Contents

Original Papers

Electronic Games for Facilitating Social Interaction Between Parents With Cancer and Their Children During Hospitalization: Interdisciplinary Game Development (e16029)
Karin Pihl, Helle Gyldenvang, Jeppe Møller, Tine Kjølsen, Jesper Juul, Helle Pappot. 3

Validation of a Portable Game Controller to Assess Peak Expiratory Flow Against Conventional Spirometry in Children: Cross-sectional Study (e25052)
Khadija Chelabi, Fabio Balli, Myriam Bransi, Yannick Gervais, Clement Marthe, Sze Tse. 8

Pupillary Responses for Cognitive Load Measurement to Classify Difficulty Levels in an Educational Video Game: Empirical Study (e21620)
Hugo Mitre-Hernandez, Roberto Covarrubias Carrillo, Carlos Lara-Alvarez. 19

Managing Game-Related Conflict With Parents of Young Adults With Internet Gaming Disorder: Development and Feasibility Study of a Virtual Reality App (e22494)
Yu-Bin Shin, Jae-Jin Kim, Hyeunj Kim, Soo-Heung Kim, Hyeo Jung, Eunju Kim. 32

Promoting Physical Activity in Japanese Older Adults Using a Social Pervasive Game: Randomized Controlled Trial (e16458)
Luciano Santos, Kazuya Okamoto, Ryo Otsuki, Shusuke Hiragi, Goshiro Yamamoto, Osamu Sugiyama, Tomoki Aoyama, Tomohiro Kuroda. 4

A Co-Designed Active Video Game for Physical Activity Promotion in People With Chronic Obstructive Pulmonary Disease: Pilot Trial (e23069)
Joshua Simmich, Allison Mandrusiak, Stuart Smith, Nicole Hartley, Trevor Russell. 56

A Bowling Exergame to Improve Functional Capacity in Older Adults: Co-Design, Development, and Testing to Compare the Progress of Playing Alone Versus Playing With Peers (e23423)
Jorge Da Silva Junior, Daiana Biduski, Ericles Bellei, Osvaldo Becker, Luciane Daroit, Adriano Pasqualotti, Hugo Tourinho Filho, Ana de Marchi. 73

Utilizing Theories and Evaluation in Digital Gaming Interventions to Increase Human Papillomavirus Vaccination Among Young Males: Qualitative Study (e21303)
Gabrielle Darville, Jade Burns, Tanaka Chavanduka, Charkarra Anderson-Lewis. 86

Perception of Game-Based Rehabilitation in Upper Limb Prosthetic Training: Survey of Users and Researchers (e23710)
Christian Garske, Matthew Dyson, Sigrid Dupan, Kianoush Nazarpour. 97
Video Game–Based Rehabilitation Approach for Individuals Who Have Undergone Upper Limb Amputation: Case-Control Study (e17017)
N Hashim, N Abd Razak, H Gholizadeh, N Abu Osman .................................................. 109

A Web-Based Game for Young Adolescents to Improve Parental Communication and Prevent Unintended Pregnancy and Sexually Transmitted Infections (The Secret of Seven Stones): Development and Feasibility Study (e23088)
Ross Shegog, Laura Armistead, Christine Markham, Sara Dube, Hsing-Yi Song, Pooja Chaudhary, Angela Spencer, Melissa Peskin, Diane Santa Maria, J Wilkerson, Robert Addy, Susan Tortolero Emery, Jeffery McLaughlin .................................................. 128

Review

Serious Games for Improving Technical Skills in Medicine: Scoping Review (e24093)
Tycho Olgers, Anne bij de Weg, Jan ter Maaten .................................................. 119
Electronic Games for Facilitating Social Interaction Between Parents With Cancer and Their Children During Hospitalization: Interdisciplinary Game Development

Karin Piil\(^1,2\), MSc, PhD; Helle Holm Gyldenvang\(^1\), BA, MSc; Jeppe Kilberg Møller\(^3\), MSc; Tine Kjoelsen\(^3\), MA; Jesper Juul\(^3\), MA, PhD; Helle Pappot\(^1,4\), MD, PhD, Prof Dr

\(^1\)Department of Oncology, Copenhagen University Hospital (Rigshospitalet), Copenhagen Ø, Denmark
\(^2\)Department of Public Health, Aarhus University, Aarhus, Denmark
\(^3\)Department of Visual Design, Schools of Architecture, Design and Conservation, The Royal Danish Academy, Copenhagen K, Denmark
\(^4\)Department of Clinical Medicine, Copenhagen University, Copenhagen, Denmark

Corresponding Author:
Karin Piil, MSc, PhD
Department of Oncology
Copenhagen University Hospital (Rigshospitalet)
Blegdamsvej 9
Copenhagen Ø, 2100
Denmark
Phone: 45 35450733
Email: Karin.piil@regionh.dk

Abstract

**Background:** Most cancer treatments today take place in outpatient clinics; however, it might be necessary for some patients to be admitted to hospital departments due to severe side effects or complications. In such situations, support from family and social relations can be crucial for the patients’ emotional well-being. Many young adolescents and children whose parents have cancer describe how they are not seen, heard, or listened to as the worried relatives they are. Within the intensive care unit, it has been recommended that early supportive interventions are tailored to include children of the intensive care patient; a similar approach might be relevant in the oncological setting. To our knowledge, no studies have explored how to involve young relatives who are visiting their parent at an oncological department. Recently, a framework for developing theory-driven, evidence-based serious games for health has been suggested. Such a process would include stakeholders from various disciplines, who only work toward one specific solution. However, it is possible that bringing together different disciplines, such as design, art, and health care, would allow a broader perspective, resulting in improved solutions.

**Objective:** This study aims to develop tools to enhance the social interaction between a parent with cancer and their child when the child visits the parent in the hospital.

**Methods:** In total, 4 groups of design students within the Visual Design program were tasked with developing games addressing the objective of strengthening relations in situ during treatment. To support their work, the applied methods included professional lectures, user studies, and visual communication (phase I); interviews with the relevant clinicians at the hospital (phase II), co-creative workshops with feedback (phase III), and evaluation sessions with selected populations (phase IV). The activities in the 4 phases were predefined. This modified user design had the child (aged 4-18 years) of a parent with cancer as its primary user.

**Results:** Overall, 4 different games were designed based on the same information. All games had the ability to make adults with cancer and their children interact on a common electronic platform with a joint goal. However, the interaction, theme, and graphical expression differed between the games, suggesting that this is a wide and fertile field to explore.

**Conclusions:** Playing a game can be an efficient way to create social interaction between a parent with cancer and a child or an adolescent, potentially improving the difficult social and psychological relations between them. The study showed that the development of serious games can be highly dependent on the designers involved and the processes used. This must be considered when a hospital aims to develop multiple games for different purposes.

