
</tr>
<tr>
<td>Farmer</td>
<td>18 (5.4)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Number of children</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>43 (38.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>27 (24.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>17 (15.3)</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>9 (8.1)</td>
<td>N/A</td>
</tr>
<tr>
<td>5 and above</td>
<td>15 (13.5)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<sup>a</sup>Including Dodoma, Iringa, Kagera, Kigoma, Kilimanjaro, Manyara, Mbeya, Mtwara, Mwanza, Pwani, Shinyanga, Songwe, Zanzibar.

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

### Topic Suggestion

The main topics which parents, caretakers, and child experts were interested in were dangers of receiving presents (83/111, 74.8%), modesty and body changes during different stages of growth and development (71/111, 64.4%), what is an abuse and who can be an abuser (70/111, 63.0%), dangerous people and environments and what to do if encountered (63/111, 57.0%), safe and healthy versus unsafe and unhealthy touches (60/111, 54.1%), abandoning bad traditions (43/111, 38.5%), private parts names (32/111, 28.9%), fear of God (16/111, 14.8%) and finally confidence (12/111, 11.1%). The locations that parents, caretakers, and child experts felt children spent the most time (social environment criteria) and could be vulnerable to abuse were the “home,” “school,” “on the way to/from school,” “playing outside with friends,” and “in religious institutions.” As a result, they felt the game should discuss the recommended topics in these settings (eg, receiving gifts at school, on the way home).
**Sociocultural Requirement**

Around 70 of the 111 parents, caretakers, and child experts (63.1%) feared that exposure to CSAPE would affect children’s innocence and thus potentially make them more sexually active. Other concerns highlighted by parents, caretakers, and child experts were that they did not have enough skills to teach their children (46/111, 41.5%) and that Tanzanian society did not favor openness to children (40/111, 39.3%) and therefore felt it prudent to not expose children to taboo subjects. The other fear included busy schedules (21/111, 19.3%), globalization (16/111, 14.1%), the CSA threat not being that great (13/111, 11.9%), poverty (11/111, 9.6%), and increase of existing temptations from society (10/111 8.9%).

**UI Design From Parents, Caretakers, and Child Experts**

Parents, caretakers, and child experts designed their ideas for the game in the first phase of the survey by drawing, writing, and telling [38] what they wanted to see as the game’s UI. The UI sketches and explanations are used as guidance when designing the game's UI and include layout and environment design. Parents, caretakers, and child experts also preferred the Tanzanian setting, characters, and objects surrounding the children; use of Swahili language; and the and less explicit content. Furthermore, a preference survey for operating systems was conducted, with Android (Google; 93/111, 83.8%) and iOS (Apple; 9/111, 14.4%) being the most preferred.

**HappyToto Game Design and Development**

HappyToto is the name given to the mobile game developed for 3 to 5-year-old children. The story script was composed in Swahili and English languages with consideration to the parents’, caretakers’ and child experts’ opinions from the surveys and the stipulated child prevention education from previous studies. The game consists of 3 levels: private parts, presents or gifts, and a safe environment (see Figure 1).

**Storyline and Storyboard Production**

The storyline is one developed based on today's Tanzania where Ibra and Elly are twins from a middle-class household. As they are playing outside, their mother teaches them different lessons related to private parts. On the second level, their mother teaches them about proper presents to receive in the flipping game and about not receiving presents from unfamiliar people. After remembering the lessons their mother taught them, the story ends with the mother teaching Ibra and Elly about safe places and safe people inside the house and how they behave outside their home. Following each lesson, there is an exercise to check understanding. If the exercise is passed, the level is complete. HappyToto contains 2 main cartoon-type characters Ibra (a boy) and Elly (a girl) with body shape, facial features, hair, brown skin, and clothes that reflect Tanzanian children that are 5 years old. Tanzanian children like to play and to listen to their parents. Other characters include a mother who is not physically present in the game but whose orders are narrated, a guest, and a stranger.

Game scenarios were sketched with paper and pencil for each setup in the levels, character, environment, and story flow for different scenes to depict Tanzanian settings. During the phase 1 survey, parents, caretakers, and child experts participated in the paper and pencil design of the game's UI (see Figure 2).
Figure 2. Sketch of the outdoor environment and 2D of the outdoor environment.

Assets Creation
The game assets included audio, interface design, and icons that are stored within the app. Background audio was obtained from open-source libraries, and narratives, instructions, and applause were recorded using a voice recorder application. The framework provided the user interface widgets, including icons for the next, back, and home buttons. The other user interface resources included images, video, buttons, and pop-up notifications. The environment and characters were created as 2D vector graphics in Adobe Illustrator using child-friendly colors like red, blue, and yellow. Adobe After Effect was used to create the game’s video segment, including merging 2D objects and storylines and animating them.

App Screens
HappyToto consists of 2 welcome screens, where the first contains information about the game and the next contains the levels of the game. There is also an information button at the top right for adults in case they want more resources on CSA prevention, and access requires users to solve a mathematical problem. Level 1 consists of 4 screens: private parts, safe touch, dressing, and self-care. Level 2 consists of 4 screens: presents, flipping, presents outdoors, and dragging. Level 3 consists of 7 screens: safe places outdoors, safe places in the market, safe places indoors, safe places in the house, who to ask for help, safe people, and reactions when someone makes you sad.

Gameplay
When the level 1 button is pressed, a video on private parts with a play button instructs the player to play. The video tells the tale of what private parts are, and the next screen contains a safe touch exercise. The next screen contains a dressing game in which a child must drag clothes to the 2 characters. The final screen of level 1 includes 2 clickable characters that, when pressed, provide a narrative of what activities Ibra and Elly may do on their own and whom to ask for assistance. When the next button is pressed it gives instructions to activate the second round. When a child lands on level 2, narratives about receiving presents begin, and the following screen features a flipping card game that shows examples of gifts that a child might receive. The subsequent screen narrative focuses on the dangerous places to receive presents and is followed by an image-dragging exercise in which after a picture is completed, after which a lesson on what a child should be mindful of is given. The level ends with instructions on the landing page on level 3, which is subsequently unlocked. Level 3 starts with a narrative on a safe environment outdoors, followed by a hide and seek exercise to locate safe and unsafe places. The fourth page in this level starts with a narrative on a safe environment indoors, followed by a hide-and-seek exercise to locate safe and unsafe places and people in the house. The end screen of the game describes what Ibra and his friends should do upon encountering dangerous people. The game has a flow mode storyline which follows the same children across different days, as this was the favored mode from the survey. There is also looping music played at each level to continue engaging a child.

Game Validation Results
Thirty-two parents in the Arusha Region participated in the phase 2 survey for validation. Of the participants, 44% (14/32) were men, 56% (18/32) were women, their ages ranged from 20 to 50 years, 31% (10/32) had received primary education, 31% (10/32) had received secondary education, and 38% (n=12/32) had received tertiary education (Table 2). Children who participated in the evaluation were between 3 and 5 years old, with 60% (3/5) being male and 40% (2/5) being female.
Parents were generally happy with the game; out of 32 interviewed, 81% (n=26) scored as a “satisfied parent” (awarding it a 5 out of 5), while a score of 4 out of 5 (“somewhat satisfied”) was given by 19% (n=6), as shown in Table 2. Two of the participants (T3 and T2) requested permission to download the game to their phones immediately. One hundred percent of the participants (n=32) would allow their children to play the game immediately if they were around and said they would recommend this game to other parents. On the features generally, 90% (n=29) of the parents liked game level 1, 81% (n=26) liked game level 2, and 78% (n=25) liked game level 3. The reasons for parents’ preference for all the levels were the way the lessons were well integrated into the game activities (15/32, 50%). The safe touch activity, dragging, story narration, and visuals used were all liked by parents. The storyline for teaching sensitive topics in a creative and appropriate way and the in-game activities were praised by all participants (n=32). For example, participant S8 said, “I love how the lessons incorporated in hiding and seek feature because children can relate with it,” and T11 commented, “It is very educative to children especially the private part lesson and the dragging features.”

The private parts topic was completely acceptable because “colloquial words used” and the names of private parts were not called explicitly, respectful and motherly language was used, and the game conformed to Tanzanian culture, with 100% (n=32) of the participants rating it 5 out of 5 (“appropriate”). Also, 94% (n=30) of the participants ranked the “receiving presents” subject 5 out of 5 (“appropriate”) and the “dangerous environment and people” subject was ranked 5 out of 5 (“appropriate”) by 84% (n=27) of all participants. The language (Swahili), not explicitly naming private parts, characters and climate, indoor and outdoor settings, and family structures were the basic features listed as conforming to the culture.

In general, the participants discussed some features including respectful, friendly, and nonexplicit language; age-appropriate lessons; inclusion of local games such as “kombolela” (hide and seek); and graphics of Tanzanian characters and environments as very useful for displaying the UI.

Although they were satisfied, when asked “What is your least favorite feature?” and “What features should be added/modified?” 28% (n=9) of participants indicated technical corrections such as the need for more instructions, icons, animation, and game analytics. For instance, participant T4 said, “I wish there were more animation in the game and more restrictions in the level where a child will play for some time before going to another level”; T6 said, “You can add game analytics to allow parents so see the progress of their children and what features they enjoy and play a lot.” Twenty-eight percent (n=9) of the participants recommended modification to the story by adding more scenes and adding detail on how to talk to children. Participants T10 and P3 suggested to “add more dangerous environment and dangerous people scenes” and “add scenes where the emphasis is put on a child to report to a parent how their day went,” respectively. Nineteen percent (n=6) recommended variation of a narrator sound incorporating a man’s and child’s voice with strictness in the unacceptable behavior. Participant P4 suggested, “You need to add more lessons related songs to make the game more fun.”

When asked about the possible effects of the game, 84% (n=27) of the participants rated it a 1 out of 5 (“not affected”). This was because the game was engaging but not to the point of addiction, and because it is offline, children cannot reach other sites. The remaining 16% (n=5) of participants were concerned about the effects of light from the mobile phone as well as addiction to the game.

### Pre- and Postconfidence and Ability Scores

Before and after using the app, the trust level in educating children about private parts, obtaining gifts, and unsafe environments of people was measured using a 5-point Likert scale ranging from 1 (uncomfortable) to 5 (comfortable), as shown in Table 2.

The ranked results were obtained when parents were asked objectively how they intended to educate a child, and the results were ranked by taking into account the key points in each topic.
and guidelines from MoHCDGECs National Parenting Education Manual for Families. The parents’ opinions of the game before and after using it are summarized in Table 3. Prior to interacting with the game, parents had different views on the topics. Some parents assumed that their children were still too young (3 years) and that they would not understand until they were approximately 5 years old or teenagers. Some parents spoke with their children in order to protect them because occurrences on the streets had given the impression that their children were not resistant to violence. Of the 32 parents who were asked whether they were familiar with the cartoons that their children enjoyed watching and playing, 30 parents indicated having seen popular Swahili language children’s cartoons (eg, *Akili and Me* and *Ubongo Kids*).

Table 3. Selected parents opinion before and after using the game.

<table>
<thead>
<tr>
<th>Participant characteristic</th>
<th>Exemplar quote</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attitude before playing the game</strong></td>
<td></td>
</tr>
<tr>
<td>31 to 40-year-old woman</td>
<td>“I know there are effects, but I do not want to be very open to the child unless they ask; I am a doctor, so I know what to do and the effects. I do normal disciplining and emphasize a safe environment and people.”</td>
</tr>
<tr>
<td>41 to 50-year-old man</td>
<td>“I provide child self-protection like no touch of the pupu [anus] and susu [vagina] parts; only receive presents from parents and no one else.”</td>
</tr>
<tr>
<td>31-40-year man</td>
<td>“I do not talk to them about these issues; their mother does, and they are still young to be taught about sexual abuse.”</td>
</tr>
<tr>
<td>41 to 50-year-old man</td>
<td>“They were not exposed to a dangerous environment, so there was no reason to tell them, but as their brother is approaching adolescence, he has started asking questions about his body changes.”</td>
</tr>
<tr>
<td>41 to 50-year-old female</td>
<td>“I feel like my child is too young to understand all dangerous environments. He understands some, but not all the scenarios are easy to explain to him.”</td>
</tr>
<tr>
<td>21 to 30-year-old female</td>
<td>“The neighboring child was abused, and there was a case I do not know how it ended; I have to scare my child so that she is not abused.”</td>
</tr>
<tr>
<td>31 to 40-year-old female</td>
<td>“These abuses happen in town areas. My children live in the interior village, so they are very safe”.</td>
</tr>
<tr>
<td><strong>Overall acceptability</strong></td>
<td></td>
</tr>
<tr>
<td>31 to 40-year-old female</td>
<td>“The language is very well understood, attractive environment, lessons are well understood, and it matches with the age of the child.”</td>
</tr>
<tr>
<td>31 to 40-year-old female</td>
<td>“The language used is not very explicit in naming the private parts; polite way.”</td>
</tr>
<tr>
<td>41 to 50-year-old female</td>
<td>“This is very creative way of teaching.”</td>
</tr>
<tr>
<td><strong>Addiction potential</strong></td>
<td></td>
</tr>
<tr>
<td>20 to 30-year-old female</td>
<td>“For what I see this game, a child can’t not be addicted because there are no features to continuously engage a child. It is just for lesson and may be repeat a few times.”</td>
</tr>
</tbody>
</table>

**Usability**

Most parents felt confident using the game after reading the instructions. It was suggested that children will also need instructions during the first-time interacting with the game (see Figure 3).
**Children Usability Results**

All 5 children enjoyed flipping, dragging, and the ability to move around the game. When the level page loaded and they could hear the background sound, all the children became curious. After playing the game, the 4 to 5 year olds were able to correctly answer questions such as “What do you do if a neighbor tells you to take off your clothes?” A researcher met a 5-year-old girl after a day, and she showed an interest in playing the game again. They all expressed the need for more animations and activities or actions in the game, similar to what they had seen in movies to hold their attention while playing.

**Discussion**

We have attempted to develop and validate a user-friendly children’s game app for prevention of sexual abuse. The distinguishing features for this study compared to other CSA prevention studies are the following: it is the first study in Tanzania to consider development and validation of the acceptability and usability of a game as a prevention education tool for children aged 3 to 5 years, Tanzania’s culture, society, and social environment were considered in the design of the game, the content of the game touched on topics that parents and child experts indicated would be suitable, and a mobile game was used to improve child’s learning experience without time or space limitations. Although other research has found significant results when using games [30], these studies considered other cultural contexts. CSAPE game assessment needs to be performed in Tanzania due to cultural differences and the increasing numbers of parents now using smartphones and allowing their children to play games on their phones. CSAPE games can provide another tool for creatively educating CSA preschool children and keeping them engaged. We discovered that parents were afraid of taking away their children’s innocence but also wanted them to learn about presents, body changes as they grow older, who an abuser is, reporting, and safe touches in a fun and culturally sensitive way. After the game was developed, parents and children were very interested in how the game was structured and showed gained skills. Herein, we discuss the usefulness and efficacy of the HappyToto game in CSA prevention.

All 32 parents were eager to introduce CSAPE to their 3 to 5 year old using HappyToto as a CSA prevention method because the lessons were age appropriate [19], but parents also believed that good parenting was the best way to protect children [20]. Proper parenting can be defined in 2 ways: (1) teaching children how to defend themselves (ie, teaching them to report to their parents when someone makes them feel uncomfortable) [20] and (2) limiting parenting for child safety, such as prohibiting children from going to specific locations where there is a high risk of seeing a perpetrator and imposing a curfew time to be back at home [20]. During the presurvey, each parent expressed an opinion about parenting in their personal situation. Without having seen a definition of parenting, each parent described parenting as one or both of the 2 definitions above. It is important to remember that even the notion of “proper parenting” is a transitory statement, as the definition of parenting has evolved as a result of the changing social environment and globalization. Most parents confessed to being aware of abuse but having difficulty discussing it openly with their young children [44] (ie, on the private parts of their bodies). Parents also expressed struggling to describe concepts such as “your uncle could abuse you.”

Previous research has shown that parents find it interesting how games can be used to teach prevention education and encourage their children to use them and agree that this idea should be used in the concept and design of programs and applications [45]; we therefore used this concept in this study and received positive feedback. It was surprising to see the parent, caretaker, and child expert participants’ contributions to CSA prevention tailored to a game, perhaps because they had not used such a game before. The role of a parent as a guiding presence while children play allows a parent to participate in CSA prevention [9] by providing additional clarification to children on concepts. HappyToto also fills the skills gap, where parents previously depended on skills learned as children [25], which were...
insufficient. HappyToto provides the skillset parents need to be more transparent and ready to overcome the taboo of not talking to children about issues of child sexual abuse. The confidence level of parents on talking about CSAPE with children significantly increased after using HappyToto (P<.001), as compared to before interacting with HappyToto, with the confidence level increasing from an average of 3.56 “neutral” before using the game to a 4.9 “confident” after using the game. Furthermore, when asked to describe how they would talk to their child before using the HappyToto game they scored an average of 5.67 out of 10 as compared to 8.8 out of 10 after using the game.

HappyToto offered several suggestions for talking to children about challenging topics in a nondirective manner that parents were willing to try. Due to its cultural appropriateness, the opportunity to connect to the game played an essential role in ensuring that it was acceptable to parents. Beginning with language (a woman in the game softly narrated in Swahili, Tanzania’s official and national language), a safe environment was created for children to feel comfortable with the content [46]. Although CSAPE recommends that children be explicitly taught about healthy and improper interaction by giving proper names to private parts, Tanzanian culture’s conservative nature makes it much more suitable to use colloquial words in the game [25,34]. The deliberate use of cartoons with brown skin color who are traditionally dressed provided commonly demanded values for society’s children. The inclusion of fathers in the CSA prevention criteria debunked the idea that mothers were solely responsible for their children's education and gave fathers a chance to participate [47]. Parents who liked the game the least indicated that it was due to a lack of details in the lessons they would like to see and expressed the need for additional functionality.

In the first phase survey, parents with a secondary level education had an average of 3.9 (“somewhat confident”) while parents with a primary level education had an average of 4 (“somewhat confident”), indicating that they were more interested in educating their children about CSAPE than were parents with a tertiary level education, who had an average of 2.83 (“neutral”). This may possibly be due to the community lifestyle where the children are more likely to play in streets [45] and where people with a higher education are more likely to live in an area with fenced perimeters, where children generally do not move outside of their parent’s view. It was okay, for example, for 3 to 5-year-old children who did not live inside the fences to receive a lesson in the game on how to behave on the walk to the shop because this is where sexual harassment occurs [9]. The fact that the game depicted real locations where abuse occurs was appealing to both parents and children, as it made the game relatable while still making the lessons easy to remember [28].

The storylike game with a narrative flow and basic activities were sufficiently fun for the children to play and help understand lessons while not being too addicting [46]. This feature increased parents’ eagerness to encourage their children to play, knowing that they are still going to balance the game’s use with other activities. Although the repetition function is necessary to ensure that children remember the material, this needs to be a balanced with the risk of addiction.

Parents were likely comfortable with the cartoon presentation and narration of HappyToto because the game was designed with the interests of preschoolers being considered [48]. The HappyToto design is also similar to culturally appropriate and relatable children educational cartoons in Swahili, such as Akili and Me and Ubongo Children. Suggested improvements from participants were to include variations in the voices of women, men, and children but also strictness in the voice when anything that should not be done was being discussed so that children can be properly informed [47].

Several limitations became evident in this study. First, only parents from 3 regions in Tanzania were included, and cultural differences between ethnic groups and communities may introduce diversity in the development and validation of this app. Tanzania has approximately 125 ethnic groups, and future studies could consider the inclusion of participants from other regions. Second, the study was carried out with no control group. Future studies should consider using a case-control model to eliminate potential confounders not evident in this study.

Additionally, although parents responded to the questionnaire immediately after using the game, there may be a disparity between the questionnaire response and parents’ behavior in real life when talking to their children. Moreover, measures of courtesy and recall bias were not performed in this study to determine if children could remember the materials. Further study would benefit from longitudinal evaluation of children’s behavior following use of the game. Finally, the COVID-19 pandemic has changed the world’s social ecosystem, including social restrictions and more indoor activities resulting in fewer interactions. The COVID-19 global pandemic also limited the total number and demographic of participants that could be included the study.

Parents are aware that they need to teach their children about sexual education, yet they are unaware of the most appropriate way to deliver this information. Hence, there is a need for game developers to design and validate different approaches to CSAPE based on the sociocultural differences of the target population. This study has contributed to the empirical discourse in designing an appropriate platform and media application against CSA using preventive games. HappyToto was designed in collaboration with game users, and was consequently able to incorporate the social cultural perspectives of the target community. Thus, the design and activities in the HappyToto game complement previous efforts in CSAPE in a culturally relevant way. This is important in ensuring that parents are comfortable with and adopt the use of CSAPE materials. There is a need for government and other appropriate authorities to support needed research on more attractive interactive features in teaching prevention and encouraging children to repeat the activities just enough to be able to remember the skills without becoming addicted to the game. Parents need to be informed that children as young as 3 years can be taught prevention education and can understand this education. This developed game will be further piloted in a wider population of children to assess its effectiveness. After the large-scale pilot and analysis...
of comments on the game design and content included in the game are completed, the game will be made available for free on the App Store and Google Play. Following this, we will develop a monitoring, assessment, and continuous improvement strategy, and increase awareness through schools and the media. Furthermore, we will perform technological demonstrations to private and governmental organizations that deal with children’s welfare on a local and national level to promote awareness of CSAP and educate on how to use mobile technology to prevent CSA.

Acknowledgments
We would like to express our gratitude to the African Development Bank for partially funding this research study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Questionnaire for parents and caretakers on child sexual abuse prevention education.
[PDF File (Adobe PDF File), 268 KB - games_v9i4e30350_app1.pdf ]

Multimedia Appendix 2
Questionnaire for child experts on child sexual abuse prevention education.
[PDF File (Adobe PDF File), 265 KB - games_v9i4e30350_app2.pdf ]

Multimedia Appendix 3
Questionnaire to parents and caretakers on the validation of the game as an educational tool for child sexual abuse prevention.
[DOCX File, 36 KB - games_v9i4e30350_app3.docx ]

References


31. Child abuse prevention-Kid Safety 3 android gameplay. YouTube. URL: https://www.youtube.com/watch?v=gPIS4yAuib0 [accessed 2021-05-08]


41. Jain S. Game development life cycle. LinkedIn.: LinkedIn; 2017 May 01. URL: https://www.linkedin.com/pulse/game-development-life-cycle-sumit-jain/ [accessed 2021-04-18]


Abbreviations

CSA: child sexual abuse
CSAP: child sexual abuse prevention
CSAPE: child sexual abuse prevention education
GDLC: game development life cycle
MoHCDGEC: Tanzania’s Ministry of Health, Community Development, Gender, Elderly and Children
UI: user interface.

©Maria Proches Malamsha, Elingarami Sauli, Edith Talina Luhanga. Originally published in JMIR Serious Games (https://games.jmir.org), 08.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Gamifying Sexual Education for Adolescents in a Low-Tech Setting: Quasi-Experimental Design Study

Hussein Haruna1,2*, PhD; Kingsley Okoye1*, PhD; Zamzami Zainuddin2, PhD; Xiao Hu2, PhD; Samuel Chu2*, PhD; Samira Hosseini1,3, PhD

1Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Monterrey, Mexico
2Faculty of Education, The University of Hong Kong, Hong Kong, Hong Kong
3School of Engineering and Sciences, Tecnologico de Monterrey, Monterrey, Nuevo Leon, Mexico
*these authors contributed equally

Corresponding Author:
Hussein Haruna, PhD
Writing Lab, Institute for the Future of Education
Tecnologico de Monterrey
Avenida Eugenio Garza Sada 2501 Sur
Monterrey, 64849
Mexico
Phone: 52 18134162370
Email: harunahussein@gmail.com

Abstract

Background: Sexual education has become increasingly important as unhealthy sexual practices and subsequent health risks become more prevalent during adolescence. Traditional sex education teaching methodologies are limiting for digital natives exposed to various digital technologies. Harnessing the power of technology applications attractive to the younger generation may be a useful approach for teaching sex education.

Objective: The aim of this study was to improve sexual health knowledge and understanding of the problems associated with unhealthy sexual practices and address sexual and reproductive health challenges experienced in a low-tech setting.

Methods: A participatory design approach was used to develop the digital gamified methodology. A sample of 120 secondary school students aged 11-15 were randomly assigned to either experimental or control group for each of the 3 teaching approaches: (1) gamified instruction (actual serious games [SG] in teaching); (2) gamification (GM; making nongames, such as game-like learning); and (3) traditional teaching (TT) methods.

Results: The SG and GM approaches were more effective than TT methods in teaching sexual health education. Specifically, the average scores across groups demonstrated an increase of mean scores from the pre- to posttest (25.10 [SD 5.50] versus 75.86 [SD 13.16]; t119=41.252; P<.001 [2 tailed]). Analysis of variance indicated no significant differences across groups for pretest scores (F2,117=1.048, P=.35). Significant differences across groups were evident in the posttest scores. Students in the SG and GM groups had higher average scores than the TT group (F2,117=83.98; P<.001). Students reported increased learning motivation, attitude, know-how, and participation in learning (P<.001) when using SG and GM approaches.

Conclusions: Digital health technologies (particularly teaching and learning through gamified instruction and other novel approaches) may improve sexual health education. These findings may also be applied by practitioners in health care settings and by researchers wishing to further the development of sex education.

(JMIR Serious Games 2021;9(4):e19614) doi:10.2196/19614

KEYWORDS
gamified instruction; serious gaming; gamification; educational innovation; teenage students; digital generation; e-learning; low-tech setting
Introduction

Background

Unhealthy sexual activity and its related diseases have increased globally. Nowhere is the effect of sexually transmitted diseases more apparent than in the countries of sub-Saharan Africa (SSA). In SSA countries, many adolescents are exposed early to sexual intercourse and sexual and emotional abuse [1]. Consequently, adolescents are vulnerable to unsafe sexual intercourse practices, sexual encounters with many partners, forced sexual contact, exploitive sexual activity and relationships, and influence from sexually active friends [2-5]. Exposure to these types of sexual practice has led to an increase in sexually transmitted infections (STIs), such as HIV/AIDS and Chlamydia [6-8], and other consequences. A plethora of research studies have supported the increase of sexual health literacy as a way to reduce the spate of unhealthy sexual practices and curb the current increase in sexually transmitted diseases [3,9-11]. A variety of sexual health education programs for adolescents have been implemented globally. However, the efficacy of pedagogy plays a crucial role in fostering sexual health knowledge acquisition. An effective pedagogy supports a host of academic achievement paradigms [7,8,12-15]. Although effective pedagogy has been given less emphasis in sexual health education than in core curriculum subjects [16], initiatives are being undertaken for more effective sexual health education in the digital era.

Digital health games designed to target sexual health practices have increasingly demonstrated their capabilities, appeal, and influence on educating digital native adolescents [17]. Gamified learning (serious games [SG] and gamification [GM]) platforms provide unique methods for delivering educational objectives, increasing knowledge, and reducing sex-related problems faced by adolescents [18-20]. The capacity of outreach for digital games is higher than that for traditional teaching (TT) methods [21]. Approximately 97% of adolescents normally engage in digital games, whereas 50% spend more than 1 hour per day on one kind of gaming equipment or platform or another. The Speak Up Project for Digital Learning revealed a higher preference for digital gamified learning platforms for instruction over traditional ways of learning [22]. When considering the sensitive nature of sexual health knowledge dissemination, digital games are attractive because they offer a discreet, interactive, and confidential environment for learning. This makes a difference for conservative societies [3,23].

Digital games facilitate role playing and offer challenging approaches to learning improving attitude and decision-making skills applicable to real-life scenarios. Because digital platforms offer an engaging approach for learners, they promote knowledge acquisition [8]. The novelty of this study is highlighted by the exploration into sexual health education in SSA countries, which have limited technology use. Only one study has investigated attitude changes and sexual health knowledge acquisition in a country with a similar low-tech environment [24]. Appeals for data supporting the use of technology to disseminate sexual health knowledge in low-tech settings have been documented [25]. This study employs a participatory research approach. It does this to design 2 digital health interventions (SG and GM) that assess 4 aspects (motivation, attitude, knowledge, and engagement [MAKE]) among adolescents in SSA countries.

Objective of This Study

The aims of this study are to (1) add to the limited existing knowledge of game-based technologies and (2) address the interest in using this novel kind of technology as the teaching approach in a low-tech setting in Africa. First, we hypothesized that the application of game elements and mechanics in learning would enhance the sexual health literacy of teenage students. Second, we hypothesized that the teenage students would develop an attitude toward gamified instruction (SG and GM) that was more favorable and receptive than that toward the traditional learning approaches. Henceforth, this study looks at how gamified instruction can improve the sexual health education of adolescents, address their sexual health challenges, and help them overcome those challenges, all of this in developing countries, which tend to be far less invested in digital technologies than developed countries [26].

Methods

Study Design

This study employed a quasi-experimental research design. The design guided the sexual health literacy interventions for students clustered in 3 classes. Sexual health education was made mandatory for all students to sanction the randomization technique [26,27]. The study was also in line with previous empirical studies and other publications that guide quasi-experimental research design [28-30]. It evaluated learning outcomes using pre- and posttest evaluations across the 3 teaching approaches (SG, GM, and TT). The students’ perceptions were compared to determine which instructional approach was the most effective in motivating students to learn, change attitudes, acquire knowledge, and to become engaged in the courses. Figure 1 presents the quasi-experimental research design employed in this study.
Participants
The study involved teenage students (n=120) aged 11-15 who were enrolled in a secondary school at the time of the study. The research was carried out in a school in Dar es Salaam, Tanzania. This school was selected because it had 2 computer laboratories, each able to accommodate around 40 students. They had internet connectivity, a power supply, and a standby generator. The 3 classes had around 40 students each. Everyone in a class was in the same grade. The participants were not chosen randomly. Each participant was assigned to either an experimental group or a control group based on their intact classroom settings. The researchers had no authority to form or annihilate the existing study population setting. The research team randomly designated 2 of the classes as experimental (digital game and GM) and 1 class as control (traditional).

Each group was unaware of the other groups. The learning materials were the same for all the 3 groups. The only differences were in the instructional approaches. The participants were from different schools that had taken part in the revision and participatory design process of the interventions. Finally, there were 5 topics, each covered in one 40-minute class per week, as reported in Table 1.

Table 1. Sexual health education lessons covered per week and their time length (n=120). The columns and individual cells represent records per lesson, per week, and per class.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Week</th>
<th>Duration (minutes)</th>
<th>Serious game</th>
<th>Gamification</th>
<th>Traditional teaching</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Personal hygiene and good manners</td>
<td>1</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>2. Sexual responsibility and decision making</td>
<td>2</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>3. Dealing with peer pressure during adolescence</td>
<td>3</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>4. Prevention of sexually transmitted infections, including HIV/AIDS</td>
<td>4</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>5. Dealing with harmful practices and sexual violence</td>
<td>5</td>
<td>40</td>
<td>40</td>
<td>40</td>
<td>40</td>
</tr>
</tbody>
</table>

Study Conditions
Interventions
SG and GM interventions were developed following “activity theory” [31], “design-based research” [32], and participatory design approach [15,33]. All of these emphasize the involvement of stakeholders in developing instructional interventions for addressing the intended needs of the study population. As this was the third round of intervention testing, the games were refined based on outcomes from the second round. Students from this group shared their comments for making further improvement. The participatory research design approach employed led to the refinement of the 2 gamified interventions with a view to addressing the challenges faced by the adolescents [18,25]. While the revisions were carried out, the intended users of the systems/learning platforms and other stakeholders (eg, pediatricians; sexual and reproductive health specialists; sexual health teachers from participating schools; computer and information science specialists, including a game designer who was a computer engineer; and targeted secondary school student end users) were all involved in the study. These stakeholders were invited to participate in a series of design workshops during the refining of the intervention. This study also reports the research conducted during the third iteration. Further details of the SG and GM design and development have
been published in another research [26]. The descriptions of each study condition are presented below.

**Traditional Teaching Class**

Students assigned to receive TT were treated as the control group. They were taught in a conventional classroom manner. Their teacher taught 1 day a week for 40 minutes for 5 weeks. Students were given hand-outs for further reading after each session. No digital technology was used.

**Serious Gaming Teaching Class**

Students in this group received sexual health education using an SG approach (Figures 2 and 3). A week before the classes started, SG students were oriented on the game in the school computer laboratory. Students played the “My Future Begins Today” game individually under the watchful eye of a teacher and the researchers after the classes have commenced. There were 5 topics arranged in chronological order. Each topic took about 40 minutes per week. The students were also allowed to use the game during free time.

The game has an introduction which presents the general learning objective. Each topic has a game scenario in which there are avatars representing a teacher and students interacting in a classroom. There were conversations between the teacher and the students’ avatars. After watching the scenarios, the students were asked to complete the quizzes online. There were 10 quizzes on each topic, to be completed within 90 minutes. Scores were provided for correct answers, and students would lose points for each wrong answer. The students also had an opportunity to repeat the gameplay within the 40-minute margin.

**Figure 2.** Representation of the third game platform and implementation.
Gamification Teaching Class

GM is the process of giving some of the characteristics of real games to activities that are not games. GM aims to make the learning activities more interactive. This is supposed to motivate students to learn in a way that is more effective because it is fun. GM is an emerging technique within education [3,34,35]. The concept is also a recent development in low-tech settings, especially in SSA countries. With GM, more actual learning tends to take place [36]. There are various types of learning management system platforms with built-in game mechanics [37,38]. This study used “Moodle” to organize and integrate the material we wanted to teach with game elements, such as badges, levels, leader boards, points, scores, competition, and quizzes (Figure 4). As with the SG students, the students who participated in the GM instruction received a 1-week orientation before the classes began.

Figure 3. The game structures.

Figure 4. Types of badges used in the gamification group.
There were 5 topics and for each topic 40 minutes were allocated per week for 5 consecutive weeks. Each student studied individually, but they could interact online via a discussion forum devoted to each topic. Students were asked to read and practice the lesson materials provided online. There were 10 questions under each section, some true/false and some multiple choice. There were 7 types of badges (Figure 4). One was automatically awarded upon completion of a lesson using the award rules as outlined in Table 2. Overall, the GM concept was used to make learning more fun, motivate the students to learn, support a change of attitude, and increase engagement. This study observed that learning in a competitive spirit increased the desire for continuous learning [39,40]. Students were also automatically positioned on the leaderboards and assigned levels based on the points they had gained after completing the learning activities.

Table 2. Criteria and rules to receive badges.

<table>
<thead>
<tr>
<th>Badge name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st Level Badge</td>
<td>Rewarded to students who complete the first topic and are moving to the second topic</td>
</tr>
<tr>
<td>2nd Level Badge</td>
<td>Rewarded to students who complete the second topic and are moving to the third topic</td>
</tr>
<tr>
<td>3rd Level Badge</td>
<td>Rewarded to students who complete the third topic and are moving to the fourth topic</td>
</tr>
<tr>
<td>4th Level Badge</td>
<td>Rewarded to students who complete the fourth topic and are moving to the fifth topic</td>
</tr>
<tr>
<td>5th Level Badge</td>
<td>Rewarded to students who complete the fifth topic</td>
</tr>
<tr>
<td>Exclusive Top Badge</td>
<td>Rewarded to a student with the highest score for a particular topic</td>
</tr>
<tr>
<td>Outstanding Achievement Badge</td>
<td>Rewarded to a student with the highest total of points from all topics</td>
</tr>
</tbody>
</table>

Procedure and Data Collection Methods

Sexual Health Literacy Tests

During the design of the interventions the selected research team and participating parties, especially the teachers and sexual/reproductive health specialists, were involved in developing a set of questions covering the 5 topics taught. This test was titled Adolescent Sexual Health Literacy Test (ASHLT). There were 50 questions, 10 per topic in the following format: Section A (multiple choice), Section B (true/false), and Section C (short answer). The ASHLT took up to 45 minutes to complete. Before initiating the actual learning, students were asked to do a pretest (using the ASHLT) to assess their sexual health knowledge (baseline). Within a week following the training, the same ASHLT questions were given to measure their level of understanding.

Students’ Perceptions of Teaching Approaches

This study used the MAKE framework [41], according to which a teaching method is regarded as effective if it shows the ability to motivate students, improve their attitude, increase their knowledge acquisition, and increase their engagement in learning. Several scholars have employed the method for evaluating the efficacy of the 4 components of MAKE independently [8,15,24-44]. This study employed a MAKE evaluation framework to evaluate and compare the efficacy of the 3 instructions by taking into account the 4 different perspectives (motivation, attitude, knowledge, and engagement) at once. The resultant MAKE instrument we used has 46 items, with the motivation construct containing 16 items and the other 3 constructs (attitude, knowledge, and engagement) having 10 each. We measured the students’ viewpoints through a self-rating method that had a 5-point Likert scale (5=strongly agree to 1=strongly disagree). The ratings took 10 minutes to complete and were all conducted within a week.

Focus Group Interviews

We conducted focus group interviews (FGIs) to yield more comments on the teaching methods [45]. The FGIs were conducted to corroborate and complement the quantitative data. A total of 21 students were requested to participate in the FGI, 7 students for each of the 3 learning instructions. These are realistic numbers for an FGI [46]. There were 3 focus group discussions, one for each of the instruction categories. A semistructured interview guide/protocol was adopted from the MAKE evaluation framework. Students were asked to share their views on the effectiveness and other aspects of their learning approach, and an audio record of the FGI data was made. Verbatim transcriptions were made using pseudonyms for data analysis. The participants were given equal time (1 hour) to provide their comments.

Quantitative Data Analysis

Overview

The collected data were imported from the Excel (Microsoft) file format to the IBM SPSS software tool for statistics (version 24) to perform the quantitative analytic tests. This was for data generated using ASHLT and the MAKE evaluation framework. A paired t test was conducted to compare the pre- and posttest average scores. This was done to determine whether there were changes in the learning scores after a series of sexual health literacy sessions. Besides, a one-way analysis of variance (ANOVA) was performed to analyze the numerical data collected from the pre- and posttest scores. This test compared the variations across the 3 learning approaches. In other words, we performed the pretest comparison across the 3 instructions to establish possible significant differences before the training. This would especially rule out any possible bias in the sexual health knowledge of students collected at baseline. The descriptive statistics was focused on determining the mean, median, and SD on the self-rating scale of the measurement using the MAKE evaluation framework pertaining to the students’ perceptions of the 3 instruction approaches. The

https://games.jmir.org/2021/4/e19614
self-ratings of the effectiveness of the 3 teaching methods using the MAKE evaluation framework for each component were tested using the one-sample Kolmogorov–Smirnov and Shapiro–Wilk tests for normality. Table 3 presents the results of the normality tests.

Although the results indicated that the data samples violated the assumption of normality ($P < .05$), as the scores are non-normally distributed in the Kolmogorov–Smirnov and Shapiro–Wilk tests (Table 3), a nonparametric Kruskal–Wallis test was consequently used to compare and contrast the responses across the 3 groups for each component. A significant value of $P < .05$ was used to determine the results of the statistical analysis.

Table 3. Normality test results for the MAKE instrument.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Kolmogorov–Smirnov</th>
<th></th>
<th>Shapiro–Wilk</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>$df$</td>
<td>$P$ value$^{a,b}$</td>
<td>Statistic</td>
</tr>
<tr>
<td>Motivation</td>
<td>.085</td>
<td>120</td>
<td>.03</td>
<td>.956</td>
</tr>
<tr>
<td>Attitude</td>
<td>.213</td>
<td>120</td>
<td>&lt;.001</td>
<td>.800</td>
</tr>
<tr>
<td>Knowledge</td>
<td>.174</td>
<td>120</td>
<td>&lt;.001</td>
<td>.896</td>
</tr>
<tr>
<td>Engagement</td>
<td>.103</td>
<td>120</td>
<td>.003</td>
<td>.961</td>
</tr>
</tbody>
</table>

$^a$All $P$ values are <.05, and thus significant.
$^b$Lilliefors significance correction.

**Measurement Reliability**

The validated instruments and reported questionnaires appeared to be satisfactory [41,47] following the factor analysis and reliability checks we conducted and documented for 120 samples. The sample size met the minimum of 100 or larger, or a ten-to-one ratio of observations per domain [48]. The motivation questionnaires showed a Cronbach $\alpha$ of .92. The attitude questionnaire showed a Cronbach $\alpha$ of .90, and the knowledge survey developed from the sexual health education syllabus showed a Cronbach $\alpha$ of .92. The engagement questionnaires developed from many sources with no existing reliability results indicated Cronbach $\alpha$ of .90. The results of the different scale reliability checks are presented in Table 4.

Table 4. Scale reliability for the MAKE evaluation instrument ($N=120$).

<table>
<thead>
<tr>
<th>Constructs and components</th>
<th>Number of items</th>
<th>Cronbach $\alpha$</th>
<th>Standardized $\alpha$</th>
<th>Kaiser–Meyer–Olkin</th>
<th>$P$ value$^{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td>4</td>
<td>.92</td>
<td>.88</td>
<td>&lt;.001$^a$</td>
<td></td>
</tr>
<tr>
<td>Relevance</td>
<td>4</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confidence</td>
<td>4</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Satisfaction</td>
<td>4</td>
<td>.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001$^a$</td>
</tr>
<tr>
<td>Affective</td>
<td>5</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>5</td>
<td>.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001$^a$</td>
</tr>
<tr>
<td>Importance</td>
<td>4</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness</td>
<td>3</td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Application</td>
<td>3</td>
<td>.89</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001$^a$</td>
</tr>
<tr>
<td>Emotional</td>
<td>6</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive</td>
<td>4</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$The mean difference is significant if $P$ value is <.05.

**Qualitative Data Reliability and Analysis**

The qualitative data collection instrument was developed using the MAKE evaluation. FGI transcriptions and records complemented the quantitative results. Membership checking, conformability, and validation were applied to the collected data to come up with critical comments on the sufficiency of the results for ensuring the reliability of the qualitative data. At the end of the data collection process, students were asked to review the transcripts to determine whether the transcripts
presented incorporated their comments. Thus, based on the MAKE evaluation instrument, 4 themes were developed (ie, motivation, attitude, knowledge, and engagement) to enable an ample analysis of the collected data. Then, a codebook was created using the 4 MAKE constructs. The students’ transcripts were merged with the quantitative data (based on the 4 MAKE themes).

Table 5. Descriptive characteristics of participants (N=120).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, n (%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>69 (57.5)</td>
</tr>
<tr>
<td>Female</td>
<td>51 (42.5)</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13.65 (0.99)</td>
</tr>
<tr>
<td>Female</td>
<td>13.65 (1.01)</td>
</tr>
<tr>
<td>Living group, n (%)</td>
<td></td>
</tr>
<tr>
<td>With both parents</td>
<td>89 (74.2)</td>
</tr>
<tr>
<td>With father only</td>
<td>7 (5.8)</td>
</tr>
<tr>
<td>With mother only</td>
<td>16 (13.3)</td>
</tr>
<tr>
<td>With guardian only</td>
<td>8 (6.7)</td>
</tr>
<tr>
<td>Economic group, n (%)</td>
<td></td>
</tr>
<tr>
<td>High class</td>
<td>14 (11.7)</td>
</tr>
<tr>
<td>Middle high class</td>
<td>47 (39.2)</td>
</tr>
<tr>
<td>Middle low class</td>
<td>57 (47.5)</td>
</tr>
<tr>
<td>Poor</td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>Access to a computer at school or home, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>118 (98.3)</td>
</tr>
<tr>
<td>No</td>
<td>2 (1.7)</td>
</tr>
<tr>
<td>Access to smart devices at school or home, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>119 (99.2)</td>
</tr>
<tr>
<td>No</td>
<td>1 (0.8)</td>
</tr>
<tr>
<td>Play of computer or mobile phone games, n (%)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>117 (97.5)</td>
</tr>
<tr>
<td>No</td>
<td>3 (2.5)</td>
</tr>
</tbody>
</table>

Results

Baseline Characteristics of Participants

In all, 120 teenage students participated in testing the interventions. Table 5 presents their demographic features and socioeconomic status, including their access to and use of digital technologies.

ASHLT Test Results

Main Findings

The study carried out a paired sample t test to assess the mean differences or effect of the 3 teaching methods based on the students’ average scores in the ASHLT. Statistically, there emerged a significant improvement in the knowledge acquisition, as the data demonstrated an increase in mean scores in the ASHLT from pretest mean of 25.10 (SD = 5.50) to posttest mean of 75.86 (SD 13.16; \( t_{119}=41.252, P<.001; \) 2-tailed). A one-way ANOVA was then used to compare and contrast pre- and posttest across the 3 instructions. The average of pretest scores indicate that participants were equally distributed in all the 3 teaching methods: \( F_{2,117}=1.048, P=.35. \) The average posttest scores stemming from the experimental instructions (SG and GM) also indicated an increase—as opposed to their counterparts in the control group (TT): \( F_{2,117}=83.98, P<.001. \) Figure 5 presents a comparison of the effectiveness of the 3 teaching groups.

Likewise, we conducted post hoc tests (which served as follow-up analysis) to establish the differences in the 3 pairs of teaching groups: 2 experimental and 1 control. Significant divergences emerged for both the traditional and game-based groups (\( P<.001 \)) and between the TT and GM groups (\( P<.001 \)). However, there was no significant difference between SG and GM groups (\( P=.19 \)). These results suggest that students assigned
to the experimental groups achieved a higher score after the lessons than the students in the control group.

**Figure 5.** Average comparison of three teaching groups.

Comparison of Average Scores Before the Series of Lessons

Table 6 presents the results for all the 5 topics. For example, the descriptive data we provide below demonstrate the average scores for the “personal hygiene and good manner during adolescence” topic as follows:

- TT group mean score of 5.45 (SD 1.73)
- SG group mean score of 5.36 (SD 2.15)
- GM mean score of 5.68 (SD 2.22)

As gathered in Table 6, a one-way ANOVA was conducted to compare each sexual health education topic taught across the 3 instructions (ie, TT, SG, and GM). The results indicated a nonstatistically significant difference in any of the 5 topics detected with $P>.05$. The results suggest that all the 3 teaching methods can be considered similar regarding sexual health knowledge in all 5 topics before the training.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Pretest, mean (SD)</th>
<th>TT</th>
<th>SG</th>
<th>GM</th>
<th>One-way ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RR</td>
<td>GM</td>
<td>F</td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5.45 (1.73)</td>
<td>5.36 (2.15)</td>
<td>5.68 (2.22)</td>
<td>$F_{2,117}=0.26, P=.76$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4.92 (1.59)</td>
<td>4.92 (1.45)</td>
<td>5.31 (1.44)</td>
<td>$F_{2,117}=0.89, P=.41$</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.85 (1.18)</td>
<td>4.87 (1.65)</td>
<td>5.40 (1.60)</td>
<td>$F_{2,117}=1.72, P=.18$</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.66 (1.16)</td>
<td>4.67 (1.57)</td>
<td>4.71 (1.15)</td>
<td>$F_{2,117}=0.01, P=.98$</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.76 (1.52)</td>
<td>4.68 (1.53)</td>
<td>5.01 (1.36)</td>
<td>$F_{2,117}=0.53, P=.58$</td>
<td></td>
</tr>
</tbody>
</table>

Comparison of Average Scores After the Series of Lessons

The students in each group also completed the same (ASHLT) quizzes as a posttest. The one-way ANOVAs were performed to evaluate the students’ average test scores across the 3 teaching methods. The descriptive data from the one-way ANOVAs and post hoc comparison tests are illustrated in Multimedia Appendix 1. The results indicated that the SG and GM groups had higher averages than the TT group.