(JMIR Serious Games 2021;9(1):e16029) doi:10.2196/16029
KEYWORDS
cancer patients; children; adolescents; social relation; emotional well-being; gamification; relatives; visual design; serious games

Introduction
Patients diagnosed with cancer often undergo aggressive oncological treatment mainly provided by outpatient clinics at hospitals [1]. The patients experience a wide range of changing symptoms that occur due to the cancer itself as well as side effects from the treatment [2]. Sometimes, there are complications, resulting in a potentially life-threatening situation for the patient, requiring admission to the hospital. The two main disadvantages of such situations are related. First, the patients may suffer due to being away from their family [3]. Second, patients find themselves in a distressing situation and are left alone with their concerns and worries. In these situations, patients benefit from receiving emotional and social support from their close relatives. Social support includes personal, informal advice, and can provide strength that helps individuals initiate and sustain self-management activities. Existing literature has identified a correlation between patients’ perceived degree of social support and quality of life [4]. Further, children and adolescents with a parent with cancer have noted that they are not seen, heard, and listened to as the worried relative they are [5]. Few studies have explored the experiences of children and adolescents across a parent’s illness trajectory; however, a recent meta-synthesis recommended early supportive interventions tailored to include the children of intensive care patients [6]. To our knowledge, no studies have explored how to include children who are visiting their parent at an oncological department. Still, there is a general acknowledgment that health care professionals need to reflect on how to approach children and young adolescents at the hospital when they are visiting a parent [5,7]. Gamification, serious gaming, and edutainment have the potential to establish enjoyable and motivating experiences when gaming elements are introduced in a nongaming context [8]. The notion of gamification and serious gaming is traditionally used to increase adherence to medication regimens, promote health education for patients, and offer apps for chronic disease management [9]. Video games and serious gaming have the ability to enchant and engage patients and their children in a different “world” and thus give them a collective short break from the otherwise rather serious context they are in together. We are not aware of other studies aimed at facilitating social and joyful interaction between a parent with cancer and their children through gamification during hospitalization in an oncological department. However, virtual emotional authenticity and feelings of empathy may be able to help patients with cancer frame the often traumatizing experiences in more positive terms [10]. The development of serious games for such purposes is, as in eHealth generally, based on scientific and design foundations [11]. However, the product developed for a given purpose might be dependent on the software provider, the processes, or the specific designer asked to address specific needs. By bringing together different disciplines and acknowledging the diversity in art and design, more solutions might be designed for the same purpose [12]. This study aims to develop tools to enhance social interaction between a parent with cancer and their children when the children visit the parent at the hospital.

Methods
Overview
This study is based on an interdisciplinary co-creation process with clinical oncology specialists from Copenhagen University Hospital (KP, HG, and HP), gaming and visual design experts from The Royal Danish Academy of Architecture, Design, and Conservation (TK, JK, and JJ), and second-year bachelor’s degree students at the Academy. The study did not need ethical approval. For this study, over 3 months, 4 different groups of design students within the Visual Design program were tasked with developing games addressing the objectives of the field of design and focusing on strengthening relations in situ during treatment. This modified user-centered design had the child (aged 4-18 years) of a parent with cancer as its primary user. Target age groups were defined in the initial phases of the project based on statistical information from the department and the already established research focus of next of kin and young adults as relatives, as well as by established researchers from the field of developmental psychology and the complexity of the developed games [12]. We approached the target groups through the lens of media use, where a concrete game is part of the broader media ecology, meaning that material from a television series can cross into a video game, and a video game can cross into children’s outdoor analog play [13]. This has a specificity that exceeds the level of detail offered by developmental psychology, and the guiding principle for design was therefore observations of the actual game and media use by children in the respective age brackets [13]. The activities in the 4 phases were structured and predefined.

Phase I
The clinical specialists presented the clinical settings and environment of the oncological department (eg, standard cancer disease and treatment trajectories) to the students, as well as the context of the study aim and ethical issues. The students signed a confidentiality agreement. Discussions on how to establish and strengthen social and joyful togetherness between the admitted parent with cancer and their children were facilitated. There was an underlying consideration of how to nurture the experience of social support that may play an important role in a multifaceted approach to preventing complicated grief among relatives. The hospital and its staff were defined as customers in the development process, and the field of developmental psychology was researched throughout the process [13].

Phase II
The students interviewed nurses (n=4) who were selected to be informers due to their extensive experiences with patients with cancer and their relatives.
Phase III
Students listened to the story of a bereaved child from The Danish National Center for Grief. Students were given a lecture by a researcher with a special interest in and knowledge of children as relatives of parents with cancer [14]. Students could ask questions of and engage in discussions with both lecturers.

Phase IV
The students presented their initial prototypes and received feedback from the multidisciplinary team. Acceptability testing was performed in an external population; a class of healthy children from the same age group was used for this purpose. Adjustments were incorporated ahead of an official presentation at the hospital and visitors at the prototype exhibition were randomly included in the usability test. Hence, the prototypes were tested for usability using this small, selected population via surveys. Tests of the prototypes with hospitalized parents with cancer and their children were not performed due to ethical considerations.

Results
In phases I-III, the facts used for the design process were documented in written reports by the 4 working groups. Information from interdisciplinary stakeholders such as patients, nurses, research specialists, clinical oncological specialists, and gaming and visual design experts was included as a design foundation for all reports. In phase IV, 4 independent prototype games were developed (Table 1). Each game included an acceptance test using nontarget populations.

Table 1. Prototypes of the games.

<table>
<thead>
<tr>
<th>Game title</th>
<th>Theme</th>
<th>Target age group (years)</th>
<th>Element of interaction</th>
<th>How the game aims to create social interaction</th>
<th>Number of players</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fun and Cozy Socks</td>
<td>Good mood gaming</td>
<td>15-18</td>
<td>Sample of social and interactive guessing minigame</td>
<td>Establishes a feel-good atmosphere with competition and laughs</td>
<td>2-3</td>
</tr>
<tr>
<td>The Witch</td>
<td>Magic team building game</td>
<td>4-8</td>
<td>Cooperative multiplayer game, book, and electronic game</td>
<td>Provides a fun and enjoyable activity that builds trust between the players</td>
<td>2</td>
</tr>
<tr>
<td>Kurt, the Hospital Cat</td>
<td>Augmented reality</td>
<td>6-12</td>
<td>The interactive experience of a real-world environment with a search for items in the physical surroundings</td>
<td>Adds information and meaning to the hospital surroundings</td>
<td>1</td>
</tr>
<tr>
<td>Bloom</td>
<td>Growing a garden and bringing it to life</td>
<td>4-18</td>
<td>Customization/minigames</td>
<td>Makes the players feel good and is enjoyable. Shows that you care for a person by growing flowers and taking care of them.</td>
<td>1 player on a social platform</td>
</tr>
</tbody>
</table>

In general, the prototypes were highly accepted by age-equivalent testing groups. A usability test among random visitors at the parallel presentation of all 4 solutions included patients (n=6), health care professionals (n=10), and young adults (n=3). The survey results included good ratings for “interesting game,” “user friendly,” and “willingness to use” in a hospital setting. In general, all solutions were positively evaluated, reflecting a fundamental interest in playing games to experience joyful moments. However, the older population (aged >70 years) seemed to have reduced digital competencies and a lower willingness to use such games.