The one-way ANOVAs revealed that the students received effective sexual health knowledge and that the acquisition rate increased for each topic in the 3 learning groups. Students from the SG and GM instruction groups showed a significant knowledge acquisition in topic 1, compared with students in the TT group ($F_{2,117}=19.04, P<.001$). Constant effects remained for the other 4 topics. The Tukey HSD post hoc multiple comparison tests indicated that the average scores for all the 5 topics significantly varied between the control and experimental groups ($P<.001$). The experimental groups did not differ significantly as $P$ values were over .05 (refer to Multimedia Appendix 1 for details).
Students’ Perceptions Toward Instruction Approaches

The study evaluated the students’ perceptions of the 3 instructional approaches using a self-rating scale and FGI. The details of both the quantitative and qualitative results are presented below.

Quantitative Component

The averages for the responses to each aspect of the MAKE evaluation framework were compared for the 3 groups. A Kruskal–Wallis test was performed to determine the existence or nonexistence of statistically significant difference among the 3 groups after rating the average scores from each MAKE evaluation framework. Statistically significant dissimilarities between the 3 groups’ averages were demonstrated (Multimedia Appendix 2). For instance, the Kruskal–Wallis test showed that there was a significant difference in motivation between the 3 instructions: SG mean of 4.51 (SD 0.25), GM mean of 4.40 (SD 0.38), and TT mean of 4.12 (SD 0.59); P<.001. Post hoc tests were also conducted to make pairwise comparisons. In the post hoc tests, we found that TT differed significantly from both GM (P=.04) and SG (P<.001). By contrast, GM and SG were not significantly different from each other (P=.79). Moreover, this effect remained consistent with the other aspects of the MAKE evaluation.

Qualitative Component

Comments were received from both the experimental and the control groups on the 3 instruction methods. Like the results from the quantitative data, students from the experimental groups commented favorably on the SG and GM instructions, whereas those in the control group commented unfavorably on TT. For example, for motivation, the students reported:

- the games were fun [SG-3]
- that learning was easy [SG-5]
- the learning offered a self-regulatory method that improved my confidence [GM-1]
- learning was [done] in a competitive [way], which helped me gain problem solving-skills [GM-7]

Students also pointed out that:

- the learning inspired me; hence I focused on learning [SG-2]
- I was extremely interested in the learning approach [GM-4]
- the availability of badges encouraged [me to learn] the subject [GM-2]

By contrast, students from the TT group were largely negative about their learning experience:

- there were no visuals [TT-5]
- [there was] no clarification on many issues [TT-2]
- [there were] limited, or no activities for concentration [TT-6]
- little or none of the critical thinking strategies were provided, including role play, demos, quizzes, team-work activities, [or] collaboration. [TT-3]

Regarding the attitude change, the FDIs revealed the opinion of the experimental group students about SG and GM:

- particularly useful in changing attitudes [GM-3]
- a non-embarrassing learning environment [SG-1]
- suitable and worthwhile for the delivery of sexual health education [SG-7]

I was excited about the activities, competitions, leader-boards, badges, avatars, and scenarios [GM-4]

By contrast, the control group students commented:

- I was bored listening to lectures [TT-1]
- [it was] hard to understand how the sexual health subject is important for changing my attitude [TT-7]
- [it is] unfriendly learning [TT-6]
- [it was an] uncomfortable learning environment due to the sensitivity of the topic taught [TT-5]
- it hides potential information for changing my attitude [TT-4]
- questions were not encouraged or not well clarified; hence I ended up with no clues that could change the myths [that produce] negative sexual health attitudes [TT-2]

Experimental group students reported having had positive interactions with the SG and GM interventions and a substantial improvement in their sexual health knowledge:

- I acquired the required knowledge for practicing healthy sexual behaviours through this learning approach [GM-6]
- I acquired potential sexual health knowledge that will help my making informed decisions [SG-6]
- I will apply the skills and understanding that are essential and applicable for curbing unhealthy sexual behaviours. [SG-4]
- from today onwards I will not participate in risky sexual activities as I am [now] knowledgeable and will apply the knowledge acquired to make informed decisions for better sexual health outcomes and future goals. [SG-4]

By contrast, control group students commented on their teaching method as follows:

- traditional teaching was less informative [TT-3]
- technical language was used that made it difficult to understand [TT-7]
- [there was a] lack of vivid examples [TT-1]
- the learning strategy narrowed the thinking capacity required for applying the knowledge and skills acquired [TT-4]

Finally, students in the experimental learning groups (SG and GM) reported that the SG and GM components were effective in engaging them:

- the learning activities made our minds active [GM-6]
- I was connected to the learning process [SG-4]
- I focused on the learning activities [SG-4]
the learning made me concentrate on learning all the time [GM-2]

Students in SG-2 and GM-7 reported that
...the learning provided opportunities for hands-on activities that made it easy to learn and remember.

By contrast, the TT students reported that
our learning was indirect [TT-3, TT-5, TT-2]
the learning was passive, as no hands-on activities were provided [TT-1, TT-4]
there was little or poor interaction [TT-7]
I lost focus during learning [TT-6]

Discussion

Findings and Interpretation

The study showed that the game elements embedded in SG and GM instruction catalyzed motivation and engagement during learning and that this contributed to attitude change, knowledge acquisition, and ultimately better learning performance.

The study results demonstrated that sexual health education taught using SG and GM approaches works better than TT methods. The SG and GM approaches resulted in higher test scores for knowledge acquisition than the TT control group. This finding conforms with previous research which found gamified learning systems to have a significant impact on sexual health education [25]. In our study, most students acknowledged several factors in their improved learning: The first factor is motivation (to learn), which consists of elements of attention, relevance, confidence, and satisfaction [49]. Indeed, motivation is a significant component for accomplishing or failing a task [50]. Students reported that their interest was caught and improved with the game elements (scenarios, quizzes, competition, challenges, scores) provided during the learning process. The game elements motivated the students to learn [45]. Furthermore, the game elements made the learning more interactive and fun, which increased the students’ motivation to learn [50].

This study found that gamified learning contents were experienced as “relevant.” For instance, the students realized that there was a common connection between what they were learning and real life. These results were consistent with an earlier study [19], which had indicated that gamifying sexual health learning approaches was promising for adolescents because the role plays and scenarios reflect the actual lifestyle of the current generation. The students felt confident while going through the self-regulated learning material provided through the gamified learning (which stimulated and sustained their learning). Perhaps this means that they would succeed in learning the subject matter to a great extent. As this paper demonstrates, such confidence enabled them to succeed and derive self-esteem from the knowledge they acquired and apply it in real life [49]. Although Keller [49] did not examine the effectiveness of the mediating role of increased knowledge in sexual health literacy, this work showed that sexual health education interventions through gamified learning are effective for the development of self-efficacy. As a result, they encourage healthy sexual practices including the digitally savvy adolescents [24]. Students reported satisfaction with their learning experience. The quality of the gamifying content gave them an experience of fun learning (thus, accomplishing learning goals). The awards and scoring mechanics also inspired them to learn with persistence and intensity [51].

The second factor accounting for the effectiveness of the gamified learning was that it changed the attitude of the students. As a similar study [7] reported, gamified learning induces positive changes in the sexual health attitude of adolescents. Essentially, the My Future Begins Today gamified learning incorporated in its design most of the known relevant features that have proved effective in transforming adolescents’ negative sexual health attitudes to positive ones and, as a consequence, curbing risky sexual behavior [10]. It considered specific settings and co-opted various stakeholders, including the targeted users (high-school students and their teachers) in the design.

During the gamified design and development, students provided input on the type of avatars and scenarios they found appealing. Their opinions were based on their different cultural settings, their level of digital literacy, their use of state-of-the-art technology, among others. This study was grounded on the social–cultural theory known as activity theory that encourages participation of different stakeholders in the development of instructional interventions [52-54]. Members of the community participated in the design process by contributing to the design of the knowledge-acquisition components useful in addressing the problems related to acquisition of sexual health knowledge among adolescents in the studied region. The resultant gamified learning elements were found to be relevant in changing the students’ attitude toward problems such as negative peer pressure, teen pregnancy, STIs (including HIV/AIDS), and sexual violence. Although this research found no participatory design being applied within the TT environment, the gamified learning instructions invited the targeted users to participate in the design. This allowed us to apprehend the participants’ relevant ideas and needs and in turn effectively deliver the sexual health information required to induce a change in attitudes of the participants [15].

The third factor accounting for the efficacy of gamified learning was that it improved the knowledge acquisition among students in the experimental groups. The students indicated that gamified learning helped them to acquire sexual health information and skills that could purportedly help them engage in healthy sexual practices. Notably, the students reported that the sexual health knowledge delivered through gamified instructions were highly effective for their current and future lives and that they now felt knowledgeable and able to resist detrimental sexual health risks or factors. According to Chu et al [15], gamified learning is effective because knowledge is acquired in a safe, nonrisk (simulated) environment. Gamified learning offers students the opportunity to experiment in their learning, play, apply decision-making skills, and test scenarios without negative consequences. Evidence from follow-up studies shows that the effect of sexual health knowledge acquired through gamified learning is compelling and persistent, but no effectiveness has been demonstrated in terms of delay in sexual initiation [7], although the determination of outcomes was based on
self-reporting. An iterative design study with the objective to
describe a methodology for developing an SG intervention for
improving sexual health education among youth in Boston [8]
demonstrated that nonidentified study participants (ie, students
and street youths) in underserved communities would have
acquired more knowledge on chlamydia because they enjoyed
the gameplay and actively participated in acquiring the
information. This may explain why the experimental groups in
our study showed better results than their counterparts in the
control group.

The fourth and final factor that boosted students’ learning was
engagement. The evaluation of the effectiveness of each of the
instructional approaches was based on the students’ engagement.
Studies have documented the effectiveness of gamified learning
in engaging students during learning [8,42,43], with some
studies related to the ability of gamified instructions to entertain
and reduce stress when learning [43,50]. The My Future Begins
Today (GM and SG) platform was designed to trigger students’
engagement by having learning tasks performed in a
problem-solving way and by having students participate in
skills-based challenges that required critical thinking. The
gamified learning platforms are useful for the students,
especially when it comes to (1) the skills that are needed to
thrive or (2) use the latest technologies of this century. The
presence of game elements (badges, score, leaderboards, levels,
immediate feedback, time pressure, and repetition) may explain
why gamified learning increased the engagement of students in
the experimental group, which, in turn, bolstered their learning.
Besides, game elements were positively commented on by the
students who saw increased engagement with the PR:EPARE
game [42]. Although Jiang et al [8] did not find a significant
correlation between the participants’ game engagement and
learning, this study shows that the 2 concepts are useful for
learning purposes.

Limitations
This study used a participatory design approach. Such an
approach is vital in designing instructions that address the needs
of the users [15,43] in their social–cultural context. It was
informed by design-based research (from a technology
perspective) [32] and grounded in sociocultural learning theory
[53,54]. As good as the foundations are, we must acknowledge
some limitations in our efforts to put them into practice. This
study evaluated knowledge acquired by the students and the
effectiveness of that knowledge in changing their attitudes
toward sexual activities. However, it is would be difficult to
determine how much of and for how long the change took place
after the study. Hence, would need to know how many students
dropped out of school due to pregnancy, were infected with
STIs, encountered sexual violence, or were peer pressured into
harmful sexual practices. A follow-up study could be conducted
when the students are about to finish their ordinary-level studies.

Conclusions
Educational gamified learning (GM and SG) has the potential
to significantly increase the sexual health literacy of adolescents.
The digital health gamified interventions designed in this study
provide a user-friendly learning environment. The designs were
influenced by a theory-driven assessment of learning. This
assessment is supported by testing the learning of the users.
This study treats serious digital health gamified instructions as
a brain activator: it keeps students active during the learning
process. The students’ participation in the learning process is
catalyzed by the motivation and engagement that are enabled
by the game elements and mechanics. The My Future Begins
Today (the digital health gamified learning interventions using
the SG and GM) design increases knowledge acquisition and
attitude change. Students reported the learning to be more
interactive through participating in the gamified learning
activities.

SG and GM methods were found to be effective and efficient
in increasing motivation, improving attitudes, increasing
knowledge acquisition, and encouraging engagement in the
learning process. Future empirical studies may verify the
efficacy of the My Future Begins Today learning platforms in
improving sexual health literacy acquisition in other countries,
especially in SSA, where the TT method is widely practiced
and tends to limit the learning with gamified digital technologies
and process [19,25]. This paper also addresses more than ever
the call from a previous study [25] to evaluate the effect of
gamifying sexual health education when different key
stakeholders are involved in the design process in low-tech
settings. This is due to the fact that in developing countries
information and communication technologies and digital literacy
are limited.

Conflicts of Interest
Non declared.

Multimedia Appendix 1
Comparison of average scores, one-way ANOVA, and post hoc comparison tests after the lessons.
[DOCX File, 17 KB - games_v9i4e19614_app1.docx ]

Multimedia Appendix 2
Evaluation of the efficacy of the instructional approaches.
[DOCX File, 16 KB - games_v9i4e19614_app2.docx ]

References
https://games.jmir.org/2021/4/e19614

JMIR Serious Games 2021 | vol. 9 | iss. 4 | e19614 | p.68
(page number not for citation purposes)


Abbreviations

ASHLT: Adolescent Sexual Health Literacy Test
AVOA: analysis of variance
FGI: focus group interview
SG: serious game
GM: gamification
MAKE: motivation, attitude, knowledge, and engagement
SSA: sub-Saharan Africa
STI: sexually transmitted infections
TT: traditional teaching

©Hussein Haruna, Kingsley Okoye, Zamzami Zainuddin, Xiao Hu, Samuel Chu, Samira Hosseini. Originally published in JMIR Serious Games (https://games.jmir.org), 12.10.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Evidence of Construct Validity of Computer-Based Tests for Clinical Reasoning: Instrument Validation Study

Tianming Zuo, MD, PhD; Baozhi Sun, MD, PhD; Xu Guan, MD; Bin Zheng, MD, PhD; Bo Qu, MD, PhD

Institute for International Health Professions Education and Research, China Medical University, Shenyang, China

Education Center for Clinical Skills Practice, China Medical University, Shenyang, China

Surgical Simulation Research Lab, Department of Surgery, University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Bo Qu, MD, PhD
Institute for International Health Professions Education and Research
China Medical University
No. 77 Puhe Road
Shenyang, 110122
China
Phone: 86 189 0091 0198
Email: qubo6666@163.com

Abstract

Background: Clinical reasoning (CR) is a fundamental skill for all medical students. In our medical education system, however, there are shortcomings in the conventional methods of teaching CR. New technology is needed to enhance our CR teaching, especially as we are facing an influx of new health trainees. China Medical University (CMU), in response to this need, has developed a computer-based CR training system (CMU-CBCRT).

Objective: We aimed to find evidence of construct validity of the CMU-CBCRT.

Methods: We recruited 385 students from fifth year undergraduates to postgraduate year (PGY) 3 to complete the test on CMU-CBCRT. The known-groups technique was used to evaluate the construct validity of the CBCRT by comparing the test scores among 4 training levels (fifth year MD, PGY-1, PGY-2, and PGY-3).

Results: We found that test scores increased with years of training. Significant differences were found in the test scores on information collection, diagnosis, and treatment and total scores among different training years of participants. However, significant results were not found for treatment errors.

Conclusions: We provided evidence of construct validity of the CMU-CBCRT, which could determine the CR skills of medical students at varying early stage in their careers.

(JMIR Serious Games 2021;9(4):e17670) doi:10.2196/17670

KEYWORDS
medical education; assessment; computer-based test; clinical reasoning; validity

Introduction

Each year, several hundred thousand students enter medical school, all of whom need to equip themselves with the necessary health care skills and knowledge [1]. Since 2014, the vast majority of Chinese medical students attend a 5-year program after high school to earn a bachelor’s degree. Then, they work in a 1-year clinical internship before taking the nation’s standardized medical licensure exams. If successful, they may register as medical practitioners. Postgraduate training in medical specialties is standardized to 3-year programs with the final credential called Master of Medicine; this is now required of all clinical practitioners. In addition to learning a broad range of medical knowledge and practicing dexterity in hands, practitioners need to learn how to collect information from patients, process this information, and make accurate diagnostic decisions, similar to the expectations from a senior physician [2,3]. Clinical reasoning (CR) is a fundamental skill that separates medical personnel from other professionals. William Osler, a legendary pioneer medical educator, emphasized proper physical examination and diagnostic reasoning while maintaining the intimate physician-patient relationship. His teachings have resonated with generations of physicians [4]. Strictly speaking, CR refers to the procedure of collecting and...
integrating patient information from various sources to arrive at a diagnosis and management plan; it is usually case specific [5]. Every medical teaching institute makes a great effort to understand the nature of CR and improve strategies for teaching CR skills to health trainees [6]. However, the conventional methods that are used in our education system today are not optimal [7-11].

Traditionally, CR is taught in the classroom (didactic lecture) and by the patient’s side (clinical clerkship) [12-15]. A recent focus of integrating problem-based learning (PBL) has significantly improved the quality of CR education [6,12,16-18]. However, PBL relies heavily on the involvement and commitment of faculty instructors, which may not always be feasible [16,19]. Fidelity of case is also a problem compared to patient-side education [12]. Acquiring patient information by reading PBL cases from charts is quite a different experience than taking information directly from patients. Although instructors are making PBL cases in collaboration with clinicians, students still report a lack of case variety [17,20]. Creating sufficient clinical cases with clinical fidelity for CR training is a difficult task. Due to the above reasons, new technology is needed to improve our CR teaching.

Table 1. Types of computer-based clinical reasoning simulations and comparison.

<table>
<thead>
<tr>
<th>Media</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text based</td>
<td>Relatively easy and rapid to develop; less expensive</td>
<td>Low level of fidelity</td>
</tr>
<tr>
<td>Graphic and animation based</td>
<td>Presents rich clinical evidence; moderate cost with enhanced fidelity</td>
<td>Replicates only part of clinical settings; low level of interactivity</td>
</tr>
<tr>
<td>Virtual reality</td>
<td>Combines highly sophisticated, life-like models with computer animations; can provide interactivity and feedback</td>
<td>Challenge to developers; often expensive</td>
</tr>
</tbody>
</table>

Sponsored by the National Medical Examination Center of China, China Medical University (CMU) started to developed computer-based CR training system in 2001. Educators and researchers at the Institute for International Health Professions Education and Research of CMU began to work with clinicians to develop cases for training CR skills and established the computer-based CR testing (CBCRT) system. Since 2002, CBCRT has been used as one part in the final comprehensive examinations of CMU to test the clinical skills of undergraduate students.

The CBCRT is composed of 5 interactive modules that allow students to interact with simulations to complete tasks: (1) history taking and physical examination, (2) writing orders and obtaining lab and medical imaging results, (3) reviewing obtained results, (4) working out diagnosis and differential diagnosis, and (5) observing the patient’s condition change at different phases and changing locations for managing different therapies. The main features of the CMU-CBCRT virtual patient are displayed in Figure 1.

To briefly summarize, the CBCRT provided clinical features of patients including history and physical and laboratory findings and then requires students to make a diagnosis as well as a treatment plan for the simulated patients. The CBCRT has also been welcomed by the examinees based on their positive feedback toward the system. Of 300 students surveyed using the questionnaire, 99.4% enjoyed participating in the CBCRT examination; 95.9% believed that the system accurately represents the real clinical environment; 72.5% agreed that the CBCRT is a better tool for teaching their clinical abilities. We can thus believe that the face validity of the CBCRT is satisfactory.

However, face validity is the weakest form of validity evidence. It can be only used at the primary stage of designing an assessment method [25]. We need to look into the structure of the CBCRT in detail to find more evidence of its validity, especially since there is a paucity of validity evidence for computer-based CR training [22]. In China, this study is the first of its type.

This study investigates construct validity of the CMU-CBCRT in medical trainees over 5 medical school. We hypothesize that the CMU-CBCRT will be able to determine CR level among different years of health trainees; specifically, senior trainees will achieve higher CBCRT scores than juniors.

In contrast to a paper- or lecture-based curriculum, computer-based CR training allows trainees to interactively take information from patients in a step-by-step process. There is also the possibility of accumulating a large volume of cases through international collaboration.

Currently, computer-based CR training can have different interfaces such as text, graphics, and animation [21]. The text-based CR training system is most widely used [22]. It is easy to create from clinical cases and deliver in the format of multiple choice questions or direct interface [23]. While medical images (including x-ray films, electrocardiograms, photos of lesions, etc) are required to give students more clinical information, graphic interface is also necessary. In several graphical models, illustrations of patients (in drawing or 2D pictures) can be used to create interactive experience for students when they collect information from patients [24]. Some computer-based CR training includes 3D animation or virtual reality technology to simulate the clinical scenario with high fidelity. However, the cost of creating 3D animation and virtual reality scenarios is much higher than the other computer-based CR models. It is difficult to create virtual patients without a team of technicians and instructional designers (Table 1).
Methods

**Ethical Statement**

Methods used for the project were reviewed and approved by the ethical review boards of the CMU (ERB 2016-027) and the 5 medical schools. Informed consent was obtained from each participant before they started the test with the CMU-CBCRT.

**Testing Sites**

From November 24 to December 8, 2016, we implemented the CMU-CBCRT system in 5 collaborative medical schools: China Medical University, Fudan University School of Medicine, Sun Yat-sen University School of Medicine, Xuzhou Medical University, and Binzhou Medical College.

**Students**

In China, medical students start their clerkships on the fifth year of medical training. The clinical training will continue for their 3 postgraduate years (PGYs). PGY 1 to PGY 3 is similar to the residency in North American medical school. We recruited students from fifth year undergraduates to PGY 3. The actual number of participants from each of 5 medical schools and their training years are shown as [Table 2](#).

![Figure 1. Screenshot displaying the main features of the China Medical University–computer-based clinical reasoning testing.](image)

<table>
<thead>
<tr>
<th>Schools</th>
<th>Fifth year medical student</th>
<th>PGY-1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PGY-2</th>
<th>PGY-3</th>
<th>Subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>China Medical University</td>
<td>40</td>
<td>18</td>
<td>7</td>
<td>2</td>
<td>67</td>
</tr>
<tr>
<td>Fudan University School of Medicine</td>
<td>17</td>
<td>41</td>
<td>16</td>
<td>18</td>
<td>92</td>
</tr>
<tr>
<td>Sun Yat-sen University School of Medicine</td>
<td>12</td>
<td>28</td>
<td>20</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>Xuzhou Medical University</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>80</td>
</tr>
<tr>
<td>Binzhou Medical College</td>
<td>20</td>
<td>19</td>
<td>19</td>
<td>16</td>
<td>74</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>125</td>
<td>82</td>
<td>69</td>
<td>385</td>
</tr>
</tbody>
</table>

<sup>a</sup>PGY: postgraduate year.
Measures
Before testing, each student was asked to watch a 5-minute presentation and get familiar with the testing interface. Demographics and level of medical training were surveyed and recorded. The computer recorded participants’ typing and computer activity, including the typing and performance times. The interaction between a learner and how data are captured is displayed in Figure 2. Once completing the testing on CMU-CBCRT, the system calculated and recorded their total score by comparing the participants’ transaction list with the scoring scheme defined by the case developers committee (Multimedia Appendix 1). Subscores on these 4 different areas: information collection, diagnosis, treatment, and treatment error are computed and recorded as well (Multimedia Appendix 2).

Figure 2. Outline of interaction flow through China Medical University–computer-based clinical reasoning testing.

Statistical Model
The known-groups technique was used to evaluate the construct validity of the CBCRT by comparing the scores among the fifth year MD, PGY-1, PGY-2, and PGY-3 participants. Testing scores, including total and subtotal, were compared over the 4 training groups using a 1-way analysis of variance (ANOVA). Results were reported as mean and standard deviation. $P \leq .50$ was considered a significant difference among testing groups.

Results

Total Score
ANOVA revealed a group difference in total score among training levels ($P < .001$). As shown in Table 3 and Figure 3, the score of the fifth year MD students (59.01 [SD 16.68]) was significantly lower than the PGY-2 (68.68 [SD 11.76]) and PGY-3 (68.06 [SD 12.67]) students; the total score of PGY-1 students was also significantly lower than the PGY-2 and PGY-3 students.
Table 3. Students from the 5 medical schools and their training years.

<table>
<thead>
<tr>
<th>Scores</th>
<th>Fifth year medical student, mean (SD)</th>
<th>PGY-1\textsuperscript{a}, mean (SD)</th>
<th>PGY-2, mean (SD)</th>
<th>PGY-3, mean (SD)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information collection</td>
<td>43.42 (12.63)</td>
<td>46.70 (11.48)</td>
<td>49.73 (9.12)</td>
<td>51.38 (9.08)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>10.90 (4.74)</td>
<td>11.24 (4.97)</td>
<td>12.76 (3.90)</td>
<td>11.25 (4.22)</td>
<td>.034</td>
</tr>
<tr>
<td>Treatment</td>
<td>4.79 (3.81)</td>
<td>4.61 (3.36)</td>
<td>6.19 (3.73)</td>
<td>5.45 (3.72)</td>
<td>.013</td>
</tr>
<tr>
<td>Treatment error</td>
<td>–0.06 (0.23)</td>
<td>–0.04 (0.20)</td>
<td>0.00 (0.00)</td>
<td>–0.01 (0.12)</td>
<td>.13</td>
</tr>
<tr>
<td>Total</td>
<td>59.01 (16.68)</td>
<td>62.50 (14.45)</td>
<td>68.68 (11.76)</td>
<td>68.06 (12.67)</td>
<td>.001</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PGY: postgraduate year.

**Figure 3.** Total score of students over training years.

ANOVA revealed group differences by training level between information collection (P<.001), diagnosis (P=.03), and treatment (P=.01) scores, but not on treatment error (P=.13) score. As shown in **Figure 4**, the information collection scores of the fifth year MD students (43.42 [SD 12.63]) were significantly lower than the PGY-1 (46.70 [SD 11.48]), PGY-2...
(49.73 [SD 9.12]), and PGY-3 (51.38 [SD 9.08]) students; information collection scores of PGY-1 students were also significantly lower than the PGY-3 students. As shown in Figure 5, the diagnosis scores of the fifth year MD (10.90 [SD 4.74]), PGY-1 (11.24 [SD 4.97]), and PGY-3 (11.25 [SD 4.22]) students were significantly lower than the PGY-2 (12.76 [SD 3.90]) students. As shown in Figure 6, treatment scores of the fifth year MD (4.79 [SD 3.81]) and PGY-1 (4.61 [SD 3.36]) students were significantly lower than the PGY-2 (6.19 [SD 3.73]) students.

Figure 4. Subscore for information collection of students over training years.
Figure 5. Subscore for diagnosis of students over training years.
Discussion

Principal Findings

Before applying an assessment tool for use with medical students, we must obtain evidence for the instrument’s reliability and validity [26-28]. Providing evidence of the validity of CBCRT will help the test management organization understand the effectiveness of the test from a broad and comprehensive perspective, clarify the aspects that the CBCRT can and cannot measure, and hence, allow for its continuous improvement. This is the goal of our current study. Our hypothesis was supported by the results obtained; specifically, senior students displayed higher testing scores than junior students (Table 3, Figure 1). In other words, the CMU-CBCRT is able to determine CR skills over different levels of medical education, especially in the early stage of the students’ medical careers.

Looking specifically into the 4 categories of skills that we tested, we found that the most significant differences were revealed in the information collection, diagnosis, and treatment scores among junior and senior medical students. This was as predicted. With years of training, their experience and ability to clinically reason are improving, and as a result, they performed better on the information collection, diagnosis, and treatment, as well as the total CBCRT score. This further suggests that the CMU-CBCRT can determine the CR skills of students at varying levels.

We also carefully studied and analyzed why there were no significant differences in treatment error scores among the 4 training groups. For a simulated case of myocardial infarction, we can observe from the test result the challenge faced by participants who have never experienced this form of examination before. When the passing score was set at 60%,
the average score in this case (59.01 [SD 16.68]) did not pass. The choice of wrong treatment is a negative item in the scoring system, so the item writing expert is very cautious in formulating the scoring standard. Only behavior that caused extreme consequences resulted in points being deducted, and the weight was set at a very low level (ie, –1%). In this test, we observed that treatment error behavior happened more with junior students than senior students, although without statistical significance.

In the absence of an available gold standard for measuring CR, evidence for construct validity is sought after in this area of research. This is an ongoing process, in which the skill measured by the assessment tool is linked to some other attribute by a hypothesis or construct. With the development of validity theory, the validity concept has a new connotation and forms a method based on multilevel evidence [22]. Validity is no longer an attribute of the measurement tool itself but rather the extent to which the evidence collected supports the interpretation, inference, and decision making of the test score [27,29].

With the positive evidence presented, we should still be aware that validity verification is a dynamic process [27] and no education instrument is 100% effective [27]. Even if the evidence indicates that the validity of a course test is significant, the validity study must continue along with the development of the CBCRT system. There are still many problems to be solved, such as the setting of the evidence framework for the specific test validity, determination of the validity criteria, feasibility of the evidence collection method, and quantification of evidence data. This will require in-depth discussion by future researchers. We aim for constant examination of these issues in the process of developing a reliable and valid CR training model. In the future, we would include more simulated cases with a wide range of case difficulties and distribute CMU-CBCRT to more students to increase sample size. We would then carefully collect data on student performance and feedback. We also plan to add graphics and animation to enhance the interface design.

Limitations

However, there were some limitations in our study to its generalizability. First, the respondents of the research were from only 5 medical institutions in China. Second, the findings of our study were limited by the representativeness and scale of the study population.

Conclusions

We provided evidence of construct validity of the CMU-CBCRT. It is able to determine CR skills over different levels of medical education, especially in the early stage of the students’ medical careers.

Acknowledgments

Authors thank the National Medical Examination Center of China, which sponsored this research through the funded project Sustainable Development of Computer-based Case Simulations Examination System. We thank educators and collaborators from China Medical University, Fudan University School of Medicine, Sun Yat-sen University School of Medicine, Xuzhou Medical University, and Binzhou Medical College who provided of testing places and expertise that greatly assisted the research. Special thanks to Ms Kritika Taparia for proofreading the manuscript and provide valuable comments.

Authors’ Contributions

TZ, XG, and BQ collected the data. TZ, BZ, and BS analyzed the data and wrote the first draft of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Scoring sheet and scoring rubrics.
[XLSX File (Microsoft Excel File), 11 KB - games_v9i4e17670_app1.xlsx ]

Multimedia Appendix 2

China Medical University–computer-based clinical reasoning testing data.
[XLSX File (Microsoft Excel File), 113 KB - games_v9i4e17670_app2.xlsx ]

References


Effects of Active Video Games on Health-Related Physical Fitness and Motor Competence in Children and Adolescents With Overweight or Obesity: Systematic Review and Meta-Analysis

Cristina Comeras-Chueca1,2,3, MSc‡; Jorge Marin-Puyalto2,3,4,5, PhD; Angel Matute-Llorente2,3,4,5,6, PhD; German Vicente-Rodriguez2,3,4,5,6, PhD; Jose Antonio Casajus1,2,3,5,6, PhD; Alex Gonzalez-Aguero2,3,4,5,6, PhD

1Department of Physiatry and Nursing, Faculty of Health Science, University of Zaragoza, Zaragoza, Spain
2GENUD Research Group (Growth, Exercise, NUtrition and Development), University of Zaragoza, Zaragoza, Spain
3EXERNET, Red de Investigación en Ejercicio Físico y Salud para Poblaciones Especiales, Zaragoza, Spain
4Department of Physiatry and Nursing, Faculty of Health and Sport Science, University of Zaragoza, Huesca, Spain
5Instituto Agroalimentario de Aragón -IA2-, CITA, Universidad de Zaragoza, Zaragoza, Spain
6Centro de Investigación Biomédica en Red de Fisiopatología de la Obesidad y Nutrición (CIBEROObn), Madrid, Spain
‡GENUD Research Group

Corresponding Author:
Alex Gonzalez-Aguero, PhD
Department of Physiatry and Nursing
Faculty of Health and Sport Science
University of Zaragoza
Despacho nº 8
Pabellón polideportivo río Isuela, Ronda de la misericordia 5
Huesca, 22001
Spain
Phone: 34 876553755
Email: alexgonz@unizar.es

Abstract

Background: Childhood obesity is one of the most important public health problems. Active video games (AVGs) have been proposed as an attractive alternative to increase energy expenditure and are being investigated to determine their effectiveness against childhood obesity.

Objective: The aim of this study is to summarize the existing research and draw conclusions about the effects of AVGs on health-related physical fitness and motor competence in children and adolescents with overweight and obesity.

Methods: The search strategy was applied to PubMed, MEDLINE, Web of Science, and SPORTDiscus, including randomized and nonrandomized controlled trials investigating the effects of AVG programs on health-related physical fitness and motor competence in children with overweight and obesity. To measure the risk of bias in randomized and nonrandomized controlled trials, 2 different quality assessment tools were used. In total, 15 articles met the inclusion criteria, and the variables of interest were BMI, body fat percentage, cardiorespiratory fitness (CRF), waist circumference, fat-free mass, muscular fitness, and motor competence. A meta-analysis was performed.

Results: Positive effects were found for BMI and body fat percentage, favoring the AVG group compared with a control group with no intervention (mean difference −0.209; 95% CI −0.388 to −0.031 vs mean difference −0.879; 95% CI −1.138 to −0.602). Positive effects seem to be observed for CRF. The effects of AVG interventions on muscular fitness, fat-free mass, waist circumference, and motor competence are unclear.

Conclusions: AVG programs showed positive effects on BMI, body fat percentage, and CRF. AVG could be a good strategy to combat childhood obesity.

(JMIR Serious Games 2021;9(4):e29981) doi: 10.2196/29981

KEYWORDS
active videogames; exergaming; BMI; body fat; motor skills; cardiorespiratory fitness; muscle
Introduction

Background
Childhood obesity is one of the most important public health problems in the 21st century in high-income societies [1]. The prevalence of overweight and obesity in childhood has acquired the status of an epidemic. The global prevalence of overweight and obesity among children and adolescents (aged 5-19 years) has risen dramatically from 4% in 1975 to over 18% in 2016. For instance, the prevalence of overweight was over 30%, and the prevalence of obesity was over 10% in European children and adolescents in 2016 [2]. Obesity has become a pandemic owing to an obesogenic environment that causes cardiovascular and cardiometabolic diseases and psychosocial problems [3,4]. Children with overweight and obesity are likely to remain obese during adulthood and are more likely to develop many other types of cardiovascular and metabolic pathologies [1]. Evidence shows that cardiovascular risk is inversely related to physical fitness [5] and the amount of physical activity (PA) [6] performed by youth. The components of health-related physical fitness are cardiorespiratory fitness (CRF), body composition, muscular strength, muscular endurance, and flexibility [7]. Childhood obesity is related to poor health-related physical fitness, such as CRF and muscular strength [8].

On the other hand, the recommendation of the World Health Organization indicates that a daily average of 60 minutes of moderate-to-vigorous PA provides any of the health benefits in young people, although daily average of beyond 60 minutes of moderate-to-vigorous PA provides additional benefits [9]. In 2016, a study including 1.6 million students aged 11-17 years showed that 81% of them did not meet this recommendation [10]. PA, especially at moderate-to-vigorous intensity, is associated with better physical fitness, independent of sedentary time [11,12].

In addition, one of the main sedentary behaviors of this population is playing electronic games, such as computer or console games [13]. The World Health Organization reported that 40.2% of children and adolescents spend at least 2 hours per day watching television or using electronic devices on weekdays, and this percentage rises to 75.8% during weekends, going further than the recommendations of maximum screen time [14]. This inactivity and excessive sedentary screen time are catastrophic for motor development in children and adolescents [15]. A recent systematic review performed by Han et al [16] showed that children and adolescents with overweight and obesity have a lower motor competence level than children and adolescents with healthy weight; therefore, low motor competence needs to be taken into consideration in children with overweight or obesity. Moreover, children with high actual and perceived motor competence will probably show higher PA and lower BMI status [17]. An improvement in motor competence may promote better perceived motor competence, which entails higher motivation and participation in extracurricular PA and sports [18,19]. In addition, evidence shows a relationship between motor competence and health-related physical fitness during childhood and adolescence [20]. Thus, improving motor skills in children with overweight and obesity is one of the main objectives.

It is well known that exercise is an effective tool to fight obesity [21], with all its associated benefits, such as improvements in BMI status or adiposity, cardiorespiratory and muscular fitness, or bone health [22]. However, the main challenge is to ensure adherence to exercise in children with overweight and obesity [23]. Therefore, the implementation of new types of exercise that are more attractive and motivational to this population is needed.

Active video games (AVGs) have been proposed as a suitable alternative to exercise and are being investigated to determine their effectiveness against childhood obesity. AVGs generally require full-body movement and therefore increase energy expenditure [24]. A systematic review showed that structured AVG sessions had the potential to increase PA in children, but there was no evidence of the benefits of conducting them in the home setting [25]. PA and energy expenditure during AVGs are a well-studied topic showing that AVGs elicit light-to-moderate PA, and also elevate energy expenditure to moderate-to-vigorous intensity, thus having a favorable influence on energy balance [26-28]. Nevertheless, energy expenditure has been found to be higher in structured programs [29]. AVGs seem to be an interesting strategic tool to encourage an active and healthy lifestyle as an alternative to sedentary behaviors [30-32]. However, according to the overview performed by Kari [33], additional high-quality research and systematic reviews concerning exergaming are needed. In addition, AVGs seem to be an effective tool for improving self-concept, self-efficacy, situational interest and motivation, enjoyment, and psychological and social well-being [34,35]. Specifically, AVGs may have positive effects on the psychological aspects and mental health of children and adolescents with overweight or obese [36,37]. Finally, AVGs may have a positive effect on motor competence and health-related physical fitness. Some studies have shown enhancements in children’s motor competence and perceived competence [38-40] or improvements in health-related physical fitness, such as cardiorespiratory and muscular fitness [41-43] and body composition [44], after an AVG intervention.

Objective
Therefore, the main aim of this systematic review is to summarize and critically appraise the existing research on the effects of AVGs on health-related physical fitness and motor competence in children and adolescents with overweight and obesity and to extract conclusions from a fair comparison of the studies included.

Methods

Data Sources and Search Strategy
This review was performed following the criteria and methodology established by the Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0) [45]. This review was performed according to the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) 2020 statement [46]. The PRISMA checklist is shown in
Multimedia Appendix 1. The protocol was registered in the International Prospective Register of Systematic Reviews, PROSPERO (CRD42020189138).

Journal articles were identified by searching electronic databases, scanning reference lists of articles, and examining tables from previous systematic reviews. The search strategy was applied to PubMed, MEDLINE, Web of Science, and SPORTDiscus up to and including March 2021.

The search strategy used to identify the articles in PubMed and MEDLINE was as follows: exergam* OR active video gam* OR active videogam* OR active gam* OR interactive video gam* OR interactive videogam* OR Wii OR Xbox OR Kinect OR PlayStation, and Species: Humans and Language: English filters were applied, along with Journal Article for MEDLINE. The search strategy applied in SPORTDiscus was as follows: TX=(exergam* OR active gam* OR active video gam* OR active videogam* OR interactive video gam* OR interactive videogam* OR Wii OR Xbox OR Kinect OR PlayStation) and document type: article and language: English filters were applied. The search strategy used in Web of Science was as follows: TS=(exergam* OR active gam* OR active video gam* OR active videogam* OR interactive video gam* OR interactive videogam* OR Wii OR Xbox OR Kinect OR PlayStation) and document type: article and language: English filters were applied.

Two reviewers (CCC and AGA) independently evaluated all studies. Titles and abstracts were examined, and relevant articles were obtained and assessed using the inclusion and exclusion criteria presented in Textbox 1. The inclusion criteria were used following the PICOS (Population, Intervention, Comparison, Outcomes and Study) format [47]. Interreviewer disagreements were resolved by consensus. A third reviewer (JAC) resolved these disagreements.

Textbox 1. Inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types of participants were children and adolescents with overweight and obesity</td>
</tr>
<tr>
<td>Trials studying the effects of active video game programs on health-related physical fitness and motor competence</td>
</tr>
<tr>
<td>Control group with no intervention or with traditional exercise intervention</td>
</tr>
<tr>
<td>Types of outcome measures included variables of health-related physical fitness, such as cardiorespiratory fitness, musculoskeletal fitness (muscular strength and muscular endurance), and body composition and variables related to motor competence</td>
</tr>
<tr>
<td>Types of studies were randomized and nonrandomized controlled trials</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies were conducted in languages other than English or Spanish</td>
</tr>
<tr>
<td>Data were unpublished</td>
</tr>
<tr>
<td>Studies were conducted with animals</td>
</tr>
<tr>
<td>Studies included participants aged 218 years</td>
</tr>
<tr>
<td>Studies included participants with disabilities, diseases, or disorders other than obesity</td>
</tr>
<tr>
<td>Studies were conducted without pre- and postassessments of the variables of interest</td>
</tr>
<tr>
<td>Studies were dissertations or abstracts from society proceedings or congresses</td>
</tr>
<tr>
<td>Studies included participants with normal weight</td>
</tr>
<tr>
<td>Noncontrolled trials were considered in the discussion of the article with the consideration of the great limitation of the lack of a control group in interpretation of the results</td>
</tr>
<tr>
<td>Noncontrolled trials were not included in the risk of bias assessments or meta-analysis</td>
</tr>
<tr>
<td>All the noncontrolled trials concerning the effects of active video games on motor competence and health-related physical fitness in children and adolescents with overweight and obesity are summarized in Multimedia Appendix 2</td>
</tr>
</tbody>
</table>

Risk of Bias

For assessing risk of bias proposed in the PRISMA 2020 statement, 2 risk of bias assessment tools were used—the Risk of Bias 2 in randomized controlled trials (RCTs) updated by Sterne et al [48] and the ROBINS-I (Risk of Bias in Nonrandomised Studies of Interventions) in nonrandomized controlled trials developed by Sterne et al [49].

Data Extraction

The following information was extracted from each included trial: name of first author, year of publication, sample size, participant characteristics including number of participants, age and sex, type of study, type of intervention, training characteristics including intervention length and frequency, variables and data sources, and outcomes. The reported variables were weight, BMI, z-score of BMI, fat mass, body fat percentage, CRF, waist circumference, fat-free mass, muscular fitness, and motor competence.
**Meta-Analyses**

Children and adolescents with overweight and obesity who underwent an AVG intervention were compared with a control group (ie, group with participants performing a PA intervention and with nonintervention participants). Effect sizes were calculated for each outcome (BMI, body fat percentage, CRF, waist circumference, fat-free mass, muscular fitness, and motor competence). Different meta-analyses were performed by stratifying the studies by type of control group (no intervention or exercise intervention without AVG). When the number of articles made it possible, analyses by subgroups were performed by dividing the studies by the length of the intervention. The free cross-platform software OpenMeta[Analyst] for advanced meta-analysis was used for data processing.

Mean differences (MD) between participants in AVG interventions and controls and their 95% CIs were calculated using a continuous random-effects model (DerSimonian-Laird method). The heterogeneity of the studies was tested using the $I^2$ statistic [48]. This statistic describes the variance between studies as a proportion of the total variance and was interpreted as follows: $I^2$=0%-25% no heterogeneity; $I^2$=25%-50% moderate heterogeneity; $I^2$=50%-75% high heterogeneity; and $I^2$=75%-100% very high heterogeneity. All analyses were performed using the OpenMeta[Analyst] software.

**Results**

**Search Summary**

A total of 13,267 relevant articles were identified using the abovementioned search strategies. Following a review of titles and abstracts and excluding duplicates, the total number of articles was reduced to 599. Of them, 15 articles met the inclusion criteria and were selected for this review. Articles were excluded for the following reasons: studies were cross-sectional (n=160); only psychological, cognitive, nutritional, and balance variables, PA, or energy expenditure were measured (n=388); participants were children with normal weight (n=31); and studies were noncontrolled trials (n=5; Figure 1).

The characteristics of each study included in this systematic review were summarized in different sections following the PICOS format [47].
Methodological Quality

Multimedia Appendix 3 summarizes the methodological quality assessment of RCTs. The risk of bias in the RCT was low.

The quality assessment for the only nonrandomized controlled trial shows a low risk of bias on preintervention, intervention and postintervention, and therefore, a low overall risk of bias [50].

AVG Interventions

There was a great deal of variety across the AVGs used. Interventions mostly ran during physical education lessons, during playtime or lunch time as extracurricular activities after school or at home. The most commonly used devices in AVG interventions were gaming consoles, such as Xbox 360 with Kinect, Nintendo Wii, Sony PlayStation 2, dance mats, and interactive video game cycling. Games included Just Dance, Wii Fit, Wii Sports, Kinect Adventures, Kinect Sport, Dance Central, Dance Dance Revolution (DDR), and EyeToy.

The length of AVG interventions ranged from 8 weeks to 6 months (mean 16.3, SD 6.7 weeks). The frequency of AVG sessions ranged from 1 day to 5 days per week (mean 126.3, SD 55.8 minutes per week). Sessions typically lasted between 30 and 90 minutes (mean 52.0, SD 11.1 minutes) and were delivered by teachers and research assistants. It is therefore complicated to establish a standard length, intensity, and duration of sessions or type of the AVG intervention.

The different control groups either performed another intervention without AVGs, such as physical education or exercise sessions, access to sedentary video games, and learning sessions or were only asked to continue their normal activities.
of daily life, the latter being the most used option for the control group.

**AVG Effects**

All the studies concerning the effects of AVG on motor competence and health-related physical fitness in children and adolescents with overweight and obesity are summarized in **Multimedia Appendix 4**.

A total of 15 randomized and nonrandomized controlled trials showed effects of AVGs on health-related physical fitness [42,44,51-62] and motor competence [50,61] in children and adolescents with overweight and obesity.

BMI, fat mass, or body fat percentage were measured in 14 studies using dual-energy x-ray absorptiometry [52,54] or bioelectrical impedance [51,58-60] to measure body fat. Waist circumference was measured in 4 studies [42,51,52,58]. CRF was evaluated in 8 studies using different tests, such as the 20-m shuttle run test [42,51,59,61], the 3-minute step test [57], and a submaximal test with a cycle ergometer [58,60]. Motor competence was only measured by Van Biljon et al [50] using the Bruininks-Oseretsky Test and by Bonney et al [61] using the Movement Assessment Battery for Children Test-Second Edition.

A quantitative analysis was performed for BMI, BMI z-score, body fat percentage, fat mass, fat-free mass, and waist circumference. Individual study results and global effects are presented in **Figures 2-6**, whereas a summary of the global results is presented in **Table 1**. Data from 2 studies [51,54] were included, taking into consideration when interpreting the data that the results obtained by these studies were adjusted for baseline outcome measures, age, and sex. Before including them, it was ascertained that the studies did not change the trend of the results without them.

**Figure 2.** BMI effect sizes for active video games compared with those for control group. AVG: active video game.

**Figure 3.** BMI z-score effect sizes for active video games compared with those for control group. Analysis by length of the intervention: subgroup 1: interventions lasting more than 12 weeks; subgroup 0: interventions lasting 12 weeks or less. AVG: active video game.
Figure 4. Body fat percentage effect sizes for active video games compared with those for control group. AVG: active video game.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Estimate (95% C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maddison et al 2011</td>
<td>-0.830 (-1.540, -0.120)</td>
</tr>
<tr>
<td>Staiano et al 2017</td>
<td>-0.690 (-0.918, -0.322)</td>
</tr>
<tr>
<td>Staiano et al 2018</td>
<td>-0.200 (-0.434, 0.034)</td>
</tr>
<tr>
<td>Overall (I²=64.59 %, P=0.059)</td>
<td>-0.462 (-0.818, -0.105)</td>
</tr>
</tbody>
</table>

Figure 5. Waist circumference effect sizes for active video games compared with those for control group. AVG: active video game.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Estimate (95% C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maddison et al 2011</td>
<td>-1.210 (-2.450, 0.030)</td>
</tr>
<tr>
<td>Christison et al 2016</td>
<td>0.010 (-4.192, 4.212)</td>
</tr>
<tr>
<td>Staiano et al 2017</td>
<td>0.000 (-0.798, 0.798)</td>
</tr>
<tr>
<td>Overall (I²=23.5 %, P=0.271)</td>
<td>-0.426 (-1.295, 0.444)</td>
</tr>
</tbody>
</table>

Figure 6. Fat-free mass effect sizes for active video games compared with those for control group. AVG: active video game.

<table>
<thead>
<tr>
<th>Studies</th>
<th>Estimate (95% C.I.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maddison et al 2011</td>
<td>-0.400 (-3.117, 2.317)</td>
</tr>
<tr>
<td>Staiano et al 2017</td>
<td>0.500 (-0.475, 1.475)</td>
</tr>
<tr>
<td>Staiano et al 2018</td>
<td>-0.300 (-1.382, 0.782)</td>
</tr>
<tr>
<td>Overall (I²=0 %, P=0.522)</td>
<td>0.106 (-0.594, 0.806)</td>
</tr>
</tbody>
</table>

Table 1. Effect sizes and heterogeneity of findings for studies comparing active video game intervention versus control group (N=15).

<table>
<thead>
<tr>
<th>Measures</th>
<th>Studies, n (%)</th>
<th>Hedge g effect size</th>
<th>Value, 95% CI</th>
<th>P value</th>
<th>P (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BMI</td>
<td>4 (27)</td>
<td>-0.209</td>
<td>-0.388 to -0.031</td>
<td>.70</td>
<td>0</td>
</tr>
<tr>
<td>BMI z-score</td>
<td>6 (40)</td>
<td>-0.066</td>
<td>-0.124 to -0.007</td>
<td>&lt;.001</td>
<td>97.55</td>
</tr>
<tr>
<td>Body fat percentage</td>
<td>3 (20)</td>
<td>-0.462</td>
<td>-0.819 to -0.105</td>
<td>.06</td>
<td>64.59</td>
</tr>
<tr>
<td>Fat-free mass</td>
<td>3 (20)</td>
<td>0.106</td>
<td>-0.594 to 0.806</td>
<td>.52</td>
<td>0</td>
</tr>
<tr>
<td>Waist circumference</td>
<td>3 (20)</td>
<td>-0.426</td>
<td>-1.295 to 0.444</td>
<td>.27</td>
<td>23.5</td>
</tr>
</tbody>
</table>

A quantitative analysis was not performed for CRF because of different measurement methods and articles with the same sample and for motor competence or muscular fitness because of the lack of articles. Some articles were excluded from the quantitative analyses given that the effect sizes could not be calculated from the information available in the papers [56,57], the sample was the same between studies [44,59], or the number of studies found was insufficient [50]. Another article was excluded from the quantitative analyses because of the control group exercised [58,60,61]. Noncontrolled trials [63-67] were
considered in the discussion and are summarized in Multimedia Appendix 1.

**Weight, BMI, and Body Fat**

A total of 13 studies evaluated changes in weight, BMI, fat mass, or body fat percentage measured. Of the 13 studies, 9 reported positive effects of an AVG intervention on body weight, BMI, or body fat in children with overweight and obesity. The first study that investigated the effect of AVG on BMI status and body composition in adolescents with overweight and obesity was performed by Adamo et al. [58], and they compared a 10-week AVG cycling intervention with a stationary bike music intervention. There were no significant group or group by time effects on body weight, BMI, fat mass, or fat-free mass, but a reduction in body fat percentage was found when groups were combined and compared with baseline. Nonsignificant results could be explained by the small sample size of the study and the shortness of the intervention period. The difference in effect sizes produced by the 2 types of training on BMI or body fat could be explained by the different energy expenditure, because AVG cycling intervention spent 576.2 kcal, whereas stationary bike music group spent 554.6 kcal.

Maddison et al. [51] investigated the effect of the EyeToy of PlayStation on the body composition of 322 children. Participants in the intervention group were encouraged to meet the recommendations of 60 minutes per day and to substitute periods of traditional inactive video games, and they received a PlayStation 2 and the EyeToy to play at home. Differences between control and intervention groups were found for BMI (0.24 kg/m²; 95% CI: -0.44 to -0.04; \( P = 0.02 \)), BMI z-score (0.06; 95% CI: -0.12 to -0.03; \( P = 0.03 \)), body weight (0.72 kg; 95% CI: -1.33 to -0.10; \( P = 0.02 \)), fat mass (0.80 kg; 95% CI: -1.36 to -0.24; \( P = 0.005 \)), and body fat percentage (0.83%; 95% CI: -1.54 to -0.12; \( P = 0.02 \)), favoring the AVG group. A year later, derived from the previous study, the authors studied the mediating effect of CRF on body composition and concluded that an AVG intervention with EyeToy can have a positive effect on body composition in children with overweight or obesity and that this effect is most likely mediated through an improvement in CRF [59].

Foley et al. [44] divided the analyses of the previous study by subgroups such as ethnicity, sex, and CRF level, and the results showed that AVG can be used to improve body composition regardless of ethnicity, sex, and CRF level.

Staiano et al. [56] studied the effect of playing Nintendo Wii on school days on the weight of adolescents and compared the effects between co-operative and competitive AVGs versus a control group. The results showed that the co-operative AVG group lost more weight than the control group, whereas the competitive AVG group did not differ from the co-operative AVG group and control group. The authors studied the effect of psychological variables on weight loss and, as expected, those who had higher peer support at baseline lost marginally more weight over time, but, unexpectedly, initial self-efficacy did not affect weight change over time, nor did higher self-esteem cause more weight loss over time. The 2 studies by Staiano et al. [52,54] investigated the effects of AVG interventions using Xbox Kinect. In the first study [52], a dancing AVG had no effects on or differences between the AVG and control groups. The small sample size and the short length of the intervention could explain the nonsignificant effects of AVG. In the other study by Staiano et al. [54], a home-based AVG intervention led to a reduction in the BMI z-score (mean: -0.06, SD: 0.03) and the weight z-score (mean: -0.09, SD: 0.05) in the AVG group in comparison with the control group (mean: 0.03, SD: 0.03 and mean: 0.07, SD: 0.04 for BMI z-score and weight z-score, respectively) when one control outlier was excluded. There was a nonsignificant intervention effect for fat mass or body fat percentage. This could be influenced by the small sample size, in addition to the fact that the performance in the sessions with AVG at home showed lower or no benefits.

Trost et al. [55] compared the effects of a 16-week weight management program with family-based theoretical sessions focused on lifestyle and the effects of the same program with AVG intervention. The overweight rate was reduced by 5.4% in the weight management program without PA and 10.9% in the AVG group, with significant pre-post and between-group differences. Both groups exhibited reductions in BMI z-score, but the AVG group showed greater reductions (mean: -0.25, SD: 0.03 vs mean: -0.11, SD: 0.03).

The most recent study was conducted by Irandoust et al. [62], and the results showed reductions in weight and BMI in the AVG and exercise groups from pretest to posttest, resulting in lower weight and BMI at posttest measures in these groups in comparison with the control group after a 6-week intervention using Xbox Kinect and Nintendo Wii.

Of the 13 studies, 4 reported no positive effects of an AVG intervention on BMI or body fat in children with overweight and obesity [42,52,53,57]. Apart from Staiano et al. [52], no other authors found effects after an AVG intervention. Christison et al. [42] compared the effects of weight management didactic sessions with an AVG intervention with Nintendo Wii and DDR. The results showed a trend of reduction in the BMI z-score in the AVG group, whereas the BMI z-score among the control group was essentially unchanged (\( P = 0.07 \)). Wagener et al. [53] showed no pre-post differences in BMI within or between conditions after 10 weeks of an AVG intervention by playing dance-based AVGs. Probably, no benefits of AVGs and even nonsignificant worse results on BMI for the AVG group were found because of the small sample size and the short length of intervention. There were same limitations in the study performed by Maloney et al. [57], which showed that there were no changes between pretest and posttest participant weight in the AVG group or control group after 12 weeks of intervention with AVGs.

Furthermore, 5 noncontrolled studies observed positive effects of an AVG intervention on BMI or body fat in children with overweight and obesity. The most recent study was performed by Argarini et al. [67], and the results showed a significant decrease in weight, BMI, and body fat percentage after a 4-week AVG intervention with Xbox Kinect. Christison et al. [66] evaluated the efficacy of an AVG intervention and the results showed a significant decrease in BMI (mean: -0.48, SD: 0.93 kg/m²) and BMI z-score (mean: -0.07, SD: 0.14) after 10 weeks of training (\( P = 0.002 \) and \( P < 0.001 \), respectively). Duman et al. [64] investigated the effects of a combination of music-accompanied...
aerobics, callisthenic exercises, and AVGs, and the results showed decrease in BMI and triceps skinfold thickness. The percentage of obese children decreased from 72% to 40%; those children who were obese became children with overweight, so the percentage of overweight children increased from 28% to 46%; the percentage of children with normal weight increased from 0% to 14%. Calceterra et al [65] demonstrated the effectiveness of a combination of circuit-based aerobics and strength and resistance exercises with AVGs, showing a significant decrease in BMI (from 32.9 to 31.9 kg/m²; P=.002) and body fat percentage (from 39.3% to 36.0%; P=.001). A very interesting result of this study was that 27.2% of the participants reported a previous negative experience with exercise, so a reduced drop-out rate during activity may be achieved with a playful aspect and adapted activities such as AVGs. Finally, Huang et al [63] investigated the effect of AVGs using Nintendo Wii and Xbox Kinect, with no effects on the percentage of body fat, probably because of the short length of the intervention and the reduced number of participants.

Systematic reviews have been performed on the effects of AVG on BMI or body fat, but they are mostly not focused on children or adolescents with overweight or obesity and including studies with children with normal weight; some limitations can be found in these studies, such as the inclusion of noncontrolled trials. The results of these studies are in line with the results of this study. The latest systematic review was performed by Gao et al [68], who included noncontrolled trials and studies with children with normal weight. Reduction in BMI after AVG interventions was found in children and adolescents. Hernández-Jimenez et al [69] performed a meta-analysis that showed a significant effect in favor of AVGs on BMI in children and adolescents, with better results achieved when the AVG intervention was applied to children with overweight or obesity. Another systematic review [70] included 4 RCTs, which are also included in this systematic review, which reported decreases in BMI or body fat after an AVG intervention. A previous systematic review performed by Gao et al [31] concluded that AVGs were a promising tool to promote PA and health as long as the AVG intervention is not home based, but this review did not focus on children with overweight or obesity. Two systematic reviews [30,71] supported the findings, although being among the first reviews on the effects of AVGs, quantitative analyses were not conducted because of a lack of articles. Lamboglia et al [30] found that AVG led to increased PA and CRF and decreased body fat, with considerable potential to fight obesity. Leblanc et al [71] found that AVG attenuated weight gain in participants with overweight and obesity, including 3 articles that are included in this systematic review. The improvement of cardiometabolic health through AVG was inconclusive because of the small number of articles at the time.

**Quality Assessment of BMI, Body Fat Percentage, and Fat Mass**

As shown in Figure 2, positive effects of the interventions were found for BMI, favoring the AVG group compared with the control group with no intervention (MD −0.209; 95% CI −0.388 to −0.031). Heterogeneity among studies for BMI was low (I²=0%; P=.70). AVG showed more positive effects on BMI z-score than on BMI (MD −0.066; 95% CI −0.124 to −0.01), but it also showed a very high heterogeneity (P=97.55%; P<.001; Figure 3). The results of the subgroup analysis by the length of the intervention showed that the decrease in BMI z-score was higher in the AVG interventions longer than 12 weeks.

As shown in Figure 4, positive effects of AVG interventions were found for body fat percentage, favoring the AVG group compared with the control group with no intervention (MD −0.462; 95% CI −0.819 to −0.105). Heterogeneity among studies for BMI was high (P=97.55%; P<.001; Figure 3). The results of the subgroup analysis by the length of the intervention showed that the decrease in BMI z-score was higher in the AVG interventions longer than 12 weeks.

As shown in Figure 4, positive effects of AVG interventions were found for body fat percentage, favoring the AVG group compared with the control group with no intervention (MD −0.462; 95% CI −0.819 to −0.105). Heterogeneity among studies for BMI was high (P=97.55%; P<.001; Figure 3). The results of the subgroup analysis by the length of the intervention showed that the decrease in BMI z-score was higher in the AVG interventions longer than 12 weeks.

As shown in Figure 4, positive effects of AVG interventions were found for body fat percentage, favoring the AVG group compared with the control group with no intervention (MD −0.462; 95% CI −0.819 to −0.105). Heterogeneity among studies for BMI was high (P=97.55%; P<.001; Figure 3). The results of the subgroup analysis by the length of the intervention showed that the decrease in BMI z-score was higher in the AVG interventions longer than 12 weeks.

As shown in Figure 5, no overall effects were found on waist circumference after AVG intervention with EyeToy (MD −0.124; 95% CI −0.209 to −0.038) and no effects were found. The first study was performed by Adamo et al [58] and showed no effects or differences between groups for waist circumference. Maddison et al [51] reported no changes in waist circumference after a 24-week AVG intervention with EyeToy performed at the participants’ homes. Christison et al [42] performed a 6-month AVG intervention with Nintendo Wii and DDR; they also did not report positive effects or differences between groups.

The most recent study on waist circumference in children and adolescents with overweight and obesity was performed by Staiano et al [52], who investigated the effects of Xbox Kinect and found no effect on waist circumference or differences between AVG and control groups. The noncontrolled trial performed by Calcaterra et al [65] demonstrated a decrease in waist circumference (-0.59 cm) and waist circumference to height ratio (-0.08) after a 12-week training program combining traditional exercise with AVGs.

**Waist Circumference**

Changes in waist circumference were evaluated by 4 RCTs [42,51,52,58], and no effects were found. The first study was performed by Adamo et al [58] and showed no effects or differences between groups for waist circumference. Maddison et al [51] reported no changes in waist circumference after a 24-week AVG intervention with EyeToy (MD −0.124; 95% CI −0.209 to −0.038) and no effects were found. The first study was performed by Adamo et al [58] and showed no effects or differences between groups for waist circumference. Maddison et al [51] reported no changes in waist circumference after a 24-week AVG intervention with EyeToy performed at the participants’ homes. Christison et al [42] performed a 6-month AVG intervention with Nintendo Wii and DDR; they also did not report positive effects or differences between groups. The most recent study on waist circumference in children and adolescents with overweight and obesity was performed by Staiano et al [52], who investigated the effects of Xbox Kinect and found no effect on waist circumference or differences between AVG and control groups.

The noncontrolled trial performed by Calcaterra et al [65] demonstrated a decrease in waist circumference (-5.9 cm) and waist circumference to height ratio (-0.08) after a 12-week training program combining traditional exercise with AVGs.

**Quality Assessment of Waist Circumference**

As shown in Figure 5, no overall effects were found on waist circumference after the AVG interventions (MD −0.426; 95% CI −1.295 to 0.444). Heterogeneity among studies for waist circumference was moderate (P=0.385; P=.027). AVGs seem not to be effective in decreasing waist circumference in children and adolescents with overweight and obesity. It is necessary to look for a way to increase the demands of the activity.

Interventions with AVG do not seem to be effective in decreasing waist circumference in children and adolescents with overweight and obesity. This result may be because of the length of the interventions in the included articles. A reduction in the waist circumference in children with obesity seems to be possible with a combination of AVG with multicomponent exercise instead of AVG exclusively, but RCTs are needed to confirm this. Waist circumference is as important as BMI or body fat percentage because they are good predictors of cardiovascular disease risk factors in children and adolescents, even better than BMI [72]. Therefore, the main aim is to...
decrease waist circumference of children and adolescents with overweight and obesity. However, the results suggest that AVG interventions do not seem to be effective in decreasing waist circumference in children and adolescents with overweight and obesity.

**Fat-Free Mass**

The effects of AVG on fat-free mass were reported in 4 articles [51,52,54,58]. To measure fat-free mass, bioelectrical impedance [51,58] or dual-energy x-ray absorptiometry [52,54] were used. Maddison et al [51] did not find any effects or differences between groups for fat-free mass. Similar results were reported by Adamo et al [58], with no changes or differences between groups. In contrast, Staiano et al [52,54] reported no effects on lean mass after an AVG intervention. Evidence on the effect of AVGs on fat-free mass is limited, and no effects have been shown.

**Quality Assessment of Fat-Free Mass**

As shown in Figure 6, no overall effects were found for fat-free mass after the AVG interventions (MD 0.106; 95% CI −0.594 to 0.806). Heterogeneity among studies for waist circumference was low (I²=0%; P=.52).

**Cardiorespiratory Fitness**

CRF assessments following their AVG interventions were included by 5 studies, and 4 of them managed to find positive results. In general, CRF was improved after an intervention with AVG, such as Nintendo Wii, DDR, EyeToy, or a cycling AVG.

The first study that reported the effects of AVG on CRF in adolescents with overweight and obesity was the study by Adamo et al [58], who observed a significant training effect over time in both AVG cycling and stationary bike interventions. Both interventions produced significant improvements in peak heart rate, peak workload, and time to exhaustion, but no significant differences were found between the exercise groups. With this same intervention, Goldfield et al [60] observed that the psychological benefits of aerobic exercises were related to improved aerobic fitness. The abovementioned study by Maddison et al [51] did not find significant increases in CRF in the AVG group, but the positive effect of AVG on body composition in children with overweight or obesity is most likely mediated through improved aerobic fitness [59]. Maloney et al [57] showed no improvements in CRF in either the AVG or control group after playing DDR for 12 weeks. Christison et al [42] showed that the number of shuttle runs did not change after a 6-month AVG intervention. The most recent study was conducted by Bonney et al [61], who investigated the effect of Wii Fit in comparison with a task-oriented functional training on the performance of the shuttle run test and positive effects on CRF in both groups, but no differences were found between the AVG and control groups performing the task-oriented functional training.

The effects of AVG on CRF in children with overweight and obesity have been studied by 2 noncontrolled trials [63,65]. Calcaterra et al [65] demonstrated in their study an improvement in CRF (3.8 mL/kg/min; P<.001) measured by a walking test on a treadmill reaching 85% of the maximal heart rate. Huang et al [63] showed no effects of AVG using Nintendo Wii and Xbox Kinect on CRF, but the heart rate demonstrated that most participants were able to achieve moderate or vigorous intensity of exercise during most AVG sessions.

The effect of AVG interventions on CRF remains unclear. Probably, the limited effects of AVG interventions on CRF of children with overweight and obesity might be because of insufficient training volume in terms of either weekly frequency or overall duration of the interventions. As mentioned earlier, interventions performed at home could be ineffective for improving health-related physical fitness, such as CRF. As it occurs in children with normal weight, only Calcaterra et al [65] used a submaximal or maximal incremental cardiopulmonary exercise test with a gas analyzer, which is widely recognized as the best single index of aerobic fitness [73,74]. Once again, science-based evidence shows that a combination of AVG with multicomponent exercise could produce more benefits on CRF than AVG exclusively, probably because of a higher volume of training. Therefore, these results must be interpreted with caution because the studies that report results from interventions using AVGs with multicomponent exercise are noncontrolled trials. RCTs are needed to confirm this finding.

A systematic review performed by Zeng and Gao [70] included only 1 RCT, also included in this systematic review, which reported positive effects of an AVG intervention in comparison with the effects of an exercise group, but these results were unclear because of the inclusion of only 1 article.

**Muscular Fitness**

Only 1 RCT [61] and 3 noncontrolled trials [63-65] investigated the effects of an AVG intervention on the muscular fitness of children and adolescents with overweight and obesity. The only RCT about the effects of AVGs on muscle fitness showed that both the AVG group trained with Wii Fit and the control group that performed a task-oriented functional training for 14 weeks improved knee extensors and ankle plantar flexors to maximal isometric strength assessed with a handheld dynamometer. Calcaterra et al [65] demonstrated an increase in muscular strength, improving from a mean of 29.6 kg (SD 9.3 kg) to 32.3 kg (SD 9.8 kg; P=.003) in a handgrip test. Duman et al [64] investigated the effects of an AVG intervention combined with traditional exercise on several physical performance tests that require muscle strength and endurance, such as time to ascend and descend 20 stairs, number of squats they can perform in 120 seconds, time to run 50 m, and rope jumps in 30 seconds. The results showed enhancements in all test performances. Finally, Huang et al [63] investigated the effect of AVG using Nintendo Wii and Xbox Kinect on the muscular strength of the quadriceps and hamstrings that were assessed using a handheld dynamometer and muscular endurance that were assessed by a 1-minute half-sit-up test consisting of completing as many half-sit-ups as possible within a minute. Nonsignificant changes in the muscular strength of the quadriceps or muscular endurance were observed. Low frequency and program duration may explain the lack of significant changes.
A combination of AVG with multicomponent exercise could enhance muscular fitness in children with overweight and obesity, but RCTs are required to confirm these results. A systematic review and meta-analysis [5] showed the importance of muscular fitness for children and adolescents and found associations between muscle fitness and bone health, total and central adiposity, and cardiovascular diseases, metabolic risk factors and self-esteem. In addition, according to Tomlinson et al [75], children and adolescents with overweight and obesity have a greater absolute maximum muscle strength than nonobese persons because increased adiposity induces a chronic overload stimulus on the antigravity muscles; however, when maximum muscular strength is normalized to body mass, individuals with obesity appear weaker, probably because of reduced mobility, neural adaptations, and changes in muscle morphology. Therefore, it is important to include exercises aimed at improving muscular fitness in programs for children and adolescents with overweight and obesity.

**Motor Competence**

Motor competence after an AVG intervention was reported in 2 articles. The first was performed by Van Biljon et al [50], and motor competence was evaluated in 30 individuals with overweight and obesity using the shorter version of the Bruininks-Oseretsky test for motor proficiency. The intervention group performed a 6-week AVG intervention for 3 days per week and 30 minutes per session using Wii; there were 2 control groups, one with access to traditional video games and the other continued with their everyday life activities with no intervention. The AVG group showed improvements in motor competence compared with both control groups, specifically in terms of agility and speed, co-ordination, and reaction time. Another more recent study by Bonney et al [61] showed that both the AVG group that trained with Wii Fit and the control group that performed a task-oriented functional training for 14 weeks improved motor co-ordination, as measured by the Movement Assessment Battery for Children Test-Second Edition. A notable difference between the controlled trials investigating the effect of AVG on motor competence is the length of the intervention; therefore, the study with the longest duration of the intervention [61] (14 weeks) showed positive effects, whereas the one with the shortest duration [50] (6 weeks) showed no effect after the AVG intervention. Another study with no control group showed an improvement in motor competence after 4 weeks of training with Xbox Kinect [67]. This scarcity of studies investigating the effect of AVG on the motor competence of children and adolescents with overweight and obesity is of great importance, given that evidence shows a relationship between motor competence and health-related physical fitness during childhood and adolescence [20], as mentioned earlier. Furthermore, low motor competence, denominated as physical illiteracy, is a component of the pediatric inactivity triad observed by Faigenbaum et al [76], together with exercise deficit disorder and pediatric dynapenia. These 3 components are closely related to each other. Children and adolescents who perceive themselves as having poor motor competence might feel less inclined to participate in PA or sports, which in turn will reduce their ability to improve their muscular fitness or motor competence. Motor competence is particularly important in children and adolescents with overweight and obesity, as they have a lower motor competence level than children and adolescents with a healthy weight [16]; this has health consequences because it is directly related to PA. Improving motor competence could increase PA, reduce sedentary behaviors, and positively impact health-related physical fitness [16,77]. Improvements in motor competence entail higher motivation and participation in extracurricular PA and sports [18,19]. AVGs could be a tool for improving motor competence, but randomized clinical trials are needed to corroborate this.

Interventions with AVG appear to be more effective in decreasing BMI and body fat in children and adolescents with overweight or obesity than in children with normal weight, but effectiveness in other health-related physical fitness parameters such as waist circumference or CRF is still unclear. Children and adolescents with overweight or obesity could benefit from AVGs to improve their motor competence, which seems to be a variable more susceptible to enhancement as improvements have been observed with shorter interventions, but further research is needed to confirm this hypothesis. Therefore, AVG programs could be a good strategy to combat childhood obesity.

**Discussion**

**Principal Findings**

This paper provides knowledge about the effects of AVG on health-related physical fitness and motor competence in children and adolescents by gathering previous scientific evidence, and it also provides the prospects for future studies. Mental health can be considered as a major influencing variable and can undoubtedly influence the effects of AVG interventions on childhood obesity; however, this systematic review has focused on the effects of AVG interventions on health-related physical fitness and motor competence in children and adolescents with overweight and obesity, discarding the psychological aspects. However, a very recent overview [37] and a systematic review [36] found a positive effect of AVG on mental health, but it also showed the need for increasing scientific research in this area. In contrast, the participants in the studies included in this review were from countries with a medium-to-high socioeconomic status; this leads to the limitation that such AVG interventions are difficult to implement or are not applicable in countries with a low socioeconomic status because of the lack of resources and lower purchasing power. It would be interesting to study the feasibility and possibilities of such interventions in societies with low socioeconomic status. Furthermore, studies investigating the effects of AVG in children with obesity with some information from geohash dashboards related to the environment where the participants live and where the school is located can be useful to increase the strength of the hypothesis on how AVG affects the features addressed, as there are several features where weak correlation has been found with AVG use, as. As stated earlier, it is necessary to establish the guidelines for an effective intervention using AVGs. It seems that the most effective application of an AVG intervention is a sufficient duration and the structuring and planning of that
intervention. In addition, most of the scientific evidence that has studied AVG interventions combined with traditional exercise lacked a control group to compare the effects; these AVG interventions seem to be promising, but RCTs are needed to investigate the effects of these AVG interventions.

Finally, few studies have examined the effects of an AVG intervention on muscular fitness and motor competence, which, as mentioned earlier, are components of the pediatric inactivity triad observed by Faigenbaum et al [76] together with exercise deficit disorder. The importance of these 2 variables lies in their close relationship with PA and sports participation, which can improve the physical fitness and body composition of children and adolescents, which is especially important in those who are overweight or obese. By improving the muscle strength and motor skill of these children and adolescents, we may make them more active and therefore healthier.

RCTs are needed to investigate the effects of AVG interventions on children's muscle strength and motor skills to learn about the possibilities of AVG interventions in stopping the vicious circle of pediatric inactivity triad.

Limitations
The limitations of this review should be acknowledged. A wide variety of AVG interventions have been included, with different devices and training interventions (duration, frequency, training setting or training dynamic, and type of AVG), which makes it difficult to analyze all the articles together and to obtain generalized results. In addition, the potential risk of bias of some studies was not considered when interpreting the results. Finally, some subgroup analyses were not performed because of the small number of controlled trials. Gender, demographics, or race influence were not deeply addressed because the studies did not show the results divided by these covariates. However, it would be interesting to investigate whether such interventions are more effective depending on them.

Strengths
This study also has several strengths. To the best of our knowledge, this is the first meta-analysis to summarize the existing research on the effects of AVG on health-related physical fitness and motor competence in children and adolescents with overweight and obesity. This analysis included not only the effects of AVG on BMI, but also those on body composition, CRF, muscular fitness, and motor competence. This study allowed us to realize that more RCTs reporting motor competence and muscular fitness results are needed.

Conclusions
AVGs could be a good strategy to fight childhood obesity. AVG programs showed positive effects on BMI and body fat percentage. Improvements in CRF have been observed after an AVG intervention. Children and adolescents could benefit from AVGs to improve motor competence, but further research is needed to confirm these results. The effects of AVG programs on muscular fitness or fat-free mass are also unclear.

In conclusion, AVGs seem to be an effective tool to improve health-related physical fitness and is a promising tool for improving motor competence in children and adolescents. AVGs can even be considered as a prospective alternative to traditional exercise for enhancing health status during childhood.

Acknowledgments
CCC received a grant from Gobierno de Aragón (DGA IIU/2023/2017).

Authors' Contributions
All authors have been actively involved in the planning and execution of this study. AGA and JAC were the main researchers in this study, and CCC is the first author. AML, JMP, and GVR were coresearchers. CCC and AGA independently evaluated all the studies. CCC drafted the document, and JMP, AML, GVR, JAC, and AGA critically reviewed the document. All authors read and approved the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 item Checklist.
[PDF File (Adobe PDF File), 231 KB - games_v9i4e29981_app1.pdf ]

Multimedia Appendix 2
Descriptive characteristics of included noncontrolled trials.
[PDF File (Adobe PDF File), 195 KB - games_v9i4e29981_app2.pdf ]

Multimedia Appendix 3
Quality assessment of randomized controlled trials.
[PDF File (Adobe PDF File), 13 KB - games_v9i4e29981_app3.pdf ]
Multimedia Appendix 4
Descriptive characteristics of included studies with children with overweight and obesity.

References


Abbreviations

AVG: active video game
CRF: cardiorespiratory fitness
DDR: Dance Dance Revolution
MD: mean difference
PA: physical activity
PICOS: Population, Intervention, Comparison, Outcomes and Study
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROBINS-I: Risk of Bias in Nonrandomized Studies-of Interventions

Edited by G Eysenbach; submitted 28.04.21; peer-reviewed by N Maglaveras; comments to author 30.05.21; revised version received 11.06.21; accepted 02.08.21; published 18.10.21.

Please cite as:

https://games.jmir.org/2021/4/e29981 Jmir Serious Games 2021 | vol. 9 | iss. 4 | e29981 | p.98 (page number not for citation purposes)
An Interactive Physical-Cognitive Game-Based Training System Using Kinect for Older Adults: Development and Usability Study

Teerawat Kamnardsiri1,2*, BSc, MSc, PhD; Kochaphan Phirom3, BSc, MSc; Sirinun Boripuntakul1*, BSc, MSc, PhD; Somporn Sungkarat3*, BSc, MSc, PhD

1Research Group of Modern Management and Information Technology, College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
2Department of Digital Game, College of Arts, Media and Technology, Chiang Mai University, Chiang Mai, Thailand
3Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand
* these authors contributed equally

Corresponding Author:
Somporn Sungkarat, BSc, MSc, PhD
Department of Physical Therapy
Faculty of Associated Medical Sciences
Chiang Mai University
9th Floor, Faculty of Associated Medical Sciences
110 Intawaroros Rd, Sripoom
Chiang Mai, 50200
Thailand
Phone: 66 53936042
Fax: 66 53935072
Email: somporn.sungkarat@cmu.ac.th

Abstract

Background: Declines in physical and cognitive functions are recognized as important risk factors for falls in older adults. Promising evidence suggests that interactive game-based systems that allow simultaneous physical and cognitive exercise are a potential approach to enhance exercise adherence and reduce fall risk in older adults. However, a limited number of studies have reported the development of a combined physical-cognitive game-based training system for fall risk reduction in older adults.

Objective: The aim of this study is to develop and evaluate the usability of an interactive physical-cognitive game-based training system (game-based exercise) for older adults.

Methods: In the development phase (Part I), a game-based exercise prototype was created by integrating knowledge and a literature review as well as brainstorming with experts on effective fall prevention exercise for older adults. The output was a game-based exercise prototype that covers crucial physical and cognitive components related to falls. In the usability testing (Part II), 5 games (ie, Fruits Hunter, Where Am I?, Whack a Mole, Sky Falls, and Crossing Poison River) with three difficulty levels (ie, beginner, intermediate, and advanced levels) were tested in 5 older adults (mean age 70.40 years, SD 5.41 years). After completing the games, participants rated their enjoyment level while engaging with the games using the Physical Activity Enjoyment Scale (PACES) and commented on the games. Descriptive statistics were used to describe the participants’ characteristics and PACES scores.

Results: The results showed that the average PACES score was 123 out of 126 points overall and between 6.66 and 7.00 for each item, indicating a high level of enjoyment. Positive feedback, such as praise for the well-designed interactions and user-friendly interfaces, was also provided.

Conclusions: These findings suggest that it is promising to implement an interactive, physical-cognitive game-based exercise in older adults. The effectiveness of a game-based exercise program for fall risk reduction has yet to be determined.

(JMIR Serious Games 2021;9(4):e27848) doi:10.2196/27848

KEYWORDS
digital game; interactive game-based training; physical-cognitive training; exergaming; Kinect sensors; older adults; falls; PACES; user-centered design; game-based exercise
**Introduction**

Declines in multiple physiological systems with ageing contribute to balance and gait deficits, leading to an increased risk of fall [1]. Fall is a serious public health problem, and its consequences have a marked adverse impact on physical and psychological aspects such as injuries, activity restriction, fear of falling, and loss of autonomy [2-4]. Given the substantial impact of falls on health, as well as their medical and economic burden, an effective strategy to prevent falls in older adults is warranted.

Several investigators have consistently reported a strong positive effect of physical exercise on fall prevention among older adults [5-9]. Researchers have also identified a critical role of cognition, especially executive function, attention, and memory, on balance and gait control [10-13]. Many examiners have demonstrated that cognitive training, which is an intervention program aimed at improving, maintaining, or restoring cognitive function via the repeated and structured practice of tasks, can improve balance and gait and reduce fall risk [14-17]. Taken together, incorporating a cognitive component into physical exercise may augment its benefits in fall prevention [11-13,18-20]. A growing number of investigators have documented the effects of combined physical-cognitive exercise training in a simultaneous form (dual-tasking) among older adults. Combined physical-cognitive exercise training programs have resulted in greater improvement in physical and cognitive performance than either type of single training alone [21-25].

As technology advances, a new alternative of rehabilitation approach targeting training of various physical and cognitive components in the form of interactive game-based exercises (exergames) is becoming available [26]. These interactive game-based exercises use technology-driven platforms that require users to move their body in order to complete assigned tasks via video game interface elements [27]. The interactive game-based exercises have advantages in terms of gamification features. Researchers have demonstrated that exergames are attractive because they provide real-time interaction and feedback to users, which enhances motivation and training adherence [28,29]. In addition, game-based exercises allow users to be active with repetitive practice and track progression, which is beneficial for training outcomes [30]. Another advantage of exergames is that they can offer an experience for daily-life task requirements based on concurrent training of physical and cognitive components [31-33]. Moreover, game-based exercises can be applied either in rehabilitation centers or community and home settings [33,34]. Several researchers have shown that game-based training using Nintendo Wii Fit, and Microsoft Kinect sensors were effective in improving physical abilities (eg, balance, gait performance), improving cognitive abilities (eg, executive function, speed of processing), and reducing the risk of falls among older adults [35-42]. However, in most existing training programs available for fall prevention, research concerning simultaneous training of physical and cognitive functions (dual tasking) with the use of exergames has remained scarce.

Among interactive game-based technology for training, the Kinect motion sensor (Microsoft Corporation) has been considered as a high-potential approach because it provides a markerless full-body 3D motion tracker and enables users to virtually interact hands-free with a computer system. Additionally, several examiners have demonstrated that among interactive game-based technology for exercising, Microsoft Kinect has an advantage in that it allows individuals to interact with games using their own body in a natural way [43,44]; this enhances the natural form of human-computer interaction [45]. In addition, the Microsoft Kinect motion sensor is an accurate input device; thus, it allows precise tracking and real-time feedback of user performance [46]. The Microsoft Kinect sensor is proposed to be a feasible and effective tool for training concurrent physical and cognitive components in older adults, with the aim of reducing the intrinsic causes (ie, physical and cognitive) of falls [47-51].

Based on the usability challenges faced by older adults, programmers should develop user interfaces that are user-friendly for older adults and specific for training purposes. A previous investigator has suggested that older adults accept innovative technology when they recognize its benefit and find it meaningful for their lives [52].

Therefore, this study focused on developing a prototype of a game-based exercise that accounts for the target user’s expectation and requirement (eg, enjoyment, attractiveness, user skill-challenge balance, and benefit for training cognitive and physical components). In particular, following a user-centered design approach, acceptability and training adherence in future empirical studies were established.

**Methods**

**Study Design**

The present study consists of two parts: (1) the development of a game-based exercise prototype, and (2) evaluation of target users’ feedback, response, and satisfaction regarding the developed game-based exercise prototype. The concept of development and evaluation of the target user’s experience for producing a final prototype followed a user-centered design (UCD) approach [27,46,53-62]. UCD is a user interface design process in which designers focus on gaining a target user’s perspective to create a product with a high degree of usability. The UCD cycle is depicted in Figure 1.
Part I: Development of a Game-Based Exercise Prototype

Development Process

In this part, the 4-phase UCD process was applied [27,46,53-62]. The first phase, the design development process of a game-based exercise prototype, was conducted in the brainstorming phase. A total of 7 team members, including 3 physical therapists and 1 physician (3-20 years of experience in geriatric and cognitive rehabilitation), 2 game programmers (5 years of experience in the Unity 3D game engine), and 1 game designer (10 years of experience in game design and game theory), participated in the brainstorming session. This phase involved generating potential core game ideas by integrating the knowledge and literature review of previous physical and cognitive training programs and interactive exergame interventions for fall prevention in older adults [16,63-65]. In the present study, the core game was composed of two training elements: (1) a physical element, including stepping and balance training, and (2) a cognitive element related to balance and falls in older adults, including executive function, attention, and memory [18,66,67]. To ensure that the difficulty of the games was appropriate for each individual user, the progression of game difficulty was considered. The game-based training program underwent critical appraisal by the physical therapists and physician. In the second phase, after consensus, proven game ideas were used to create a game-based digital exercise game prototype using the Unity 3D game engine software with Kinect Sensor V2 for Windows. In the third phase, user feedback to improve the game-based exercise prototype was provided by end users using a think-aloud method [68]. Finally, the fourth phase concentrated on assessing physical activity enjoyment during game engagement. The Physical Activity Enjoyment Scale (PACES) questionnaire [69,70], an 18-item scale questionnaire, was used to analyze physical activity enjoyment. Moreover, the usability of the game-based exercise prototype was determined with a structured interview (feedback about the game, themes, user interface, sound effect, graphics, and interaction). Feedback from all participants was considered by the research team to improve the game-based exercise prototype.

Characteristics of the Game-Based Exercise Prototype

The characteristics of the game-based exercise prototype are presented in Table 1. The game can be described as individual interactive game-based training using Kinect. The prototype of the game-based exercise comprised 5 games, including (1) Fruits Hunter, (2) Where Am I ?, (3) Whack a Mole, (4) Sky Falls, and (5) Crossing Poison River. The games had three different levels: beginner, intermediate, and advanced. The level of game complexity progressed by increasing the difficulty of physical demand (ie, movement speed, distance, duration, base of support) and cognitive demand (ie, number of stimuli, complexity of the game’s rules, and amount of cognitive load). The estimated play time was 45 to 60 minutes.
Table 1. Summary of the characteristics of the developed game-based exercise prototype.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Health topic</td>
<td>A game-based exercise prototype</td>
</tr>
<tr>
<td>Targeted age group</td>
<td>Older adults (age ≥ 65 years)</td>
</tr>
<tr>
<td>Short description of the game idea</td>
<td>The game-based exercise is a virtual, interactive game-based training system using Microsoft Kinect motion sensor technology. The game-based exercise comprises 5 games that include physical and cognitive components associated with balance and falls in older adults.</td>
</tr>
<tr>
<td>Target player</td>
<td>Individual</td>
</tr>
<tr>
<td>Behavior change procedure used</td>
<td>A game-based exercise is used to enhance motivation and engagement in older adults.</td>
</tr>
<tr>
<td>Clinical support needed</td>
<td>Physical therapist and geriatric physicians</td>
</tr>
<tr>
<td>Data shared with clinician</td>
<td>Data are saved and stored in the hard disk. However, reaction time, error, and score are given as feedback on the display screen (ie, the rubber mat) at the end of each game.</td>
</tr>
<tr>
<td>Type of game</td>
<td>Physical, action, real-time strategy</td>
</tr>
<tr>
<td><strong>Game components</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Player’s game goal/objective</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Physical components         | • Improve static and dynamic balance  
• Improve stepping reaction and response time  
• Improve lower limb muscle strength                                                                                                                                                  |
| Cognitive components        | • Fruits Hunter: improves response ability and speed of processing via a stepping task.  
• Where am I ?: improves semantic memory and visuospatial ability via visual sense  
• Whack a Mole: improves selective attention ability, visual attention performance, speed of processing, and inhibition ability  
• Sky Falls: improves sequencing and planning ability  
• Crossing Poison River: improve episodic memory via auditory sense                                                                                                                                 |
| Rules                       | • Fruits Hunter: step on the presented fruits as fast as possible within a limited time.  
• Where am I ?: step to the presented objects and remember as many as of them possible. The recall questions are provided at the end of the game.    
• Whack a Mole: respond correctly to different rules of this game as follows:  
  • Mole or rabbit: steps on the target 1 time  
  • Mole or rabbit with helmet: steps on the target 2 times  
  • Bomb: do not step on the target  
• Sky Falls: step with alternating feet to collect as many dropping objects in the basket as possible  
• Crossing Poison River: listen to a short story and remember the content of the story while standing on one leg                                                                                                                                 |
| Game mechanics              | The game-based exercise system allows users to interact with the virtual games by stepping on the presented targets in different directions in pursuit of the game’s goals. The game-based exercise also provides audio and visual feedback to the users while they are playing the games. |
| Virtual environment         | A forest with fruits, animals, vegetables, and a river                                                                                                                                                       |
| Setting                     | The game-based exercise can be set in a room environment                                                                                                                                                     |
| Device requirements         | Personal computer/notebook/laptop with LED projector                                                                                                                                                        |
| Sensors used                | Microsoft Kinect Sensor V2                                                                                                                                                                                 |
| Estimated play time         | 45-60 minutes                                                                                                                                                                                               |

Part II: Evaluation of Target User’s Experience

Recruitment and Participants

A total of 5 community-dwelling older adults were enrolled as representative target users. The inclusion criteria were (1) age 65 years or older, (2) had normal cognitive function (determined by a Mental State Examination T10 [71] score ≥24 points or depending on the level of education), (3) ability to walk without an assistive device for at least 10 m, and (4) ability to step in all directions independently and safely. Exclusion criteria were (1) depressive symptoms (determined by a Thai Geriatric Depression Scale-15 [72] score >6 points), (2) orthopedic deficits, neurological deficits, and/or other significant health problems that precluded the participant from completing the testing protocol, and (3) uncorrected visual and hearing
impairment. The study protocol was approved by the Human Ethical Review Board of the principal investigator’s institute (AMSEC-61EX-078). All participants gave written informed consent prior to participating in the study. The demographic data of the participants, which consisted of age, height, weight, medication use, and history of falls in the previous 12 months, were recorded.

**Hardware Configuration**

To set up the system, the capture volume of the system was configured using the three main devices: Kinect Sensor V2 [73], LED projector, and laptop computer. In the present study, the Kinect Sensor V2 was used because it provides greater precision and more stable results compared to the Kinect Sensor V1 [74]. The Kinect Sensor V2 is a depth sensor camera manufactured by Microsoft that provides information about the depth, color, and skeleton of a user who is standing in front of the sensor. The Kinect sensor and LED projector were set on a portable metal storage rack at a height of 0.8 m and 2.0 m from the floor, respectively. The laptop computer that contained the developed game software was set near the Kinect sensor. The game was projected on a rubber mat (2.0 m width × 1.2 m height) that was placed on the floor; thus, the participants could virtually interact by stepping. The center of the rubber mat was set 2.5 m from the Microsoft Kinect sensor and LED projector. The configuration of hardware for playing the game-based exercise is illustrated in Figure 2.

**Figure 2.** Environment configuration of the game-based exercise system.

The Kinect Sensor V2 is a depth sensor camera manufactured by Microsoft that provides information about the depth, color, and skeleton of a user who is standing in front of the sensor. After completing the game-based exercise, participants were asked to rate their enjoyment using the PACES questionnaire [69,70]. The PACES is an 18-item scale questionnaire that assesses physical activity enjoyment during game engagement with a 7-point Likert scale (1, strongly disagree, to 7, strongly agree). A higher PACES score reflects a greater level of enjoyment. Moreover, using a structured interview, participants were interviewed about their impressions of the game-based exercise features in terms of rules, mechanics, interfaces, and scoring, as well as their physical and cognitive involvement while playing the games.

**Protocol**

The game-based exercise was connected with the Microsoft Kinect sensor, an LED projector, and the laptop computer. After that, the system was calibrated by moving four markers in the game-based training system over the four corner marks on the rubber mat. For individualized body position calibration, each participant was asked to perform a T-pose stand for approximately 5 seconds at the center of the rubber mat (Figure 2). After the calibration process, each participant received a comprehensive description of the rules of the games, including a demonstration, and was requested to play the game-based exercise. The American College Sport of Medicine exercise guidelines recommend that older adults should participate in aerobic activities for a minimum of 30 minutes per session to promote and maintain their health-related outcomes [75]. In this study, the exercise duration ranged between 30 and 50 minutes, with a rest interval of between 10 and 15 minutes, depending on the participant performance; this resulted in a total time of 45 to 60 minutes.