The prototypes (Figure 1) were tested on a small, selected population via a paper-based survey. Physical contact testing was not possible due to ethical concerns.

Discussion
Principal Findings
Based on the same information and evaluation, 4 different working groups skilled in visual design developed games with different themes and interactions, for different target age groups, to solve the same challenge. This study shows that gamification has the potential to solve newly revealed needs within health care, such as improving interactions between a parent with cancer and their child. This study suggests that several types of designs can solve the same problem, but also underscores that...
solutions are highly dependent on the software developer and designer. These factors should be considered when, for example, a hospital aims to develop multiple games for different purposes, or when the best game for a specific purpose is requested.

It has previously been described how patient and family involvement is valuable for health [15]; however, little is known about the involvement of children as relatives. In the adult setting, eHealth tools have been developed for adult relatives [16] to improve their psychological well-being, but not to improve their interactions with the patient. It is known that web-based tools and games can be used to improve adolescents’ and young adults’ adherence to vaccination programs and therapy [17,18]. It seems that eHealth tools can improve psychological outcomes [19]. It might be argued that the development of serious games for health should be based on a framework suggested by Verschueren et al [11]. However, to improve the quality of serious games for health, this study tries to encourage interdisciplinary cooperation between art and health science experts to elicit a greater number of suggestions of game-based solutions for the same purpose, followed by a selection of the best feasible solution for a specific need considering the different themes and graphic design elements of the developed games.

Limitations
A limitation of this study might be that the tools were developed by students; however, they were supervised by experts in gamification and evaluated and given continuous feedback during the process by user design and gamification experts.

A modified concept of user-centered design was applied to this tool developing process, where experiences from one young relative and other stakeholders were the basis for the development. However, this one representative was used to represent a group of children and young relatives of patients with cancer.

All 4 working groups performed an acceptance test of their product in a nonhospital environment (eg, a day care facility or school). The usability test involved the parallel testing of all 4 tools in a hospital environment and included children/young people, patients/relatives, and hospital staff. None of these tests were performed with a larger group of children who are relatives of patients with cancer; thus, the generalizability of the findings must be further explored.

Conclusion
A game can be an efficient method for fostering social interaction between a parent with cancer and a child or an adolescent, and games likely have the power to improve the difficult social and psychological relations between parents and children in this context. When creating games for specific purposes such as this, it must be remembered that games designed based on the same assignment might use different themes and graphic design elements, indicating that the development of serious games may be highly dependent on the designer involved. The described games remain to be investigated more thoroughly regarding their feasibility and efficacy in the context of qualitative oncoligical research. If shown to be feasible, implementation among patients with cancer and their relatives will be carried out.

Acknowledgments
Thanks to the students at The Royal Danish Academy of Architecture, Design, and Conservation for taking part in this study. Thanks to Annemarie Dencker, University of Southern Denmark, and The Danish National Center for Grief for participating in phase III.

Conflicts of Interest
None declared.

References


©Karin Piil, Helle Holm Gyldenvang, Jeppe Kilberg Møller, Tine Kjoelsøen, Jesper Juul, Helle Pappot. Originally published in JMIR Serious Games (http://games.jmir.org), 21.01.2021. 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 http://games.jmir.org, as well as this copyright and license information must be included.
Validation of a Portable Game Controller to Assess Peak Expiratory Flow Against Conventional Spirometry in Children: Cross-sectional Study

Khadijja Chelabi1, DCS; Fabio Balli2,3, MAS; Myriam Bransi4,5, MD; Yannick Gervais2, GradD; Clement Marthe2, MSc; Sze Man Tse6,7, MDCM, MPH, FRCPC

1Faculty of Medicine, McGill University, Montreal, QC, Canada
2Breathing Games Association, Geneva, Switzerland
3Milieux Institute, Concordia University, Montreal, QC, Canada
4Faculty of Medicine, Laval University, Quebec, QC, Canada
5Department of Pediatrics, Centre mère-enfant Soleil du CHU de Québec, Quebec, QC, Canada
6Division of Respiratory Medicine, Department of Pediatrics, Sainte-Justine University Hospital Center, Montreal, QC, Canada
7Faculty of Medicine, University of Montreal, Montreal, QC, Canada

Corresponding Author:
Sze Man Tse, MDCM, MPH, FRCPC
Division of Respiratory Medicine
Department of Pediatrics
Sainte-Justine University Hospital Center
3175 Chemin de la Côte-Sainte-Catherine
Montreal, QC, H3T 1C5
Canada
Phone: 1 (514) 345 4931 ext 5409
Email: sze.man.tse@umontreal.ca

Abstract

Background: International asthma guidelines recommend the monitoring of peak expiratory flow (PEF) as part of asthma self-management in children and adolescents who poorly perceive airflow obstruction, those with a history of severe exacerbations, or those who have difficulty controlling asthma. Measured with a peak flow meter, PEF represents a person’s maximum speed of expiration and helps individuals to follow their disease evolution and, ultimately, to prevent asthma exacerbations. However, patient adherence to regular peak flow meter use is poor, particularly in pediatric populations. To address this, we developed an interactive tablet-based game with a portable game controller that can transduce a signal from the user’s breath to generate a PEF value.

Objective: The purpose of this study was to evaluate the concordance between PEF values obtained with the game controller and various measures derived from conventional pulmonary function tests (ie, spirometry) and to synthesize the participants’ feedback.

Methods: In this cross-sectional multicenter study, 158 children (aged 8-15 years old) with a diagnosis or suspicion of asthma performed spirometry and played the game in one of two hospital university centers. We evaluated the correlation between PEF measured by both the game controller and spirometry, forced expiratory volume at 1 second (FEV1), and forced expiratory flow at 25%-75% of pulmonary volume (FEF25-75), using Spearman correlation. A Bland-Altman plot was generated for comparison of PEF measured by the game controller against PEF measured by spirometry. A post-game user feedback questionnaire was administered and analyzed.

Results: The participants had a mean age of 10.9 (SD 2.5) years, 44% (71/158) were female, and 88% (139/158) were White. On average, the pulmonary function of the participants was normal, including FEV1, PEF, and FEV1/forced vital capacity (FVC). The PEF measured by the game controller was reproducible in 96.2% (152/158) of participants according to standardized criteria. The PEF measured by the game controller presented a good correlation with PEF measured by spirometry (r=0.83, P<.001), with FEV1 (r=0.74, P<.001), and with FEF25-75 (r=0.65, P<.001). The PEF measured by the game controller presented an expected
mean bias of –36.4 L/min as compared to PEF measured by spirometry. The participants’ feedback was strongly positive, with 78.3% (123/157) reporting they would use the game if they had it at home.