After completing the game-based exercise, participants were asked to rate their enjoyment using the PACES questionnaire [69,70]. The PACES is an 18-item scale questionnaire that assesses physical activity enjoyment during game engagement with a 7-point Likert scale (1, strongly disagree, to 7, strongly agree). A higher PACES score reflects a greater level of enjoyment. Moreover, using a structured interview, participants were interviewed about their impressions of the game-based exercise features in terms of rules, mechanics, interfaces, and scoring, as well as their physical and cognitive involvement while playing the games.

**Data Analysis**

Descriptive statistics were used to describe both the participants’ characteristics and their scores on the 18-item PACES questionnaire. All data were analyzed using SPSS 21.0 (IBM Corporation).
**Results**

**Part I: Development of a Game-Based Exercise Prototype**

The framework of the game-based exercise comprised 6 components (Figure 3), including:

1. The Microsoft Kinect sensor: the depth sensor that was used to track and monitor full-body movements in 3D coordinates (ie, the x-, y-, and z-axes). The tracking data were then converted to the 24 points of the Joint ID Map (body skeleton model). In this study, 4 points of the Joint ID Map [43] (ANKLE_RIGHT, FOOT_RIGHT, ANKLE_LEFT and FOOT_LEFT) were used for the interaction between the user and the game.

2. Game programmers: the specialists who generated the digital game using the computer programming language and game engine software (the Unity 3D game engine software with Microsoft Kinect sensor V2 for Windows).

3. Domain knowledge: the experts having core knowledge and experience of physical and cognitive training programs for fall prevention in older adults.

4. Game-based training system: the digital game system, which comprised 5 games (Games I-V) with 3 levels (levels 1-3) and feedback (ie, score, response time, and error).

5. User and laptop computer: a system operator who was responsible for controlling the game-based training system while participants were playing the games.

6. Graphical user interface: a form of user interface that allows participants to interact with the game-based exercise (Figure 4). An example of a participant training with the game-based exercise prototype is displayed in Figure 5.
Figure 3. Framework of the game-based exercise system. GUI: graphical user interface.
Figure 4. Screenshots of the 5 games in the game-based exercise system: (A) Fruits Hunter, (B) Where Am I ?, (C) Whack a Mole, (D) Sky Falls, and (E) Crossing Poison River.

Figure 5. Examples of a participant playing the game-based exercise: (A) Whack a Mole, (B) Sky Falls, and (C) Crossing Poison River.

Part II: Evaluation of the Target Users’ Experience

Participant Characteristics

A total of 5 community-dwelling older adults participated in the usability testing phase. Their mean age was 70.40 (SD 5.41) years (range 65-79 years). No participants had any experience with using exergames. They had low incidence rate of falls in the past 12 months, and they either did not take medication or took only one type. The participants’ characteristics are summarized in Table 2.

Table 2. Characteristics of the study participants.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
<th>Median</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>70.40 (5.41)</td>
<td>68</td>
<td>65-79</td>
</tr>
<tr>
<td>Height (cm), mean (SD)</td>
<td>155.20 (6.06)</td>
<td>156</td>
<td>149-164</td>
</tr>
<tr>
<td>Weight (kg), mean (SD)</td>
<td>53.40 (8.62)</td>
<td>55</td>
<td>39-60</td>
</tr>
<tr>
<td>BMI (kg/m²), mean (SD)</td>
<td>22.08 (2.63)</td>
<td>22.21</td>
<td>17.60-24.30</td>
</tr>
<tr>
<td>Education (years), mean (SD)</td>
<td>14.40 (5.90)</td>
<td>16</td>
<td>4-18</td>
</tr>
<tr>
<td>Types of medication, mean (SD)</td>
<td>0.20 (0.45)</td>
<td>0</td>
<td>0-1</td>
</tr>
<tr>
<td>Falls in the past year, n</td>
<td>1</td>
<td>0</td>
<td>0-1</td>
</tr>
</tbody>
</table>
**User Experience in Using the Game-Based Exercise System**

All 5 participants completed the game-based exercise and answered the PACES questionnaire. The average score on each item was between 6.66 and 7.00, which indicated greater levels of enjoyment. The 18-item PACES scores are illustrated in Table 3.

**Table 3.** Physical Activity Enjoyment Scale (PACES) rating scores (n=5). All items were rated on a 7-point scale from 1, strongly disagree, to 7, strongly agree.

<table>
<thead>
<tr>
<th>Question</th>
<th>Participants</th>
<th>Response rating, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I enjoy it; I hate it</td>
<td>S01 7 S02 6 S03 7 S04 7 S05 6</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>2. I feel interested; I feel bored</td>
<td>S01 7 S02 6 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>3. I like it; I dislike it</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>4. I find it pleasurable; I find it unpleasurable</td>
<td>S01 7 S02 7 S03 6 S04 6 S05 6</td>
<td>6.66 (0.24)</td>
</tr>
<tr>
<td>5. I am very absorbed in this activity; I am not at all absorbed in this activity</td>
<td>S01 6 S02 7 S03 7 S04 7 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>6. It’s a lot of fun; it’s no fun at all</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>7. I find it energizing; I find it tiring</td>
<td>S01 7 S02 6 S03 7 S04 7 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>8. It makes me happy; it makes me depressed</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>9. It’s very pleasant; it’s very unpleasant</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 6</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>10. I feel good physically while doing it; I feel bad physically while doing it</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>11. It’s very invigorating; it’s not at all invigorating</td>
<td>S01 7 S02 7 S03 7 S04 6 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>12. I am not at all frustrated by it; I am very frustrated by it</td>
<td>S01 7 S02 7 S03 7 S04 7 S05 7</td>
<td>7.00 (0.00)</td>
</tr>
<tr>
<td>13. It’s very gratifying; it’s not at all gratifying</td>
<td>S01 7 S02 7 S03 7 S04 6 S05 7</td>
<td>6.66 (0.24)</td>
</tr>
<tr>
<td>14. It’s very exhilarating; it’s not at all exhilarating</td>
<td>S01 7 S02 7 S03 7 S04 6 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>15. It’s very stimulating; it’s not at all stimulating</td>
<td>S01 7 S02 7 S03 6 S04 7 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>16. It give me a strong sense of accomplishment; it does not give me any sense of accomplishment</td>
<td>S01 7 S02 7 S03 7 S04 6 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
<tr>
<td>17. It’s very refreshing; it’s not at all refreshing</td>
<td>S01 7 S02 6 S03 7 S04 7 S05 7</td>
<td>6.66 (0.24)</td>
</tr>
<tr>
<td>18. I felt as though there was nothing else I would rather be doing; I felt as though I would rather be doing something else</td>
<td>S01 7 S02 7 S03 7 S04 6 S05 7</td>
<td>6.80 (0.20)</td>
</tr>
</tbody>
</table>

Rating scale of all items (total points: 126) 125 123 125 120 120 123.00 (1.26)

**User Feedback and Suggestions**

The feedback and suggestions provided by the users are presented in Table 4.


**Discussion**

**Principal Findings**

In this study, we aimed to develop and test the usability of a virtual, interactive game-based training system that is focused on simultaneously training the physical and cognitive function (dual-tasking) of community-dwelling older adults. The core games were formulated by integrating the principal knowledge and existing evidence from the literature related to effective fall prevention exercise programs for older adults as well as by subjecting the games to critical appraisal from experts. We also assessed older adults’ experiences in terms of enjoyment and game features using the PACES questionnaire and a structured interview.

Several intervention studies have reported high dropout rates and limited use of technology-supported platforms for delivering exercise training programs [76-78]. Interactive game-based training may be considered as a more efficient approach for empowering user engagement, which contributes to positive outcomes of an intervention. However, many older adults tend to be less engaged with modern digital technology, and not all are accepting of it. To overcome this limitation, the UCD approach, which incorporates game design principles (ie, goals, rules, feedback, points, time, reward structures, levels, and aesthetics), was used in the process of designing and developing a game-based exercise prototype with an aim to motivate and engage older adults in exercising [54-60]. Researchers have identified the potential benefits of using the UCD concept for developing exergames for older adults with and without health-related problems. For example, Hemingway et al [53] used the UCD process to develop a mobile game to influence the behavior of HIV service uptake among a key population. Lange et al [27] established an interactive game-based program constructed on a UCD design process for training the dynamic balance of individuals who have previously experienced a stroke. Howes et al [79] also used UCD to develop the bespoke Active Computer Gaming system to deliver strength and balance exercise programs for older adults. Together, the present and previous findings consistently suggest that the UCD approach is a core process that should be embedded in health games to ensure usability and acceptability. Therefore, the target users may benefit fully from exergames technology.

To our knowledge, our game-based exercise is the first game prototype that was mainly designed to support older adults in combined physical-cognitive exercising using the Microsoft Kinect motion sensor. In particular, the game-based exercise focused on the core impairment aspects that are related to falls in older adults, including balance and stepping performance as well as executive function, attention, and memory. Several investigators have consistently reported that the most important component of exercise programs for fall prevention is balance training [6,7]. In addition, stepping training, a form of highly specific balance training, has shown to be an effective fall prevention strategy [80]. Thus, balance and stepping training were included in the game-based exercise. Regarding cognition, declined executive function, attention, and memory have been identified as crucial contributors to falls [11-13,19]. Thus, adding these cognitive components to physical training may potentially enhance the efficacy of fall prevention programs for older adults. In addition, age-related perception and sensation decline in older adults were considered. Therefore, visual and

---

**Table 4. Feedback and suggestions provided by the study participants for the game prototype.**

<table>
<thead>
<tr>
<th>Type of feedback or suggestion</th>
<th>Feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>“The games’ feature and appearance were very attractive and enhanced my motivation to complete the games.”</td>
</tr>
<tr>
<td></td>
<td>“The games provided my performance with visible outcomes and scores which motivated me to try harder to get a better score in the next trial.”</td>
</tr>
<tr>
<td></td>
<td>“The difficulty of each game was optimal; it was not too easy and not too difficult.”</td>
</tr>
<tr>
<td></td>
<td>“The games had variety of forms and rules that challenged my physical and cognitive abilities.”</td>
</tr>
<tr>
<td></td>
<td>“The games had a meaningful sound effect which helped me to identify my right or wrong responses.”</td>
</tr>
<tr>
<td></td>
<td>“The games’ systems were quite simple to set up and easy to manage, thus it appeared to be feasible to use in the community or home settings.”</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>“In the Whack a Mole, sometimes I did not step on the bomb, but it eventually blew up.”</td>
</tr>
<tr>
<td></td>
<td>“In the Sky Falls, sometimes it was quite hard to control the movement of the bamboo basket even though I tried to alternate my stepping rhythmically.”</td>
</tr>
<tr>
<td><strong>Suggestions</strong></td>
<td>Game rules</td>
</tr>
<tr>
<td></td>
<td>Clearly state the game instructions and rules at the beginning</td>
</tr>
<tr>
<td>Level design</td>
<td>Reduce the speed of dropping objects in the beginner level of Sky Falls</td>
</tr>
<tr>
<td></td>
<td>Use different types of animals and vegetables for each difficulty level of Where Am I?</td>
</tr>
<tr>
<td>Graphics/look and feel</td>
<td>Adjust the distance of each presented object in Whack a Mole</td>
</tr>
<tr>
<td>Audio</td>
<td>Increase the display volume</td>
</tr>
<tr>
<td></td>
<td>Use different background music for each game</td>
</tr>
</tbody>
</table>
audio presentations, such as the size and distance of target objects as well as the level of sound volume, were included. Findings from the study demonstrated that the target users viewed the game-based exercise prototype as an enjoyable and practical intervention approach for their physical and cognitive training at home and in community settings. This may be because the development of the game-based exercise prototype incorporated the current knowledge regarding the key contributing factors for training continuation by using exergames among older adults. These factors gradually increase the level of game difficulty, provide clearer feedback, and offer a simple setup [81]. In this way, we expected that the newly developed exergames would overcome the barrier to exercise in older adults. This enjoyment (determined by PACES scores) and positive feedback from the users may be, at least in part, due to the fundamental elements of the game-based exercise, which feature a real-time interface display and feedback. Consistent with previous studies, our games system provides feedback, including scores and performance outcomes (ie, response time, error), which enhances the motivation of the users [82-84]. Moreover, the positive responses from older learners who are unfamiliar with new technologies can be attributed to the design features of the game-based exercise prototype, such as having a user-friendly interface and providing optimal levels of task difficulty [85]. Nevertheless, some comments indicated that further refinements are required prior to implementation of the game-based exercise among older adults in a realistic context.

Limitations
This study has certain limitations that need to be acknowledged. This study is a preliminary study that involved a small number of participants who were all female. Further, results on the enjoyment and experiences of using the game-based exercise prototype were obtained from a single training session. Thus, the findings should be considered preliminary and interpreted with caution. Future studies with larger sample sizes, a balanced gender ratio, and data obtained from multiple training sessions are warranted. Another limitation concerns the hardware specification of the Microsoft Kinect sensor. In this study, the capture volume of the Kinect sensor was restricted to 0.5 to 4.5 m, which partly limited the design of the configurations of the games. Further studies should consider using multiple Kinect sensors to cover a greater capture volume. Moreover, our game system was designed for the individual player. Exergames systems that allow group players should be considered for promoting social interaction. Finally, this study investigated the enjoyment during game engagement using the 18-item PACES questionnaire. In future studies, the 8-item version of PACES would be an appropriate questionnaire to reduce the completion time.

Conclusions
This preliminary study demonstrated a prototype of a game-based exercise for older adults using the Microsoft Kinect sensor. The game-based exercise prototype contained combined physical and cognitive training elements with different levels of difficulty. The developed game-based exercise was well accepted by the target users, with prominent enjoyment and positive feedback. Thus, the game-based exercise appears to be a promising tool for enhancing older adults’ motivation to engage in physical-cognitive exercise with the aim to reduce the risk of falls.

Acknowledgments
This research was supported by the Research and Researchers for Industries (RRI) Project, Thailand Science Research and Innovation (TSRI) Grant MSD61I0015 and TSRI Grant RSA6180023 (SS). Moreover, this research was partially supported by Chiang Mai University and with the collaboration of a research group of Modern Management and Information Technology, College of Arts, Media and Technology, and the Department of Physical Therapy, Faculty of Associated Medical Sciences, Chiang Mai University, Thailand. In addition, the authors would like to thank K Intanon, P Sirinual, T Kunthadech, and S Korpraphan for their contributions to the development of the games.

Authors’ Contributions
TK, SB, and SS share first authorships, conducted a major part of the methods and experimental design, developed the software, and contributed to the majority of the writing and reviewing of the manuscript. KP conducted a major part of the methods and experimental design, and commented on the manuscript and reviewed the final manuscript.

Conflicts of Interest
None declared.

References


73. Kinect for Windows Human Interface Guidelines v2.0. Microsoft Corporation. 2014. URL: http://download.microsoft.com/download/6/7/6/676611b4-1982-47a4-a42e-4cf84e1095a8/kinecthig.2.0.pdf [accessed 2021-08-18]


Abbreviations

PACES: Physical Activity Enjoyment Scale
UCD: user-centered design
Immersion Experiences in a Tablet-Based Markerless Augmented Reality Working Memory Game: Randomized Controlled Trial and User Experience Study

Bo Zhang¹, BSc, MSc, PhD; Nigel Robb², PhD

¹Department of Education Information Technology, Faculty of Education, East China Normal University, Shanghai, China
²Research Faculty of Media and Communication, Hokkaido University, Sapporo, Japan

Corresponding Author:
Bo Zhang, BSc, MSc, PhD
Department of Education Information Technology
Faculty of Education
East China Normal University
No 3663, North Zhongshan Road
Putuo District
Shanghai, 200062
China
Phone: 86 13505194318
Email: bzhang@ed.ecnu.edu.cn

Abstract

Background: In recent years, augmented reality (AR), especially markerless augmented reality (MAR), has been used more prevalently to create training games in an attempt to improve humans’ cognitive functions. This has been driven by studies claiming that MAR provides users with more immersive experiences that are situated in the real world. Currently, no studies have scientifically investigated the immersion experience of users in a MAR cognitive training game. Moreover, there is an observed lack of instruments on measuring immersion in MAR cognitive training games.

Objective: This study, using two existing immersion questionnaires, investigates students’ immersion experiences in a novel MAR n-back game.

Methods: The n-back task is a continuous performance task that taps working memory (WM) capacity. We compared two versions of n-back training. One was presented in a traditional 2D format, while the second version used MAR. There were 2 experiments conducted in this study that coordinated with 2 types of immersion questionnaires: the modified Immersive Experiences Questionnaire (IEQ) and the Augmented Reality Immersion (ARI) questionnaire. Two groups of students from two universities in China joined the study, with 60 participants for the first experiment (a randomized controlled experiment) and 51 participants for the second.

Results: Both groups of students experienced immersion in the MAR n-back game. However, the MAR n-back training group did not experience stronger immersion than the traditional (2D) n-back control group in the first experiment. The results of the second experiment showed that males felt deeply involved with the AR environment, which resulted in obtaining higher levels of immersion than females in the MAR n-back game.

Conclusions: Both groups of students experienced immersion in the MAR n-back game. Moreover, both the modified IEQ and ARI have the potential to be used as instruments to measure immersion in MAR game settings.

Trial Registration: UMIN Clinical Trials Registry UMIN000045314; https://upload.umin.ac.jp/cgi-open-bin/ctr_e/ctr_view.cgi?recptno=R000051725

(JMIR Serious Games 2021;9(4):e27036) doi:10.2196/27036

KEYWORDS
augmented reality; markerless augmented reality; immersion experience; cognitive training games; working memory; markerless augmented reality n-back game
Introduction

In the past decade, studies have reported significant improvements in cognitive functions through game-based cognitive training in different groups [1-3]. In recent years, research shows that the application of augmented reality (AR) technology in games is related to the user’s cognitive functionality, as it has a positive impact on mental processes and psychological reactions [4,5]. Many studies also started to explore the affordances of using AR technology to create cognitive training tasks. There is some evidence that AR-based cognitive training is effective in improving users’ cognitive functions, such as attention and memory [6-8]. Other research shows that AR-based tasks play a significant role in increasing students’ motivation and attention during the learning process [9].

By overlapping computer-generated 3D graphics onto a real-world environment, AR enables borderless interactions between the digital and physical worlds [10]. AR has the following key features: combination of virtual content and real context, real-time interaction between virtual objects and real contexts, and the registered 3D virtual object [11]. Through mobile devices (e.g., smartphones and tablets), AR apps capture the real scene with the camera and then present the digital contents to the real context surrounding the user. This means of interaction creates a new mode of learning and training, which can be easily conducted even by students who have no experience using digital devices [12]. AR apps enable users to have ubiquitous access to surrounding real contextual environments while playing AR games on a mobile interface [13].

According to how AR apps represent virtual graphics in a real context through device camera screens, AR games are divided into 2 types: markerless AR (MAR) games and marker-based AR games [10]. Generally, marker-based AR requires a fiducial or artificial marker placed in a real context to implement AR experiences. These markers, such as printed barcodes and QR codes on cards, generally have specific geometric or color properties, which makes them easy to extract and identify in a video frame. Marker-based AR uses an external marker to calibrate the camera’s pose and then achieves a successful projection of virtual content into the captured real environment [14]. However, this approach has been shown to have many drawbacks, such as virtual content being easily lost in the tracking process if a live camera moves quickly and visual markers have to remain in sight due to a limited range [13,15]. Moreover, AR with fiducial markers is not scalable for outdoor scenes and mobile learning [16].

By contrast, the tracking system (based on a simultaneous localization and mapping technique) of MAR can rely on natural features instead of using external markers to trigger augmentations in the real environment [17]. Through an eligible mobile device, markerless tracking can manipulate any part of a real location as a marker to place virtual objects [18,19]. MAR games are similar to location-aware AR games [20], which are geo-based augmented realities that do not need any special markers for identifying where to place a virtual object in the real environment. In this way, these games can provide more realistic and interactive AR experiences for users, accurately track the real context, and detect locations through smart devices [21]. A MAR app brings a markedly different experience for its users, which enhances interactive involvement with and perception of the real surroundings without physical constrains [5,18,22]. Thus, MAR apps have been argued to provide users with more immersive and interactive experiences, which can be beneficial for promoting enjoyment and engagement in learning and training [21]. As AR technology can enable users to experience a sense of immersion [23,24], MAR game playing may provide users with a stronger sense of immersion experience.

Immersion has been widely discussed in the context of AR games [10,20], especially ones related to cognitive training [6,25]. Immersion, a form of cognitive and emotional absorption, has been used to promote enjoyment and engagement in tasks and learning [26,27]. Immersion is widely considered as a desired outcome of the gaming experience [27]. Performing and training related to immersive digital experiences are assumed to be reliant on the degree of achieved immersion, specifically the degree to which users become cognitively and emotionally engaged with a given digital app [28,29]. These statements indicate that measuring the immersion in the context of an AR game developed for cognitive training is a primary tool to evaluate the effects of AR games. Hence, to discover the potential of MAR games in the field of cognitive training, we need to investigate users’ immersion experiences in MAR games playing. However, there is an observed lack of studies investigating immersion in MAR cognitive training games.

Immersion in the field of digital games has been widely discussed in recent years [20,27,29-31]. The definition of immersion was first proposed in the field of virtual environments. The research by Georgiou and Kyza [20] argues that immersion can be seen as a reliable dimension to objectively access the properties of a virtual environment. On the other hand, Wittmer and Singer [32] described the concept of immersion as a “psychological state characterized by perceiving oneself to be enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences.” In addition, Brooks [28] argued that an immersive experience can not only occur in virtual reality (VR) technologies but also appear when playing simple computer games, such as video games. The definition of immersion also considers the different degrees of cognitive and affective absorption when playing a digital game, as they are closely related to task enjoyment and engagement [20,33]. Moreover, many studies suggest that gender affects the immersion level in digital games [34]. Specifically, men and women report different levels of immersion when interacting with virtual environments [35] and when playing video games [36].

However, the virtual environments described in the definitions were generally created by VR technology. By wearing head-mounted displays, users were immersed in virtual environments, isolating completely from real contexts. While AR games are played using mobile devices, the games can also provide users with immersive experiences by interacting with a continuous stream of digital-based stimuli situated in the real

https://games.jmir.org/2021/4/e27036
contexts. Hence, users may perceive a different immersive experience when playing AR apps, compared with virtual or other digital environments.

Previously, in the absence of valid immersion measurements, some researchers discussed AR immersive experiments through the evaluation of flow and presence [37,38]. Csikszentmihalyi [39] defined flow as “the state in which individuals are so involved in an activity that nothing else seems to matter.” As for the immersion state, flow can be described as people so absorbed in their activities that irrelevant thoughts and perceptions are filtered out from their minds [27]. The definition of presence is a psychological sense of being in a virtual environment [40]. The degree to which the virtual environment mimics real-world experiences is impacted by the degree of participants’ presence in the virtual environment [32].

Some studies argued that the sense of flow and presence could barely be created and maintained in AR games [38], as potential external distractions exist in the context of an AR system. For instance, many factors such as temperature, light and noise can be easily controlled in the virtual environment, while these parameters are hardly controlled by designers of AR apps in a real context [41]. These potential and uncontrolled external distractions may prevent users from giving their attention and thus disrupt their immersive experiences in AR games [38,41].

In addition, Benyon [42] stated that immersion in AR experiences is different from feelings of presence, in terms of being in a virtual space. Yet, immersion in AR experiences is the feeling of involvement in a blended space of real and digital elements. Moreover, a good AR game should be able to immerse its users into playing the game and decrease their focus on external distractions situated in real contexts. Hence, the evaluation of immersion provides a valuable option to describe the user experiences in the context of AR (or MAR) games.

There are several previously developed validated instruments for evaluating immersion in digital games [20,27,29,30]. Although these existing instruments are validated in some games, they may be invalidated in other types of games (eg, AR games) when measuring the immersion experience. As described above, we can see MAR games are technically and functionally different from other digital software (eg, 2D games), as they can present 3D elements and software assets in real environments without external markers.

It is uncertain which existing valid immersion questionnaire is valid for measuring user immersion in emerging MAR apps. As a potential benefit of MAR technology in developing cognitive training games, identifying valid instruments for evaluating immersion in MAR games is significant for the researchers and developers in this field. Therefore, this study aims to explore students’ immersion experiences in a novel MAR cognitive training task by using 2 existing immersion instruments for digital games.

**Methods**

**Development of a MAR n-Back Task**

An n-back task is used extensively in the literature as a working memory (WM) training task [43,44]. Participants in an n-back task are presented with a series of stimuli. They are required to retain some aspects of each stimulus in WM (eg, the location of the stimulus in a grid). During the n-back task, the trainee is instructed to respond whenever a current stimulus is presented on-screen that matches the one presented n positions back in the sequence [43]. In this study, a MAR n-back game and a 2D n-back game were developed (Figure 1) based on the same principles of the n-back task described in previous studies [45-47]. The MAR n-back game was developed using a combination of the Unity game engine and Apple’s ARKit database, based on the MAR game framework developed by Chen et al [14] that could achieve real-time 3D context reconstruction by using mobile devices. The main differences between the 2 versions of the game were how the player perceived the location of the stimuli: whether these stimuli were more stimulating graphics and whether these stimuli were perceived as being located on the screen of the device (non-MAR) or as being in real-world surroundings (MAR).

Figure 1. Examples of (a) markerless augmented reality n-back game and (b) traditional (2D) n-back game.

Figure 2 shows 6 successive trials from the MAR n-back game. Each trial consisted of a stimulus that was presented for 2 seconds (s), followed by an interstimulus interval of 2.5 s and the next stimulus [43]. In this example, with n (n=1, 2, or 3) set...
to 2 (2-back), the correct responses would be to indicate a match on trial (c) because the location matches the trial’s previous 2 steps (a), and to indicate a match on trial (f) because the location matches the trial’s previous 2 steps (d).

**Figure 2.** Six successive trials from the augmented reality version of the n-back game.

Hence, by playing the MAR n-back game, participants can interact with stimulus within the real context via smart devices. We assume that MAR n-back games could provide users with more immersive and interactive experiences, thereby being potentially beneficial for promoting enjoyment and engagement in cognitive training. As AR technology has been increasingly applied to various domains, it is urgent to investigate the essential theories and practices related to the key aspects of AR experiences, such as immersive features. In this study, both versions of the n-back games were created with 3 load levels (1-back, 2-back, and 3-back). **Figure 3** illustrates the app settings of both versions of the n-back game, while **Figure 4** displays the flowchart of a full gaming session (spanning 3 minutes). Hence, the motivation of n-back game training is based on the participants responding to the correct targets as quickly and accurately as possible, while the difficulty increases by raising the value of n. Through this way, participants were motivated to take the n-back game training seriously and perform effectively, as their scores were cumulatively displayed against other players.
Figure 3. App settings of both versions of the n-back game. MAR: markerless augmented reality.
Immersion Measurements

Background

In this study, two existing validated immersion instruments: the Immersive Experiences Questionnaire (IEQ) and the Augmented Reality Immersion (ARI) questionnaire have been selected as the measurements with some reasons.

In 2004, Brown and Cairn [30] created a famous model of immersion with the 3 sequential levels of engagement, engrossment, and total immersion to describe the degrees of involvement with digital games. The level of engagement includes access and investment, which means that the player needs to invest time and effort in learning how to play the game and get familiar with the game’s controls. From the engagement level, participants may be able to become more interested in and further involved with the game and then move into the engrossment level. During this level, gamers’ attention spans and emotions were directly affected by the game and game controls became invisible because gamers became less aware of their surroundings [27]. Total immersion occurred when gamers reached a sense of presence, became lost in the game’s world, and achieved a sense of feeling that the game was all that mattered.

On the basis of immersion model [30], flow [39], cognitive absorption [33] and presence [32], the IEQ was developed for measuring immersion in digital games [27]. In particular, the validity of the IEQ was proven through experiments with 2 different types of games: a video (2D) and a VR game. As virtual environments, although AR differs from VR in displaying digital content, both technologies are claimed to represent effective immersive technology [48]. Moreover, the focus of the IEQ covers 5 factors: cognitive involvement, emotional involvement, real-world dissociation, challenge, and control [27]. These factors are more related to the purpose of measuring immersion in a cognitive training game (eg, n-back game). Hence, using the IEQ, this study sets a similar experiment to explore differences in students’ immersion experiences with 2 different types of games: a video (2D) game and a MAR game.

The ARI questionnaire, based on the immersion theory of Brown and Cairns [30], was developed particularly for the purpose of...
measuring immersion in location-aware AR settings, which, as explained earlier, is similar to MAR games. It has also been used for measuring immersion in other types of AR apps used for training [31]. Georgiou and Kyzza [20] emphasize that immersion in AR apps is not about getting disengaged from the real world but shifting attention toward AR games, which in turn results in a decreased focus on any potential distractors. Hence, to investigate students’ immersion experiences in a MAR game, the ARI questionnaire is necessary for this study.

**Immersive Experiences Questionnaire**

The IEQ includes 33 items; 32 questions (Q1-Q32) were all designed on 5-point scales (1 = strongly disagree and 5 = strongly agree). One question used a 10-point scale, asking participants to indicate how immersed the felt overall (1 = not at all and 10 = very much).

However, as discussed in the previous section, users perceive a feeling of being surrounded by a blended world of real and digital elements in AR games, unlike virtual environments in which users can totally immerse themselves [36].

Therefore, to make the IEQ suitable for assessing immersion in AR game settings, some questions were omitted:

- Q19: It was as if I could interact with the world of the game as if I was in the real world.
- Q20: Interacting with the world of the game did not feel as real to me as it would be in the real world.
- Q23: I felt detached from the outside world.
- Q30: I still felt as if I was in the real world while playing.

Moreover, certain potential external distractions, such as weather and noises, exist in the context of an AR system. Therefore, the Q21 “I was unaware of what was happening around me” should also be removed from the IEQ [20]. Hence, a modified IEQ, with 27 5-point scale questions (omitting questions Q19, Q20, Q21, Q23, Q30) and one 10-point scale question, was used to measure participant immersion experiences in this study.

**ARI Questionnaire**

The ARI questionnaire is a 21-item, 7-point Likert-type instrument ranging from totally disagree (1) to totally agree (7) [20]. The ARI questionnaire has a 3-level construct (with each construct broken down further into 2 subcategories) according to the following: engagement (interest and usability), engrossment (emotional investment and focus of attention), and total immersion (presence and flow). These factors measure the immersion level of students while using AR apps.

**First Experiment**

**Participants**

Recruitment of participants began in October 2019. A total of 60 participants took part in the first experiment (average age 18.97 [SD 1.09] years; 42 females). They were a group of undergraduate students from a university in China. No eligibility criteria were specified. Written informed consent forms were collected from these students before they began the WM training. Participants were informed that they could drop out of the study at any time. Prior to beginning the study, ethics approval was obtained from the ethics committee at East China Normal University (HR 055-2019). The trial was registered at the University Hospital Medical Information Network Clinical Trials Registry [UMIN000045314]. All participants were paid US $15.

**Procedure**

At the beginning of the first experiment, participants were randomly divided into 2 groups, with 30 participants in the control group (who trained on the video n-back game) and 30 in the experimental group (who trained on the MAR n-back game). Participants were not blinded. Random allocation was based on random numbers generated in Microsoft Excel by BZ. During the training, six 10.5-inch iPad Pro tablets (Apple Inc) were used to present the n-back game (see Figure 5). The instructions for the n-back task were demonstrated to the participants prior to the WM training. Two groups of participants then received separate WM training in 2 versions of n-back tasks for 8 sessions (2 sessions a day, with each session lasting 3 minutes) within 4 days. All training sessions were conducted in a laboratory. At this stage, the aim of the training was not to investigate its effects on WM but to focus on comparing the level of immersion between the 2 groups. After all the training sessions were completed, each participant completed a modified IEQ through a link to an online questionnaire using their mobile phones.

Figure 5. Visual depictions of the first experiment.
Second Experiment

Participants

A total of 51 participants took part in the second experiment (average age 21.22 [SD 1.69] years; 30 females). Participants, studying in the major of Educational Technology, were recruited from another university in China. To complete the assignments on the evaluation of an innovative software, these students need to test an innovative app. All participants were fully debriefed at the end of the experiment. Written informed consent forms were also collected from these students before they completed the survey.

Procedures and Materials

In this experiment, 51 participants practiced in a MAR n-back game for once a day, 3 minutes each time, using 6 iPad tablets for 1 week. Participants were allowed to take the tablets with them outside of the classes and could play the MAR n-back game at any location within the university’s campus. The instructions for the WM training with MAR n-back game were explained to the participants and several practice trials were conducted before the initiation of the study. Some visuals illustrating the practice process are presented in Figure 6.

Results

First Experiment

Following the example of Jennett et al [27], we first examined correlations between the 2 immersion measures of 27 5-point scale questions and one 10-point scale question in the modified IEQ. Taking both groups (control and experimental groups) together, the 2 immersion measures were positively, significantly correlated (Pearson \( r=0.466, \ P<.001 \)). This finding appears reliable, as variables such as participant age, gender, and education level were all found to have nonsignificant correlations with both immersion measures. Taking the groups individually, the correlation was stronger in the MAR n-back group (Pearson \( r=0.518, \ P=.003 \)) than the 2D n-back group (Pearson \( r=0.362, \ P=.049 \)). Based on a total 28 items of the modified IEQ, while the MAR n-back group had a higher mean value on all immersion measures (see Table 1) than the 2D n-back group, the differences between the 2 groups were not significant on either measure.
Table 1. Mean and standard deviation for the measures of immersion in the modified Immersive Experiences Questionnaire.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>5-point scale (27 items)</th>
<th>10-point scale (1 item)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAR(^a) (n=30)</td>
<td>2D(^b) (n=30)</td>
</tr>
<tr>
<td>Mean</td>
<td>3.584</td>
<td>3.421</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.619</td>
<td>0.516</td>
</tr>
</tbody>
</table>

\(^a\)MAR: markerless augmented reality n-back game.
\(^b\)2D: n-back game.

Second Experiment
Cronbach alpha for the ARI scale was 0.783 overall and ranged from 0.711 to 0.836 for the 6 subdimensions (ie, interest, usability, emotional investment, focus of attention, presence and flow), indicating acceptable reliability. The mean values for engagement, engrossment, and total immersion with the subdimensions of interest, usability, emotional investment, focus of attention, presence, and flow (see Table 2) were similar to the research by Georgiou and Kyza [20] using the ARI to investigate immersion in AR experiences but noticeably lower than that found in the research of Salar et al [31].

Table 2. Mean values for interest, usability, emotional investment, focus of attention, presence and flow.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Value, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>5.853 (0.156)</td>
</tr>
<tr>
<td>Usability</td>
<td>3.995 (0.961)</td>
</tr>
<tr>
<td>Emotional investment</td>
<td>4.580 (1.697)</td>
</tr>
<tr>
<td>Focus of attention</td>
<td>4.954 (0.300)</td>
</tr>
<tr>
<td>Presence</td>
<td>4.299 (0.388)</td>
</tr>
<tr>
<td>Flow</td>
<td>4.848 (0.150)</td>
</tr>
</tbody>
</table>

However, males reported higher mean values than females on all 3 constructs of the immersion model (see Figure 7). These differences were significant in the case of engagement (Student \(t\) test \(P=0.048\), \(P<0.05\)) and total immersion (Student \(t\) test \(P=0.033\), \(P<0.05\); see Table 3 and Figure 7).

Figure 7. Mean values for engrossment, engagement, and total immersion for males and females. Significant differences (\(P<0.05\)) are marked with an asterisk (*).
It may be that for MAR apps to be more immersive, more is influenced by real-world surroundings, such as sound and light, during the MAR n-back game training in Figure 5 [38,41,42].

As Figure 1 shows, the traditional (2D) n-back game was played on the screen of a device, and the MAR n-back game was performed in the real environment. Although the MAR n-back game has more simulating graphics, both games have the same motivation rules. Moreover, the participants may be influenced by real-world surroundings, such as sound and light, during the MAR n-back game training in Figure 5 [38,41,42].

It may be that for MAR apps to be more immersive, more is required than simply translating an existing 2D game into an AR version, as we did in this study. On the other hand, participants’ scores on the modified IEQ, in both the 2D and MAR conditions, were similar to the scores in previous research using immersive experiences [27]. This suggests that participants were experiencing immersion in both versions of the n-back training games.

Furthermore, Table 2 shows that the mean values of the second experiment for 6 subdimensions with 51 students all showed relatively higher results compared with the average of the 7-point (average 3.500). We can see participants paid attention to the MAR n-back game and felt it was interesting. These results illustrate that a good level of immersion within participants has been established. However, there is still potential for improvement, as the mean value of the 6 dimensions (see Table 2) are all slightly less than those measured in the study by Salar et al [31]. At this stage, the results of using the ARI questionnaire with participants in the second experiment proved that a MAR n-back game can establish participants’ immersion experiences.

Previous studies argue that immersion is a desired outcome of the gaming experience and can promote participants’ enjoyment and engagement in tasks if their level of involvement increases in the games [27]. Performing and training related to experiences based on immersive technology are assumed to be reliant on the degree of achieved immersion in the games [28,29]. MAR apps bring a different experience for its users, which enhances interactive involvement with the real surroundings without physical constrains [5,18,22]. The findings show that the MAR n-back game (WM training task) can provide users with an immersive and interactive experience, which is beneficial for promoting enjoyment and engagement in training [21]. As the potential affordances of AR-based tasks can improve cognitive functions, such as attention and memory [6-8], this study indicates that the MAR games also have the potential to promote cognition functions (eg, WM in this study), as all the participants experienced immersion during the MAR n-back training.

Unexpectedly, the findings of the second experiment revealed that males gained higher mean values than females on all 3 constructs of the immersion model (see Figure 7), with significantly stronger values for engagement and total immersion (see Table 3 and Figure 7). The immersion theory of Brown and Cairn [30] defines that engagement is the first level of immersion, which indicate that player is willing to invest time and make effort in learning and controlling the game. With the same assignment of manipulating the MAR n-back game, we can see males were willing to make more efforts in controlling the MAR games than females at the first learning stage.

### Discussion

#### Principal Findings

Recently, many researchers have paid increasing attention to the application of AR games for cognitive development. However, with the lack of validated instruments for measuring users’ immersion in AR apps, the scientific evaluation of emerging AR games is still missing in the literature. Therefore, this study aimed to investigate students’ immersion experiences in a novel MAR n-back game for WM training by using 2 existing immersion questionnaires. As explained earlier, IEQ has the potential to compare users’ immersion with 2 different types of games [27], such as a traditional (2D) n-back and a MAR game; therefore, we used the modified IEQ as the immersion instrument in a randomized controlled experiment. This aims to explore students’ immersion experience with both versions of n-back games and investigate whether students can obtain stronger immersion in the MAR n-back game than the 2D n-back game. However, as the ARI questionnaire has been widely accepted as a valid tool for assessing immersion in the context of AR apps [20], it was used to measure only a single group of students, with the MAR n-back game in the second experiment. The purpose of this experiment was to further explore students’ immersion experiences with the MAR n-back game.

The results show that the correlation between the 2 immersion measures of 27 5-point scale questions and one 10-point scale question of the modified IEQ was positive and significant. This indicates that these 2 different sets of questions were still strongly consistent and connected with each other, which proves the reliability of the modified IEQ. The findings show that while the MAR n-back group had a higher mean value on all immersion measures (see Table 1) than the 2D n-back group, there were no significant differences in the immersion measures between groups on the modified IEQ (27 5-point scale items and one 10-point scale item). The lack of significant differences in immersion may imply that the 2 versions of the n-back games used in this research, although being visually different in many aspects, did not produce obvious different levels of immersion in players. As Figure 1 shows, the traditional (2D) n-back game was played on the screen of a device, and the MAR n-back game was performed in the real environment. Although the MAR n-back game has more simulating graphics, both games have the same motivation rules. Moreover, the participants may be influenced by real-world surroundings, such as sound and light, during the MAR n-back game training in Figure 5 [38,41,42].

Table 3. Differences between males and females on mean values for engrossment, engagement and total immersion.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Male</th>
<th>Female</th>
<th>Student t test</th>
<th>df(^a)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engrossment</td>
<td>5.565 (0.849)</td>
<td>5.108 (0.750)</td>
<td>−2.029</td>
<td>49</td>
<td>.048</td>
</tr>
<tr>
<td>Engagement</td>
<td>4.976 (0.775)</td>
<td>4.622 (1.063)</td>
<td>−1.301</td>
<td>49</td>
<td>.20</td>
</tr>
<tr>
<td>Total immersion</td>
<td>4.994 (1.307)</td>
<td>4.303 (0.950)</td>
<td>−2.189</td>
<td>49</td>
<td>.03</td>
</tr>
</tbody>
</table>

\(^a\)df: degree of freedom.

#### Differences between males and females on mean values for engrossment, engagement and total immersion.

- **Engrossment**: Males scored higher (5.565) compared to females (5.108) with a statistically significant difference (t = −2.029, df = 49, P = .048).
- **Engagement**: Similarly, males (4.976) had a higher mean value than females (4.622) (t = −1.301, df = 49, P = .20).
- **Total Immersion**: The mean for males (4.994) was also higher than females (4.303), but the difference was not statistically significant (t = −2.189, df = 49, P = .03).

### Table 3.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Male Mean (SD)</th>
<th>Female Mean (SD)</th>
<th>t</th>
<th>df(^a)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engrossment</td>
<td>5.565 (0.849)</td>
<td>5.108 (0.750)</td>
<td>−2.029</td>
<td>49</td>
<td>.048</td>
</tr>
<tr>
<td>Engagement</td>
<td>4.976 (0.775)</td>
<td>4.622 (1.063)</td>
<td>−1.301</td>
<td>49</td>
<td>.20</td>
</tr>
<tr>
<td>Total immersion</td>
<td>4.994 (1.307)</td>
<td>4.303 (0.950)</td>
<td>−2.189</td>
<td>49</td>
<td>.03</td>
</tr>
</tbody>
</table>

\(^a\)df: degree of freedom.
However, males and females experienced no significant differences in the engrossment level, which involves more emotional investment on the games. This result may be influenced by the fact that all the students were required instead of self-driven to play the MAR n-back game, hence the emotional input process is short but directly enters the total immersion level. The results also show that male students researched significantly stronger sense of becoming lost in the MAR games (total immersion level) than female students. These results coincide with the finding that male players experienced deeper involvement with the AR environment, which in turn resulted in feeling higher levels of immersion than female players [34].

In conclusion, by conducting 2 experiments by using the modified IEQ and ARI questionnaires, we found that both groups of students were experiencing immersion in the MAR n-back game. As discussed in the Introduction section, MAR games are very different from other digital games as they can present 3D elements and software assets in real environments without external markers. However, there is an observed lack of studies on investigating immersion in MAR games. This study shows that the modified IEQ and ARI questionnaire have the potential to be used as an instrument to measure immersion in MAR settings. Also, as results show that MAR cognitive games can provide users with an immersive and interactive experience, this suggests the affordance of MAR games in improving users’ cognitive functions. However, the MAR n-back game did not produce obvious different levels of immersion in players than the traditional (2D) n-back game in this study, this may imply that the design of the MAR games should have obvious differences from 2D versions or other type of games in order to increase immersion levels. Finally, we found that male players experienced stronger immersion than female players in the MAR n-back game, which further supports that males and females perceive different levels of immersion when interacting with virtual environments [36]. In the end, we believe these findings can contribute to the development of MAR games in the field of cognitive training for AR designers and also for the researchers who are interested in exploring users’ immersion in MAR settings.

Limitations and Future Study

Although this study has a number of practical implications, there are a number of limitations that must be considered. First, in both experiments, the immersion questionnaire was completed by participants after the practice sessions were completed. Although students finished immersion questionnaires immediately after the game practices, their immersion experiences may not be accurate, as immersion is an instantaneous state that would disappear after the activity is finished. Second, in the second experiment, only a single group of students completed the ARI questionnaire to measure their immersion with the MAR n-back game. Further, a randomized controlled experiment needs to be considered to explore whether the ARI questionnaire can be used to measure immersion within the non-AR games. Third, an experiment with the modified IEQ should be repeated with an app that really differs in an AR version from its 2D version game to see whether still no obvious different levels of immersion in players. It is important to further investigate the affordances of AR technology in building games. Finally, as only one MAR game has been tested using the modified IEQ and ARI questionnaires in this study, more MAR games need to be tested in the future to strengthen the findings of this study.

Conclusion

In this study, we used two questionnaires, the IEQ and ARI, to investigate immersion in an AR working memory game. In addition, we compared immersion levels between the AR game and a 2D version of the same game. We found that the AR working memory game produced reasonable levels of immersion in players (on both questionnaires), which may contribute to its effectiveness as a cognitive training program. However, there were no significant differences between immersion levels in the AR game and the 2D version (on the IEQ). We also found different levels of immersion experience in men and women on two constructs of the ARI, with men showing significantly higher levels of engagement and total immersion than women in the AR game.

Acknowledgments

This study was sponsored by grant CCA210256 from the National Youth Project of National Education Sciences Planning Project: the Intervention Study of Markerless Augmented Reality Cognitive Training Game on Working Memory of Children with Autism Spectrum Disorder in China. The authors would like to thank the reviewers for providing valuable comments that helped us to greatly improve this manuscript.

Authors’ Contributions

BZ designed and conducted the study, contributed to the design of the n-back games, collected the data, performed statistical analyses, and wrote the manuscript. NR advised on the design of the n-back games, assisted with analyses, and read and approved the final manuscript.

Conflicts of Interest

None declared.
This randomized study was only retrospectively registered, explained by authors as follows: "The funders of the research ... did not require the registration." The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1074 KB - games_v9i4e27036_app1.pdf ]

References

https://games.jmir.org/2021/4/e27036


Serious Games Are Not Serious Enough for Myoelectric Prosthetics

Abstract

Serious games show a lot of potential for use in movement rehabilitation (eg, after a stroke, injury to the spinal cord, or limb loss). However, the nature of this research leads to diversity both in the background of the researchers and in the approaches of their investigation. Our close examination and categorization of virtual training software for upper limb prosthetic rehabilitation found that researchers typically followed one of two broad approaches: (1) focusing on the game design aspects to increase engagement and muscle training and (2) concentrating on an accurate representation of prosthetic training tasks, to induce task-specific skill transfer. Previous studies indicate muscle training alone does not lead to improved prosthetic control without a transfer-enabling task structure. However, the literature shows a recent surge in the number of game-based prosthetic training tools, which focus on engagement without heeding the importance of skill transfer. This influx appears to have been strongly influenced by the availability of both software and hardware, specifically the launch of a commercially available acquisition device and freely available high-profile game development engines. In this Viewpoint, we share our perspective on the current trends and progress of serious games for prosthetic training.