**Conclusions:** The game controller we developed is an interactive tool appreciated by children with asthma, and the PEF values measured by the game controller are reproducible, with a good correlation to values measured by conventional spirometry. Future studies are necessary to evaluate the clinical impact this novel tool might have on asthma management and its potential use in an out-of-hospital setting.

**KEYWORDS**
asthma; pediatrics; serious game; peak expiratory flow; pulmonary function test, adherence, self-management

**Introduction**

Asthma is a chronic pulmonary disorder that affects more than 339 million people globally [1]. It is the world’s most prevalent chronic disease in children, affecting 7.5% of children in the US [2], and can have a significant impact on one’s life. Indeed, asthma attacks result in 146,000 annual visits to emergency departments and are the primary cause for school absence [3]. Characterized by reversible airway narrowing and chronic inflammation, asthma is a heterogeneous disease that can present with diverse respiratory symptoms, which can be periodically exacerbated by personal triggers [4,5]. While recognition of asthma symptoms is key to managing and preventing asthma exacerbations, self-perception and self-reporting of symptoms by the affected individuals is often difficult and unreliable [6]. This is particularly problematic in children. As such, international guidelines [7-9] recommend self-monitoring of PEF in the asthmatic population starting at 5 years of age, more specifically in children and adolescents who poorly perceive airflow obstruction, those with a history of severe exacerbations, or those who have difficulty controlling asthma. While symptoms-based asthma action plans have been shown to be superior to peak flow–based plans, peak flow may be used as a complementary measure to monitor symptoms in self-management of asthma. In fact, studies [10-12] have effectively shown that daily measurement of PEF in children as a self-monitoring tool can increase adherence to medications and decrease asthma exacerbations in children with poor symptoms perception. Despite this evidence, there is a low adherence to PEF monitoring using current peak flow meters in the pediatric population, with a reported 15% adherence to the Mini Wright Peak Flow Meter in children only 3 weeks after initial use [13]. Some of the identified barriers to PEF monitoring are the self-perception that one’s asthma is well-controlled, the perceived burden of taking frequent measurements with seemingly low direct benefit to one’s general health, and the financial barriers to purchasing a peak flow meter, with the cost ranging from 20-50 Canadian dollars per device (US $15.73-$39.33) [14]. This issue of low adherence to PEF monitoring in pediatric asthma needs to be addressed in order to engage and empower children early in healthy management of their chronic illness, improve their ability to perceive symptoms, and potentially decrease exacerbations.

Serious games, defined as games with a primary purpose other than pure entertainment [15], can be a means to increase adherence to treatments or therapies in chronic diseases and have been associated with positive health outcomes [16-18]. Serious games may be used to actively engage patients in their health management in a playful setting and are of particular use to children, given the gaming approach. Although several serious games have addressed the goal of knowledge acquisition in asthma [19], none has yet suggested a serious game approach to self-monitoring of asthma symptoms, including the measurement of PEF. Importantly, airflow measurements taken through serious games need to be validated against standard measures before the games can be implemented clinically.

In this study, we evaluated the validity of an airflow-based game controller paired with the TikiFlow game for portable electronic devices to monitor PEF in children with asthma using conventional spirometry as the comparison standard. We evaluated the concordance between PEF measurements obtained with the game controller and several spirometry measurements, and the reproducibility of PEF measured by the game controller without coaching. Furthermore, we assessed the participants’ feedback both quantitatively and qualitatively to guide the continuous development of the game.

**Methods**

**Game Controller and Game Design**

We designed a 3D-printed game controller to transduce the user’s breath into a digital signal and to meet the calibration standards of the Omron PF9940 peak flow meter. The game controller uses Bluetooth to connect to a portable electronic device and allows control of the TikiFlow game by the means of airflow through the game controller (Figure 1). The TikiFlow game challenges the participant to exert a forceful expiration in order to propel plant seeds, trigger a rainfall, and repopulate an island destroyed by pollution. PEF results measured by the game controller appear at the end of the game. The game can be previewed online [20].
Design of the Study

We conducted a cross-sectional study on children who presented to the pulmonary function test laboratory of two pediatric university health centers in Quebec, Canada (Centre Hospitalier Universitaire Sainte-Justine and Centre Hospitalier Universitaire de Québec – Université Laval) over a 3-week period in November 2019. The respective institutional research ethics board approved this study (Centre Hospitalier Universitaire Sainte-Justine #MP-21-2020-2310, Centre Hospitalier Universitaire de Québec-Université Laval #MEO-21-2020-4711 and Concordia University #30011659).

Study Population

We included children aged 8-15 years old with a diagnosis of asthma or a suspicion of asthma presenting to pulmonary function test laboratory for a spirometry, with or without bronchodilator reversibility testing, at their physician’s request. We excluded children with cystic fibrosis, bronchopulmonary dysplasia, and other conditions where the indication for pulmonary function test was unrelated to asthma (eg, follow-up of oncology patients or esophageal atresia). We also excluded children who were unable to cooperate for either conventional spirometry or assessment via the game controller according to American Thoracic Society guidelines. We identified eligible participants through the pulmonary function test laboratory database of both pediatric university health centers.

Measurements

We asked eligible and consenting participants to play TikiFlow with the game controller without coaching within 10 minutes of their conventional spirometry. Specifically, participants were given a brief overview of the game and its purpose and were asked to play on their own without instructions to exert more forceful exhalations or to increase the effort of breathing. Following the game, they also filled out a self-reported general health questionnaire and a satisfaction questionnaire.

Conventional spirometry measurements were obtained from a Jaeger MasterScope spirometer (Cardinal Health, Dublin, OH), using the Global Lung Initiative 2012 reference values [21] and interpreted according to American Thoracic Society guidelines [22]. Specifically, we retained the forced expiratory volume at 1 second (FEV₁), the forced expiratory flow at 25%-75% of pulmonary volume (FEF₂₅-₇₅), and the forced vital capacity values from the participants’ best curve, and the highest PEF value (PEF measured by spirometry). For PEF measured by spirometry, we evaluated the reproducibility of the spirometry in accordance with the American Thoracic Society guidelines, defined as the 2 highest values falling within 40L/min of each other [22]. For PEF measured by the game controller, we established a minimum of 3 reproducible forced exhalations as necessary, of which 2 must meet the American Thoracic Society reproducibility criteria, in order to end the game. The game was stopped if reproducibility could not be achieved after 5 exhalations. If the child was prescribed a reversibility testing with bronchodilators, which consists of repeating the spirometry 15 minutes following the administration of the bronchodilator salbutamol, we repeated the PEF game controller measurements following the post-bronchodilator spirometry. We collected the participant’s appreciation of the game through a 5-question Likert-scale questionnaire and narrative feedback.