Background

Adherence of patients to interventions (eg, home exercises) remains a key challenge in rehabilitation medicine [1]. Patients complain that exercises often feel tedious and tiring and that progress, if any, is slow and incremental [1]. Delivering virtual training in the form of games can help overcome issues related to nonadherence (or noncompliance) of patients to their exercise regimen [1]. The use of serious games has been recommended to motivate patients in performing their prescribed exercises consistently and completely [2-4].

The stroke rehabilitation literature includes a large number of publications that use serious games. Koutisiana et al [5] identified 96 publications between the years 1999 and 2019. The serious games used in stroke rehabilitation are showing significant benefits for the users, most notably an increased number of repetitions performed, which is a prime goal for this kind of rehabilitation [6]. Supported by this academic evidence, rehabilitation programs like Reliability (Imaginary srl), which has grown out of the Rehab@Home project [7], are being incorporated in clinical practice.

Although serious games have found their way into a multitude of areas of everyday life, industry, and research, including prosthetic training [8], academic results supporting the efficacy of serious games in myoelectric prosthetic training are scarce, if not nonexistent. Using games in virtual rehabilitation has...
been a part of research for 30 years [9], but they have only gained proper traction in the field in the last decade. This rise coincided with the commercialization of a range of game-related technologies (eg, motion tracking cameras and game controllers with inertial measurement sensors [6]). In this paper, we will offer our perspective on the efficacy of virtual training in general and serious games specifically for myoelectric prosthetics training.

Current research claims that the use of serious games in myoelectric prosthetics training has promise to improve training. Examples include faster learning [10], reduction of fatigue and irritation while training [11], and increased muscle control [12]. In addition, serious games can offer a faster route to myoelectric training after limb loss [11], as a game would likely not rely on socket fitting or full wound closure. Furthermore, it can make the training more enjoyable and engaging [11], as well as affordable and accessible for the home environment [10]. It also has the potential to assist the user with their body image [11], decrease phantom limb pain [11], and let the user feel more in charge of their own rehabilitation [8,10], while at the same time make it feel less like rehabilitation [8].

The prevailing view is that this combination of positive effects has the potential to significantly add to the existing prosthetic training and lead to a reduction in prosthesis abandonment, which has been linked to a lack of motivation and engagement [13] and poor training [10]. The performance of virtual prosthetic training at home can also offer benefits to the therapists. As a supplement to existing training regimes, it can offer an objective measure of how diligently the patient is doing their exercises at home and of their improvements [10]. It also has the potential to decrease rehabilitation times and the time necessary for each patient, thereby reducing the workload for therapists [8].

We investigated papers that included any virtual training or assessment for upper limb prosthesis control using myoelectric signals as input. The included papers were identified during investigation of the literature and has been augmented with systematic searches in multiple databases, including PubMed, Web of Science, and Google Scholar. This led to the inclusion of 55 journal articles and conference papers, with a total of 59 different virtual training programs. CAG classified these programs into two categories, namely serious games and simulators (Table 1), according to Narayanasamy et al [14]. Both training simulators and serious games are interactive simulations in a virtual environment with the purpose of skill development. Simulators often duplicate real-world scenarios, require standard operational procedures, are not designed for entertainment, have no secondary purpose, and usually do not have an obvious final state. Conversely, a serious game is set in a fictitious scenario, provides various challenges, allows for entertainment, and allows the user to develop gameplay patterns while trying to achieve game-specific goals. This can include an end state. Therefore, some of the programs are classified as “simulators,” even when the authors identified them as “games.” The programs were further classified by the type of task the user was given, the type of control scheme the program used, and the input and output devices that were used. This more detailed table can be found in Multimedia Appendix 1.
<table>
<thead>
<tr>
<th>Names</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Serious games</strong></td>
<td></td>
</tr>
<tr>
<td>Air-Guitar Hero (rhythm game)</td>
<td>[15]</td>
</tr>
<tr>
<td>WiiEMG (sports game)</td>
<td>[16]</td>
</tr>
<tr>
<td>Sonic Racing (racing game)</td>
<td>[17]</td>
</tr>
<tr>
<td>MyoBox (dexterity game)</td>
<td>[18,19]</td>
</tr>
<tr>
<td>Flappy Bird (sidescroller)</td>
<td>[20]</td>
</tr>
<tr>
<td>Space Invaders (fixed shooter)</td>
<td>[20]</td>
</tr>
<tr>
<td>MyoBeatz (rhythm game)</td>
<td>[21]</td>
</tr>
<tr>
<td>Falling of Momo (vertical scroller)</td>
<td>[22-25]</td>
</tr>
<tr>
<td>Volcanic Crush (reaction game)</td>
<td>[10]</td>
</tr>
<tr>
<td>Dino Sprint (endless runner)</td>
<td>[10]</td>
</tr>
<tr>
<td>Dino Feast (dexterity game)</td>
<td>[10]</td>
</tr>
<tr>
<td>Space ARMada (fixed shooter)</td>
<td>[11]</td>
</tr>
<tr>
<td>SuperTuxKart (racing game)</td>
<td>[2,12,26,27]</td>
</tr>
<tr>
<td>Step Mania 5 (rhythm game)</td>
<td>[2,12,26,27]</td>
</tr>
<tr>
<td>Pospos (dexterity game)</td>
<td>[2,12,26,27]</td>
</tr>
<tr>
<td>Who nose?/Nose Picker (simple game)</td>
<td>[28,29]</td>
</tr>
<tr>
<td>Smash Bro/Bash and Debris (sidescroller)</td>
<td>[28,29]</td>
</tr>
<tr>
<td>Sushi Slap (action game)</td>
<td>[28,29]</td>
</tr>
<tr>
<td>Crazy Meteor (multidirectional shooter)</td>
<td>[28,29]</td>
</tr>
<tr>
<td>Dog Jump/Beeline Border Collie (sidescroller)</td>
<td>[28,29]</td>
</tr>
<tr>
<td>Breakout-EMG (arcade game)</td>
<td>[30]</td>
</tr>
<tr>
<td>Training Game Prototype</td>
<td>[31]</td>
</tr>
<tr>
<td>Dino Claw (dexterity game)</td>
<td>[10]</td>
</tr>
<tr>
<td>Training, TAC&lt;sup&gt;c&lt;/sup&gt; test, and Crossbow Game</td>
<td>[32]</td>
</tr>
<tr>
<td>UpBeat (rhythm game)</td>
<td>[4]</td>
</tr>
<tr>
<td>Rhythm Game&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>[13]</td>
</tr>
<tr>
<td>Crate Whacker (tech demo)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[33]</td>
</tr>
<tr>
<td>Race the Sun (endless runner)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[33]</td>
</tr>
<tr>
<td>Fruit Ninja (dexterity game)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[33]</td>
</tr>
<tr>
<td>Kaiju Carnage (action game)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[33]</td>
</tr>
<tr>
<td><strong>Simulators</strong></td>
<td></td>
</tr>
<tr>
<td>UVa Neuromuscular Training System</td>
<td>[34,35]</td>
</tr>
<tr>
<td>Commercial software PAULA</td>
<td>[36]</td>
</tr>
<tr>
<td>Virtual training</td>
<td>[36]</td>
</tr>
<tr>
<td>Virtual training environment</td>
<td>[37]</td>
</tr>
<tr>
<td>Mixed reality training&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[38]</td>
</tr>
<tr>
<td>Virtual box and beans test&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[39]</td>
</tr>
</tbody>
</table>
Different Approaches

The categorization of the publications in this field and the software presented therein has shown a significant split of the approaches of researchers into roughly two groups, as can be seen in Figure 1. This divide is most noticeable with regard to whether the software is classified as a game or a simulator and which type of tasks are implemented. The first approach focuses on the engagement and motivation of the user and seems to have grown in popularity in recent years. Researchers develop serious games that often have an explicit or implicit focus on game design elements to keep the user engaged in the game and therefore in the rehabilitation or training. The majority of these myogames (21/30) incorporate abstract tasks not resembling a real-life scenario. These games attempt to train the user in the use of a myoelectric prosthesis by focusing on different aspects of muscle control, including proportional control, independent control, and others. Only two games feature a task that is somewhat activities of daily living–related, both consisting of variants of a pick-and-place task, one stationary [10] and one moving in a 3D environment [31]. In a further seven games, the user is tasked with reproducing specific postures in two rhythm games [4,13], a virtual reality crossbow game [32], and four open-source games [33].
Publications introducing and assessing these games show that they are engaging and enjoyable to the participants. Some studies involving people with limb difference showed their willingness to use them in a home environment [8,21,27]. With regard to skill acquisition, a general increase in in-game performance is shown for a number of these games [12,25,26]. However, it was rarely tested whether these myogames increase prosthetic ability or speed up the learning process of acquiring that skill. The one research group that tested for an increase in prosthetic ability did not find evidence of a significant increase following the playing of a myogame for different control schemes [19,30].

The second approach focuses on skill transfer and therefore involves the development and investigation of simulators that mostly show the user a representation of the real world and require the performance of activities of daily living–related or posture reproduction tasks. Only two training programs that were classified as simulators used abstract tasks; these tasks were embedded in a sterile software environment and lacked distinctive game traits. In this type of research, the focus is on the effectiveness of the skill transfer from the virtual training to actual prosthetic ability. The prescribed tasks can involve recreations or tasks inspired by tests used in the assessment of prosthetic ability, like the Southampton Hand Assessment Procedure (SHAP) test [68], the Target Achievement Control (TAC) test [69], and others. The focus on task specificity for learning prosthetic skills seems like a promising approach as the results of one study indicated that skill transfer occurred. Performing a virtual task resembling the control of a prosthetic hand led to an increase in prosthetic ability [49] as opposed to when the task was to play a classic arcade game [30]. The task specificity therefore seems to have an influence on the effectiveness of virtual training; however, further research must be done to substantiate this.

The effectiveness of virtual training in increasing prosthetic ability is without doubt one of the necessary requirements for any adoption into clinical rehabilitation; however, consensus on a universal measure of effectiveness is not available. A criticism of the myogames in the game-focused group is that they work on the implied assumption that an improvement in skill performing any myoelectric task will lead to an improvement in prosthetic skill [49]. Although it has been shown that the user increases their skill in different aspects on the muscular level [10,25], it is not clear whether that influences the way or speed at which a person might acquire prosthetic skill.

**Other Influences**

Figure 2 shows another interesting development regarding the first research approach. It clearly features a joint spike in more recent years in the development of serious games incorporating an abstract task and presented in a nonimmersive environment using traditional media. The development of simulators and software using other task types and environment experiences has remained comparatively steady over the same time frame. The start of this spike in publications coincides with the release of the Myo Gesture Control Armband (Thalmic Labs), a dry surface electrode armband, on the commercial market in the year 2015 [70]. The spike in publications started to decrease when the company stopped selling this product in 2018. However, even though it is no longer sold, the Myo armband is still in widespread use in research as there is currently no
commercial alternative available. In total, 30% of the programs analyzed use the Myo armband and the most recent work using it was published in 2021 [61]. The uptake in development has likely been further boosted due to a few professional commercial game development engines being free for personal and low-profit use alongside the provision of an application programming interface (API) for the Myo Armband; there are numerous open-source examples of the use of this API in these engines. The main example of one of these engines is the Unity game engine (Unity Technologies) made freely available in the year 2009 [71]; this game engine has been used in at least one-third of the published programs. Figure 2. Number of training software introduced by task, software, and environment type. ADL: activities of daily living; AR: augmented reality; VR: virtual reality.

These two factors indicate that the development of serious games intended for prosthetic training was strongly influenced by the emergence of readily available software and hardware technology. However, the enthusiastic embrace of the newly available technology tended toward research exploring the engagement aspects of game design. This is likely due to the low barrier to entry of this approach as there are a multitude of resources for game development available and the study of engagement does not require the involvement of people with limb difference. Investigation of these serious games confirmed that people are more willing to engage in learning a task if it is an intrinsically enjoyable and motivating experience [2]. Such research has also shown that with these games, participants are able to quickly master fine control of their muscles [10,25]. However, this research often tacitly assumed an efficacy in skill transfer by this virtual muscle training, which has yet to be substantiated. As such, it is not clear whether this increase in motor control would lead to enhanced prosthesis control and which types of games might be more conducive to learning how to use a prosthetic device. Therefore, at this point in time, serious games are not serious enough to train upper limb prosthesis use effectively.

**Where Do We Go From Here?**

In the research targeting other conditions, such as stroke rehabilitation, the main target is to get the user to move their respective body part more to regain a substantial degree of control over it. The reason for the strong focus on the engagement and motivation of users to increase repetitions of a movement is therefore clear. However, using a myoelectric upper limb prosthesis requires the user to acquire a completely new set of skills. This can mean to either retrain or newly train muscles and their associated uses, depending on whether the limb difference is acquired or congenital. Therefore, a necessary requirement for a serious game in this field to be considered for clinical adoption would be evidence of a benefit to prosthetic ability (ie, evidence that the skill learned in the game transfers to the use of an actual myoelectric prosthesis). So far this kind of skill transfer has only been shown for software that we classified as simulators. It is hypothesized that the task specificity of the actions performed virtually allows the transfer to the real world to occur [49].

Research in this field needs to establish viable paths for transfer to occur before focusing on the topic of engaging and motivating the target user group. Serious games intended for prosthetic training need to show their benefit for prosthetic ability, be it direct or indirect. Hence, a sensible approach for the development of such a serious game could be to first demonstrate which types of tasks allow transfer at all and then to develop the engaging and motivating game structure around it. As with other conditions, researchers employ the attractiveness of games to actively engage users; however, the clinical benefit cannot be neglected or compromised. The two different approaches in this field encourage separate habit loops when they should merge and form a single loop more beneficial for the user, as shown in Figure 3. Engagement should not be viewed separately but in conjunction with transfer-enabling
training to enhance the habit formation of the user to adhere to the training.

Therefore, the research should establish one or multiple tasks that either directly enable skill transfer to prosthetic use or show evidence of supporting the acquisition of prosthetic skill. Built on these tasks, an engaging and motivating platform should be implemented, which should enable users to increase their prosthetic skill while having fun. The positive reinforcement of the skill increase combined with the fun experienced while training should have a positive effect on adherence to the training and therefore on the long-term success of the intervention.

*Figure 3.* Diagram of the existent and recommended habit loops.

Furthermore, the effect of the introduction of myogames on therapists’ workloads should be determined; this is highly dependent on the nature of the training and whether there is a need for the direct involvement of the therapist, which could potentially result in a similar or even larger workload [72]. As this factor depends strongly on the design of the program, it emphasizes the importance of smart development including the input and feedback of all parties involved, including clinicians and therapists, to lead to a product that benefits everyone.

In conclusion, research on prosthetic training has confirmed that myoelectric skills can be acquired with serious games. However, for the development of a viable serious game intended for prosthetic training, it is necessary to validate the “serious” part of the game, namely the tasks that would allow for skill transfer. Serious games for prosthetic training can only be expected to yield fruitful results beyond engagement when they incorporate tasks that are found to facilitate prosthetic skill. We recommend that the research community investigates which types of myogame tasks might facilitate transfer, as the only existing results at the time of writing this paper indicate a lack of effectiveness [19,30,73]. This lack does not necessarily hold true for all tasks that are not related to activities of daily living, however, and ignoring abstract tasks entirely would exclude a wide range of possible avenues for prosthetic game development.

It would be beneficial to be more accurate regarding the terminology used in the field and, if the term “game” is used, to specify the incorporated game design elements explicitly. More long-term and ideally home-based experiments are needed to conclusively test for any prosthetic skill transfer that might occur with the consistent use of prosthetic gaming devices. Even though previous studies indicate that no change in prosthetic ability occurs after training with a myogame [19,30,73], these only tested the effect of comparatively short training sessions with able-bodied people or very small groups of prosthesis users. It should also be tested whether prosthetic gaming has the potential to support traditional prosthetic training by allowing for supplementary practice sessions between visits to the prosthetist.
Acknowledgments

This work was supported by the Leverhulme Doctoral Scholarship Programme in Behaviour Informatics (DS-2017-015), the National Institute of Health Research (NIHR) via Devices 4 Dignity Starworks STWK-006, and the Engineering and Physical Sciences Research Council (EPSRC) via grant EP/R004242/1.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed categorization of the virtual training programs.

[PDF File (Adobe PDF File), 39 KB - games_v9i4e28079_app1.pdf]

References


33. Hashim N, Abd Razak NA, Gholidzadeh H, Abu Osman NA. Video Game-Based Rehabilitation Approach for Individuals Who Have Undergone Upper Limb Amputation: Case-Control Study. JMIR Serious Games 2021 Feb 04;9(1):e17017 [FREE Full text] [doi: 10.2196/17017] [Medline: 33538698]


41. Dhawan D, Barlow M, Lakshika E. Prosthetic rehabilitation training in Virtual Reality. In: 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH). 2019 Presented at: 2019 IEEE 7th International Conference on Serious Games and Applications for Health (SeGAH); August 5-7, 2019; Kyoto, Japan. [doi: 10.1109/segah.2019.8882455]

42. Nissler C, Nowak M, Connan M, B...


70. Thalmic Labs Myo armband hits consumer release, for sale on Amazon. IT Business Canada. URL: https://www.itbusiness.ca/news/thalmic-labs-myo-armband-hits-consumer-release-for-sale-on-amazon/54056 [accessed 2021-02-16]

71. Unity 2.6 Released And Now Free!. Unity Technologies. URL: https://unity.com/our-company/newsroom/unity-2-6-released-and-now-free [accessed 2021-02-16]

Abbreviations

**EMG:** electromyography
**SHAP:** Southampton Hand Assessment Procedure
**TAC:** Target Achievement Control

---

©Christian Alexander Garske, Matthew Dyson, Sigrid Dupan, Graham Morgan, Kianoush Nazarpour. Originally published in JMIR Serious Games (https://games.jmir.org), 08.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Original Paper

Dissecting Digital Card Games to Yield Digital Biomarkers for the Assessment of Mild Cognitive Impairment: Methodological Approach and Exploratory Study

Karsten Gielis¹, MSc, PhD; Marie-Elena Vanden Abeele², MD; Robin De Croon³, MSc, PhD; Paul Dierick⁴, MA, PhD; Filipa Ferreira-Brito⁵, MA; Lies Van Assche⁵,⁷, MA, PhD; Katrien Verbert³, MA, PhD; Jos Tournoy⁸,⁹, MD, PhD; Vero Vanden Abeele¹, MA, PhD

¹e-Media Research Lab, Katholieke Universiteit Leuven, Leuven, Belgium
²Memory Clinic, Jessa Hospital, Hasselt, Belgium
³Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium
⁴Department of Gerontopsychiatry, University Psychiatric Center, Duffel, Belgium
⁵Instituto de Saúde Ambiental, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal
⁶Section of Geriatric Psychiatry, University Hospital Leuven, Katholieke Universiteit Leuven, Leuven, Belgium
⁷Department of Psychiatry, University Hospital Leuven, Katholieke Universiteit Leuven, Leuven, Belgium
⁸Department of Geriatric Medicine, University Hospital Leuven, Leuven, Belgium
⁹Department of Public Health and Primary Care, Gerontology and Geriatrics, Katholieke Universiteit Leuven, Leuven, Belgium

Corresponding Author:
Karsten Gielis, MSc, PhD
e-Media Research Lab
Katholieke Universiteit Leuven
Andreas Vesaliusstraat 13
Leuven, 3000
Belgium
Phone: 32 16376866
Email: karsten.gielis@kuleuven.be

Abstract

Background: Mild cognitive impairment (MCI), the intermediate cognitive status between normal cognitive decline and pathological decline, is an important clinical construct for signaling possible prodromes of dementia. However, this condition is underdiagnosed. To assist monitoring and screening, digital biomarkers derived from commercial off-the-shelf video games may be of interest. These games maintain player engagement over a longer period of time and support longitudinal measurements of cognitive performance.

Objective: This paper aims to explore how the player actions of Klondike Solitaire relate to cognitive functions and to what extent the digital biomarkers derived from these player actions are indicative of MCI.

Methods: First, 11 experts in the domain of cognitive impairments were asked to correlate 21 player actions to 11 cognitive functions. Expert agreement was verified through intraclass correlation, based on a 2-way, fully crossed design with type consistency. On the basis of these player actions, 23 potential digital biomarkers of performance for Klondike Solitaire were defined. Next, 23 healthy participants and 23 participants living with MCI were asked to play 3 rounds of Klondike Solitaire, which took 17 minutes on average to complete. A generalized linear mixed model analysis was conducted to explore the differences in digital biomarkers between the healthy participants and those living with MCI, while controlling for age, tablet experience, and Klondike Solitaire experience.

Results: All intraclass correlations for player actions and cognitive functions scored higher than 0.75, indicating good to excellent reliability. Furthermore, all player actions had, according to the experts, at least one cognitive function that was on average moderately to strongly correlated to a cognitive function. Of the 23 potential digital biomarkers, 12 (52%) were revealed by the generalized linear mixed model analysis to have sizeable effects and significance levels. The analysis indicates sensitivity of the derived digital biomarkers to MCI.
Conclusions: Commercial off-the-shelf games such as digital card games show potential as a complementary tool for screening and monitoring cognition.

Trial Registration: ClinicalTrials.gov NCT02971124; https://clinicaltrials.gov/ct2/show/NCT02971124

(KEYWORDS) mild cognitive impairment; Klondike Solitaire; card games; generalized linear mixed effects analysis; expert study; monitoring; screening; cognition; dementia; older adults; mobile phone

Introduction

Assessing Cognitive Performance

Mild cognitive impairment (MCI) is a clinical entity defined as a transitional state between normal and pathological aging, where one or more cognitive domains are significantly impaired, yet activities of daily living are still preserved [1]. Early detection of MCI is important for signaling possible prodromes of dementia, monitoring the progression of possible decline, taking supportive measures, and detecting any possible underlying causes. However, cognitive impairment is still underdiagnosed [2-4]. In response, governmental bodies have called for novel, scalable, and longitudinal tools to assist in the early screening and monitoring of dementia [5-7]. To answer this call, researchers have explored the use of digital games as a suitable medium for assessing cognitive impairment [8-11]. Games are autotelic in nature, tapping into the intrinsic motivation to play [12,13], hence captivating a player’s attention. Furthermore, digital games are a natural source of information on player behavior, cognitive performance, motor skills, social conduct, and affective experiences [14].

As such, digital games may help by providing digital biomarkers of cognitive performance. Biomarkers, defined as “objectively measured and evaluated indicators of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention [15]” have a longstanding tradition in dementia research [16,17]. Complementary to their biological counterparts, digital biomarkers are “user-generated physiological and behavioral measures collected through connected digital devices to explain, influence and/or predict health-related outcomes [18].” User interaction with digital games produces dense and detailed behavioral traces that may inform on the users’ social health, praxis, and cognition.

Today, the focal point of research assessing cognitive performance is serious games, that is, games intentionally designed and built for a serious purpose and not solely to entertain [19]. Although serious gaming interventions show potential, they are typified by less funding, shorter development cycles, and missing know-how compared with traditional video games, which affect the in-game hallmarks of quality, such as graphics, music, and storytelling [20,21]. This may lead to frustrating player experiences, a lack of engagement, and less attention being paid during gameplay, which may lower the reliability and validity of any findings and possibly cause attrition in longitudinal studies [22-24]. Therefore, most recently, the study by Mandryk and Birk [14] argued in favor of turning to commercial off-the-shelf (COTS) video games instead [14]. Instead of spending limited resources on building a serious game, researchers can devote themselves to investigating existing games already enjoyed by the target population. Although not designed to measure cognition, COTS games are woven into the fabric of everyday life and may be able to provide digital biomarkers of cognitive performance that are reflective of cognitive status [14,25].

This study aims to explore the possibility of using COTS card games to screen cognition among patients living with MCI. This study involved 11 experts in the domain of MCI coming together to craft 23 candidate digital biomarkers from the digital card game Klondike Solitaire. Subsequently, a data acquisition campaign was set up involving 46 participants: 23 (50%) healthy older adults and 23 (50%) older adults with MCI. The participants were asked to play 3 games on a tablet. We examined the game data for differences at a group level for the candidate digital biomarkers using a generalized linear mixed model (GLMM) analysis. The results show that of the 23 candidate digital biomarkers, 12 (52%) differed significantly between both groups, taking age, tablet experience, and performance into account. By providing a methodological approach and an exploratory study for crafting digital biomarkers, as well as articulating the rationale and the different steps taken, we hope to inform future researchers who aim to leverage the use of COTS video games to yield digital biomarkers.

MCI Classification

Persons diagnosed with MCI show a deficit in cognition in at least one cognitive domain that cannot be attributed to age or any other disease; yet, they do not fulfill the diagnosis of dementia [26]. Persons with MCI, however, have a higher risk of progressing to a form of dementia such as Lewy body dementia [27], vascular dementia [28], or the most common form of dementia, Alzheimer disease [29]. Depending on the early symptoms, persons with MCI can be classified into 2 groups: amnestic MCI (aMCI) and nonamnestic MCI (naMCI). Persons in the aMCI group show a significant memory deficit, whereas in persons with naMCI, mainly a nonmemory impairment (eg, language) is present [30]. For both aMCI and naMCI, a further distinction can be made between persons with 1 cognitive domain impaired (single domain MCI) and those with multiple cognitive domains impaired (multiple-domain MCI). Although no treatments exist with the current state of modern medicine to cure the neuronal damage of these progressive forms of dementia [31,32], early diagnosis matters [33] because there are several measures that can be taken to slow down disease progression [34], including starting nonpharmacological treatment for delaying symptoms.
Detecting MCI

Typically, the process leading to a diagnosis of MCI is set into motion by a cognitive complaint from the older adult, relative, or (informal) caregiver, followed by a presumptive identification through a screening test. The cognitive screening tests used most often to detect MCI are the Montreal Cognitive Assessment (MoCA) [36] and the Mini-Mental State Examination (MMSE) [37]. These neuropsychological assessment tools are designed to evaluate cognitive functions and level of impairment more thoroughly. In addition, the assessment may include a semiguided interview with a relative or caregiver to evaluate the change in symptoms over time, such as in the Clinical Dementia Rating (CDR) Scale [40]. However, this neuropsychological assessment is laborious and time-intensive, requiring skilled test administrators, who, despite their training, are still subject to interassessor variability [41]. In addition, from a patient perspective, the process has been described as bewildering, highly stressful, and even humiliating [42,43], contributing to malperformance. This in turn can make patients self-aware of impairment, leading to feelings of distress or helplessness, possibly spiraling into even worse performance [44,45]. Although biological and imaging biomarkers are becoming more common to support diagnosis, they remain expensive and invasive, making them equally unfit for high-frequency measurements [41]. As a result, health professionals and policy makers welcome additional tools that support the monitoring of cognition [1,46-49] while reducing patient-level barriers and are more considerate of patients’ experiences [44].

Serious Games for the Assessment of Cognitive Functions

Serious (digital) games are “games that do not have entertainment, enjoyment, or fun as their primary purpose [19].” An early and longstanding tradition [50] is the use of serious games for cognitive evaluation [51]. Space Fortress [19,50] is perhaps the first research game to measure and train cognitive control and related cognitive functions. Ever since, the popularity of creating serious games and game-based interventions to measure, detect, and train cognition has only increased, as indicated by systematic reviews on this topic by Ferreira-Brito et al [52], Lumsden et al [53], and Valladares-Rodriguez et al [51].

Serious games may provide certain advantages for the assessment of cognitive performance compared with standard cognitive assessment tests. First, by offering an interactive and immersive audiovisual experience, serious games can be considered to be more engaging than classical tests [23,51,53]. As ensuring the full attention of the participant is paramount in neuropsychological testing, such increased engagement may also result in more reliable research results; previous research has linked effort to cognitive test performance in healthy undergraduate students [54]. Second, games allow for embedding of cognitive tasks in a virtual, audiovisual world, which more closely mimics the actual lived-in world, allowing for better transfer of task results and providing higher ecological validity [55]. However, it has to be noted that the skills learned through serious games might still be difficult to generalize to the skills needed in a real-life context [55]. Third, serious games can be designed in such a manner that they minimize the need for the presence of a trained administrator. Setting a pace, reading out loud, or cueing can be integrated into the game itself. In this manner, test administration bias is reduced, and white-coat effects can be minimized [56,57]. If assessments are possible with less supervision and manual effort, they also become more scalable because testing becomes less resource intensive [55]. However, this lack of supervision while performing measurements has an important caveat. When measurements are performed in a personal environment, it becomes more difficult to prevent distractions that influence gameplay behavior.

Although serious games show promising results and have merit for both patient and physician, serious games are at risk of being dismissed as “chocolate-covered broccoli”: neuropsychological tests embellished with a thin layer of gameplay [58-60]. This can lead to games that are suboptimal in terms of esthetic quality and game mechanics [20], negatively affecting gameplay [58]. A meta-analysis of serious games [22] shows that although they can be more effective and improve retention compared with conventional methods, they are not found to be more motivating. Similar signs of lack of motivation have been noted in game-based interventions designed to train cognitive functions [23,24].

This lack of sustained engagement contrasts with surveys on gameplay among older adults. A large-scale (N=3737) survey of older adults’ attitude toward video games, conducted in 2019 by the American Association of Retired Persons [61], highlights that older adults enjoy playing digital games. Of the nine reasons to play, to have fun was indicated to be the top reason (78%) to play video games, with to help stay mentally sharp coming in second (69%). In the 70-years-or-older age category, this difference became marginal, with 73% indicating to have fun and 72% indicating to stay mentally sharp as the reasons to play. Therefore, to increase engagement and to tap into intrinsic motivation, popular COTS video games may present an interesting alternative. These games are already woven into the daily life of the older adult, providing meaningful play [62,63].

COTS Video Games for Mental Health

COTS games may have the ability to retain players over a longer period and to support continuous measurements of cognitive performance. As frequent measurements are more sensitive to detecting small deviations in the cognitive performance of older adults [64], this could lead to a better interpretation of the patient’s cognitive trajectory. Furthermore, fluctuations in
cognitive performance [65], a common feature of dementia, may be more easily detected. In addition, this continuous monitoring enables establishing an intraindividual cognitive baseline [66]. This cognitive baseline can be used to compare patients with themselves, as opposed to comparisons with normed references. In turn, this can lead to improved management and care [1]. Nevertheless, a prominent disadvantage of COTS games is that researchers have less control over which cognitive functions are measured in the game [60].

Recent studies on using COTS games to measure cognitive impairment have generated promising results. The study by Jimison et al [8] used FreeCell, a Solitaire variant, to compare cognitive performance between a group of people living with MCI and a healthy control group by means of an optimized solver. The results indicated that based on gameplay, the group with MCI could be discerned from the healthy control group [8]. In the case of sudoku, another popular game among older adults, the study by Grabbe [67] showed that performance in the game was significantly related to measures of working memory. Using a set of smartphone-based puzzle games, which also contained sudoku, the study by Thompson et al [68] explored smartphone-based games as a means of portable cognitive assessment and monitoring. The participants’ performance in these games correlated to several measures of cognition, among which were visual memory, verbal learning, and reasoning. Finally, the study by Wallace et al [11] developed a word search game and sudoku above other games because of certain properties such as the percentage of successful deals. The study by Thompson et al [68] chose games based on face validity with regard to targeting cognitive functions. Although these reasons are valid arguments for choosing a game, it should be noted that these studies have no arguments rooted in empirical evidence for their game of choice.

**Klondike Solitaire**

One of the most popular card games among older adults is Klondike Solitaire, also known as Patience, Fascination, or even just Solitaire [70]. The popularity of Klondike Solitaire among older adult gamers was recently noted in the study by Boot et al [71]. For 1 year, participants had access to a computer where 11 games were installed, among which were sudoku, Klondike Solitaire, and crossword puzzles. The study noted that “There was a strong, clear preference for Solitaire […] After Solitaire, there was no clear second choice, and on average participants infrequently played the other games.” In addition, the results showed that of all 11 games, 1 (9%)—Solitaire—was being played most consistently.

This popular card game is played with a standard 52-card deck, with 28 (54%) cards dealt in 7 build stacks and the other 24 (46%) cards placed in a pile, as can be seen in Figure 1. The goal of the game is to order all cards from the ace to the king on the 4 corresponding suit stacks. Cards can be moved on top of other build stacks if their rank is one lower than the current top card and of the opposite color. Cards can be requested from the pile to be placed on the talon (Figure 1).

**Figure 1.** Klondike Solitaire. The seven build stacks can be seen at the bottom, the suit stacks are at the top left. The pile of undealt cards can be seen in the top right.
Study Objective

Given the popularity of Klondike Solitaire among the older population, and given the need for engaging, ecologically valid, scalable tools to assist in the screening and monitoring of MCI, this paper aims to investigate the feasibility of Klondike Solitaire game sessions to yield digital biomarkers of MCI. To this end, the study comprised the following investigations:

1. an exploration of the digital biomarkers of cognitive performance, based on the player actions (PAs) of Klondike Solitaire and
2. an evaluation of the candidate digital biomarkers captured in Klondike Solitaire to measure the differences between healthy older adults and older adults living with MCI.

Methods

Crafting Candidate Digital Biomarkers in Klondike Solitaire

To explore the potential digital biomarkers of cognitive performance in Klondike Solitaire, we first conducted an expert consensus study, involving 11 experts. In this first part of the paper, we discuss the 3 steps taken to compile a final list of 23 candidate digital biomarkers.

Step 1: Defining PAs

To transform gameplay into PAs, a methodical approach was applied. In all, 4 researchers in the field of human-computer interaction carried out the following tasks. First, the literature on Klondike Solitaire was studied, ranging from scientific work [72-78] to more general sources [79-81]. Afterward, the game was played for a minimum of 10 sessions of 30 minutes by each of the researchers. Combining the theoretical background with this practical experience, the experts independently drafted a list of game events, which was then reviewed as a team. This list was iterated 3 times until no more game events were found. The game events included, but were not limited to, game outcomes (eg, the game was won or lost), correct player moves (eg, the player moves a card among the build stacks), and erroneous player moves (eg, the player places cards on other cards such that they are not in descending order on the build stack).

These game events were then converted to PAs; they were described as an action of the player rather than as an event of the game. Next, all these PAs were transformed into their negative equivalents, for example, The player takes little time to think of a move was reworded as The player takes a lot of time to think of a move. The reason for this was 2-fold. It enabled inverse PAs (the positive and negative equivalents, eg, moving cards fast or moving them slowly) to be combined, reducing the rating complexity for the professionals. Furthermore, the negative equivalent aimed to facilitate the rating process because impaired cognition leads to reduced performance in gameplay. After this step was completed, 21 PAs (Table 1) were defined for evaluation.
Table 1. Average of the experts’ ratings for each player action and cognitive function.

<table>
<thead>
<tr>
<th>Player actions</th>
<th>Mental flexibility, mean (SD)</th>
<th>Inhibitory control, mean (SD)</th>
<th>Working memory, mean (SD)</th>
<th>Selective attention, mean (SD)</th>
<th>Visuospatial ability, mean (SD)</th>
<th>Object recognition, mean (SD)</th>
<th>Apraxia, mean (SD)</th>
<th>Cognitive planning, mean (SD)</th>
<th>Processing speed, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA 1. Player takes a lot of time to think of a move.</td>
<td>1.64 (1.12)</td>
<td>0.73 (1.01)</td>
<td>1.82 (0.4)</td>
<td>1.55 (0.82)</td>
<td>1.18 (0.98)</td>
<td>1.27 (0.79)</td>
<td>0.27 (0.47)</td>
<td>2.45 (0.92)</td>
<td>2.55 (0.82)</td>
</tr>
<tr>
<td>PA 2. Player takes a lot of time to move the card.</td>
<td>0.73 (1.01)</td>
<td>0.73 (1.1)</td>
<td>0.64 (0.5)</td>
<td>0.64 (0.67)</td>
<td>1.45 (1.04)</td>
<td>0.64 (0.67)</td>
<td>1.64 (0.92)</td>
<td>0.91 (1.04)</td>
<td>2.09 (1.04)</td>
</tr>
<tr>
<td>PA 3. Player does not move a suitable card from the talon to the suit stack.</td>
<td>2.27 (0.79)</td>
<td>0.73 (0.9)</td>
<td>2.07 (0.77)</td>
<td>2.55 (0.69)</td>
<td>1.45 (0.93)</td>
<td>1.64 (0.67)</td>
<td>0.73 (0.75)</td>
<td>2.18 (0.75)</td>
<td>0.91 (1.04)</td>
</tr>
<tr>
<td>PA 4. Player does not move a suitable card from the build stack to the suit stack.</td>
<td>1.82 (0.98)</td>
<td>0.91 (0.83)</td>
<td>1.82 (0.75)</td>
<td>2.36 (0.67)</td>
<td>1.36 (0.92)</td>
<td>1.36 (0.92)</td>
<td>0.27 (0.47)</td>
<td>1.73 (0.79)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>PA 5. Player does not move a suitable card from the talon to the build stack.</td>
<td>2.18 (0.75)</td>
<td>1.27 (0.9)</td>
<td>1.91 (0.7)</td>
<td>2.64 (0.67)</td>
<td>1.55 (1.04)</td>
<td>1.73 (0.79)</td>
<td>0.18 (0.4)</td>
<td>2.09 (0.94)</td>
<td>0.82 (1.08)</td>
</tr>
<tr>
<td>PA 6. Player does not move a suitable card from 1 build stack to another build stack.</td>
<td>2.36 (0.81)</td>
<td>1.09 (0.94)</td>
<td>2.07 (0.77)</td>
<td>2.45 (0.69)</td>
<td>1.45 (0.93)</td>
<td>1.64 (0.92)</td>
<td>0.18 (0.4)</td>
<td>2.18 (0.98)</td>
<td>0.73 (1.01)</td>
</tr>
<tr>
<td>PA 7. Player does not place an ace immediately on an empty suit stack.</td>
<td>1.27 (1.01)</td>
<td>0.73 (1.01)</td>
<td>2.18 (0.4)</td>
<td>2.36 (0.92)</td>
<td>1 (0.77)</td>
<td>1.18 (1.17)</td>
<td>0.45 (0.69)</td>
<td>2.09 (0.7)</td>
<td>1.09 (1.04)</td>
</tr>
<tr>
<td>PA 8. Player does not place a king on an empty build stack.</td>
<td>1.45 (1.13)</td>
<td>0.73 (0.9)</td>
<td>2.07 (0.77)</td>
<td>2.27 (0.9)</td>
<td>1 (0.77)</td>
<td>1.55 (1.13)</td>
<td>0.36 (0.67)</td>
<td>2.09 (0.7)</td>
<td>1 (0.89)</td>
</tr>
<tr>
<td>PA 9. Player moves cards without benefit (eg, moving a jack from 1 queen to another).</td>
<td>1.45 (1.04)</td>
<td>0.82 (0.98)</td>
<td>2.18 (1.17)</td>
<td>2.18 (1.03)</td>
<td>0.82 (0.75)</td>
<td>1.45 (1.21)</td>
<td>0.18 (0.4)</td>
<td>2.27 (1.01)</td>
<td>0.45 (0.52)</td>
</tr>
<tr>
<td>PA 10. Player flips a lot through the pile.</td>
<td>2.09 (0.89)</td>
<td>2.55 (0.69)</td>
<td>1.73 (1.01)</td>
<td>1.82 (1.08)</td>
<td>1 (0.89)</td>
<td>1.45 (1.13)</td>
<td>1 (1)</td>
<td>2.09 (0.7)</td>
<td>0.91 (0.7)</td>
</tr>
<tr>
<td>PA 11. Player moves a card onto a card with the same color.</td>
<td>1.73 (1.1)</td>
<td>2.55 (0.52)</td>
<td>2.18 (0.98)</td>
<td>2.18 (0.98)</td>
<td>1 (1)</td>
<td>1.82 (1.17)</td>
<td>0.27 (0.65)</td>
<td>1.36 (0.81)</td>
<td>0.45 (0.69)</td>
</tr>
<tr>
<td>PA 12. Player moves a card to another card with the wrong number (eg, placing a 5 on a 10).</td>
<td>1.18 (1.08)</td>
<td>2 (1)</td>
<td>2.27 (0.9)</td>
<td>1.91 (0.94)</td>
<td>1.09 (1.04)</td>
<td>2.09 (0.94)</td>
<td>0.45 (0.69)</td>
<td>1.45 (0.93)</td>
<td>0.36 (0.5)</td>
</tr>
<tr>
<td>PA 13. Player selects the cards with a very bad precision (taps on edge of card or next to the card).</td>
<td>0.45 (0.69)</td>
<td>0.73 (0.79)</td>
<td>0.27 (0.47)</td>
<td>0.64 (0.81)</td>
<td>2.27 (0.9)</td>
<td>0.82 (0.75)</td>
<td>2.27 (0.79)</td>
<td>0.45 (0.82)</td>
<td>0.45 (0.69)</td>
</tr>
<tr>
<td>PA 14. Player starts tapping on the playfield with no apparent target (with short intervals, fidget tapping).</td>
<td>0.73 (0.79)</td>
<td>2.27 (1.01)</td>
<td>0.27 (0.47)</td>
<td>0.82 (0.87)</td>
<td>0.73 (0.9)</td>
<td>0.45 (0.52)</td>
<td>1.55 (1.29)</td>
<td>0.91 (1.04)</td>
<td>0.73 (0.9)</td>
</tr>
<tr>
<td>PA 15. Player presses the undo button a lot.</td>
<td>1.82 (0.6)</td>
<td>2.45 (0.69)</td>
<td>1.73 (1.1)</td>
<td>1.36 (1.12)</td>
<td>0.64 (0.67)</td>
<td>0.64 (0.67)</td>
<td>0.73 (1.01)</td>
<td>2.27 (1.01)</td>
<td>1.27 (1.1)</td>
</tr>
<tr>
<td>PA 16. Player requests a lot of hints.</td>
<td>1.91 (1.04)</td>
<td>1.73 (1.01)</td>
<td>2 (1)</td>
<td>1.45 (0.93)</td>
<td>0.64 (0.81)</td>
<td>1 (0.77)</td>
<td>0.45 (0.69)</td>
<td>2.27 (1.01)</td>
<td>1.18 (0.75)</td>
</tr>
<tr>
<td>PA 17. Player takes a very long time to finish games.</td>
<td>2.18 (1.25)</td>
<td>1 (1.34)</td>
<td>2.18 (0.75)</td>
<td>1.64 (1.21)</td>
<td>1.09 (0.83)</td>
<td>1.18 (0.98)</td>
<td>0.91 (0.83)</td>
<td>2.64 (0.5)</td>
<td>2.91 (0.3)</td>
</tr>
<tr>
<td>PA 18. Player does not have a high score in the game.</td>
<td>2.18 (0.98)</td>
<td>2 (1)</td>
<td>2.36 (1.03)</td>
<td>1.91 (1.04)</td>
<td>1.45 (0.93)</td>
<td>1.36 (0.92)</td>
<td>0.91 (0.94)</td>
<td>2.27 (0.9)</td>
<td>1.55 (1.04)</td>
</tr>
<tr>
<td>PA 19. Player does not win a lot of games (low win ratio).</td>
<td>2.36 (0.67)</td>
<td>1.82 (1.08)</td>
<td>2.64 (0.5)</td>
<td>2 (1)</td>
<td>1.36 (0.92)</td>
<td>1.18 (0.87)</td>
<td>1 (0.89)</td>
<td>2.82 (0.4)</td>
<td>1.64 (0.81)</td>
</tr>
<tr>
<td>PA 20. Player’s scores in different games vary greatly.</td>
<td>2.27 (1.1)</td>
<td>1.64 (1.12)</td>
<td>2.27 (0.79)</td>
<td>2.36 (1.12)</td>
<td>0.73 (0.9)</td>
<td>0.73 (0.9)</td>
<td>0.64 (0.92)</td>
<td>2.18 (1.08)</td>
<td>1.82 (1.08)</td>
</tr>
<tr>
<td>PA 21. Player’s win ratio decreases rapidly as the game’s level of difficulty increases.</td>
<td>2.36 (0.67)</td>
<td>1.91 (0.94)</td>
<td>2.64 (0.67)</td>
<td>2.18 (0.87)</td>
<td>1.18 (0.87)</td>
<td>1.09 (0.94)</td>
<td>0.82 (0.87)</td>
<td>2.64 (0.81)</td>
<td>1.64 (1.03)</td>
</tr>
</tbody>
</table>
Step 2: Defining Cognitive Functions

A set of cognitive functions was drafted in 5 phases (Figure 2). A first draft was prepared (phase 1), beginning with the cognitive functions measured using the screening tests used most often for MCI [36,37,82]. In phase 2, during a trial with a psychologist, we replaced abstraction with object recognition to more clearly indicate the problems with finding suitable cards, based on key articles on cognitive aging and cognition [83-87]. Next (phase 3), to better delineate attention, it was specified as selective attention. In phase 4, a pilot study was conducted with an expert on memory and age-related disorders (with 23 years of clinical and research experience). On the basis of this pilot testing, it was decided to split executive functioning into inhibitory control, cognitive planning, and mental flexibility. Memory was further specified as working memory and lack of motor skills as apraxia. In the final iteration, cognitive functions ostensibly not present in Klondike Solitaire, that is, orientation in time and space, as well as language, were removed to reduce the rating complexity. This resulted in a set of 9 cognitive functions (Figure 2, phase 5).

Figure 2. The 5 phases through which the cognitive functions were defined. CDR: Clinical Dementia Rating, MoCA: Montreal Cognitive Assessment, MMSE: Mini-Mental State Examination.

Protocol for Rating Functions and Actions

Overview

In the next step, the experts were asked to rate the extent to which each PA was related to a specific cognitive function. These experts were recruited from 2 leading memory clinics in Belgium using a snowball sampling method. Of the 11 experts, 3 (27%) held Doctor of Medicine degrees and were experienced in treating cognitive decline, whereas the other 8 (73%) were neuropsychologists; 7 (64%) were women; and 4 (36%) were men. The average age of the experts was 45 (SD 13.3) years, and their average working experience was 20 (SD 14) years. In all, 3 coauthors of this paper (LVA, PD, and FFB) also participated as experts. None of the experts were compensated for participating in the study.

Before they began the rating process, the experts received a standardized introduction comprising a video that explained all concepts of the game [88], a video that visualized all 21 PAs [89], and a document that provided explanatory notes for all 9 cognitive functions. The aim of providing this introduction was to prevent confusion regarding game terminology, interpretation of PAs, and cognitive functions. The introduction also included a delineation of the target group to persons living with multiple-domain aMCI. The experts could revisit the videos and document at any time.
Next, each expert received a coding sheet in which they could map the 21 PAs to the 9 cognitive functions. Each cell had to be filled in according to the following 4-point scale:

0: This cognitive function has no significant correlation to the PA. 1: This cognitive function correlates weakly to the PA. 2: This cognitive function correlates moderately to the PA. 3: This cognitive function correlates strongly to the PA.

Finally, they were also given the choice to explain their train of thought in an optional clarification column.

**Expert Agreement on PAs and Cognitive Functions**

The intraclass correlation (ICC) for each PA as variables of interest with cognitive functions was computed. In addition, we computed ICCs for each of the cognitive functions as variables of interest, with all PAs considered as observations [90]. All calculations were executed using SPSS software (version 23.0; IBM Corp) [91]. The ICC was calculated to verify the rater agreement [92] on PAs and cognitive functions based on a 2-way, random, fully crossed design with type consistency [93]. According to the criteria described in the study by Koo et al [94], ICCs below 0.5 are indicative of low reliability, ICCs between 0.5 and 0.75 are indicative of moderate reliability, ICCs between 0.75 and 0.9 are indicative of good reliability, and ICCs above 0.9 are indicative of excellent reliability.

We found that the ICCs for all PAs scored higher than 0.75, suggesting good to excellent reliability according to the study by Koo et al [94]. With the exception of 4 cognitive functions (ie, mental flexibility, visuospatial ability, object recognition, and apraxia with scores of 0.68, 0.42, 0.66, and 0.71, respectively), the ICCs of the cognitive functions scored higher than 0.75, suggesting good to excellent reliability.

**Cognitive Functions Present in Klondike Solitaire**

An overview of the associations between individual PAs and cognitive functions, according to the expert mapping, is presented in Table 1. In addition, an overview of all ICCs with 95% CIs is presented in Multimedia Appendix 1. Each PA was related by the experts to one or more cognitive functions with an average association above two, which indicated a moderate to strong relationship with the cognitive functions. Similarly, each cognitive function was associated with at least one PA, with an average association of more than two.

**Step 3: Defining Candidate Digital Biomarkers**

These PAs were captured through the game as potential digital biomarkers, that is, measurable factors of the game, such as score, duration of the game, and detailed moves. These candidate digital biomarkers were enriched with additional information about the game. This contextualization is important to ensure an unambiguous interpretation of the cognitive information derived from the gameplay. For example, whereas a game played with many moves on the pile may indicate that a player progressed in the game, it may indicate equally that the player did not realize that they were stuck. By calculating the percentage of pile moves by dividing the number of pile moves by the total number of moves, a more informative metric can be obtained. In this manner, 23 potential digital biomarkers were defined that we further classified into 1 of 6 categories: time-based, performance-based, error-based, execution-based, auxiliary-based, and result-based. Time-based digital biomarkers are biomarkers related to the speed of PAs. Performance-based digital biomarkers are biomarkers related to optimal gameplay (ie, if the game was played according to strategies that ensure optimal performance). Error-based digital biomarkers relate to making incorrect moves according to the rules of Klondike Solitaire. Auxiliary-based digital biomarkers are interactions that are not part of the core gameplay (ie, requesting undo moves and hints). Execution-based digital biomarkers relate to the accuracy in moving cards and the presence of accidental taps. Finally, result-based digital biomarkers are biomarkers that evaluate the final outcome of the game (eg, how far did the participant get in the game). An overview of all digital biomarkers and their contextualizations is presented in Table 2.
### Table 2. Digital biomarkers related to the player actions in Klondike Solitaire\(^b\).

<table>
<thead>
<tr>
<th>Related PA(^b)</th>
<th>Digital biomarker</th>
<th>Description</th>
<th>Contextualization</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 1</td>
<td>Think time</td>
<td>Time spent thinking of a move. Defined as the time necessary to find and touch a suitable card</td>
<td>Average (SD)</td>
<td>Number (ms)</td>
</tr>
<tr>
<td>PA 2</td>
<td>Move time</td>
<td>Time spent moving cards. Defined as the time necessary to move a suitable card to the destination</td>
<td>Average (SD)</td>
<td>Number (ms)</td>
</tr>
<tr>
<td>PA 1, PA 2</td>
<td>Total time</td>
<td>Total time to make a move. Defined as the combination of think time and move time</td>
<td>Average (SD)</td>
<td>Number (ms)</td>
</tr>
<tr>
<td><strong>Performance-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 3</td>
<td>Final (\beta) error</td>
<td>Whether there were still moves possible when quitting the game</td>
<td>None</td>
<td>Boolean</td>
</tr>
<tr>
<td>PA 3, PA 4, PA 5, PA 6</td>
<td>(\beta) error</td>
<td>Number of pile moves made with moves remaining on the board divided by the total number of pile moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 7</td>
<td>Ace (\beta) error</td>
<td>Number of missed opportunities to place an ace on the suit stacks divided by the total number of game moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 8</td>
<td>King (\beta) error</td>
<td>Number of missed opportunities to place a king on an empty spot divided by the total number of game moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 10</td>
<td>Pile move</td>
<td>Number of pile moves divided by the total number of board moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td><strong>Error-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 11, PA 12</td>
<td>Successful move</td>
<td>Number of successful moves divided by the total number of board moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 11, PA 12</td>
<td>Erroneous move</td>
<td>Number of erroneous moves divided by the total number of board moves</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td><strong>Execution-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 13</td>
<td>Accuracy</td>
<td>Accurateness of selecting a card, defined by how close to the center a card was touched</td>
<td>Average (SD)</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 14</td>
<td>Taps</td>
<td>Number of actuations on nonuser interface elements</td>
<td>None</td>
<td>Number</td>
</tr>
<tr>
<td><strong>Auxiliary-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 15</td>
<td>Undo move</td>
<td>Number of undo moves requested</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td>PA 16</td>
<td>Hint move</td>
<td>Number of hints requested</td>
<td>Percentage</td>
<td>0%-100%</td>
</tr>
<tr>
<td><strong>Result-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA 17</td>
<td>Game time</td>
<td>Total time spent playing a game</td>
<td>None</td>
<td>Number (ms)</td>
</tr>
<tr>
<td>PA 18</td>
<td>Score</td>
<td>Final score of a game</td>
<td>None</td>
<td>Number</td>
</tr>
<tr>
<td>PA 19</td>
<td>Solved</td>
<td>Whether the game was completed. Indicator of how successfully the game was played</td>
<td>None</td>
<td>Boolean</td>
</tr>
</tbody>
</table>
Evaluating Digital Biomarkers

The aim of this second study is to explore the potential of these candidate digital biomarkers of cognitive performance. Relying on 46 participants, we captured data and performed a GLMM analysis to examine the differences between healthy participants and those diagnosed with MCI.

Participants

In total, 46 participants—23 (50%) healthy participants and 23 (50%) with MCI—were enrolled. The older adults with MCI were recruited by 2 leading memory clinics in Belgium. Healthy participants were recruited from multiple senior citizen organizations, using a snowball sampling method. All healthy participants were aged 65 years or older, were fluent in written and verbal Dutch, had 20/20 (corrected) vision, no motor impairments, and lived independently or semi-independently at home, in a service flat, or at a care home. The exclusion criteria for healthy participants were subjective-memory concerns expressed by the participant, caretaker, or clinician. In addition, they were screened using the MMSE, MoCA, and CDR Scale. To minimize the risk of including potential individuals with MCI among the healthy participants, cutoff scores of 27 on the MMSE, 26 on the MoCA, and 0 on the CDR Scale were enforced. The participants living with MCI had been formally diagnosed with multiple-domain aMCI by 1 of the 2 collaborating memory clinics, based on the diagnostic criteria described in the study by Petersen [95]. Participants with MCI were excluded if they scored lower than 23 on the MMSE to avoid including participants who could be on the borderline between the diagnosis of MCI and dementia. All recruited participants had prior experience with Klondike Solitaire. This familiarity with the rules was imperative because participants with MCI may have problems with memorizing and recalling new game rules because of short-term memory issues. Moreover, the rationale underlying this study is that it is imperative to draw from games already played and enjoyed by participants and where the rules are crystallized in their memory. Demographic and basic neuropsychological data of both groups are presented in Table 3.
Table 3. Demographic and neuropsychological data (N=46).

<table>
<thead>
<tr>
<th></th>
<th>Healthy participants (n=23)</th>
<th>Participants diagnosed with MCI (n=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years), mean (SD)</strong></td>
<td>70 (5.4)</td>
<td>80 (5.2)</td>
</tr>
<tr>
<td><strong>Education (ISCED(^b) level) [96], n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Levels 1-2</td>
<td>5 (22)</td>
<td>4 (17)</td>
</tr>
<tr>
<td>Levels 3-4</td>
<td>7 (30)</td>
<td>13 (57)</td>
</tr>
<tr>
<td>Levels 5-6</td>
<td>11 (48)</td>
<td>6 (26)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>11 (47)</td>
<td>13 (57)</td>
</tr>
<tr>
<td>Male</td>
<td>12 (53)</td>
<td>10 (43)</td>
</tr>
<tr>
<td><strong>Tablet proficiency, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>12 (52)</td>
<td>3 (13)</td>
</tr>
<tr>
<td>Weekly</td>
<td>2 (9)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Monthly</td>
<td>0 (0)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Yearly or less</td>
<td>2 (9)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Never</td>
<td>7 (30)</td>
<td>15 (65)</td>
</tr>
<tr>
<td><strong>Klondike Solitaire proficiency, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>3 (13)</td>
<td>7 (30)</td>
</tr>
<tr>
<td>Weekly</td>
<td>6 (26)</td>
<td>8 (35)</td>
</tr>
<tr>
<td>Monthly</td>
<td>3 (13)</td>
<td>2 (9)</td>
</tr>
<tr>
<td>Yearly or less</td>
<td>11 (47)</td>
<td>6 (26)</td>
</tr>
<tr>
<td>Never</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>MMSE(^c) score, mean (SD)</strong></td>
<td>29.61 (0.65)</td>
<td>26.17 (1.75)</td>
</tr>
<tr>
<td><strong>MoCA(^d) score, mean (SD)</strong></td>
<td>28.09 (1.28)</td>
<td>N/A(^e)</td>
</tr>
<tr>
<td><strong>CDR(^f) Scale score, mean (SD)</strong></td>
<td>0 (0)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)MCI: mild cognitive impairment.
\(^b\)ISCED: International Standard Classification of Education.
\(^c\)MMSE: Mini-Mental State Examination.
\(^d\)MoCA: Montreal Cognitive Assessment.
\(^e\)N/A: not applicable.
\(^f\)CDR: Clinical Dementia Rating.

Data Collection Tools

All game sessions were completed on a Lenovo Tab3 10 Business tablet (Lenovo Group Ltd) running Android 6.0 (Google LLC). A Solitaire app created by Bielefeld [97] under the Lesser General Public License 3 was modified to capture and store game metrics that served as building blocks for the digital biomarkers of cognitive performance.

Data Collection Procedure

Each observation was made between 9 AM and 5 PM in the home environment of the participant to ensure a familiar and comfortable environment. Each observation took between 2 and 3 hours and consisted of 2 main parts:

1. a game session where game-based digital biomarkers of Klondike Solitaire were collected on a tablet and
2. a neuropsychological examination where cognitive information was collected.

Each game session started with a standardized 5-minute introduction during which the tablet, the game mechanics, and possible touch interactions were explained. This was followed by a practice game, identical for all participants, where questions to the researcher were allowed and the participant could get used to the touch controls. Data from this practice game were not used for analysis. After this practice game, each participant played 3 games in succession. The order and games were identical across all participants. All games were purposefully chosen through prior playtesting, in that they were solvable and varied in difficulty level. During these 3 games, no questions were allowed, and gameplay continued until the participants either finished the game or indicated that they deemed further progress impossible. All game sessions and cognitive
evaluations were conducted by the same researcher to avoid differences arising from researcher bias.

**Ethical Statement**

This study is in accordance with the declaration of Helsinki and General Data Protection Regulation compliant. Ethical approval was provided by the ethics committee of UZ/KU Leuven, Belgium (CTC S59650). Because of the fragile nature of our participants’ health, utmost care was taken when providing information to them about the game sessions. The tests were conducted only after we received written informed consent.

**Statistical Analysis**

To assess the difference between the healthy participants and those diagnosed with MCI, a GLMM analysis was performed using R software (The R Foundation for Statistical Computing) [98] with the lme4 library [99]. Concerning the design of our GLMM, the fixed effects consisted of MCI, age, tablet proficiency, and Klondike Solitaire proficiency. The random effects were modeled as random intercepts for game seed and participant. In addition, by-participant random slopes for the effect of MCI were modeled.

Continuous digital biomarkers (eg, think time average) were modeled using a GLMM with the identity link function. Binary outcomes (eg, solved or not solved) were modeled using a GLMM with the logit link function. The significance of the effect of MCI was determined using the likelihood ratio test, which compares the model with a model without the effect of MCI, both estimated without restricted maximum likelihood [100,101]. Assumptions of homoscedasticity and normality were visually inspected using residual plots. To provide supplemental information on the fit of the models, the marginal $R^2$ and the conditional $R^2$ were given as specified in the study by Nakagawa and Schielzeth [102]. Given the exploratory nature of this study, we did not correct for family-wise inflation error [103].

**Results**

**Overview**

The results of the GLMM analysis on the effect of MCI are presented below. A visualization of digital biomarker performance for all groups across all games is presented in Figures 3-8. A summary is presented in Table 4.
Figure 3. Performance on time-based digital biomarkers for both groups. MCI: mild cognitive impairment.
Table 4. Generalized linear mixed model analysis results for each digital biomarker.

<table>
<thead>
<tr>
<th>Digital biomarker</th>
<th>Value, constant (SD)</th>
<th>Value, $\beta$ (SD)</th>
<th>$P$ value ($\chi^2$)</th>
<th>Value, $R^2 m$ ($R^2 c$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Think time average</td>
<td>-1371.778 (1415.444)</td>
<td>1119.947 (405.815)</td>
<td>.006</td>
<td>0.416 (0.904)</td>
</tr>
<tr>
<td>Think time SD</td>
<td>-814.527 (1720.073)</td>
<td>1112.533 (490.53)</td>
<td>.02</td>
<td>0.211 (0.655)</td>
</tr>
<tr>
<td>Move time average</td>
<td>-508.575 (373.89)</td>
<td>156 (95.547)</td>
<td>.1</td>
<td>0.257 (0.579)</td>
</tr>
<tr>
<td>Move time SD</td>
<td>-856.605 (847.852)</td>
<td>323.599 (202.032)</td>
<td>.1</td>
<td>0.137 (0.419)</td>
</tr>
<tr>
<td>Total time average</td>
<td>-912.419 (2149.177)</td>
<td>1278.263 (573.839)</td>
<td>.02</td>
<td>0.318 (0.870)</td>
</tr>
<tr>
<td>Total time SD</td>
<td>206.569 (2670.062)</td>
<td>1315.598 (673.665)</td>
<td>.04</td>
<td>0.176 (0.715)</td>
</tr>
<tr>
<td><strong>Performance-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final $\beta$ error</td>
<td>-7.233 (4.131)</td>
<td>0.435 (0.922)</td>
<td>.65</td>
<td>0.096 (0.068)</td>
</tr>
<tr>
<td>$\beta$ error percentage</td>
<td>-7.203 (33.849)</td>
<td>6.108 (6.879)</td>
<td>.36</td>
<td>0.089 (0.371)</td>
</tr>
<tr>
<td>Ace $\beta$ error percentage</td>
<td>-0.132 (0.629)</td>
<td>0.051 (0.137)</td>
<td>.73</td>
<td>0.023 (0.209)</td>
</tr>
<tr>
<td>King $\beta$ error percentage</td>
<td>-3.682 (5.918)</td>
<td>0.907 (1.323)</td>
<td>.48</td>
<td>0.028 (0.230)</td>
</tr>
<tr>
<td>Pile move percentage</td>
<td>71.759 (24.052)</td>
<td>13.333 (4.88)</td>
<td>.06</td>
<td>0.097 (0.513)</td>
</tr>
<tr>
<td><strong>Error-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Successful move percentage</td>
<td>87.486 (9.443)</td>
<td>-8.913 (3.595)</td>
<td>.02</td>
<td>0.104 (0.795)</td>
</tr>
<tr>
<td>Erroneous move percentage</td>
<td>9.486 (6.529)</td>
<td>3.624 (1.651)</td>
<td>.03</td>
<td>0.081 (0.466)</td>
</tr>
<tr>
<td><strong>Execution-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy average</td>
<td>92.134 (9.167)</td>
<td>-3.817 (1.903)</td>
<td>.04</td>
<td>0.246 (0.805)</td>
</tr>
<tr>
<td>Accuracy SD</td>
<td>4.519 (3.746)</td>
<td>0.137 (0.772)</td>
<td>.85</td>
<td>0.056 (0.196)</td>
</tr>
<tr>
<td>Taps</td>
<td>-5.113 (10.704)</td>
<td>5.334 (2.762)</td>
<td>.05</td>
<td>0.098 (0.500)</td>
</tr>
<tr>
<td><strong>Auxiliary-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undo move percentage</td>
<td>0.228 (0.955)</td>
<td>0.135 (0.205)</td>
<td>.49</td>
<td>0.008 (0.151)</td>
</tr>
<tr>
<td>Hint move percentage</td>
<td>-0.58 (0.87)</td>
<td>-0.311 (0.204)</td>
<td>.12</td>
<td>0.046 (0.491)</td>
</tr>
<tr>
<td><strong>Result-based</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gametime</td>
<td>-167427.7 (187325.5)</td>
<td>93211.27 (53699.25)</td>
<td>.08</td>
<td>0.198 (0.690)</td>
</tr>
<tr>
<td>Score</td>
<td>29.03 (1389.752)</td>
<td>-744.433 (286.576)</td>
<td>.009</td>
<td>0.105 (0.612)</td>
</tr>
<tr>
<td>Solved</td>
<td>-2.954 (4.578)</td>
<td>-2.63 (1.007)</td>
<td>.008</td>
<td>0.186 (0.152)</td>
</tr>
<tr>
<td>Cards moved average</td>
<td>1.111 (0.262)</td>
<td>-0.119 (0.054)</td>
<td>.03</td>
<td>0.061 (0.093)</td>
</tr>
<tr>
<td>Cards moved SD</td>
<td>0.135 (0.705)</td>
<td>-0.38 (0.147)</td>
<td>.009</td>
<td>0.072 (0.152)</td>
</tr>
</tbody>
</table>

Time-Based Digital Biomarkers

For time-based digital biomarkers (Figure 3), MCI significantly affected think time average ($\chi^2=7.7; P=.006$), increasing it by 1119.947 ms (SD 405.81). Equally significantly, MCI affected think time SD ($\chi^2=5.1; P=.02$), increasing it by 1112.533 ms (SD 490.53). However, MCI did not significantly affect move time average ($\chi^2=2.7; P=.10$) or move time SD ($\chi^2=2.6; P=.10$). MCI significantly affected total time average ($\chi^2=5.2; P=.02$), increasing it by 1278.263 ms (SD 573.84), and total time SD ($\chi^2=4.1; P=.04$), increasing it by 1315.598 ms (SD 673.67).

Performance-Based Digital Biomarkers

For performance-based digital biomarkers (Figure 4), MCI did not significantly affect final $\beta$ error percentage ($\chi^2=0.2; P=.65$). Equally, MCI did not significantly affect $\beta$ error percentage ($\chi^2=0.8; P=.36$), ace $\beta$ error percentage ($\chi^2=0.1; P=.73$), or king $\beta$ error percentage ($\chi^2=0.5; P=.48$). MCI significantly affected pile move percentage ($\chi^2=7.5; P=.006$), increasing it by 13.333% (4.88).
**Figure 4.** Performance on performance-based digital biomarkers for both groups. MCI: mild cognitive impairment.

Error-Based Digital Biomarkers

For error-based digital biomarkers (Figure 5), MCI significantly affected successful move percentage, \( \chi^2 = 5.9; P = .02 \), lowering it by 8.913% (SD 3.6). MCI also significantly affected erroneous move percentage, \( \chi^2 = 4.8; P = .03 \), increasing it by 3.624% (SD 1.65).
Figure 5. Performance on error-based digital biomarkers for both groups. MCI: mild cognitive impairment.

**Execution-Based Digital Biomarkers**
For execution-based digital biomarkers (Figure 6), MCI significantly affected accuracy average ($\chi^2 = 4.1; P = .04$), lowering it by 3.817% (SD 1.9). MCI did not significantly affect accuracy SD ($\chi^2 = 0.04; P = .85$) or taps ($\chi^2 = 3.8; P = .05$).

Figure 6. Performance on execution-based digital biomarkers for both groups. MCI: mild cognitive impairment.

**Result-Based Digital Biomarkers**
For result-based digital biomarkers (Figure 7), MCI did not significantly affect gametime ($\chi^2 = 3.1; P = .08$). MCI significantly affected solved ($\chi^2 = 6.9; P = .008$), lowering it by 2.63 (SD 1.01). MCI also significantly affected cards moved average ($\chi^2 = 4.9; P = .03$), lowering it by 0.119 cards (SD 0.05).
and cards moved SD ($\chi^2_1 = 6.7; P = .009$), lowering it by 0.38 cards (SD 0.15).

**Figure 7.** Performance on result-based digital biomarkers for both groups. MCI: mild cognitive impairment.

### Auxiliary-Based Digital Biomarkers

With regard to auxiliary-based digital biomarkers (**Figure 8**), none of these candidate biomarkers reached significance: undo move percentage ($\chi^2_1 = 0.4; P = .49$) and hint move percentage ($\chi^2_1 = 2.4; P = .12$).
Figure 8. Performance on auxiliary-based digital biomarkers for both groups. MCI: mild cognitive impairment.

**Discussion**

**Overview**

MCI is a neurological disorder that is linked to an increased risk of developing dementia. As such, early detection of cognitive deterioration is essential for timely diagnosis and for allowing tailored care and treatment. Collecting digital biomarkers through COTS games may help by providing cognitive information through behavior traces of activities already integrated into the daily life of older adults. In this study, we investigated in particular whether Klondike Solitaire game sessions could yield digital biomarkers. In the paragraphs below, we discuss our findings and reflect on the different potential digital biomarkers, their relationship with cognitive functions, and the ethical implications of their use for cognitive assessment purposes.

**Dissecting Digital Biomarkers**

Of the 23 candidate digital biomarkers, 12 (52%) differed significantly between older adults with MCI and a healthy control group. This supports the use of digital card games for monitoring cognitive performance and possibly detecting differences in cognitive performance caused by MCI.

Although the overall findings are promising, not all candidate biomarkers performed equally. In the case of time-based digital biomarkers, the biomarkers related to coming up with a move—think time average and think time SD—were significantly affected by MCI. In contrast, the biomarkers related to the actual physical movement of cards—move time average and move time SD—were not significantly affected. Total time average ($P=.02$), which consists of move time as well as think time, was significantly affected; yet, it was less significant than think time average ($P=.006$). These results indicate that segmenting in-game actions can be beneficial because they can more accurately isolate cognitive functions such as praxis and cognitive planning.

In the case of performance-based digital biomarkers, in contrast with expectations, none of the biomarkers related to $\beta$ errors were proven to differ significantly. Upon rewatching gameplay, it became clear that there were two different types of $\beta$ errors: strategic and unintentional. However, because of the current configuration of the app, it was impossible to discriminate between the two types. This is discussed further in the **Limitations** section. In contrast, pile move percentage was proven to differ significantly. This may indicate that older adults with MCI may not recognize the same cards being returned as quickly as their healthy counterparts.

Equally, the results indicated that participants with MCI made more mistakes because both error-based digital biomarkers (ie, successful move percentage and erroneous move percentage) were significant. In contrast, none of the auxiliary-based digital biomarkers differed significantly. Upon inspecting the data, it was noted that neither group consistently used these functionalities, which may have contributed to the lack of significance.

Finally, of the 5 digital biomarkers in the result-based category, 4 (80%) were significant, 3 (60%) with $P<.01$ (ie, score, solved, and cards moved SD). The outcome of these measures is the result of a series of consequent moves, each of them being potentially crucial to complete the game. For example, a lapse in attention or executive functioning can cause important moves to be overlooked, in turn making the game unsolvable. Although overall gametime was not significant, this can be explained by the fact that time spent in the game by itself does not indicate a lesser performance. Time-based digital biomarkers, which are equally measures of time but contextualized with the number of moves made, show more significant results (ie, think time average, think time SD, total time average, and total time SD), stressing the importance of contextualization.

In sum, our findings are in accordance with those of the study by Jimison et al [8], which used FreeCell, a Solitaire variant, to compare cognitive performance between a group of people living with MCI and a healthy control group. Using card gameplay, we can discern older adults with MCI from a healthy control group. Moreover, the results gathered from this study are in line with those of previous studies by Bankiqued et al [60] and Angeles Quiroga et al [104]. The study by Bankiqued et al [60] found that casual games that tap working memory and reasoning can be robustly related to performance on working memory and fluid intelligence. Similar research on commercial
video games by Ángeles Quiroga et al [104] found a strong relationship between performance in video games and general intelligence test performance. Our results confirm these findings at a finer granularity and show that when scrutinizing PAS, time-based, error-based, and result-based biomarkers yield promise in particular.

**Future Work**

In this study, the participants with MCI were diagnosed with multiple-domain aMCI based on the diagnostic criteria described in the study by Petersen [95]. As MCI is a multidimensional clinical entity, it would be interesting to explore whether Klondike Solitaire game sessions are suitable for monitoring the cognitive status of participants with non-aMCI as well. The focus on executive functioning can be useful for identifying both MCI subtypes because it has been shown that both have a similar decrease in executive functioning [105]. Although we acknowledge that the evaluation of other cognitive functions such as anterograde memory, retrograde memory, orientation, and language is paramount to obtain a complete overview of the patient’s cognitive profile, these cognitive functions were not identified by the experts and were thus not included in our analysis.

**Reflections on the Use of COTS Games to Assess Cognitive Performance**

COTS games also have their limitations. First, neuropsychological assessments are typically designed to assess a broad yet targeted spectrum of cognitive functions. Moreover, different tests are devised to measure 1 cognitive function in particular. COTS games, and more particularly digital card games, were found to be more limited in terms of the cognitive functions that they can specifically assess. When using COTS games, it may be hard to separate the evaluations of specific cognitive functions. In this study, experts judged every single PA to be moderately to strongly related to at least one cognitive function.

Second, using COTS games as an instrument to measure cognitive performance and possibly flag MCI necessitates ethical reflection. We envisioned that COTS games would be used only in accordance with the informed consent of the patient, with the positive aspiration that this could aid in the longitudinal monitoring of cognitive deterioration, more accurately measuring cognitive performance and variance. This project grew out of an ambition to escape the limitations of serious games and provide meaningful play to older adults. Nonetheless, we have to acknowledge that we may have transformed an activity previously considered enjoyable and innocent into an instrumental activity that may even trigger a sense of being under health surveillance [106]. Observational notes taken during this study did not reveal any verbal remarks of stress from the participants diagnosed with MCI. However, such remarks were made by some of the healthy participants because they felt pressure to outperform the participants living with MCI. Further research is needed to understand how the instrumentalization of COTS games affects the playing experience of patients.

Third, it has to be noted that deriving digital biomarkers from digital games may not be relevant for all older adults. Not everyone is an avid gamer, and even avid gamers may have preferences for different game genres. In addition, these preferences might change over time [14]. Although digital card games such as Klondike Solitaire are in general a popular pastime for the population susceptible to MCI [61,71,107-109], they might not be so for the coming generations. Therefore, it is important to identify other accessible games suitable for cognitive monitoring with a broad appeal.

Finally, the interaction between health care professional and patient, often stimulating and motivating in and of itself, is crucial for full assessment. Hence, we argue that COTS games for screening and monitoring of MCI should not be used as a replacement for current neuropsychological examination but rather as a source of additional information.

**Limitations**

**Fine-Tuning β Errors**

In contrast with expectations, β error–related digital biomarkers proved to be insignificant. Upon inspecting the games of both groups, it became clear that there are two types of β errors: build stack β errors and suit stack β errors. The former represents missed moves among the build stacks. These errors were rarely made on purpose and occurred fewer times in the healthy participants’ group, based on observation. In contrast, the latter represent missed moves between build stacks and suit stacks. We observed that this latter category was used strategically to prevent the inability to place future cards. Our observations suggest that these occurred more often in the healthy participants’ group. However, because of the current configuration of the app, it was impossible to discriminate between these two types of β errors. Hence, this points to the importance of further contextualization and refinement of the measurement of β errors, and biomarkers in general, which should be addressed in future work.

**Limited Sample Size**

An a priori power analysis [110] estimated an adequate sample size to be between 32 and 88 participants (assuming comparable effect sizes as cognitive screening instruments to detect MCI [111]). Because of the strict inclusion criteria, only 46 participants were eligible. Although this strict protocol was designed with data quality in mind, the sample size may have affected the effects estimated in this study. It could be that our study was underpowered, leading to some digital biomarkers to be wrongly found insignificant. Future studies should therefore critically inspect the different digital biomarkers and the results obtained.

In addition, because of the average age difference between the 2 groups, we chose a GLMM for our statistical analysis because it can factor in confounding effects. A side exploration included using trained machine learning models on the same data set to predict age instead of MCI. These models were found to be less performant than the ones modeling MCI, underscoring that the effect of MCI was greater than the effect of age in our data set. Nevertheless, it is a limitation that we have to acknowledge and take into account while interpreting the results.
Conclusions
This study provides insight into the cognitive functions addressed while playing digital card games and assesses the potential of digital card game sessions for screening for MCI. To this end, 11 experts in neuropsychology or geriatrics mapped the associations of PAs in Klondike Solitaire with cognitive functions. On the basis of this exercise, which showed that the experts agreed that PAs were related to cognitive functions, 23 potential digital biomarkers of cognitive performance were crafted. A GLMM analysis, taking the effects of age, tablet experience, and Klondike Solitaire experience into account, compared digital biomarker performance between a group consisting of people living with MCI and a healthy control group. We found that of the 23 digital biomarkers, 12 (52%) had a significant and sizeable effect, despite the strict inclusion criteria and natural variations in human cognition. These exploratory results support the notion of detecting MCI through Klondike Solitaire game sessions.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Intraclass correlations of the cognitive functions and player actions with 95% CIs.

References


38. Pinto TC, Machado L, Bulgacov TM, Rodrigues-Júnior AL, Costa ML, Ximenis RC, et al. Is the montreal cognitive assessment (MoCA) screening superior to the mini-mental state examination (MMSE) in the detection of mild cognitive...
improvement (MCI) and Alzheimer’s disease (AD) in the elderly? Int Psychogeriatr 2019 Apr;31(4):491-504. [doi: 10.1017/S1041610218001370] [Medline: 30426911]


80. Winning at klondike. Semicolon Software. URL: https://semicolon.com/Solitaire/Articles/Klondike.html [accessed 2019-10-03]


Abbreviations

- **aMCI**: amnestic mild cognitive impairment
- **CDR**: Clinical Dementia Rating
- **COTS**: commercial off-the-shelf
- **GLMM**: generalized linear mixed model
- **ICC**: intraclass correlation
- **MCI**: mild cognitive impairment
- **MMSE**: Mini-Mental State Examination
- **MoCA**: Montreal Cognitive Assessment
- **naMCI**: nonamnestic mild cognitive impairment
- **PA**: player action

©Karsten Gielis, Marie-Elena Vanden Abeele, Robin De Croon, Paul Dierick, Filipa Ferreira-Brito, Lies Van Assche, Katrien Verbert, Jos Tournoy, Vero Vanden Abeele. Originally published in JMIR Serious Games (https://games.jmir.org), 04.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Effectiveness and Utility of Virtual Reality Simulation as an Educational Tool for Safe Performance of COVID-19 Diagnostics: Prospective, Randomized Pilot Trial

Tanja Birrenbach\textsuperscript{1,2}, MME, MD; Josua Zbinden\textsuperscript{1}; George Papagiannakis\textsuperscript{3,4,5}, PhD; Aristomenis K Exadaktylos\textsuperscript{1}, MD, PhD; Martin Müller\textsuperscript{1}, MD, PhD; Wolf E Hautz\textsuperscript{1}, MME, MD; Thomas Christian Sauter\textsuperscript{1}, MME, MD

\textsuperscript{1}Department of Emergency Medicine, Inselspital, University Hospital Bern, Bern, Switzerland
\textsuperscript{2}Centre for Health Sciences Education, Faculty of Medicine, University of Oslo, Oslo, Norway
\textsuperscript{3}ORamaVR SA, Geneva, Switzerland
\textsuperscript{4}Institute of Computer Science, Foundation for Research and Technology, Hellas, Heraklion, Greece
\textsuperscript{5}Department of Computer Science, University of Crete, Heraklion, Greece

Corresponding Author:
Tanja Birrenbach, MME, MD
Department of Emergency Medicine, Inselspital
University Hospital Bern
Freiburgstrasse 16C
Bern, 3010
Switzerland
Phone: 41 31 6322111
Email: tanja.birrenbach@insel.ch

Abstract

Background: Although the proper use of hygiene and personal protective equipment (PPE) is paramount for preventing the spread of diseases such as COVID-19, health care personnel have been shown to use incorrect techniques for donning/doffing of PPE and hand hygiene, leading to a large number of infections among health professionals. Education and training are difficult owing to the social distancing restrictions in place, shortages of PPE and testing material, and lack of evidence on optimal training. Virtual reality (VR) simulation can offer a multisensory, 3-D, fully immersive, and safe training opportunity that addresses these obstacles.

Objective: The aim of this study is to explore the short- and long-term effectiveness of a fully immersive VR simulation versus a traditional learning method regarding a COVID-19–related skill set and media-specific variables influencing training outcomes.

Methods: This was a prospective, randomized controlled pilot study on medical students (N=29; intervention VR training, n=15, vs control video-based instruction, n=14) to compare the performance of hand disinfection, nasopharyngeal swab taking, and donning/doffing of PPE before and after training and 1 month later as well as variables of media use.

Results: Both groups performed significantly better after training, with the effect sustained over one month. After training, the VR group performed significantly better in taking a nasopharyngeal swab, scoring a median of 14 out of 17 points (IQR 13-15) versus 12 out of 17 points (IQR 11-14) in the control group, \( P=.03 \). With good immersion and tolerability of the VR simulation, satisfaction was significantly higher in the VR group compared to the control group (median score of User Satisfaction Evaluation Questionnaire 27/30, IQR 23-28, vs 22/30, IQR 20-24, in the control group; \( P=.01 \)).

Conclusions: VR simulation was at least as effective as traditional learning methods in training medical students while providing benefits regarding user satisfaction. These results add to the growing body of evidence that VR is a useful tool for acquiring simple and complex clinical skills.

(JMIR Serious Games 2021;9(4):e29586) doi:10.2196/29586

KEYWORDS

virtual reality; VR; simulation; medical education; hand hygiene; COVID-19; PPE; nasopharyngeal swab; protection; effectiveness; utility; diagnostic; testing; pilot study
Introduction

The COVID-19 pandemic is a global health emergency that places massive demands on health systems and health care workers [1]. Proper use of hygiene and personal protective equipment (PPE) is paramount to prevent spreading of disease and contamination of health care workers. One possible reason for the high infection rate of COVID-19 is ineffective use of PPE. In Italy, up to 20% of health care workers were initially infected with the disease [2].

PPE recommendations from international organizations are largely consistent (eg, those from the US Centers for Disease Control and Prevention [CDC] and World Health Organization [WHO]) [3-6]; however, the actual use of PPE is not. Health care personnel in all professions and at all levels of training have been shown to use incorrect techniques for donning and doffing of PPE and hand hygiene [7-11]. The main reason appears to be inadequate training in correct PPE technique and lack of assessment of proficiency [7,11,12].

Simulation proves to be a powerful tool to test the accurate use of hygiene skills that are relevant for the treatment of patients with COVID-19 (ie, PPE and hand hygiene) [13]; nevertheless, there is still ambiguity regarding which training method works best. A recent Cochrane review of evidence relating to PPE and protection of health care staff exposed to contaminated body fluids highlights the lack of robust evidence in this area [11].

Furthermore, education and training of health care personnel is difficult with social distancing restrictions in place and shortages of both PPE [2] and testing materials.

Virtual reality (VR) uses computer systems to generate realistic pseudoenvironments that provide users with visual, tactile, and auditory sensations, with the possibility of realistic interaction with the virtual environment [14]. Milgram and Kishino [15] referred to mixed reality (MR) as the technologies which involve the merging of real and virtual worlds.

VR simulation with the use of head-mounted devices (HMDs) can offer a multisensory, 3-D, fully immersive, and safe training opportunity, avoiding the restrictions of social distancing and material shortages [16,17]. Through the concept of immersion, sense of presence, and interaction with the virtual environment in a real-time and realistic manner, VR simulation can create emotional experiences that facilitate experiential learning, exceeding other 2-D learning modalities [18].

The value of VR in medical education has already been demonstrated for various tasks [16,19-30]. VR is often used for training skills of varying complexity, ranging from simple nursing skills (Foley catheter placement, gaining venous access [31]) to laparoscopic/endoscopic/endovascular skills [23,32] or complex surgical procedures [26-29,33,34]. It is suggested that skills training as a first step to acquiring a competency can be better taught with VR than with traditional learning methods, because VR allows for a more active and immersive learning experience [21].

Recent studies suggest that VR improves postintervention knowledge and skills of health professionals better than traditional education or other types of digital education [21,31,35]. VR offers several advantages, as it provides possibilities for flexible learning and self-learning, providing standardization, reproducibility, and stimuli control; it enables automated generation of data about the details of simulations, including performance measurements that can be used for research or to provide automated individualized feedback [24]. The simple novelty of interactive technologies themselves, such as VR, can improve student motivation [36]. The initial cost and effort of creating the program can easily be compensated by broad distribution [16,37], as VR training is gradually finding its way into the medical curriculum [38].

Only a few virtual or mixed reality simulations exist for training hand hygiene [39-42]. In addition, high-quality studies evaluating the effectiveness and long-term retention of a VR simulation compared to conventional training methods are lacking.

Therefore, we hypothesized that a VR simulation could be an effective and useful tool with high student satisfaction to teach COVID-19 diagnostics, and we performed a randomized pilot study in medical students to explore (1) the effectiveness of a fully immersive VR simulation versus a traditional learning method regarding a COVID-19–related skill set (ie, proper hand hygiene, proficiency in PPE use, and correct acquisition of a nasopharyngeal specimen) tested in a simulated clinical scenario, and, as a secondary outcome, its long-term effectiveness 1 month after training; and (2) media-specific variables influencing training outcomes, such as usability, satisfaction, simulator sickness, and the experience of presence and immersion.

Methods

Study Design, Setting, and Ethical Approval

This is a prospective randomized controlled pilot study, taking place at the emergency department of the Inselspital, University Hospital Bern, Switzerland [43], from September to November 2020.

The study population consisted of a convenience sample of medical students at the University of Bern. All participants attended on a voluntary basis; no remuneration was provided. Informed consent was obtained. Data were collected, analyzed, and stored in anonymized form.

The local ethics committee deemed our study exempt from full ethical approval (Business Administration System for Ethics Committees Req-2020-00889).

Inclusion/Exclusion

The inclusion criteria were as follows: medical students (years 3-6 out of a 6-year curriculum) at the University of Bern. The exclusion criteria consisted of unwillingness to participate or to provide informed consent.

Baseline Investigations

Baseline Survey

A brief survey about sociodemographic factors; prior training; and experience in hand hygiene and PPE use, taking of
Assessments/Measurements
We evaluated the performance of hand disinfection, taking a nasopharyngeal swab on a manikin, and contamination during doffing of PPE (Figure 1).

Figure 1. Flowchart of the study. VR: virtual reality.

Hand Disinfection
Hand disinfection performance was evaluated using a fluorescent marker (Visirub conc and Sterillium, Hartmann AG) and UV-light scanning was performed using the Derma Litecheck UV Multimedia device (KBD Ltd) at the time of enrollment (pretest), directly after the intervention (posttest 1), and 1 month after the intervention (posttest 2). Participants were blindfolded during the assessment and were unable to assess their results. A performance analysis scheme (documentation of missed locations; n=38 locations for each hand, with a total of 76 areas investigated) was developed by the institution’s infection control and medical educator team as adapted from Pan et al [44]. The outcome was the number of missed locations (range from 0 to 76; a lower number is better). Performance was supervised by an independent and trained rater, and the images were electronically recorded and analyzed according to the predefined scheme by a rater blinded to the intervention (Figures 2 and 3).
Figure 2. UV-light scanning of a perfectly disinfected palmar surface of the right hand.

Figure 3. UV-light scanning of right hand dorsal surface. Missing areas of hand disinfection, right hand: digitus I, dorsal, distal phalanx; digitus II, dorsal, distal, and middle phalanx; digitus III, dorsal, distal, and middle phalanx; digitus IV, dorsal, distal phalanx; digitus V, dorsal, distal, and middle phalanx (total number of missed locations=8).
**Obtaining the Nasopharyngeal Swab and Evaluating Contamination During Doffing**

A simulation setup for conducting a nasopharyngeal swab for COVID-19 testing on a manikin (Little Anne, Laerdal Medical) using proper hand hygiene and PPE was installed.

The correct procedure of taking a nasopharyngeal swab sample as well as possible contamination while doffing were evaluated directly after the intervention (posttest 1) and 1 month after the intervention (posttest 2).

An independent and trained rater blinded to the intervention assessed each participant’s performance using a 17-item checklist adapted from [8,10] based on the CDC guidelines for PPE [3,6], WHO guidelines for hand hygiene [5], and international recommendations for taking a nasopharyngeal swab [6]; this checklist was developed by the institution’s infection control and medical educator team (Multimedia Appendix 1). The outcome was the number of points achieved on the checklist (range 0-17; a higher number of points indicated a better result).

Contamination during the procedure was evaluated using fluorescent lotion (Dermalux Testlotion S, KBD Ltd), which was applied to the participants’ hands, forearms, and torso before the donning of PPE. After doffing, 10 areas (right hand, right forearm, right upper arm, left hand, left forearm, left upper arm, torso ventral, torso dorsal, neck, head/ears) were analyzed by UV lighting for contamination by an independent rater. The outcome was the number of contaminated areas (range from 0 to 10; a lower number indicated a better result).

**Intervention**

Participants were randomized to either the intervention group (VR simulation) or control group in a 1:1 ratio using a computer-generated system.

**VR Simulation**

The intervention group was trained in COVID-19–related skills using the VR simulation (the Covid-19 VR Strikes Back (CVRSB) module, version 1.1.6), a software platform developed by ORamaVR SA, and the Oculus Rift S head mounted device and hand controllers (Facebook Inc). The proprietary ORamaVR software medical VR training application is available free of charge for all VR desktop and mobile HMDs [45] (Figure 4).

The participants performed two runs in the simulation using the single player modus.

**Control Group**

The control group was trained using traditional learning methods: printed instructions and local instruction videos on COVID-19–related skills, ie, PPE donning and doffing, as well as formal videos on proper hand hygiene according to the WHO and on taking a correct nasopharyngeal sample [46].

**Intervention Survey**

Both groups were evaluated regarding variables of media use according to established questionnaires.

Usability for both training modules was assessed using the After-Scenario Questionnaire (ASQ) [47], which assesses the ease of task completion, satisfaction with completion time, and satisfaction with supporting information on a 7-point Likert scale (total score ranges from 1, full satisfaction, to 7, poor satisfaction).

The User Satisfaction Evaluation Questionnaire (USEQ) [48] contains 6 questions with a 5-point Likert scale to evaluate user satisfaction (total score ranges from 6, poor satisfaction, to 30, excellent satisfaction).

---

**Figure 4.** Screenshot of the virtual reality (VR) application, Covid-19 VR Strikes Back, showing the taking of a nasopharyngeal swab.
For the VR simulation, “visually induced motion sickness” was assessed with 4 items (nausea, headache, blurred vision, dizziness) according to the Simulator Sickness Questionnaire (SSQ) adapted from Kennedy et al [49] (total score ranges from 1, no simulator sickness, to 5, strong simulator sickness).

Presence and immersion in the virtual world was determined according to the 6-item questionnaire developed by Slater-Usoh-Steed [50] (total score ranges from 1, no immersion, to 7, full immersion).

**Statistical Analysis**

Data were analyzed using SPSS, version 22 (IBM Corporation), and Stata 16.1 (StataCorp).

Baseline characteristics are presented as numbers and percentages or medians and interquartile ranges using descriptive statistics as appropriate.

The intervention and control group were compared regarding the baseline characteristics by chi-square test and Wilcoxon rank sum test as applicable.

The Wilcoxon rank sum test was used at a specific time point between the study groups for the comparison of all four outcome groups: (1) the number of missed areas during hand disinfection, (2) achieved items from the 17-item checklist during nasopharyngeal swab acquisition, (3) the number of contaminated areas during doffing, and (4) variables of media use. Within-group differences for different time points were tested using the Wilcoxon matched-pairs signed-rank test.

For all tests, a $P$ value <.05 was considered significant. For this pilot study, no adjustment for multiple testing was performed. Furthermore, pairwise comparisons were favored over more complex analyses, such as mixed linear regression analysis.

**Results**

**Sample Characteristics**

In total, 29 students completed the study (control group, n=14; intervention group, n=15) (Figure 1). All students included completed the whole study. There were no dropouts. The baseline characteristics of the participants are detailed in Table 1. No significant differences were found regarding gender, mean age, educational level in medical school, need to wear glasses, previous experience with computer games, or previous experience with VR. Likewise, previous education and experience regarding hand disinfection, use of PPE, and taking nasopharyngeal swabs did not show any significant differences.
Table 1. Baseline characteristics of the study sample (N=29).