Statistical Analysis

We portrayed a descriptive analysis of the participants’ baseline health and demographics based on the results of the self-reported questionnaire. A Bland-Altman plot was used to compare the game controller method to that of conventional spirometry and to evaluate the concordance of both methods to measure PEF. We calculated the spearman correlation between PEF measured by the game controller and by conventional spirometry measures. Statistical significance was set at $P<0.05$. Analyses were performed with R software, version 3.5.0 [23]. Participant’s appreciation of the game was evaluated quantitatively based on the satisfaction questionnaire administered and results were plotted in a bar graph.

Results

Population

We identified 337 children aged 8-15 years old presenting to both centers for a spirometry, with or without bronchodilator reversibility testing, 187 of whom were eligible based on inclusion criteria. We included 158 participants, as 9 were unable to perform spirometry and 20 refused to take part in the study, mostly due to time constraints (see Multimedia Appendix 1).

Participants had a mean age of 10.9 (SD 2.5) years (Table 1), and most were male (88/158, 55.7%), a proportion that reflects the sex distribution in pediatric asthma in the general population [24]. The majority were White (139/158, 88.0%). Up to 68%
of participants presented an atopic profile, defined as also having eczema or food or environmental allergies. The characteristics of participants were similar between both centers (see Multimedia Appendix 2).

Almost all participants (156/158, 98.7%) performed a reproducible spirometry. On average, the baseline lung function of participants was normal, with a mean percent predicted value of 103.7% (SD 11.7%) for forced vital capacity, 99.6% (SD 13.2%) for FEV₁, and 86.2% (SD 25.5%) for FEF₂₅-₇₅. However, varying degrees of airway obstruction are represented in our participants as demonstrated by the range of percent predicted values and z scores (Table 2; see Multimedia Appendix 3).

Table 1. Selected baseline characteristics of participants (n=158).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean years (SD)</td>
<td>10.9 (2.5)</td>
</tr>
<tr>
<td>Male, n (%)</td>
<td>88 (55.7)</td>
</tr>
<tr>
<td>BMI percentile, median (IQR)</td>
<td>49 (33.3-90.8)</td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>139 (88.0)</td>
</tr>
<tr>
<td>Black</td>
<td>11 (6.8)</td>
</tr>
<tr>
<td>Other</td>
<td>8 (5.1)</td>
</tr>
<tr>
<td>Diagnosis of asthma, n (%)</td>
<td>131 (82.9)</td>
</tr>
<tr>
<td>Age at diagnosis, mean years (SD)</td>
<td>4.2 (3.2)</td>
</tr>
<tr>
<td>Daily controller therapy, n (%)</td>
<td></td>
</tr>
<tr>
<td>ICS® only</td>
<td>133 (84.2)</td>
</tr>
<tr>
<td>ICS-LABA®</td>
<td>58 (36.7)</td>
</tr>
<tr>
<td>ICS-LTRA®</td>
<td>33 (20.9)</td>
</tr>
<tr>
<td>ICS-LABA-LTRA®</td>
<td>11 (7.0)</td>
</tr>
<tr>
<td>LTRA® only</td>
<td>27 (17.1)</td>
</tr>
<tr>
<td>Eczema, n (%)</td>
<td>45 (28.5)</td>
</tr>
<tr>
<td>Food allergy, n (%)</td>
<td>38 (24.1)</td>
</tr>
</tbody>
</table>

ICS: inhaled corticosteroids.
ICS-LABA: inhaled corticosteroids and long acting beta-agonist.
ICS-LTRA: inhaled corticosteroids and leukotriene receptor antagonist.
ICS-LABA-LTRA: inhaled corticosteroids, long acting beta-agonist and leukotriene receptor antagonist.
LTRA: leukotriene receptor antagonist.
Table 2. Baseline lung function assessed by conventional spirometry (n=158).

<table>
<thead>
<tr>
<th>Function</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FVC</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>% predicted, mean (SD)</td>
<td>103.7 (11.7)</td>
</tr>
<tr>
<td>% predicted, range</td>
<td>66.5 to 142.9</td>
</tr>
<tr>
<td>z score, mean (SD)</td>
<td>0.3 (1.0)</td>
</tr>
<tr>
<td>z score, range</td>
<td>−3.0 to 3.5</td>
</tr>
<tr>
<td>**FEV&lt;/sub&gt;1&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>% predicted, mean (SD)</td>
<td>99.6 (13.2)</td>
</tr>
<tr>
<td>% predicted, range</td>
<td>59.3 to 130.2</td>
</tr>
<tr>
<td>z score, mean (SD)</td>
<td>0.0 (1.1)</td>
</tr>
<tr>
<td>z score, range</td>
<td>−3.4 to 2.3</td>
</tr>
<tr>
<td>**FEF&lt;/sub&gt;25-75&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>% predicted, mean (SD)</td>
<td>86.2 (25.5)</td>
</tr>
<tr>
<td>% predicted, range</td>
<td>20.8 to 184.8</td>
</tr>
<tr>
<td>z score, mean (SD)</td>
<td>−0.7 (1.2)</td>
</tr>
<tr>
<td>z score, range</td>
<td>−4.1 to 3.6</td>
</tr>
<tr>
<td>FEV&lt;/sub&gt;1/FVC&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>mean (SD)</td>
<td>83.6 (6.8)</td>
</tr>
<tr>
<td>PEF&lt;/sub&gt;spiro&lt;sup&gt;e&lt;/sup&gt; (L/min), median percentile (IQR)</td>
<td>266.7 (230.1-336.9)</td>
</tr>
</tbody>
</table>

<sup>a</sup>FVC: forced vital capacity.

<sup>b</sup>FEV<sub>1</sub>: forced expiratory volume in 1 second.

<sup>c</sup>FEF<sub>25-75</sub>: forced expiratory flow at 25%-75% of the pulmonary volume.

<sup>d</sup>FEV<sub>1</sub>/FVC: forced expiratory volume in 1 second/forced vital capacity.

<sup>e</sup>PEF<sub>spiro</sub>: PEF measured by spirometry.

**Comparison of the Game Controller and Conventional Spirometry**

Despite the lack of coaching, 96.2% (152/158) of PEF values measured by the game controller were reproducible. Spearman correlations indicated that PEF measured by the game controller presented a good correlation with PEF measured by spirometry (r=0.83, P<.001), with FEV<sub>1</sub> (r=0.74, P<.001), and with PEF<sub>25-75</sub> (r=0.65, P<.001) (Figure 2). The Bland-Altman plot showed that the PEF measured by the game controller presented a mean bias of −36.4 L/min (95% CI −43.2 to 29.6) compared to PEF measured by spirometry. The 95% limits of agreement, defined as the values within which 95% of differences between the two measurement methods lay, were −121.0 and 48.3 L/min (Figure 3). We identified 5 outliers, whose exclusion in a sensitivity analysis did not result in significant change in the concordance or correlation measures.