<table>
<thead>
<tr>
<th>Sociodemographic factors</th>
<th>VR&lt;sup&gt;a&lt;/sup&gt; group (n=15)</th>
<th>Control group (n=14)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female gender, n (%)</td>
<td>9 (60)</td>
<td>9 (64)</td>
<td>.81</td>
</tr>
<tr>
<td>Age (years), median (IQR)</td>
<td>23 (22-25)</td>
<td>22.5 (22-24)</td>
<td>.56</td>
</tr>
<tr>
<td>Year of medical school, n (%)</td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>3</td>
<td>1 (7)</td>
<td>2 (14)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14 (93)</td>
<td>11 (79)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>1 (7)</td>
<td></td>
</tr>
<tr>
<td>Wears glasses, n (%)</td>
<td>8 (53)</td>
<td>8 (57)</td>
<td>.84</td>
</tr>
<tr>
<td>Plays computer games regularly (Likert scale response&lt;sup&gt;b&lt;/sup&gt;), n (%)</td>
<td></td>
<td></td>
<td>.95</td>
</tr>
<tr>
<td>1</td>
<td>10 (67)</td>
<td>10 (71)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4 (27)</td>
<td>3 (21)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 (7)</td>
<td>1 (7)</td>
<td></td>
</tr>
<tr>
<td>Uses VR regularly (Likert scale response&lt;sup&gt;b&lt;/sup&gt;), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>15 (100)</td>
<td>14 (100)</td>
<td></td>
</tr>
<tr>
<td>Previous education and experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous education in hand disinfection, n (%)</td>
<td>14 (93)</td>
<td>11 (79)</td>
<td>.25</td>
</tr>
<tr>
<td>Previous education in PPE&lt;sup&gt;c&lt;/sup&gt;, n (%)</td>
<td>6 (40)</td>
<td>3 (21)</td>
<td>.28</td>
</tr>
<tr>
<td>Previous no. of swabs, median (IQR)</td>
<td>0 (0-0)</td>
<td>0 (0-0)</td>
<td>.54</td>
</tr>
<tr>
<td>Uses PPE regularly (Likert scale response&lt;sup&gt;b&lt;/sup&gt;), n (%)</td>
<td></td>
<td></td>
<td>.36</td>
</tr>
<tr>
<td>1</td>
<td>10 (67)</td>
<td>8 (57)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 (20)</td>
<td>2 (14)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0 (0)</td>
<td>3 (21)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 (7)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1 (7)</td>
<td>1 (7)</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> VR: virtual reality.
<sup>b</sup> Likert scale: 1, completely disagree, to 5, completely agree.
<sup>c</sup> PPE: personal protective equipment.

**Hand Disinfection**

There was no significant difference in the number of missed areas during hand disinfection at baseline (intervention group: median 21, IQR 11-27, vs control group: median 20, IQR 14-21; P=.47) (Table 2). Both groups performed significantly better after training without a significant group difference (posttest 1) (median 7, IQR 4-14, in the intervention group vs median 10, IQR 6-14, in the control group; P=.34). For the secondary outcome, at posttest 2, again, no significant difference was noted between the intervention and control groups (median 14, IQR 8-17, in the intervention group vs median 11, IQR 7-16, in the control group; P=.74). In both groups, no significant difference was found between posttests 1 and 2 (intervention group, P=.11; control group, P=.25) (Table 3).
Table 2. Comparison between the VR group and the control group regarding hand disinfection, nasopharyngeal swab testing, and contamination during doffing.

<table>
<thead>
<tr>
<th></th>
<th>Values, median (IQR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VR group (n=15)</td>
<td>Control group (n=14)</td>
</tr>
<tr>
<td>Number of missing areas during hand disinfection (out of 76 possible areas)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>21 (11-27)</td>
<td>20 (14-21)</td>
</tr>
<tr>
<td>Posttest 1</td>
<td>7 (4-14)</td>
<td>10 (6-14)</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>14 (8-17)</td>
<td>11 (7-16)</td>
</tr>
<tr>
<td>Nasopharyngeal swab test score (out of 17 points)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest 1</td>
<td>14 (13-15)</td>
<td>12 (11-14)</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>14 (14-16)</td>
<td>14 (14-15)</td>
</tr>
<tr>
<td>Number of contaminated body areas during doffing (out of 10 possible areas)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest 1</td>
<td>2 (2-4)</td>
<td>3 (1-4)</td>
</tr>
<tr>
<td>Posttest 2</td>
<td>1 (0-2)</td>
<td>0 (0-1)</td>
</tr>
</tbody>
</table>

VR: virtual reality.

Table 3. Comparison between posttests 1 and 2 for the VR group and the control group.

<table>
<thead>
<tr>
<th></th>
<th>Values, median (IQR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posttest 1</td>
<td>Posttest 2</td>
</tr>
<tr>
<td>Hand disinfection (number of missing areas)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR group</td>
<td>7 (4-14)</td>
<td>14 (8-17)</td>
</tr>
<tr>
<td>Control group</td>
<td>10 (6-14)</td>
<td>11 (7-16)</td>
</tr>
<tr>
<td>Swab test (score out of 17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR group</td>
<td>14 (13-15)</td>
<td>14 (14-16)</td>
</tr>
<tr>
<td>Control group</td>
<td>12 (11-14)</td>
<td>14 (14-15)</td>
</tr>
<tr>
<td>Doffing (number of contaminated areas)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VR group</td>
<td>2 (2-4)</td>
<td>1 (0-2)</td>
</tr>
<tr>
<td>Control group</td>
<td>3 (1-4)</td>
<td>0 (0-1)</td>
</tr>
</tbody>
</table>

VR: virtual reality.

Nasopharyngeal Swab Acquisition

At posttest 1, the intervention group performed significantly better in taking a nasopharyngeal swab, scoring a median of 14 points on the 17-item checklist (IQR 13-15) versus 12 points (IQR 11-14) in the control group (P=.03) (Table 2). No significant differences between the groups were found after 1 month at posttest 2 (P=.79) for long-term retention as the secondary outcome.

The number of actual nasopharyngeal swabs performed in real life between posttests 1 and 2 did not differ between the groups (VR group, median no. of swabs 0, IQR 0-0; control group, median swabs 0, IQR 0-0; P=.56).

Contamination During Doffing

No significant difference between the number of contaminated areas during doffing was found between the groups at both time points (Table 2). However, in both groups, a significant reduction of contamination was noted at posttest 2 compared to posttest 1 (intervention group: posttest 1, median contaminated areas 2, IQR 2-4; posttest 2, median contaminated areas 1, IQR 0-2, P=.005; control group: posttest 1, median contaminated areas 3, IQR 1-4; posttest 2, median contaminated areas 0, IQR 0-1, P=.003).

Variables of Media Use

The results of the intervention survey regarding usability, satisfaction, simulator sickness, sense of presence, and immersion are detailed in Table 4.
Table 4. Variables related to media use.

<table>
<thead>
<tr>
<th></th>
<th>Values, median (IQR)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VR group (n=15)</td>
<td>Control group (n=14)</td>
</tr>
<tr>
<td>After-Scenario Questionnaire score(^b)</td>
<td>1 (1-3)</td>
<td>3 (2-4)</td>
</tr>
<tr>
<td>User Satisfaction Evaluation Questionnaire(^c)</td>
<td>27 (23-28)</td>
<td>22 (20-24)</td>
</tr>
<tr>
<td>Simulator Sickness Questionnaire, abbreviated(^d)</td>
<td>1 (1-3)</td>
<td>N/A(^e)</td>
</tr>
<tr>
<td>Presence and immersion according to Slater-Usoh-Steed(^f)</td>
<td>5 (5-5)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\(^a\)VR: virtual reality.
\(^b\)Range 1 to 7 (1=full satisfaction).
\(^c\)Range 6 to 30 (30=optimal satisfaction).
\(^d\)Range 1 to 5 (1=no simulator sickness).
\(^e\)N/A: not applicable.
\(^f\)Range 1 to 7 (7=full presence and immersion).

The ASQ revealed a significantly better result for the VR module (median score in the intervention group: 1, IQR 1-3; median score in the control group: 3, IQR 2-4; \(P=.002\)), as did the USEQ (median score in the intervention group: 23/30, IQR 23-28; median score in the control group: 22/30, IQR 0-24; \(P=.01\)).

The median score in the 4-item SSQ in the intervention group was 1 (IQR 1-3), thus revealing good tolerability of the VR simulation.

Presence and immersion in the virtual world according to the questionnaire of Slater, Usoh, and Steed was high (median 5 on a scale from 1 to 7, IQR 5-5).

**Discussion**

In this study, we demonstrated that our VR simulation was at least as effective as traditional learning methods (video and written instruction) in training medical students in COVID-19-related skills—namely, the correct performance of hand hygiene, use of PPE, and taking of a nasopharyngeal swab specimen—and that it provides a benefit in user satisfaction.

**Effectiveness of Training**

For the most investigated steps of the training, both educational methods improved performance to a similar extent without a significant difference. However, students in the VR group performed significantly better in acquiring a nasopharyngeal swab specimen on a manikin directly after the intervention than the control group, but this finding may be a result of multiple testing. Furthermore, the medians of the two groups only differed by 2 points on a 17-point outcome scale, and the IQRs overlapped.

There are few VR and MR simulations for training hand hygiene, and high quality evidence of their effectiveness is limited [39,40,42]. Shimada and colleagues [41] targeted preschool children, using a data glove instead of hand-held controllers to obtain the posture of user’s hand as a VR device. They used Leap Motion to obtain the posture of a hand, in contrast to our setting, which used only commercially available standard hardware. They found that their VR system was more effective than conventional hand hygiene instruction in a small group (n=12) of young children.

Performance of correct hand hygiene in our study was poor at baseline, in accordance with the existing evidence, stressing the need for effective instruction [7,11]. However, the performance significantly improved in both groups after the instruction, despite the limited technical possibilities to simulate the hand movements of disinfection with the need to hold the standard controllers in both hands while performing the correct movements. To minimize this limitation, participants could see the correct movements of their avatar’s hands (without the controllers) in a mirror.

This study adds to the body of evidence that VR simulation can help with the acquisition of simple skills, in combination with increased user satisfaction. Furthermore, our study highlights the necessity of strong collaborations between developers, users, and educators to ensure that these new technologies can complement and enhance existing educational curricula.

**Variables of Media Use**

Satisfaction is considered to be one of the key components of usability [48]. The satisfaction of participants in the VR group measured by the USEQ was significantly higher than that of the participants in the control group. As most of our students were inexperienced in VR, the novelty effect may have added to the results. This effect consists of an increase in perceived usability of a technology due to its newness or the tendency for performance to initially improve when new technology is instituted, not because of any actual improvement in learning or achievement but in response to increased interest in the new technology. However, Huang and colleagues [51] state that novelty does not necessarily increase learning achievement. According to them, the increase of learning achievement is more dependent on a match between the learning content and the learning method. The embodied learning method in VR is particularly appropriate for instructing difficult knowledge and spatial knowledge [51]. It remains to be elucidated whether satisfaction with and efficiency of VR simulations will decrease over time as the technology becomes more widespread.
However, with potentially increasing technological advances, a certain novelty effect will remain.

To make our VR training available to as many users as possible, we avoided using specialized hardware such as that used in many other studies of VR skill training. It could be speculated that this use of off-the-shelf controllers may reduce the realism of the VR simulation. However, we were able to demonstrate that a high degree of immersion and satisfaction could be achieved with our simulation even with standard hardware. This may be even more pronounced with future developments, such as hand tracking without the need for traditional controllers.

This “experience of presence” in VR, which could be demonstrated in our study, is known to correlate positively with training effectiveness [52,53].

**Skill Retention**

In the follow-up after 1 month, there were no significant differences between the groups regarding any outcome. This cannot be explained by different exposures, as there was no difference in the mean number of swabs performed in real life in the meantime. One possible explanation is that the participants prepared more deliberately for the second appointment, as they might have suspected that it would involve a repetition of the first assessment (“assessment drives learning”) [54]. Most motor skills are lost over time, or at least the level of performance deteriorates, starting as soon as 1 day after training [55]. Maagaard et al [56] detected that the laparoscopic skills of novices acquired in VR simulator training deteriorated in a period between 6 and 18 months without further training.

We were able to show that the observed learning effect was maintained over the observed time frame of 1 month in both learning groups. Whether there will be a difference in skill decay between the two learning methods in the long run remains an important open question.

**Strengths and Limitations**

Our study has several strengths to point out. First, the study assesses outcomes of direct clinical relevance, not surrogates such as performance on multiple-choice tests or user satisfaction only. Second, the study not only assesses the effect of training on performance gains but also includes a quantification of skill retention over time. Third, we compared the novel VR intervention to an established educational alternative rather than to no intervention.

In addition, attrition bias was nonexistent because all participants completed the study protocol without dropping out.

As a practical benefit, our VR simulation program is available for free; thus, program directors and educators are able to enhance their existing curricula with an effective novel adjunct or alternative, or replicate the study setting.

This study has several limitations, including its single center design, which restricts external validity. The number of participants in our study was limited due to the large logistical and human resources required to conduct the study during a pandemic. Therefore, the detection of small differences between training modalities is not possible with our study design. There is the possibility of selection bias, based on volunteer convenience sampling of medical students, as well as a possible performance bias, with allocation to the interventional group leading to higher motivation, satisfaction, and performance. The need to use hand-held controllers instead of hand-tracking might have further impacted the efficiency of the VR simulation; however, we wanted to apply technological equipment that is widely available.

Furthermore, the correlation of these findings to clinical, patient-oriented outcomes remains to be validated.

**Conclusion**

To our knowledge, this is the first study using a VR simulation to train health care personnel in the correct use of hand hygiene, PPE, and taking a nasopharyngeal swab specimen and to compare the effectiveness to established traditional training (video and written instructions). VR simulation was at least as effective as traditional learning methods in training medical students while providing a benefit in user satisfaction. These results add to the growing body of evidence that VR is a useful tool for acquiring and maintaining simple and complex clinical skills.

**Conflicts of Interest**

WEH has received research funding from the European Union, the Swiss National Science Foundation, the Zoll Foundation, Dräger Medical Germany, Mundipharma Research UK, MDI International Australia, and Roche Diagnostics Germany, all outside the submitted work. WEH has provided paid consultancies to AO Foundation Switzerland, outside the submitted work. WEH received financial support for a congress he chaired from EBSCO Germany, Isabel Healthcare UK, Mundipharma Medical Switzerland, and VisualDx USA, all outside the submitted work. TCS holds an endowed professorship supported by the Touring Club Switzerland. The sponsor has no influence on the research conducted, in particular on the results or the decision to publish. GP is Chief Technical Officer of ORamaVR, the company that developed the software for this study. The company had no influence on the study protocol, the analysis, the results, or the decision to publish. This study was funded in part by a grant from the Swiss National Science foundation (grant number 31CA30_196615) for the project “Mixed-method evaluation of an online forward triage tool within the COVID-19 pandemic” to TCS and WEH. The other authors report no conflict of interest.

**Editorial note:** This randomized study was not registered, explained by authors as follows: “As our trial was not a clinical trial and did not involve patients or patient data, but instead was a pilot study involving medical students, it was not registered as a clinical trial.” The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, given
that the subjects were medical students. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness.

Multimedia Appendix 1
Checklist used for performance assessment.

[DOCX File, 13 KB - games_v9i4e29586_app1.docx ]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V. 1.6.1).

[PDF File (Adobe PDF File), 693 KB - games_v9i4e29586_app2.pdf ]

References
15. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst 1994 Dec 24;E77-D(12):9 [FREE Full text]


Abbreviations

ASQ: After-Scenario Questionnaire
CDC: US Centers for Disease Control
CVRSB: Covid-19 VR Strikes Back
HMD: head-mounted device
MR: mixed reality
PPE: personal protective equipment
SSQ: Simulator-Sickness Questionnaire
USEQ: User Satisfaction Evaluation Questionnaire
VR: virtual reality
WHO: World Health Organization
Please cite as:
Birrenbach T, Zbinden J, Papagiannakis G, Exadaktylos AK, Müller M, Hautz WE, Sauter TC
Effectiveness and Utility of Virtual Reality Simulation as an Educational Tool for Safe Performance of COVID-19 Diagnostics: Prospective, Randomized Pilot Trial
JMIR Serious Games 2021;9(4):e29586
URL: https://games.jmir.org/2021/4/e29586
doi:10.2196/29586
PMID:34623315

©Tanja Birrenbach, Josua Zbinden, George Papagiannakis, Aristomenis K Exadaktylos, Martin Müller, Wolf E Hautz, Thomas Christian Sauter. Originally published in JMIR Serious Games (https://games.jmir.org), 08.10.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Software Engineering Frameworks Used for Serious Games Development in Physical Rehabilitation: Systematic Review

Abstract

Background: Serious games are a support in the rehabilitation process for treating people with physical disabilities. However, many of these serious games are not adapted to the patient’s needs because they are not developed with a software engineering framework with a set of activities, actions, and tasks that must be executed when creating a software product. Better serious games for rehabilitation will be developed if the patient and therapist requirements are identified, the development is planned, and system improvements and feedback are involved. The goal is that the serious game must offer a more attractive environment, while maintaining patient interest in the rehabilitation process.

Objective: This paper submits the results of a systematic review of serious games in physical rehabilitation identifying the benefits of using a software engineering framework.

Methods: A systematic research was conducted using PubMed, PEDro (Physiotherapy Evidence Database), IEEE Xplore, ScienceDirect, ACM Digital Library, Mary Ann Liebert, Taylor & Francis Online, Wiley Online Library, and Springer databases. The initial search resulted in 701 papers. After assessing the results according to the inclusion criteria, 83 papers were selected for this study.

Results: From the 83 papers reviewed, 8 used a software engineering framework for its development. Most of them focused their efforts on 1 or more aspects, such as data acquisition and processing, game levels, motivation, therapist supervision.

Conclusions: This systematic review proves that most of the serious games do not use a software engineering framework for their development. As a result, development systems overlook several aspects and do not have a standardized process, eventually omitting important implementation aspects, which impact the patient’s recovery time.

(JMIR Serious Games 2021;9(4):e25831) doi:10.2196/25831

KEYWORDS

serious game; physical rehabilitation; framework; methodology
Overview

According to the World Health Organization, over 1 billion people have some form of disability [1], with up to 200 million people having loss or decrease in movement, which limits their ability to perform activities of daily living. To overcome it, they must undergo a rehabilitation program to gradually regain movement and consequently, improve their quality of life.

However, the traditional rehabilitation process is often slow and presents problems such as lack of motivation, boredom, and others; as a result, many patients consider the exercises stressful, and therefore abandon the therapy [2].

To avoid these situations, new ways of conventional therapy support have been used in recent years, such as medicinal treatments, robotics, video games (known as serious games), and others [3], which have contributed to faster rehabilitation when performing exercises in a fun way, allowing the patients to forget their conditions and concentrate on the game.

For this reason, new interaction modes, such as serious games [4], have the potential to provide more attractive, motivating, and enriching experiences for patients who suffer from decreases in movement. Currently, serious game–based physical rehabilitation is an area of research in constant evolution, and therefore, there is the need for developing guidelines adapted from other research fields.

Despite the potential benefits of serious games in physical rehabilitation, many available platforms are inflexible and limited in their scope. Many developments do not follow a process involving a set of activities, actions, or tasks that must be executed when a software product is to be created. As a result, essential elements to the patient’s improvement process are ignored within the video game. Some of these elements are motivation, play levels, player commitment, challenges according to the patient’s level, clinical evaluation, assessment scales, among others [5,6].

This work aims to describe the software engineering frameworks used in serious games development and their benefits in the physical rehabilitation process.

Background

A Note on Frameworks

The term framework has several meanings depending on the field. For example, it may refer to a model, prescription, guidelines underlying a design and analysis, among others. The concept of framework is widely used in the field of computer science. However, there is some confusion between the software engineering framework and the application framework. The former provides a skeletal abstraction of a solution to several problems that have some similarities. A software engineering framework will generally outline the steps or phases that must be followed in implementing a solution without getting into the details of what activities are done in each phase [7]. The goal is for developers to use the framework as a guide to creating software systems by applying “building blocks” depending on the problem domain; by contrast, application framework is an integrated set of software artifacts (such as classes, objects, and components) that collaborate to provide a reusable architecture for a family of related applications [8]. They are used to facilitate the development process of applications, reducing time, effort, and costs.

Software engineering framework and application framework should not be confused. The latter is composed of pre-established source codes (eg, data access routines, form validation, templates) that the programmer uses to reduce workload and do not start the project from scratch.

One of the main motivations for applying a software engineering framework in serious game development is to design an efficient and satisfactory system for the patient.

Software Engineering Frameworks and Serious Games

The use of software engineering frameworks for the development of serious games allows the application of a variety of concepts, models, techniques, and artifacts at a high level of abstraction. Being an interdisciplinary field, an orientation on the developed tasks is required. Besides, it is flexible to adapt to changing conditions or personalization according to the final approach of the video game (rehabilitation, education, etc.).

Serious games like other software developments require a “systematic, disciplined, and quantifiable” approach. Every aspect of production, from early stages of system specification to maintenance after its operation, must be established. Below is a set of related activities that lead to the development of a software product [9-12].

Structural Activities in Software Development

In software engineering, 5 generic structural activities are used during software development [9-12]: communication, planning, modeling, construction, and deployment. The software process details will be different in each case, but the structural activities are the same. The definitions of the structural activities are presented inTextbox 1.
Textbox 1. Definitions of the structural activities in software development.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Defining the software characteristics and functions is particularly important to communicate and collaborate with the client and other participants. This activity aims to understand the project objectives of the participants and meet the requirements.</td>
</tr>
<tr>
<td>Planning</td>
<td>Once the requirements are obtained, this activity presents an estimate of the resources; establishes a software project plan; and describes technical tasks, probable risks, and program activities.</td>
</tr>
<tr>
<td>Modeling</td>
<td>Its objective is to help understand the requirements through models. The models’ aim is to affirm the understanding of the work and give technical guidance to those who will implement the software, establishing, for example, the database model, the software architecture, user screen prototypes, and others. In some developments, this activity is the equivalent of the design stage.</td>
</tr>
<tr>
<td>Construction</td>
<td>This activity consists of the code generation and tests required to discover bugs in the software product.</td>
</tr>
<tr>
<td>Deployment</td>
<td>Once the software is created (completely or an increment), it is delivered to the client who will evaluate it and give feedback for system improvement.</td>
</tr>
</tbody>
</table>

Gamification

According to Kumar [13] gamification applies game design principles and mechanics to nongame environments. In the rehabilitation process, gamification can increase motivation and engagement through rewards, game levels, accessibility, feedback, and challenge. Therefore, the software engineering framework for serious game development must incorporate gamification. Various gamification elements include immersion, support for different roles, flow enhancement, visual enhancement, support for different learning stages and experience levels, design for interactivity, and progress [14]. By contrast, Vermeir et al [15] identified the following elements: avatar, challenge, competition, difficulty adjustment, feedback loops, levels, progress, rewards, social interaction, sound effects, and story/theme.

Benefits of Gamification in Rehabilitation

de Castro-Cros et al [16] analyzed the effects of gamification on the mental imagery brain–computer interface in rehabilitation functional assessments in 10 patients with stroke with hemiparesis in the upper limb and 6 healthy individuals. The authors concluded that user opinions about the game level of entertainment, clarity of rules, narrative, and visual attractiveness were all positive. The patients were consensus about the interest in gamifying stroke rehabilitation sessions. By contrast, Steiner et al [17] performed a scoping review of gamification in the rehabilitation of patients with musculoskeletal disorders of the shoulder. They concluded that gamification is essential in health care to enhance motivation and support therapy in general, especially in chronic diseases and rehabilitation. Other advantages are motivation, avoiding boredom, and distraction from pain and anxiety.

Related Works

A systematic review of literature is a method to identify, evaluate, and interpret all available and relevant research of a particular research question, subject area, or phenomenon of interest. The individual studies that contribute to the systematic review are called primary studies. A systematic review is also considered a form of secondary study [18].

This systematic review includes literature work on developing serious games in physical rehabilitation using a software engineering framework. To identify existing secondary research in the same field, we searched the following electronic databases: IEEE Xplore, ACM Digital Library, Wiley Digital Library, PubMed, ScienceDirect, Taylor & Francis, Mary Ann Liebert, and Springer. Besides, we used Google Scholar as a web source to broaden our results.

The search was realized using the following search string: A1 AND B1 AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6). Textbox 2 shows the terms included in the search string.
When this search was performed in the electronic databases, no related secondary studies were identified. Therefore, we sought systematic reviews focused on software engineering frameworks in any field. Table 1 summarizes the secondary studies found.

Mubin et al [19] performed a review on gamification design framework and its application for children with autism. This review aimed to offer gamification solutions for interaction skills. They identified the framework phases in 5 papers and target users/audience/focus. The authors concluded that frameworks have been analyzed from an in-game context but did not emphasize on children with autism. In the literature, studies show that gamification is very effective in the areas of therapy and education for children with autism. The most important contribution of this review is the development of interaction skills. This review identified phases of the development process in some studies (eg, planning, designing). However, it does not explain how users benefit from the process interaction.

Vargas et al [20] developed a systematic mapping study on serious game quality. The aim was to discover the current state of serious games quality initiatives. One of the research questions focused on discovering if quality has been constant throughout the software development cycle or in some stages. The authors showed that 97% of the literature reviewed applied quality in the final phase (product). Only 7.14% focused on quality in the design phase and 1.79% in the requirement phase. This study was included because it identified the phases in which quality was applied: requirement, design, code, and final product.

Tomalá-Gonzáles et al [21] reported on methodologies, game engines currently used in serious games development in various areas (education, cognitive disabilities, and physical rehabilitation), and criteria for game engine selection. From the 27 papers, 8 used a defined methodology such as XP, Cascade, and others, while 3 proposed their own model. The authors concluded that although several software development methodologies can be adapted to serious game development, the best option was the SUM methodology because it is based on Scrum (fast, precise, optimized, and adaptable programming characteristics). However, this review did not make distinctions between framework and methodology. It also did not identify methodology phases nor the benefits of applying a methodology in the learning or rehabilitation process.

Although our work shares similarities with the aforementioned studies, the literature review presented in this paper is different because this review (1) focuses on serious games for physical rehabilitation, (2) identifies the software development stages in each software engineering framework according to the structural activities proposed by Pressman [9], who states that “The software process details will be different in each case, but the structural activities are the same”; (3) identifies contributions of software engineering frameworks to the rehabilitation process; and (4) identifies if the proposed software engineering framework provides objective monitoring of the rehabilitation process.
Methods

Research Methodology

The systematic literature review process proposed by Brereton et al [22] was applied for this systematic review. Figure 1 shows the process and steps for each phase. The process consists of 3 main phases: plan review, conduct review, and document review. The first phase consists of the following steps: (1) describe the main reasons for the literature review, (2) specify a set of research questions, and (3) review the protocol. The second phase comprises 4 steps: (1) identify important research, (2) select primary studies, (3) extract data from primary studies, and (4) synthesize data. Finally, the third phase consists of 3 steps: (1) obtain results, (2) identify the validity threats, and (3) conclusions. Figure 1 shows the literature review process. In the following subsections, we describe the activities carried out in each phase of this systematic literature review.

Research Questions

In this subsection, we present the 9 research questions that guided this study through the investigation to meet the objectives of the systematic review. Table 2 presents these questions. The research questions can be classified into 4 fields of interest. RQ1 and RQ2 study serious games evaluated in software engineering. These questions identify the number of serious games developed with a software engineering framework and the set of activities, actions, and tasks required.

RQ3 and RQ4 describe framework contributions to the rehabilitation process and implementation of gamification elements. It allows transforming obstacles into positive and fun reinforcements, thereby encouraging patients.

RQ5 and RQ6 are centered on applicability and serious game characteristics for rehabilitation using a software engineering framework. These questions identify relevant data such as target audience, interaction technology for data acquisition, main modalities, among others.

Finally, RQ7, RQ8, and RQ9 studied important aspects to evaluate and provide follow-up of rehabilitation progress depending on the type of exercise.
Table 2. Research questions.

<table>
<thead>
<tr>
<th>Research question</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>What framework is used in the development of the serious game?</td>
</tr>
<tr>
<td>2</td>
<td>What are the generic structural activities used in frameworks?</td>
</tr>
<tr>
<td>3</td>
<td>How the framework contributes to the rehabilitation process?</td>
</tr>
<tr>
<td>4</td>
<td>What gamification elements does the framework use?</td>
</tr>
<tr>
<td>5</td>
<td>What is the targeted disability contemplated in the frameworks?</td>
</tr>
<tr>
<td>6</td>
<td>If the framework includes a case study, which part of the body is rehabilitated? What is the modality of the serious game? Which interaction technology is used?</td>
</tr>
<tr>
<td>7</td>
<td>What type of evaluation and number of patients are involved in the clinical trials?</td>
</tr>
<tr>
<td>8</td>
<td>Does the framework contemplate a standardized scale to evaluate the patient’s rehabilitation progress?</td>
</tr>
<tr>
<td>9</td>
<td>Does the framework contemplate adaptability?</td>
</tr>
</tbody>
</table>

Search Strategy

The objective of the search strategy was to identify all relevant primary studies. A literature search was conducted to answer the proposed research questions.

The search strategy is an adaptation of Guidelines for Performing Systematic Literature Reviews in Software Engineering [18] and Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [23]. Relevant papers were identified by searching in the following databases: PubMed, PEDro (Physiotherapy Evidence Database), IEEE Xplore, ScienceDirect, ACM Digital Library, Mary Ann Liebert, Taylor & Francis Online, Wiley Online Library, and Springer. To build the search string, a list of keywords and their synonyms were identified. Logical operators (AND and OR) and words related to rehabilitation, serious games, and framework were used. The final search strings consisted of the following Boolean expressions: “(A1 AND (B1 OR B2)) AND (C1 OR C2 OR C3) AND D1”. The search terms are shown in Textbox 3.

Textbox 3. Search terms for the final search string.

A term
- A1. Serious

B term
- B1. Game
- B2. Games

C term
- C1. Rehabilitation
- C2. Disability
- C3. Disabilities

D term
- D1. Framework

Inclusion Criteria

The systematic review is focused on serious games for physical rehabilitation; clear inclusion criteria were established to determine the eligibility of papers for inclusion in the review. Only studies with the following criteria were considered eligible for inclusion: serious game papers for physical rehabilitation, papers published in English, and all serious games regardless of the year of development.

Exclusion Criteria

Papers duplicated, papers regarding opinion pieces, existing literature reviews, papers that are not related to rehabilitation using serious games, serious games for educational purposes, and serious games for cognitive rehabilitation were excluded from the study.

Study Selection

First, the search string was used in different databases. Potentially relevant papers were identified after reading the title and abstract. Duplicate papers were removed. Subsequently, an exhaustive verification of compliance with the inclusion and exclusion criteria was carried out to select the papers. Figure 2 shows the item selection process. In the systematic review, 701 papers were included. Table 3 shows the number of documents retrieved from each database.
Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)–based flowchart.
Table 3. Search results.

<table>
<thead>
<tr>
<th>Databases</th>
<th>Results, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>14</td>
</tr>
<tr>
<td>PEDro</td>
<td>12</td>
</tr>
<tr>
<td>IEEE Xplore</td>
<td>103</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>88</td>
</tr>
<tr>
<td>ACM Digital Library</td>
<td>166</td>
</tr>
<tr>
<td>Mary Ann Liebert</td>
<td>27</td>
</tr>
<tr>
<td>Taylor &amp; Francis Online</td>
<td>50</td>
</tr>
<tr>
<td>Wiley Online Library</td>
<td>43</td>
</tr>
<tr>
<td>Springer</td>
<td>198</td>
</tr>
</tbody>
</table>

Extract Data From Primary Studies

After identification, the primary papers were rigorously analyzed in accordance with the following considerations: (1) only the authors of this review can participate in the data collection process; (2) each primary paper should be reviewed with at least two reviewers; (3) each reviewer will collect a set of data from each primary study, then meet with another reviewer to reach an agreement on the data obtained.

Table 4. Bibliographic data of the primary papers.

<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Country</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranyi et al [24]</td>
<td>2013</td>
<td>Austria</td>
<td>IEEE Xplore</td>
</tr>
<tr>
<td>Pirovano et al [25]</td>
<td>2016</td>
<td>Italy</td>
<td>ScienceDirect</td>
</tr>
<tr>
<td>Amengual Alcover et al [26]</td>
<td>2018</td>
<td>Spain</td>
<td>PubMed</td>
</tr>
<tr>
<td>Baranyi et al [27]</td>
<td>2019</td>
<td>Austria</td>
<td>IEEE Xplore</td>
</tr>
<tr>
<td>Zain et al [28]</td>
<td>2012</td>
<td>Malaysia</td>
<td>IEEE Xplore</td>
</tr>
<tr>
<td>Noveletto et al [29]</td>
<td>2018</td>
<td>Brazil</td>
<td>ScienceDirect</td>
</tr>
<tr>
<td>Maggiorini et al [31]</td>
<td>2012</td>
<td>Switzerland</td>
<td>ACM Digital Library</td>
</tr>
</tbody>
</table>

Results

RQ1: What Framework Is Used in the Development of the Serious Game?

Only 8 (10%) out of the 83 papers related to physical rehabilitation using a software engineering framework (Multimedia Appendix 2).

In Baranyi et al [24,27], the proposed studies were based on the user-centered design framework. The physiotherapist is important because s/he identifies the needs and limitations of the patients in the rehabilitation process. There are 3 phases: research, design, and evaluation. In research, a physiotherapist conducts brainstorming with the work team and identifies the requirements. Afterward, in the design phase, the team creates mock-ups and a prototype. Finally, the physiotherapist evaluates the application.

Pirovano et al [25] proposed a 4-step procedure to create safe exergames for rehabilitation therapies: exercise definition, virtualization, game design, and secondary goals. In exercise definition, a set of exercises is proposed to cover therapy needs. Each exercise is structured into primary and secondary goals. During virtualization, the team identifies primary goals, and they are implemented into a virtual exercise by defining input (tracking) and output (feedback) requirements through simple graphical elements and specifying interaction mechanisms. Through game design, the virtual exercise is converted into a true exergame. In the last step, there are 2 functionalities. The first is to analyze motion data and identify wrong movements. The second provides feedback to the patients.

In Amengual Alcover et al [26], the serious game development framework follows an iterative process flow structured into 2 dimensions: activities and incremental development. The first dimension is based on 3 approaches: Scrum, the web application development model, and a clinical trial. The activities dimension includes a project initiation activity, an iterative flow composed of 4 basic development activities (planning and control, modeling, construction, and evaluation), and a final clinical
study to evaluate the rehabilitation process of the patient through the serious game. Incremental development includes 3 different increments: interaction mechanism, interaction elements, and serious game. In the first increment, an existing device on the market is identified to capture the movements of patients according to their needs. In the second increment, the development team must design the interaction elements that force patients to perform the therapy correctly. The final increment is aimed at designing a serious game that motivates the patient to perform therapy to obtain the best results.

Zain et al [28] proposed a conceptual framework for people with motor impairment, so they can enjoy the experience of playing serious games. The framework’s main elements were player skills, challenge, concentration, feedback, immersion, learning opportunities, accessibility, and adaptivity. The proposed framework will help the game designer and developer create a serious game that combines the game’s technology with the learning environment. This framework is based on the game flow model.

Noveletto et al [29] presented a conceptual model for the design or development of serious games to rehabilitate people with stroke. The framework establishes a relationship between experts and patients to obtain the requirements, considering that the biomedical device and the video game score are used to design serious games.

Afyouni et al [30] proposed a framework consisting of a therapy-driven 3D environment augmented with a natural user interface based on movement. The framework incorporated different adaptation techniques to adjust patient’s needs. Patient preferences and limitations were considered key parameters for changing the game, thereby creating personalized games for each patient.

Maggiorini et al [31] presented a framework for serious game development that allows the therapist to remotely control the video game home activities. The objective was to create a more attractive game for the elderly with easily adjustable parameters for therapy adaptability. The framework includes 3 phases of serious game development: requirements definition, empirical validation of requirements list, and design and prototyping.

RQ2: What Are the Generic Structural Activities Used in Frameworks?

The objective of this research question was to identify generic structural activities in primary studies (see the “Background” section). Table 5 summarizes the structural activities and Multimedia Appendix 3 shows the frequency of occurrence of each structural activity in primary studies.

Every study established a communication activity to obtain the requirements. Baranyi et al [24] brainstormed with a physiotherapist to identify relevant problems and needs for patients undergoing rehabilitation. Pirovano et al [25] defined exercises addressing the primary and secondary objectives of rehabilitation. To achieve maximum effectiveness, the exercises are defined in collaboration with therapists. In Amengual Alcover et al [26], the communication began by identifying the context, operational objectives, restrictions, and requirements. Baranyi et al [27] established communication with the therapist to obtain the requirements. Zain et al [28] identified the user abilities, limitations, and behavior, which become requirements for the serious game. Noveletto et al [29] considered experts in the field (health personnel, therapists, etc.) and patients to obtain the requirements. Afyouni et al [30] established the type of game through patient needs, preferences, and limitations, allowing custom game features. Finally, Maggiorini et al [31] analyzed the most diffused issues present in elders’ homes (eg, size of rooms, habits) to explore requirements and limitations through an immersive approach.

The planning activity was implemented in Amengual Alcover et al [26]. The goal of this activity was to determine the tasks to perform during the development by identifying the end products and the people who will do the work. The activity includes 3 tasks: planning, scheduling, and tracking.

The modeling activity was performed in several papers. For example, Baranyi et al [24] called it design, elaborating basic models discussed with a therapist. Pirovano et al [25] transformed the exercise requirements into a true exergame by adding all the elements and characteristics of a game and a good game design for the patients. Amengual Alcover et al [26] created models that helped the development team to understand the requirements obtained and the game design. By contrast, Baranyi et al [27] contemplated the use of prototypes to refine user requirements. Finally, Maggiorini et al [31] established a list of technical characteristics (desired) for the prototype creation.

The construction activity was implemented in every study. Developments produce executable software units that will be used by users, through the creation of prototypes to improve the software [24-27,30,31], or the final product [28,29].

Finally, the user evaluates and provides feedback on the serious game in the deployment stage. In the primary papers, Pirovano et al [25] and Baranyi et al [27], patients were asked to give their opinion to improve the game design and change some aspects of the application.
Table 5. Structural activities in primary studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Communication</th>
<th>Planning</th>
<th>Modeling</th>
<th>Construction</th>
<th>Deploy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranyi et al [24]</td>
<td>X</td>
<td>_a</td>
<td>X</td>
<td>X</td>
<td>—</td>
</tr>
<tr>
<td>Pirovano et al [25]</td>
<td>X</td>
<td>—</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Amengual Alcover et al [26]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>—</td>
</tr>
<tr>
<td>Baranyi et al [27]</td>
<td>X</td>
<td>—</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Zain et al [28]</td>
<td>X</td>
<td>—</td>
<td>—</td>
<td>X</td>
<td>—</td>
</tr>
<tr>
<td>Noveletto et al [29]</td>
<td>X</td>
<td>—</td>
<td>—</td>
<td>X</td>
<td>—</td>
</tr>
<tr>
<td>Afyonu et al [30]</td>
<td>X</td>
<td>—</td>
<td>—</td>
<td>X</td>
<td>—</td>
</tr>
<tr>
<td>Maggiorini et al [31]</td>
<td>X</td>
<td>—</td>
<td>X</td>
<td>X</td>
<td>—</td>
</tr>
</tbody>
</table>

*Not available.

**RQ3: How the Framework Contributes to the Rehabilitation Process?**

Baranyi et al [24,27] applied a user-centered design approach to establish constant communication with the physiotherapist who has the experience to identify the needs and limitations of the final user. Serious games are developed with entertainment elements such as levels, rewards, challenges, and adaptability to the patient need, considering special conditions.

Pirovano et al [25] proposed the creation of safe exergames, identifying the needs of real exercise besides therapy goals. These needs are incorporated into a video game considering the primary objectives (what a user should do) and secondary objectives (how user actions should be carried out). The former is easily integrated into the gameplay, while the latter aids the patient with corrections or prevention of compensatory movement through analysis of the flow of movement data and wrong movements in real time, thereby providing immediate feedback to patients to correct themselves during the exercise.

Amengual Alcover et al [26] proposed an iterative, prototype-oriented, systematized serious game development process. The proposed process guarantees that products based on this framework are developed and validated by following a coherent and systematic method that leads to high-quality serious games.

For users with motor impairment, Zain et al [28] used flow theory [32] to propose user interface design factors that make their experience enjoyable when playing serious games. This framework includes user interface design factors and aims to establish a conceptual model that can be used by a game designer for efficient game development or an educational practitioner when designing enjoyable serious games for users with motor impairment.

Noveletto et al [29] established a relationship among key stakeholders (experts and patient) and elements (biomedical device and game score) for serious game design. The framework states that a correlation between the game score and clinical tests can aid treatment and evaluation through the biomedical system.

Afyounu et al [30] proposed a framework for video game development with an adaptive and user-centered approach. The framework embeds different adaptation techniques to tailor to patients’ needs. The video game adapts to the difficulty level based on the patient’s profile and performance in real time. Other aspects such as patient preferences and constraints are considered as key game-changing parameters.

Finally, in Maggiorini et al [31], the framework allowed serious game development with telerehabilitation allowing the therapist to remotely control the video game home activities. It supports parameter adjustments for therapy adaptability. Table 6 summarizes framework contributions.
Table 6. Framework contributions in primary studies.

<table>
<thead>
<tr>
<th>Framework contribution to rehabilitation</th>
<th>Utility</th>
<th>Primary studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication with health expert</td>
<td>A physiotherapist establishes communications with patients undergoing rehabilitation to identify the problems and needs.</td>
<td>Baranyi et al [24], Pirovano et al [25], Baranyi et al [27], Noveletto et al [29]</td>
</tr>
<tr>
<td>Exercise definition</td>
<td>Exercise can be defined as a sequence of different actions needed to complete it to achieve maximum effectiveness.</td>
<td>Pirovano et al [25]</td>
</tr>
<tr>
<td>Analyzes the stream of motion data and identifies in real time wrong movements</td>
<td>Provides immediate feedback to the patients for correct exercising.</td>
<td>Pirovano et al [25]</td>
</tr>
<tr>
<td>Iterative and prototyping</td>
<td>Visualize prototypes of serious games from early stages. The therapist or patients identify additional requirements or modify them.</td>
<td>Baranyi et al [24], Pirovano et al [25], Baranyi et al [27], Noveletto et al [29], Afyouni et al [30], Maggiorini et al [31]</td>
</tr>
<tr>
<td>User interface design factors</td>
<td>Motivation and immersion</td>
<td>Baranyi et al [24], Pirovano et al [25], Amengual Alcover et al [26], Baranyi et al [27], Zain et al [28], Noveletto et al [29]</td>
</tr>
<tr>
<td>The correlation between game score and clinical tests</td>
<td>Aids in patient treatment and evaluation</td>
<td>Noveletto et al [29]</td>
</tr>
<tr>
<td>Adaptive approach</td>
<td>Adapts difficulty level according to the patient’s profile and performance in real time</td>
<td>Baranyi et al [24], Pirovano et al [25], Zain et al [28], Afyouni et al [30], Maggiorini et al [31]</td>
</tr>
<tr>
<td>Telerehabilitation</td>
<td>Therapists can remotely control the video game for home activities and provide adjustable parameters to improve therapy</td>
<td>Maggiorini et al [31]</td>
</tr>
</tbody>
</table>

RQ4: What Gamification Elements Does the Framework Use?

Overview
Gamification allows the transformation of obstacles into positive and fun reinforcement, encouraging users to make the right decisions for their health and well-being [33]. It is essential to keep the patient motivated in physical rehabilitation. For this reason, the software engineering framework is required to use gamification elements. The papers identified the following elements: feedback, motivational factor, adaptability, challenge, levels, immersion, rewards, concentration, and avatar. Table 7 shows the gamification elements in primary studies, and Multimedia Appendix 4 shows the frequency of occurrence of each gamification element.

The gamification elements of primary studies are described below.

Table 7. Gamification elements in primary studies.

<table>
<thead>
<tr>
<th>Gamification element</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback</td>
<td>Pirovano et al [25], Amengual Alcover et al [26], Baranyi et al [27], Zain et al [28], Noveletto et al [29], Afyouni et al [30], Maggiorini et al [31]</td>
</tr>
<tr>
<td>Motivational factor</td>
<td>Baranyi et al [24], Pirovano et al [25], Amengual Alcover et al [26], Baranyi et al [27], Noveletto et al [29]</td>
</tr>
<tr>
<td>Adaptability</td>
<td>Baranyi et al [24], Pirovano et al [25], Zain et al [28], Afyouni et al [30], Maggiorini et al [31]</td>
</tr>
<tr>
<td>Challenge</td>
<td>Baranyi et al [24], Zain et al [28], Afyouni et al [30]</td>
</tr>
<tr>
<td>Levels</td>
<td>Baranyi et al [24], Amengual Alcover et al [26], Baranyi et al [27], Afyouni et al [30]</td>
</tr>
<tr>
<td>Immersion</td>
<td>Zain et al [28]</td>
</tr>
<tr>
<td>Rewards</td>
<td>Pirovano et al [25]</td>
</tr>
<tr>
<td>Concentration</td>
<td>Zain et al [28]</td>
</tr>
<tr>
<td>Avatar</td>
<td>Pirovano et al [25]</td>
</tr>
</tbody>
</table>

Feedback
In Pirovano et al [25], the feedback mechanisms were designed to show the outcome of actions to patients. For instance, whether a target is met or a movement has been successfully performed. Amengual Alcover et al [26] used “mirror feedback,” which consists of projecting the user onto the screen and simulating a mirror in such a way that the users can see themselves on the screen at all times. In Baranyi et al [27], the feedback provided was either visual, aural, or haptic. In Zain et al [28], users with motor impairment received feedback on their progress, and when they lose the game, feedback is provided to continue in the right direction. Noveletto et al [29] established that serious same should reward players with feedback on progress. Afyouni et al [30] used a scoring system that was designed to keep track of the number of times the patient successfully passed through
the targets. Finally, in Maggiorini et al [31], a skeleton wireframe is drawn in red to provide immediate visual feedback, and an alarm is raised on the screen.

**Motivational Factor**

Baranyi et al [24] used “goals.” The gameplay was based on achieving goals that should act as motivation factors. Pirovano et al [25] established that extrinsic motivational effects can be achieved through careful use of verbal praise, scoring mechanisms, and virtual reward systems. In Amengual Alcover et al [26], the development of new serious games allowed the inclusion of motivational elements to increase engagement. Baranyi et al [27] used rewards in serious games for the user. Finally, Noveletto et al [29] used the “motivational score” to improve attention during rehabilitation sessions.

**Adaptability**

Baranyi et al [24] proposed an adaptive system with the opportunity to adapt the game difficulty. Pirovano et al [25] established that virtual exercises should use dynamic difficulty adaptation, thus further increasing the flexibility of serious games. For Zain et al [28], an adaptive factor was important to design and develop serious games for users with motor impairment because the application, aware of the users’ current cognitive load and physical limitations, can change its response, presentation, and interaction flow to improve users’ experience and their task performance. In Afyouni et al [30], the framework embeds different adaptation techniques to adapt to the patients’ needs. Key game-changing parameters such as patient preferences and constraints are considered. This allows the creation of personalized game features for every patient. Maggiorini et al [31] proposed that remotely controlled serious games may also provide easily tunable parameters to better adapt the game therapy to the actual patient recovery.