**Figure 2.** Correlation between PEFGC and selected conventional spirometry measurements. PEFGC shows good correlation with PEFspiro (a; r=0.83, P<.001), FEV<sub>1</sub> (b; r=0.74, P<.001) and PEF<sub>25-75</sub> (c; r=0.65, P<.001). PEFGC: peak expiratory flow measured by the game controller. PEFspiro: peak expiratory flow measured by conventional spirometry. FEV<sub>1</sub>: forced expiratory volume in 1 second. PEF<sub>25-75</sub>: forced expiratory flow at 25%-75% pulmonary volume.
Figure 3. Bland-Altman plot comparing the game controller and conventional spirometry for PEF measurements. The game controller presents a mean bias of –36.4 L/min (blue; 95% CI –43.2 to –29.6) with a 95% upper limit of agreement of 48.3 (green; 95% CI 36.6 to 60.0) and a 95% lower limit of agreement of –121.0 (red; 95% CI –132.7 to –109.4) when compared to conventional spirometry. PEF: peak expiratory flow. PEF_{GC}: peak expiratory flow measured by the game controller. PEF_{spiro}: peak expiratory flow measured by conventional spirometry.

Of the 84 participants who underwent a spirometry pre- and post-bronchodilator administration, 7 were considered to have reversible airway obstruction as defined by a change in FEV\(_1\) of ≥12% post-bronchodilator [25]. Using the game controller, 4 of those 7 participants were considered to have reversible obstruction as defined by a change in PEF measured by the game controller of ≥15%[26]. Based on these limited numbers, we could not reliably assess the responsiveness of PEF measured by the game controller to bronchodilator as compared to conventional spirometry.

**Appreciation Feedback of the Game Controller**

Overall, users highly appreciated the game (Figure 4). No participant reported “Strongly disagree” to any of the 5 statements, and 78.3% (123/157) of participants reported they would use the game controller and play the game if they had it at home. 107 participants provided positive narrative feedback from which we noted that the main positive aspects of the game were the game concept (n=31), its educational aspect coupled with its entertaining aspect (n=14), and the ease of use of the game controller (n=25). Amongst the 28 youngest participants, aged 8-11 years old, who provided feedback on areas of improvement, the most recurrent comment was the desire for a longer duration to the game.

Of the 21 participants that gave negative narrative feedback, we noted the older users (aged 12-15 years) sometimes felt that the game content was intended for a younger population (n=6).
Figure 4. Game appreciation as reported by the participants.

Discussion

Principal Findings

This study shows that the PEF measurements obtained through our game controller present a good correlation with conventional spirometry measurements and have a high reproducibility. Additionally, the participants reported overall strongly positive feedback on the game controller and the game, with 78.3% (123/157) reporting that they would play the game if they had it at home. These findings suggest that our game and controller could provide an objective measure of airway obstruction and that its use in an out-of-hospital setting could be explored.

PEF self-monitoring has been shown to reduce asthma exacerbations through the development of asthma awareness and control [10–12]. However, despite the international asthma guidelines recommendations for its adoption in poor symptoms perceivers and those with difficult to control asthma, few physicians educate and recommend PEF monitoring to their patients [27,28]. Two of the reasons for clinicians’ low advocacy for peak flow meter use are the wide intraindividual variability in PEF values [29] and the low adherence to their use, particularly in pediatric patients with asthma.

Specifically, peak flow meters do not seem to consistently report PEF values that correlate with conventional spirometry, and a different measurement bias is associated with each peak flow meter [30,31]. Our game controller underestimated PEF measured by spirometry by 36.4 L/min, which is comparable to other peak flow meters [31-33]. This bias is expected as differing breathing techniques are used for PEF (child breathes out forcefully without coaching) and spirometry (child breathes out forcefully and completely with coaching) [31]. Reassuringly, our findings suggest a good correlation between PEF measured by the game controller and various spirometry measures across different degrees of baseline airway obstruction, suggesting that PEF measured by the game controller could be used as a surrogate to monitor trends in lung function. Furthermore, game controller measures were reproducible in 96.2% (152/158) of participants despite the lack of coaching, likely facilitated by the intuitiveness of the game narrative and mechanics. This finding suggests that PEF self-monitoring with our game controller would be feasible in a nonclinical environment without the supervision of a health care professional. Although we are not advocating for peak flow to replace symptoms-based asthma action plans, we believe that the addition of this objective measure in select patients can help with asthma control. This feature could have potentially significant benefits, namely by reducing barriers to health care access and reducing costs and transportsations, all of which need to be further studied.

One suggested approach to increase adherence to medical treatments and to engage and empower the population in management of diseases is the use of serious games. Serious games have found their utility in different disciplines of medicine, whether to distract patients from a painful procedure, to engage adolescents in psychotherapies, or to educate patients on complex medical topics. Importantly, studies report a high adherence of serious games whether at home or in clinical settings [17-19]. Only 5 of 187 (2.7%) eligible children refused to participate due to a lack of interest, suggesting that TikiFlow appealed to the vast majority of children (see Multimedia Appendix 1). Additionally, more than 75% of the participants reported that they enjoyed the game (150/156), that it was fun to play (131/157), easy to use (139/157), and that they would use it at home (123/157). These findings were independent of age or sex.

Another approach to increase adherence specifically to PEF monitoring in asthma is the automatic generation of diaries as opposed to the traditional handwritten personal diaries of PEF.
measures. This approach decreases the effort required by the user and reduces the risk of false data generation. It also offers the possibility to depict measures graphically over time to identify trends. A recent study in the United Kingdom [34] suggests smartphone self-generated PEF diaries alone increase compliance with up to 32% of adults taking at least one PEF measurement daily for 3 months after the initiation of the study and 28% after 6 months. Our game takes advantage of this approach and integrates an automatically generated diary of PEF measures.

Serious games may be particularly beneficial among children with chronic conditions, where they were found to improve knowledge level and self-management [18]. The role of serious games in asthma self-management has been little explored. Asthma-related serious games, to date, have focused heavily on asthma education, and while they seem to improve knowledge transfer, they have been found to have little to no effect on behavior modification [19]. The novelty in our game, TikiFlow, is the use of a game controller with which the child uses their breath to control the game, bringing the focus to the play, rather than education alone. The acquisition of good asthma self-care habits early in childhood and adolescence may translate into long-lasting benefits in adulthood [35].

One notable strength of our game is its accessibility. Asthma is a global disease and the development of innovative tools to help patients and their families in their asthma management must be accessible to individuals in low- and middle-income countries. Smartphones and tablets represent ideal mediums to help patients and our controller is 3D-printed, which makes it easy to reproduce at low cost and with minimal equipment, reducing access barriers. Furthermore, its small size and its portable format makes it easy to carry and incorporate in daily activities and, importantly, easy to disinfect.

There are several limitations to this study. First, our study does not include children aged 5-8 years old despite the international asthma guidelines recommending PEF monitoring in children aged ≥5 years old [8,37]. Our decision was to limit the confounders of poor effort and lack of cooperation in young children in the assessment of the game controller’s validity. Further studies are necessary to evaluate the feasibility of PEF game controller measurements in this age group. Second, our study was conducted in tertiary care centers with a relatively healthy population with average normal functions at baseline. Thus, our results cannot be generalized to individuals with lower lung function or to those who are in acute exacerbations. Third, all game controller measurements were done systematically after conventional spirometry. Therefore, it is possible that fatigue contributed to skewing the game controller’s measurements to lower values in our study. Participants were given instructions for the conventional spirometry prior to the game, which may have motivated a better effort on the game controller than could be expected with home measurements. However, we explicitly did not give instructions for the use of the game in order to reproduce the home environment as much as possible. Because we only had a few participants with documented reversibility to bronchodilator, we were not able to evaluate the responsiveness of PEF measured by the game controller to bronchodilator. Future studies are needed to explore the characteristics of PEF measured by the game controller and whether it can help predict asthma exacerbations.

Conclusions

Our study evaluated the use of a novel airflow-based game controller coupled with a smart device serious game in children with asthma. Our game controller produced reproducible PEF measures that correlated well with several conventional spirometry measures and was highly appreciated by participants. Thus, our game controller, coupled with a serious game, may have the potential to increase adherence to PEF self-monitoring in children. Further studies are necessary to evaluate users’ adherence over time, the validity of the game controller in different situations, and its impact on clinical outcomes, specifically its role in the prevention of asthma exacerbations.

Acknowledgments

We would like to thank Dr Alena Valderrama and Dr Valentin Gomez, who initiated the collaboration on games for asthma at Sainte-Justine hospital; Renaud Ory, Lucas Delvalle, Aminata Pierson, and Levan Jannneret, who developed the initial version of TikiFlow in Switzerland; Collin Gallacher and Steve Ding, who developed the game controller building on the work of Tiberius Brastaviceanu and Jim Anastassiou in Canada; the teams and patients of Dr Isabelle Sermet-Gaudelus (Necker hospital), Dr Pierre-Régis Burgel (Cochin hospital), and Odile Flez (Fondation Arc-en-Ciel), who tested the game controller in France; all the families who participated in this study; the research support staff at each of two recruiting centers, specifically Mary-Ellen French, Louise Gosselin, Vincent Laguè, and Mylène Leblanc; and all other contributors to the initiative. This study was funded by a grant from the Canadian Institute of Health Research (CIHR), Strategy for Patient-Oriented Research program (#151755). Publication fees were paid by Dr. Sze Man Tse’s research funds and the Breathing Games Association (Switzerland). The participatory design of the game and controller were funded by grants from CIHR (#151755), the French Hospitals Federation - Fonds HFF (#2017), Concordia University Council of Student Life (#63), Concordia University Sustainability Action Fund (#PYPF); Sainte-Justine health promotion and pulmonology services; Quebec university hospital; and Breathing Games Association. TikiFlow and the controller were initially developed at three cocreation events that were supported by Open Geneva, Swiss Game Center, Genevan Foundation against Cystic Fibrosis, Lift, University of Geneva, La Suite Necker, Hacking Health Besançon, Fablab Grand Besançon, and Maison des parents Besançon. The source code of TikiFlow [38] and the design of the controller [39] can be found online.
Authors' Contributions

KC participated in the data collection, analyses and interpretation, and wrote the manuscript. FB participated in the design of the study and of TikiFlow. YG and CM participated in the design of TikiFlow. MB participated in the study design, data collection, and analyses. SMT participated in the design of the study, TikiFlow, data analyses and interpretation. FB and SMT secured funding for the study. All authors reviewed and approved the manuscript.

Conflicts of Interest

FB and YG cofounded the Breathing Games commons. The development of the game and game controller were led by Breathing Games. All other authors report no competing interests in this study.

Multimedia Appendix 1

Flowchart of patient enrolment in the study. [PDF File (Adobe PDF File), 58 KB - games_v9i1e25052_app1.pdf ]

Multimedia Appendix 2

Baseline characteristics of participants. [PDF File (Adobe PDF File), 99 KB - games_v9i1e25052_app2.pdf ]

Multimedia Appendix 3

Baseline lung function assessed by conventional spirometry. [PDF File (Adobe PDF File), 15 KB - games_v9i1e25052_app3.pdf ]

References

2. Most Recent National Asthma Data | CDC Internet. Center for Disease Control and Prevention. URL: https://www.cdc.gov/asthma/most_recent_national_asthma_data.htm [accessed 2020-07-31]


23. R project. URL: www.r-project.org [accessed 2020-08-13]


34. Antalfy T, De SA, Griffiths C. Promising peak flow diary compliance with an electronic peak flow meter and linked smartphone app. npj Primary Care Respiratory Medicine. 00 May 8;? 2020 May 08;30(1):2. [doi: 10.1038/s41533-020-0178-y]


38. Source code of TikiFlow. URL: https://gitlab.com/breathinggames/bg_tikiflow

39. Design of the game controller for TikiFlow. URL: https://breathinggames.net/hardware

Abbreviations

\( \text{FEF}_{25-75} \): forced expiratory flow at 25%-75% of pulmonary volume

\( \text{FEV}_1 \): forced expiratory volume at 1 second

\( \text{FVC} \): forced vital capacity

\( \text{PEF} \): peak expiratory flow
Please cite as:
Chelabi K, Balli F, Bransi M, Gervais Y, Marthe C, Tse SM
Validation of a Portable Game Controller to Assess Peak Expiratory Flow Against Conventional Spirometry in Children: Cross-sectional Study
JMIR Serious Games 2021;9(1):e25052
URL: http://games.jmir.org/2021/1/e25052/
doi:10.2196/25052
PMID:33512226
Original Paper

Pupillary Responses for Cognitive Load Measurement to Classify Difficulty Levels in an Educational Video Game: Empirical Study

Hugo Mitre-Hernandez¹, PhD, MSc, BSc; Roberto Covarrubias Carrillo¹, BSc, MSc; Carlos Lara-Alvarez¹,², PhD, MSc, BSc

¹Center for Research in Mathematics, Zacatecas, Mexico
²Center for Research and Advanced Studies of the National Polytechnic Institute, Tamaulipas, Ciudad Victoria, Mexico

Corresponding Author:
Carlos Lara-Alvarez, PhD, MSc, BSc
Center for Research in Mathematics
Calle Lasec y Andador Galileo Galilei
Quantum, Ciudad del Conocimiento
Zacatecas, 98160
Mexico
Phone: 52 4929980 ext 1105
Email: c.alberto.lara@gmail.com

Abstract

Background: A learning task recurrently perceived as easy (or hard) may cause poor learning results. Gamer data such as errors, attempts, or time to finish a challenge are widely used to estimate the perceived difficulty level. In other contexts, pupillometry is widely used to measure cognitive load (mental effort); hence, this may describe the perceived task difficulty.