**Challenge**

Baranyi et al [24] proposed the challenge as a “key fact.” They considered that the game should not be too easy nor too hard to manage. The game should be sufficiently challenging and match the player’s skill level. Zain et al [28] proposed that serious games should also vary the level of difficulty and keep an appropriate pace. Afyouni et al [30] generated therapy-aware navigational movements with multiple levels of difficulty.

**Levels**

Baranyi et al [24] stated that the purpose of the serious game developed is to have a rehabilitation system containing different levels that were adapted and created for the individual needs of the patients and to fit their impairments. Amengual Alcover et al [26] considered that serious games must have a definition of different levels in the game. In Baranyi et al [28], when the game is started for the first time, a diagnostic routine is performed; using these data, a baseline for the exercises can be defined by the therapist to get an initial idea about how easy or complex a level might be for a patient. Afyouni et al [30] presented different levels of difficulty based on therapeutic gestures and patient performance.

**Immersion**

Zain et al [28] considered that immersive games draw players into the game and affect their senses through elements such as audio and narrative.

**Reward**

Pirovano et al [25] used a scoring system, and at the end of each exergame, a virtual reward is presented to the patients.

**Concentration**

Zain et al [28] considered that the more concentration a task requires in terms of attention and workload, the more absorbing it will be. The games should grab the player’s attention quickly and maintain it throughout the game.

**Avatar**

Pirovano et al [25] used an avatar for feedback on wrong movements, changing the color of the associated avatar segments. When wrong movements persist for a long time, the game is paused, and a virtual therapist avatar pops up to advise patients.

**RQ5: What Is the Targeted Disability Contemplated in the Frameworks?**

This specifies whether a study focuses on a particular pathology with loss or decrease in movement. The papers established the following target pathology: 4 defined strokes [24,25,27,29], 2 defined neuromotor disorder [26,30], 1 defined users with motor impairment [28], and 1 defined rehabilitation of the elderly [31]. Stroke is mainly targeted in studies because it is the second cause of death and the third cause of disability worldwide [34].


As Table 8 reports, Baranyi et al [24] presented a prototype that rehabilitates patients with lower limb disabilities with balance and strength exercises using the Wii Fit Balance Board. Pirovano et al [25] developed serious games for upper limb motor rehabilitation therapy using Microsoft Kinect and lower limb with the Wii Fit Balance Board. Amengual Alcover et al [26] also rehabilitated the lower limb by allowing patients to perform repetitions in a video game controlled with Microsoft Kinect, with each repetition varied according to the participant’s tolerance and the physiotherapist’s recommendations. Baranyi et al [27] performed hand rehabilitation using gesture exercises, touch, and patient movement levels using mobile phone sensors. Zain et al [28] and Noveletto et al [29] did not report any case studies. Afyouni et al [30] developed a serious game for hand rehabilitation using leap motion. Game instructions can be visual (shown on the screen) or voice, depending on the perception capacity of the patient. Finally, Maggiorini et al [31] developed a prototype for rehabilitation using Microsoft Kinect. It only presents the skeleton tracking by a sensor and does not mention whether the video game implements another form of communication with the patient.
The modality is a way in which information is transmitted from the computer to the participants [35]. Baranyi et al [24,27], Pirovano et al [25], Amengual Alcover et al [26], Afyouni et al [30], and Maggiorini et al [31] used a visual modality, presenting a graphical interface for user interaction. Pirovano et al [25], Amengual Alcover et al [26], Baranyi et al [27], and Afyouni et al [30] used audio effects such as music or voice instructions. Baranyi et al [27] used haptic modality to control the video game through a touch screen. Zain et al [28] and Noveletto et al [29] did not report modalities.

Table 8. Rehabilitated limb, serious game modality, and data-acquisition device in primary studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Rehabilitation/extremity</th>
<th>Modality</th>
<th>Interaction technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranyi et al [24]</td>
<td>Lower limbs</td>
<td>Visual</td>
<td>Wii Fit Balance Board</td>
</tr>
<tr>
<td>Pirovano et al [25]</td>
<td>Lower and upper limbs</td>
<td>Visual, auditory</td>
<td>Wii Fit Balance Board and Microsoft Kinect</td>
</tr>
<tr>
<td>Amengual Alcover et al [26]</td>
<td>Lower limbs</td>
<td>Visual, auditory</td>
<td>Microsoft Kinect</td>
</tr>
<tr>
<td>Baranyi et al [27]</td>
<td>Hand</td>
<td>Visual, auditory, haptic</td>
<td>iOS platform sensors</td>
</tr>
<tr>
<td>Zain et al [28]</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Open</td>
</tr>
<tr>
<td>Noveletto et al [29]</td>
<td>Not reported</td>
<td>Not reported</td>
<td>Open</td>
</tr>
<tr>
<td>Maggiorini et al [31]</td>
<td>Full body</td>
<td>Visual</td>
<td>Microsoft Kinect</td>
</tr>
</tbody>
</table>

RQ7: What Type of Evaluation and Number of Patients Are Involved in the Clinical Trials?

The objective of this research question was to identify clinical validation of the studies and the number of patients involved. In clinical trials, participants receive specific interventions according to the research plan or protocol created by the researchers to determine the safety and efficacy of the interventions through the measurements of the outcomes [36]. Table 9 shows these data. Amengual Alcover et al [26] conducted a clinical trial and observed a significant difference between before and after scores. They used the Berg Balance Scale and their results showed a significant functional improvement (P=.002) in comparison with assessments before (mean 29.5 [SD 3.9] and after (mean 34.1 [SD 2.2]) the intervention. The Functional Reach Test revealed significant differences in functional balance before and after the intervention: right upper limb, before (mean 8.6 [SD 1.4]) and after intervention (mean 10.1 [SD 2.0]; P=.007); and left upper limb, before (mean 8.3 [SD 2.0]) and after intervention (mean 10.1 [SD 3.7]; P=.052). Finally, a significant difference between the pre- and post-assessment scores for the Tinetti Balance Test was observed at the end of the 24-week intervention period. The average score rose from 16 to 21 points on a scale of 28 points. Afyouni et al [30] reported that patients showed improved hand movement using a range of motion. They were able to document 66% of the elements in the video game. No other study reported a clinical trial.

Table 9. Type of evaluation and number of patients in the primary studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Evaluation</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baranyi et al [24]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
<tr>
<td>Pirovano et al [25]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
<tr>
<td>Amengual Alcover et al [26]</td>
<td>Clinical trial</td>
<td>9</td>
</tr>
<tr>
<td>Baranyi et al [27]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
<tr>
<td>Zain et al [28]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
<tr>
<td>Noveletto et al [29]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
<tr>
<td>Afyouni et al [30]</td>
<td>Clinical trial</td>
<td>5</td>
</tr>
<tr>
<td>Maggiorini et al [31]</td>
<td>No clinical validation</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*NA: not applicable.

RQ8: Does the Framework Contemplate a Standardized Scale to Evaluate the Patient’s Rehabilitation Progress?

An assessment instrument allows to objectively quantify the disability degree of the patient and measure the progress of rehabilitation [37,38]. The evaluation scales in the framework are used to quantify the improvement in rehabilitation depending on the type of exercise applied. During the analysis of primary papers, we identified 3 studies with assessment instruments: Pirovano et al [25], Amengual Alcover et al [26], and Afyouni et al [30].

RQ9: Does the Framework Contemplate Adaptability?

Adaptability is the ability to dynamically adapt difficulty in a video game according to the patient’s performance [39]. Five primary studies use this characteristic. In Baranyi et al [24], the physiotherapist designed the level of difficulty of the video
game. Pirovano et al [25] established that for every exercise, quality parameters are necessary to define movement properties. This will allow one to determine the challenge degree of the exercises and adapt the difficulty to the patient’s needs. Zain et al [28] mentioned that adaptability must consider the following elements: (1) user motivation, (2) experience and abilities, and (3) detection, which identifies necessary changes. Afyouni et al [30] adapted the difficulty level based on the patient’s profile and performance in real time. In Magnagiorini et al [31] the therapist can remotely adapt the game therapy to the patient’s actual recovery. Amengual Alcover et al [26], Baranyi et al [27], and Noveletto et al [29] did not specify how adaptability is incorporated into their game. Multimedia Appendix 6 shows the percentage of frameworks contemplating adaptability.

Threats to Validity of Primary Studies Selected

Although we used search strategies and techniques to systematically find papers by using keywords in the selected databases, these words may vary within papers, so some relevant studies may have been omitted.

Discussion

Preliminary Findings

We found only few studies that used a systematic process for serious game development. Each framework analyzed in the primary papers highlighted a different feature.

Planning was the structural activity least implemented. This activity is essential because it allows goal definition, objectives, and path to follow in the software development [9,10,40,41].

Regarding applicability, most studies focused on the treatment of stroke sequelae using various modalities such as visual and auditory. The latter should also be implemented to provide feedback on patient performance. Lastly, test cases directly use playable commercial platforms such as Microsoft Kinect and Leap motion as interaction technology.

There were a few clinical trials, and the type of improvement reported varies from one study to another. Amengual Alcover et al [26] used the Berg Balance Scale and Tinetti Balance Test measurements and reported significant functional improvement from previous results. Afyouni et al [30] also reported improvements using range of motion evaluation in hand movement. No other studies used clinical trials to evaluate the framework. Clinical evaluation is essential to objectively validate the patient’s rehabilitation progress [36].

Pirovano et al [25], Amengual Alcover et al [26], and Afyouni et al [30] used an evaluation scale to assess the patient’s progress. It should also be used as an alternative to adaptability, which is essential for progress and motivation [42]. It is also a technique that can be used to advance game levels [5]. Game levels help engage in the game and could increase treatment compliance.

Conclusions

The objective of this study was to identify the software engineering frameworks used in the development of serious games through a literature review of 8 primary studies. The conclusions of this study are as follows:

About 75% (6/8) of the primary papers proposed a framework [25,26,28-31], whereas the rest were adaptations of the user-centered design framework (RQ1). Regarding the structural activities, 100% (8/8) of the papers applied the communication and construction activity [24-31], 63% (5/8) used modeling (known as a design in some developments) [24-27,31], 25% (2/8) considered user feedback to improve the serious games [25,27], and only 13% (1/8) included the planning phase [26] (RQ2).

Each primary study contributes in one or more aspects to the rehabilitation process. Baranyi et al [24,27] applied a user-centered design using which the physiotherapist can personalize individual needs in the serious game. Pirovano et al [25] proposed ease of play and assisted help during the rehabilitation exercise. Amengual Alcover et al [26] developed a framework for motor rehabilitation therapies using a systematized process. Zain et al [28] embraced immersion and fun in the game to maintain flow interest. Noveletto et al [29] used game scores for patient assessment. Afyouni et al [30] developed games with dynamic adaptability that were patient centered. Finally, Magnagiorini et al [31] incorporated telerehabilitation and adaptability for the elderly to perform rehabilitation exercises at home (RQ3). Every study applies gamification elements that allow patients to transform rehabilitation obstacles into positive and fun reinforcements. Feedback was the gamification element most applied (7/8, 88%) [25-31]. Other elements frequently implemented were adaptability [24,25,28,30,31] and motivational factor [24-27,29] (5/8, 63%) for both; RQ4.

Stroke is the primary pathology on which serious games are focused. This pathology is the third cause of disability worldwide, and a characteristic symptom is the sudden, generally unilateral, loss of muscle strength in the arms, legs, or face (RQ5). Regarding the case studies of limb rehabilitation, 2 studies [24,26] included the lower limb, 1 [25] included lower and upper limbs, 2 [27,30] included hand, 1 [31] full body, and 2 [28,29] did not report case studies. The most used video game modality was visual (6/8, 75%) [24-27,30,31], followed by auditory (4/8, 50%) [25-27,30]. Although each case study used a different motion acquisition technology, every framework allowed a wide variety of the interaction style to obtain the patient’s movement and control the serious game (RQ6).

Of the primary papers, 25% (2/8) applied a clinical evaluation to assess patient improvement when the serious game is used [26,30] (RQ7). To objectively evaluate progress and identify abilities and deficits, only 38% (3/8) of the primary studies used an assessment instrument [25,26,30] (RQ8). The assessment used standardized procedures indicating how a patient of any given age and intelligence level would perform. Adjusting the video game difficulty to the patient’s rehabilitation needs is essential to avoid frustration or boredom, and 63% (5/8) of the primary studies used adaptability [24,25,28,30,31] (RQ9).

Finally, we recommend that all serious games have to be developed with a framework or methodology. If for some reason this is not possible, they should at least involve the therapist to
define requirements. It is also important to include evaluation scales to measure the patient’s progress and gamification elements. Besides, the video game development must be an iterative and incremental process based on generic structural activities and the patient should be considered in the validation and feedback phases.

We propose the following recommendations for future studies:

- Carry out a study of the papers that propose a methodology for serious game development.
- Study software engineering framework proposals in serious games from other fields, such as education.
- Develop a software engineering framework applying all the structural activities and gamification elements for the creation of serious games for physical rehabilitation.

Acknowledgments

The authors thank the National Council of Science and Technology (CONACyT) for their support of this study. Additionally, this work was supported by Secretariat of Public Education (SEP) through PRODEP (Programa para el Desarrollo Profesional Docente).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Percentage of primary studies provided by each electronic database.

Multimedia Appendix 2
Percentage of serious games that used a software engineering framework.

Multimedia Appendix 3
Frequency of occurrence of structural activities in primary studies.

Multimedia Appendix 4
Frequency of occurrence of each gamification element.

Multimedia Appendix 5
Target disability contemplated in frameworks.

Multimedia Appendix 6
Frameworks contemplating adaptability.

References


Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

©Jorge Fernando Ambros-Antemate, María Del Pilar Beristain-Colorado, Marciano Vargas-Treviño, Jaime Gutiérrez-Gutiérrez, Pedro Antonio Hernández-Cruz, Itandehui Belem Gallegos-Velasco, Adriana Moreno-Rodríguez. Originally published in JMIR Serious Games (https://games.jmir.org), 11.11.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.

https://games.jmir.org/2021/4/e25831 JMIR Serious Games 2021 | vol. 9 | iss. 4 | e25831 | p.197

(page number not for citation purposes)
Changes in Self-Reported Web-Based Gambling Activity During the COVID-19 Pandemic: Cross-sectional Study

Emma Claesdotter-Knutsson, MD, PhD; Anders Håkansson, MD, PhD

1 Child and Adolescent Psychiatry, Department of Clinical Sciences, Lund University, Lund, Sweden
2 Psychiatry, Department of Clinical Sciences, Faculty of Medicine, Lund University, Lund, Sweden

Corresponding Author:
Emma Claesdotter-Knutsson, MD, PhD
Child and Adolescent Psychiatry
Department of Clinical Sciences
Lund University
Barav 1
Lund, 221 85
Sweden
Phone: 46 768871765
Email: emma.claesdotter-knutsson@med.lu.se

Abstract

Background: The COVID-19 pandemic has affected not only somatic health with over 3.7 million deaths worldwide, but also has had a huge impact on psychological health, creating what amounts to a mental health crisis. The negative effect of the pandemic on traditional addictions is well described and concerning, and the same has been seen for gambling.

Objective: This study explores self-reported web-based gambling behavior during the COVID-19 pandemic in Sweden. We investigated overall changes, but also changes in specific web-based gambling types, and whether they are associated with certain risk factors or lifestyle changes.

Methods: Our study is based on an anonymous web-based survey of web panel participants in Sweden (N=1501) designed to study a range of behavioral changes during the COVID-19 pandemic. Increases in gambling were analyzed using logistic regression models against sociodemographic data and psychological distress.

Results: The majority of the respondents who gambled reported no changes in their gambling habits during the COVID-19 pandemic. We found significant associations with the problem gambling severity index (PGSI), the Kessler score (indicating psychological distress), employment status, changes in alcohol habits, and self-exclusion when looking at overall changes in gambling activity in the pandemic. In the subgroup that reported an increase in gambling activity, we found an association with both the PGSI and Kessler scores. The PGSI score was also an independent predictor for all specific web-based gambling (horses, sports, poker, and casino) whereas the Kessler score only had a significant impact on changes in casino gambling. In addition, male gender was an independent predictor for gambling on sports and casino gambling.

Conclusions: The majority of respondents who gambled reported no changes in their gambling activity during the COVID-19 pandemic. The group that reported an increase in overall gambling activity during the COVID-19 pandemic represent a group with gambling problems and psychological distress. The group that reported increased sports and casino gambling were often male, and this group seemed to experience more psychological distress.

(Keywords: COVID-19; pandemic; web-based gambling; psychological distress; gender)

Introduction

The first case of COVID-19 was detected in December 2019 in Wuhan, China [1] and then the virus spread rapidly, reaching Europe in early 2020 and Sweden in January 2020 [2]. The COVID-19 pandemic has had an enormous effect on mental health [3-5]. Research has confirmed increasing rates of stress, depression, addiction, anxiety, and other psychiatric disorders during the pandemic [6,7].

(JMIR Serious Games 2021;9(4):e30747) doi:10.2196/30747
Historically in economic crises, when people experienced stress due to, for example, isolation, gambling activity per se increased, and so did gambling problems [15-17], but recent studies on potential changes in gambling activity during the COVID-19 pandemic have reported different changes in behavior [18]. One possible explanation might be the restrictions in place in the field of study, along with differences in study populations. Auer et al [19] and Lindner et al [20] found a substantial decrease in overall gambling activity, especially in gambling, where there were far fewer betting opportunities because of cancelled or postponed sports events such as football leagues.

This was also reported in an Italian study conducted during the lockdown, when gamblers reported not only a decrease in gambling but also the relief they felt because the restrictions meant they could not gamble as before [21]. Auer et al [19] confirmed this situation in a population of Swedish, German, Finnish, and Norwegian citizens, who reported no increase in gambling activity. Our research group, focusing on COVID-19–related changes in gambling in a Swedish setting, observed increased gambling activity in a subgroup of high-risk gamblers [22,23]. Lindner et al [20] observed the same pattern among high-risk gamblers and Frisone et al [24] observed this among moderate-risk gamblers.

Early during the pandemic, the Swedish government voiced its concerns about an anticipated increase in gambling owing to rising anxiety, boredom, and depression prevalence expected as the pandemic accelerated [20]. Social distancing to prevent virus transmission was thought likely to encourage gambling [20]. To prevent this, the Swedish government introduced a temporary legislation (Förordning 2020:495) to govern the Swedish gambling market, including a deposit limit at web-based casinos, of Swedish kronor (SEK) 5000 (US $580.48) per gambling provider and obligatory time limits because of concerns in the early stages of the outbreak that social distancing would increase web-based activities such as gambling and, by extension, the problems associated with gambling. Other countries followed suit [20,23].

Gambling is strongly associated with comorbid disorders, including alcohol use disorder, depression, substance use disorders, nicotine dependence, anxiety disorders, and antisocial personality disorder [25,26]. Gambling is a potentially addictive behavior; gambling disorder is an addiction diagnosis given the same status as alcohol and drug addiction [27,28]. Web-based gambling, known for its autonomy, accessibility, and speed of play, is considered a higher-risk form of gambling, which easily leads to addiction [29,30].

In our study, we explored self-reported web-based gambling behavior during the COVID-19 pandemic in Sweden. We considered overall changes, along with changes for specific types of gambling, and whether potential changes in gambling may be associated with specific risk factors such as level of education, employment status, disposable income, and psychological distress.

Methods

Setting
The first wave of the COVID-19 pandemic in Sweden was in spring 2020. After a decline in virus transmission in the summer, there was a second wave in the autumn of 2020. In the first few months of 2021, there was a third wave of virus transmission and a further increase in the disease burden of hospitalizations (Swedish Public Health Agency 2021). This cross-sectional study is based on a self-report web-based survey carried out in Sweden in March 2021, during the third wave of the COVID-19 pandemic in the country. At the time of this study, upper-secondary schools in Sweden had resumed in-person teaching, but the nationwide COVID-19 restrictions on leisure activities and gatherings of more than 8 people were still in effect.

Participants
We consulted the market survey company Ipsos and used its internet-based web survey panel. We have previously used the Ipsos web panel for internet-based surveys for research purposes [30]. Respondents from the general population, aged 16 years and above, from the Ipsos web panel were invited with the information that the survey would address “computer gaming, gambling for money, and other behavioural patterns in Sweden during Covid-19—Association of mental health, social situation, and attitudes to the pandemic.” The language used in the survey was Swedish. Respondents could only access the survey once they had provided electronic informed consent. Most Ipsos surveys are worth 1 point to respondents, and each point is worth about €1 (US $1.16) when redeemed as goods or services. In this study, invitations were sent until 1500 complete answers were obtained. In addition, in this study, the final distribution of age groups, gender, and geographical location (region) was compared by Ipsos to the general population, such that the data set was weighted in accordance with a summarized weighting score derived from these 3 variables. The survey was halted when the final sample consisted of 1501 individuals. The study was carried out between March 19 and 29, 2021. The study was reviewed and approved by the Swedish Ethical Review Board (File: 2021/00369).

Measures
The basic sociodemographic variables included gender (female or male), age (divided into 2 age groups of 16-24 years and ≥25 years), monthly income (divided into 3 groups of SEK 10,000-20,000 [US $1164.40-$2328.81], SEK 20,000-40,000 [US $2328.81-$4657.62], or >SEK 40,000 [>US $4657.62]), level of education (university, upper-secondary school [ages 16-19 years], compulsory school [ages 6-16 years], or other), and employment status (studying, employed, unemployed,
The questionnaire continued with questions about changes in personal behavior during the COVID-19 pandemic (“since these changes in Sweden started”): the time they spent at home (much more, slightly more, unchanged, or less time at home) and how much alcohol they consumed (more, less, unchanged, or “I don’t drink, now nor before”). The questionnaire continued with a general question about their web-based gambling on horse racing, sports, poker, and casino games (more, less, unchanged, or “I don’t gamble, now or before”). This was followed by questions for each type of web-based gambling in turn (horses, sports, poker, and casinos). A question was asked about self-exclusion from gambling (gambling breaks) using spelpaus.se, a national self-exclusion scheme that covers all forms of licensed gambling, which has been available since January 1, 2019 (yes, no, or prefer not to answer) [30].

Psychological distress was measured using the Kessler-6 scale to describe symptoms of depressed mood and anxiety [31]. It is a 6-item scale of symptoms perceived in the preceding 6 months, with 6 questions about depressive and anxiety-related symptoms with options ranging from “not at all” to “all the time,” including a “cannot answer/prefer not to say” response. For the Kessler-6 scale, the scores of 0-4 for each question were summed, and a total score of ≥5 was classified as at least moderate psychological distress.

The level of potential gambling problems was measured with the 9-item problem gambling severity index (PGSI) [32], where each statement addresses the preceding 12 months, with options including “never,” “sometimes,” “most of the time,” and “almost always.” For the PGSI scale, the scores of 0-3 for each question were summed: a total score of 0 indicated no problem with gambling; 1-2, a certain risk of gambling problems; 3-7, a moderate risk of gambling problems; and ≥8 indicated gambling problems.

Statistical Analysis
The reporting of prevalence measures and group-wise comparisons related to the weighted data and statistical tests were applied using the chi-square test, whereas binary logistic regression analyses were carried out using the unweighted data. Binary logistic regressions were carried out with each increased change in gambling (yes/no) as the dependent variable to study potential–independent variables associated with the outcome. In a subsequent step, greater changes in specific forms of gambling (horses, sports, poker, and casino) were established as outcomes. For all models, odds ratios with 95% CIs are presented. SPSS (version 25.0) was used for all statistical analyses.

Availability of Data and Materials
The data sets used and analyzed in the current study are available from the corresponding author on reasonable request.

Ethics Approval and Consent to Participate
All procedures performed in this study involving human participants were in accordance with the ethical standards of the national research committee and with the tenets of the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Patient Consent for Publication
Informed consent was obtained from all individual participants included in the study.

Informed Consent
The manuscript does not contain any individual person’s data in any form.

Results

Descriptive Data on Changes in Web-Based Gambling Activity and Type of Gambling
The descriptive data of the sample are shown in Table 1 (weighted data). Of the 1501 respondents, 29% (n=437) responded they had never gambled and were therefore excluded from further analysis. Of the remaining 1064 who were further assessed, 56% (n=596) were male and 44% (n=468) were female. The distribution regarding the different types of web-based gambling activity is described in Table 1.

<table>
<thead>
<tr>
<th>Changes in gambling activity</th>
<th>Total, n (%)</th>
<th>Horses, n (%)</th>
<th>Casino, n (%)</th>
<th>Poker, n (%)</th>
<th>Sports, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never gambled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study population</td>
<td>1064 (70.9)</td>
<td>428 (28.5)</td>
<td>379 (25.2)</td>
<td>237 (18.8)</td>
<td>494 (32.9)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>468 (44.0)</td>
<td>150 (35.1)</td>
<td>123 (32.3)</td>
<td>56 (23.5)</td>
<td>140 (28.4)</td>
</tr>
<tr>
<td>Male</td>
<td>596 (56.0)</td>
<td>278 (64.9)</td>
<td>256 (67.7)</td>
<td>181 (76.5)</td>
<td>354 (71.6)</td>
</tr>
</tbody>
</table>

Based on the findings, approximately 10% (n=111) of the participants reported an increase in overall gambling, 82% (n=876) reported unchanged gambling activity, and 7% (n=77) reported a decrease in gambling (Table 2). The figures for increased activity by gambling type were similar at approximately 12% (11.0%-12.9%). The same was seen for the self-reported decrease in activity by type of gambling on horses, sports, and casinos at approximately 13% (13.2%-14.2%), but self-reported decrease in poker was 21% (Table 2).
Table 2. Changes in web-based gambling activity overall and by gambling type (weighted).

<table>
<thead>
<tr>
<th>Changes</th>
<th>Change in gambling, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overall (n=1064)</td>
</tr>
<tr>
<td>Increased</td>
<td>111 (10.4)</td>
</tr>
<tr>
<td>Unchanged</td>
<td>876 (82.3)</td>
</tr>
<tr>
<td>Decreased</td>
<td>77 (7.2)</td>
</tr>
</tbody>
</table>

Changes in Web-Based Gambling Activity by Demographic and Socioeconomic Factors

Table 3 presents the weighted data for changes in gambling by demographic and socioeconomic factors. We found significant differences in employment status ($P<.001$), PGSI ($P<.001$), Kessler score ($P<.001$), gambling paus ($P<.001$), as well as changes in alcohol habits ($P<.001$) (Table 3).
Table 3. Changes in gambling activity by demographic and socioeconomic characteristics (weighted).

<table>
<thead>
<tr>
<th></th>
<th>Increased (n=111), n (%)</th>
<th>Unchanged (n=876), n (%)</th>
<th>Decreased (n=77), n (%)</th>
<th>P value</th>
<th>Total (N=1064), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>39 (34.9)</td>
<td>398 (54.4)</td>
<td>32 (41.4)</td>
<td>.10</td>
<td>468 (44.0)</td>
</tr>
<tr>
<td>Male</td>
<td>72 (65.1)</td>
<td>478 (54.6)</td>
<td>45 (58.6)</td>
<td></td>
<td>596 (56.0)</td>
</tr>
<tr>
<td><strong>Age groups (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td>.07</td>
<td></td>
</tr>
<tr>
<td>16-24</td>
<td>19 (17.3)</td>
<td>96 (11.0)</td>
<td>13 (16.4)</td>
<td></td>
<td>128 (12.0)</td>
</tr>
<tr>
<td>≥25</td>
<td>92 (82.7)</td>
<td>780 (89.0)</td>
<td>64 (83.6)</td>
<td></td>
<td>936 (88.0)</td>
</tr>
<tr>
<td><strong>Level of education</strong></td>
<td></td>
<td></td>
<td></td>
<td>.51</td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>55 (49.2)</td>
<td>402 (46.0)</td>
<td>41 (52.9)</td>
<td></td>
<td>498 (46.8)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>45 (40.4)</td>
<td>364 (41.5)</td>
<td>28 (36.5)</td>
<td></td>
<td>437 (41.0)</td>
</tr>
<tr>
<td>University</td>
<td>9 (7.7)</td>
<td>82 (9.3)</td>
<td>3 (4.5)</td>
<td></td>
<td>94 (8.8)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (2.7)</td>
<td>28 (3.2)</td>
<td>5 (6.1)</td>
<td></td>
<td>36 (3.3)</td>
</tr>
<tr>
<td><strong>Employment status</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
</tr>
<tr>
<td>Studying</td>
<td>21 (18.9)</td>
<td>86 (9.9)</td>
<td>12 (16.1)</td>
<td></td>
<td>120 (11.3)</td>
</tr>
<tr>
<td>Employed</td>
<td>63 (56.3)</td>
<td>499 (57.0)</td>
<td>50 (65.0)</td>
<td></td>
<td>612 (57.5)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>16 (14.2)</td>
<td>44 (5.1)</td>
<td>4 (4.6)</td>
<td></td>
<td>63 (6.0)</td>
</tr>
<tr>
<td>Retired</td>
<td>10 (8.9)</td>
<td>225 (25.7)</td>
<td>11 (14.3)</td>
<td></td>
<td>246 (23.1)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (1.9)</td>
<td>21 (2.4)</td>
<td>0 (0)</td>
<td></td>
<td>23 (2.2)</td>
</tr>
<tr>
<td><strong>Disposable income</strong></td>
<td></td>
<td></td>
<td></td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>≥SEK 40,000 (US $4657.62)</td>
<td>35 (31.8)</td>
<td>241 (27.5)</td>
<td>29 (38.1)</td>
<td></td>
<td>305 (28.7)</td>
</tr>
<tr>
<td>SEK 20,000-40,000 (US $2328.81-$4657.62)</td>
<td>60 (54.1)</td>
<td>474 (54.2)</td>
<td>37 (48.2)</td>
<td></td>
<td>572 (53.7)</td>
</tr>
<tr>
<td>≤SEK 10,000-20,000 (US $1164.40-$2328.81)</td>
<td>16 (14.2)</td>
<td>161 (18.3)</td>
<td>11 (13.7)</td>
<td></td>
<td>187 (17.6)</td>
</tr>
<tr>
<td><strong>Problem Gambling Severity Index</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
</tr>
<tr>
<td>No problem with gambling</td>
<td>25 (24.5)</td>
<td>600 (80.0)</td>
<td>41 (61.2)</td>
<td></td>
<td>666 (62.6)</td>
</tr>
<tr>
<td>Certain risk of gambling problems</td>
<td>17 (17.2)</td>
<td>86 (11.4)</td>
<td>14 (21.6)</td>
<td></td>
<td>117 (11.0)</td>
</tr>
<tr>
<td>Increased risk of gambling problems</td>
<td>34 (33.9)</td>
<td>55 (7.3)</td>
<td>7 (10.7)</td>
<td></td>
<td>96 (9.1)</td>
</tr>
<tr>
<td>Gambling problems</td>
<td>25 (24.4)</td>
<td>10 (1.4)</td>
<td>4 (6.5)</td>
<td></td>
<td>39 (3.7)</td>
</tr>
<tr>
<td><strong>Kessler score</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
</tr>
<tr>
<td>Score 0-4; no psychological distress</td>
<td>19 (16.7)</td>
<td>443 (51.1)</td>
<td>25 (34.7)</td>
<td></td>
<td>487 (45.8)</td>
</tr>
<tr>
<td>Score 5-24; psychological distress</td>
<td>92 (83.2)</td>
<td>423 (48.9)</td>
<td>48 (65.3)</td>
<td></td>
<td>563 (52.9)</td>
</tr>
<tr>
<td><strong>Gambling pause</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;.001a</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>21 (18.2)</td>
<td>24 (2.8)</td>
<td>12 (14.9)</td>
<td></td>
<td>51 (5.2)</td>
</tr>
<tr>
<td>No</td>
<td>91 (81.7)</td>
<td>846 (97.2)</td>
<td>65 (85.1)</td>
<td></td>
<td>1000 (94.8)</td>
</tr>
<tr>
<td><strong>Time at home</strong></td>
<td></td>
<td></td>
<td></td>
<td>.11</td>
<td></td>
</tr>
<tr>
<td>Much more</td>
<td>76 (67.8)</td>
<td>496 (5.7)</td>
<td>44 (57.3)</td>
<td></td>
<td>614 (57.8)</td>
</tr>
<tr>
<td>Slightly more</td>
<td>29 (25.8)</td>
<td>258 (29.5)</td>
<td>20 (25.8)</td>
<td></td>
<td>307 (28.8)</td>
</tr>
</tbody>
</table>
Comparison of Increased Gambling in Different Outcomes

Multivariable analysis using binary logistic regression models of the potential predictors of increased gambling (yes vs no) is presented in Table 4 (unweighted data). Increased gambling was associated with all 3 categories of PGSI score, whether “Certain risk of gambling problems” (OR 4.40, 95% CI 2.17-8.93), “Increased risk of gambling problems” (OR 12.53, CI 6.54-24.04), or “Gambling problems” (OR 32.42, 95% CI 13.77-76.35). The PGSI for “Gambling problems” was also correlated with increased gambling in all specific forms of gambling: horses (OR 12.64, 95% CI 3.94-40.68), sports (OR 38.81, 95% CI 11.65-129.28), poker (OR 10.53, 95% CI 2.24-49.43), and casinos (OR 26.17, 95% CI 7.09-96.62). An increased change in gambling overall was associated with the >5 Kessler score (OR 2.62, 95% CI 1.39-4.91) and with increased casino gambling (OR 4.47, 95% CI 1.15-17.3). In addition, male gender is correlated with increased sports betting (OR 2.58, 95% CI 1.09-6.09) and casino gambling (OR 2.73, 95% CI 1.06-7.07). The correlation between “Does not drink” and increased sports betting was also significant (OR 3.95, 95% CI 1.02-15.33).
Table 4. Multivariable logistic regressions model with increased gambling as the outcome of interest (yes vs no).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Increased gambling on horses, odds ratio (95% CI)</th>
<th>Increased gambling on casino, odds ratio (95% CI)</th>
<th>Increased gambling on poker, odds ratio, (95% CI)</th>
<th>Increased gambling on sports, odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Male</td>
<td>1.46 (0.86-2.48)</td>
<td>0.84 (0.39-1.81)</td>
<td>2.58(^a) (1.09-6.09)</td>
<td>2.73(^a) (1.06-7.07)</td>
</tr>
<tr>
<td>Age groups (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-24</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>≥25</td>
<td>1.09 (0.48-2.50)</td>
<td>2.09 (0.35-12.52)</td>
<td>0.43 (0.14-1.34)</td>
<td>0.91 (0.18-4.54)</td>
</tr>
<tr>
<td>Level of education</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Secondary school</td>
<td>0.73 (0.27-1.99)</td>
<td>0.80 (0.19-3.47)</td>
<td>1.62 (0.26-9.93)</td>
<td>—</td>
</tr>
<tr>
<td>University</td>
<td>1.18 (0.44-3.19)</td>
<td>0.85 (0.20-3.69)</td>
<td>1.69 (0.27-10.43)</td>
<td>—</td>
</tr>
<tr>
<td>Other</td>
<td>1.31 (0.27-6.36)</td>
<td>0.72 (0.05-9.64)</td>
<td>2.62 (0.15-45.97)</td>
<td>—</td>
</tr>
<tr>
<td>Employment status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Studying</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Employed</td>
<td>0.61 (0.22-1.69)</td>
<td>0.38 (0.07-2.03)</td>
<td>0.41 (0.09-1.96)</td>
<td>0.49 (0.05-5.00)</td>
</tr>
<tr>
<td>Unemployed</td>
<td>1.79 (0.62-5.19)</td>
<td>1.59 (0.28-9.11)</td>
<td>2.84 (0.63-12.85)</td>
<td>0.89 (0.10-8.18)</td>
</tr>
<tr>
<td>Retired</td>
<td>0.64 (0.21-1.98)</td>
<td>0.73 (0.12-4.53)</td>
<td>0.51 (0.04-5.82)</td>
<td>—</td>
</tr>
<tr>
<td>Other</td>
<td>2.11 (0.37-12.06)</td>
<td>7.64 (0.70-83.40)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Disposable income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥SEK 40,000 (US $4657.62)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>SEK 20,000-40,000 (US $2328.81-$4657.62)</td>
<td>0.90 (0.42-1.95)</td>
<td>0.51 (0.19-1.36)</td>
<td>0.59 (0.15-2.33)</td>
<td>1.23 (0.27-5.58)</td>
</tr>
<tr>
<td>≤SEK 10,000-20,000 (US $1164.40-2328.81)</td>
<td>0.73 (0.26-2.04)</td>
<td>0.30 (0.08-1.21)</td>
<td>0.41 (0.06-2.65)</td>
<td>0.44 (0.04-5.01)</td>
</tr>
<tr>
<td>Problem Gambling Severity Index</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No problem with gambling</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Certain risk of gambling problems</td>
<td>4.40 (2.17-8.93)</td>
<td>4.01(^a) (1.26-12.81)</td>
<td>4.54(^a) (1.13-18.30)</td>
<td>7.49(^a) (1.52-36.78)</td>
</tr>
<tr>
<td>Increased risk of gambling problems</td>
<td>12.5(^a) (6.54-24.04)</td>
<td>10.93(^a) (3.88-30.82)</td>
<td>6.02(^a) (1.68-21.56)</td>
<td>3.84 (0.80-18.51)</td>
</tr>
<tr>
<td>Gambling problems</td>
<td>32.4(^a) (13.77-76.35)</td>
<td>12.64(^a) (3.93-40.68)</td>
<td>26.17(^a) (7.09-96.62)</td>
<td>10.53(^a) (2.24-49.43)</td>
</tr>
<tr>
<td>Kessler score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Score 0-4; no psychological distress</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Score 5-24; psychological distress</td>
<td>2.62(^a) (1.39-4.91)</td>
<td>2.06 (0.84-5.07)</td>
<td>4.47(^a) (1.15-17.30)</td>
<td>2.80 (0.69-11.74)</td>
</tr>
<tr>
<td>Gambling pause</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>No</td>
<td>1.14 (0.43-3.02)</td>
<td>0.77 (0.22-2.72)</td>
<td>1.66 (0.47-5.94)</td>
<td>0.92 (0.18-4.82)</td>
</tr>
<tr>
<td>Time at home</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Much more</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Slightly more</td>
<td>0.56 (0.30-1.04)</td>
<td>0.86 (0.36-2.05)</td>
<td>0.81 (0.29-2.20)</td>
<td>0.27 (0.07-1.12)</td>
</tr>
<tr>
<td>Unchanged</td>
<td>0.25 (0.08-0.77)</td>
<td>0.13 (0.01-1.18)</td>
<td>0.28 (0.05-1.76)</td>
<td>—</td>
</tr>
</tbody>
</table>

\(a\) Significantly different from reference (95% CI = confidence interval).
behavior in the initial stages of the pandemic, similar to other
explanations for increased gambling during the pandemic. This might
result from increased opportunities—were more readily available, thus affecting
web-based gambling. The seemingly steady rise in self-reported
gambling is something to follow as the pandemic continues.

**Discussion**

**Principal Findings**

This cross-sectional study considers self-reported changes in
web-based gambling in a general population aged 16 years and
over during the third wave of the COVID-19 pandemic in
Sweden. We set out to study possible changes in gambling behavior and relate them to risk factors such as gender, level of
education, employment status, disposable income, alcohol consumption, and psychological distress.

The majority of the respondents who gambled reported no
changes in their gambling activity during the COVID-19 pandemic. Regarding overall changes in gambling, we found
significant associations with the PGSI (indicating the severity
of the problem), the Kessler score (indicating psychological
distress), employment status, changes in alcohol habits, and self-exclusion (gambling breaks). In the subgroup that reported
an increased gambling activity, we found an association with
both the PGSI score and the Kessler score. The PGSI score was
also an independent predictor for all web-based gambling activities (horses, sports, poker, and casinos) whereas the Kessler
score only had a statistically significant impact on changes in
casino gambling. In addition, male gender was an independent
predictor for sports betting and casino gambling.

**Changes in Web-Based Gambling Activity During the
Pandemic**

In a previous study conducted during the first wave of the
COVID-19 pandemic in 2020, we found a 4% self-reported increase in gambling [23], whereas in another study 7 months
later, we found a 6% self-reported increase in gambling [22].
Surprisingly, Lindner et al [20], who also studied changes in
gambling activity in Sweden during the first wave, found an
overall decrease in gambling [20]. In this study conducted in
March 2021 the majority (82%) of the respondents who reported
that they gambled had not changed their gambling type during
the pandemic, but we found a self-reported increase in gambling activity of 10%. Even though the studies are not fully compared
because they use different subjects and this study is limited to
web-based gambling only, one can nevertheless discern a
consequential rise in gambling during the pandemic. This might
be explained by people being more likely to change their
behavior in the initial stages of the pandemic, similar to other
times of crisis, but with time people tended to return to a stable,
“normal” level, whether in gambling or other activities.

In the early stages of the pandemic, all sports betting events
such as the top football leagues were cancelled [19]. Gradually,
sporting events restarted, making sports betting easier. Another explanation for the increase in betting during the pandemic
could be changes in social behavior, such as restrictions on social events and restaurant opening hours, which implies that people were
more likely to stay at home, where their screens—and gambling
opportunities—were more readily available, thus affecting
web-based gambling. The seemingly steady rise in self-reported
gambling is something to follow as the pandemic continues.

**Characteristics Among Those Reporting an Increase
in Specific Types of Gambling**

The PGSI, Kessler score (psychological distress), employment
status, gambling breaks, and changes in alcohol habits were all
individually associated with increases in gambling—something
that was not true for gender, age group, level of education,
disposable income, and changes in time spent at home. The only
factors that remained significantly associated with an overall
increase in gambling were self-reported gambling problems and
psychological distress.

Our findings are in line with those of previous findings that the
minority who reported increased gambling during the COVID-19 pandemic also reported increased gambling problems and greater
psychological distress [18,20,23,33]. Problem gambling is known to be associated with mental health problems, adding to
the impression that individuals who gambled more during the
COVID-19 pandemic were a vulnerable subgroup of the
population. Increased casino gambling was the only form of
gambling associated with psychological distress. Casino
gambling is considered potentially highly addictive and is
associated with problem gambling and personal debt [30,34,35].
Notably, at the onset of the COVID-19 pandemic, both Lindner
et al [20] and Håkansson [23] found that respondents who
gambled using web-based casinos, rather than gambling less
(as was the case for other types of gambling), increased their
gambling activity.

Among respondents reporting an increase in sports betting and
casino gambling, we found the traditionally male gender
correlation [36,37]. This had not been true of sports betting in
our previous study in April 2020, early in the COVID-19

<table>
<thead>
<tr>
<th>Change in alcohol habits</th>
<th>Increased gambling on horses (odds ratio (95% CI))</th>
<th>Increased gambling on casino (odds ratio (95% CI))</th>
<th>Increased gambling on poker (odds ratio (95% CI))</th>
<th>Increased gambling on sports (odds ratio (95% CI))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes more</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>Unchanged</td>
<td>0.66 (0.30-1.44)</td>
<td>0.28 (0.09-0.86)</td>
<td>0.72 (0.18-2.81)</td>
<td>0.18* (0.03-0.98)</td>
</tr>
<tr>
<td>Less alcohol</td>
<td>0.77 (0.34-1.76)</td>
<td>0.53 (0.17-1.67)</td>
<td>1.45 (0.36-5.72)</td>
<td>0.24 (0.04-1.48)</td>
</tr>
<tr>
<td>Does not drink</td>
<td>1.31 (0.51-3.34)</td>
<td>1.22 (0.33-4.37)</td>
<td>1.32 (0.25-6.95)</td>
<td>2.07 (0.34-12.69)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*aSignificant correlations observed.

*b: not determined.

https://games.jmir.org/2021/4/e30747

JMJIR Serious Games 2021 | vol. 9 | iss. 4 | e30747 | p.205

(page number not for citation purposes)
pandemic [23]. At that stage the sports betting market had shrunk—a plausible explanation for not seeing any gender differences in self-reported increased sports betting.

As mentioned earlier, gambling is strongly associated with comorbid disorders and is also considered a potentially addictive behavior that may result in related harms [26,27,38]. Problem gambling is expensive for not only gamblers and their personal networks but also for society in terms of health care and legal costs, lost productivity owing to unemployment, and poorer quality of life [39].

Limitations
This study has some limitations. First, the data are self-reported, with the inevitable risk of recall bias. Respondents who gamble might be more likely to answer questions about it. Findings based on self-reporting would resist generalization were it not for the representative sociodemographic and geographic distribution. The advantage of our study setting is thus our use of a weighted sample: we continued to include respondents until we had a sample that was both geographically and sociodemographically representative. Another study limitation is that Sweden differs to most other countries in regard to its COVID-19 regulations [40]. There were no lockdowns and no shopping bans because policies to prevent transmission were centered on recommendations rather than regulations, making the results difficult to transfer to other settings [40].

Conclusions
Compared to our previous studies from earlier during the pandemic, we found that a higher proportion of the general population had increased their gambling activity during the COVID-19 pandemic. The group that reported increased overall gambling was small but characterized by gambling problems and psychological distress. The group that reported an increase in sports betting and casino gambling were predominantly male. The group that reported an increase in casino gambling were independently related to psychological distress.

It is concerning that those who reported increased gambling activity during the pandemic are a growing group. Moreover, they are a vulnerable group that needs to be addressed by caregivers, the gambling industry, and policy makers.

We have yet to see the full effects of the COVID-19 pandemic on gambling. More studies are needed to chart gambling patterns in possible subsequent waves of the pandemic and later in its aftermath.

Acknowledgments
The authors acknowledge Region Skåne, Yngre ALF (a Swedish government research grant), the Crafoord Foundation, the Fanny Ekdahl Foundation for Paediatric Research, and Svenska Spel’s Research Council.

Authors’ Contributions
All authors have contributed significantly and agree with the content of the paper.

Conflicts of Interest
AH is employed as a researcher in addiction medicine at Lund University sponsored by AB Svenska Spel (the Swedish state-owned gambling operator) as part of its responsible gambling policies. He also has research funding from Svenska Spel’s Research Council, and from the research council of Systembolaget AB (the Swedish alcohol retailing monopoly) and from the Swedish Sports Confederation. ECK has research funding from Svenska Spel’s Research Council. None of these organizations have been involved in or had any influence on any part of the present work.

References


https://journals.jmir.org/2021/4/e30747

JMI SERIOUS GAMES
Claesdotter-Knutsson & Håkansson


Abbreviations

OR: odds ratio
PGSI: Problem Gambling Severity Index
SEK: Swedish kronor