Objective: This study aims to assess the use of task-evoked pupillary responses to measure the cognitive load measure for describing the difficulty levels in a video game. In addition, it proposes an image filter to better estimate baseline pupil size and to reduce the screen luminescence effect.

Methods: We conducted an experiment that compares the baseline estimated from our filter against that estimated from common approaches. Then, a classifier with different pupil features was used to classify the difficulty of a data set containing information from students playing a video game for practicing math fractions.

Results: We observed that the proposed filter better estimates a baseline. Mauchly’s test of sphericity indicated that the assumption of sphericity had been violated ($\chi^2_{14}=0.05; P=.001$); therefore, a Greenhouse-Geisser correction was used ($\epsilon=0.47$). There was a significant difference in mean pupil diameter change (MPDC) estimated from different baseline images with the scramble filter ($F_{5,78}=30.965; P<.001$). Moreover, according to the Wilcoxon signed rank test, pupillary response features that better describe the difficulty level were MPDC ($z=-2.15; P=.03$) and peak dilation ($z=-3.58; P<.001$). A random forest classifier for easy and hard levels of difficulty showed an accuracy of 75% when the gamer data were used, but the accuracy increased to 87.5% when pupillary measurements were included.

Conclusions: The screen luminescence effect on pupil size is reduced with a scrambled filter on the background video game image. Finally, pupillary response data can improve classifier accuracy for the perceived difficulty of levels in educational video games.

(JMIR Serious Games 2021;9(1):e21620) doi:10.2196/21620

KEYWORDS
video games; pupil; metacognitive monitoring; educational technology; machine learning

Introduction

Overview

An educational video game (EVG) is a video game that provides learning or training value to the player. Potential contributions of video games cover each of the three main fields of psychology: the affective (awakening feelings), the connate (aggressive or impulsive behavior), and the cognitive (learning-related skills) [1].
Video games have been demonstrated to be effective for improving working memory, mental rotation skills, and geometry performance [2]. Some of the effective features of educational video games include a clear goal, an adequate level of difficulty, quick-moving stimuli, and integrated instructions [3].

Several works have used EVGs to foster fraction understanding and to assess students [4,5]. However, our research focuses on the cognitive load (mental effort) generated by reasoning tasks [6] about math fractions; this is a direct way to measure the difficulty perceived by the EVG's player.

Video game difficulty refers to the amount of skill required by the player to progress through the game experience. Studying how to set an adequate difficulty level has attracted particular attention in the educational video games field [7,8]. Basic approaches to setting difficulty include allowing users to manually select levels and increasing the difficulty at a steady rate over the course of the game, with earlier levels being easier and later levels being harder [9]. Manually adapting difficulty or designing an incremental-difficulty solution could cause serious problems; for instance, the player may not know how they will perform before playing a given level, or the predefined change rate could be slower or faster than required by the player.

On the other hand, dynamic difficulty adjustment or dynamic difficulty balancing changes the game behavior according to the skill level of the players. For this purpose, the dynamic difficulty adjustment requires evaluation of the player's performance (through game scores, time, number of errors, player's decisions, etc) and adjustment of a set of game variables that regulate difficulty [10]. It has been shown that a dynamic approach that uses gamer behavior data presents better learning outcomes than an incremental difficulty approach [7].

As a step toward finding an imperceptible difficulty control, this paper proposes to use pupil dilation to detect very easy (or hard) activities. It is known that pupil dilation reflects activity in the brain as cognitive load—that is, the total amount of mental effort (information processing) induced by reasoning tasks or involving memory resources [6,11].

**Background**

**The Impact of Difficulty on Learning**

The flow experience model, proposed by Csikszentmihalyi [12], marks an achieved balance of arousal-increasing and arousal-decreasing processes. As shown in Figure 1, the flow model describes this balance in terms of the fit between perceived challenges and skills: an activity wherein challenges predominate increases arousal; an activity wherein skills predominate reduces arousal. Thus, a synchrony of challenges and skills permits a state of deep involvement, while the pitfalls of either over- or under-arousal (ie, anxiety or boredom) are avoided [12].

![Figure 1. Flow experience model of Mihaly Csikszentmihalyi [12].](image)

The dynamic flow passing through states $a \rightarrow c \rightarrow e$ shown in Figure 1 is the optimal path for increasing difficulty. However, $b_1 \rightarrow d_1$ are states of anxiety that demand new learning skills to return to optimal flow. Moreover, $b_2 \rightarrow d_2$ are states of boredom that need more challenges to return to optimal flow.

Several studies have supported that the rate of change of pupil diameter is related to task difficulty.

**Pupillary Responses**

The eye can be seen as a camera, with the pupil as the eye aperture, and it involves the iris activity [13]. The iris movement is controlled by the activity of two muscles, the dilator and the
Cognitive Load and Pupillary Response

The cognitive load (mental activity) imposed by tasks has a pupillary response, known as a task-evoked pupillary response (TEPR) [20]. TEPR occurs shortly after the onset of a task and subsides quickly after the mental activity is terminated. The TEPR depends on several factors; for instance, the response is greater for novice participants doing an arithmetic task than for an expert because novices require more mental effort [21]. Then, through pupillometry (measuring the pupil diameter), one can decide whether a challenge is adequate for the skills of a learner (Figure 1); that is, we can balance a video game to maximize the learning outcomes.

Pupil diameter is widely used to study cognitive load. Researchers have studied this relationship in different tasks, such as driving a vehicle while listening to a dialog, reasoning through math exercises, memorizing numbers, and perceiving visual stimuli [6,22,23].

Concerning industrial areas, cognitive load has been used in automotive and healthcare applications to optimize user's decision-making tasks [21,24]. Most studies in these fields are oriented to discover how to preserve attention and mental work on primary tasks and how to reduce it on secondary tasks to avoid critical errors. In addition, cognitive load has been used in video game studies without significant results, mainly due to changes in screen luminescence.

Playing EVG involves memorization and reasoning tasks that are associated with cognitive load. This paper uses pupillary response data to assess cognitive load in educational video games.

Beatty [6] points out that pupillary responses occur at short latencies following the onset of mental processing and subside quickly once processing is terminated. Most of the latency is due to slow iris muscle constriction. Different features have been used to evaluate cognitive load with pupillary responses such as mean pupil diameter change (MPDC), average percentage change in pupil size (APCPS), peak dilation (PD), and latency to peak (LP) [13,24-26].

Estimating Pupillary Responses

Individual differences in pupil size have been well documented; for example, pupil size decreases linearly as a function of age at all illuminance levels, and students high in cognitive ability have a larger pupil size [27,28]. These differences must be considered when studying factors that dilate the pupil; for this purpose, researchers calculate a pupil baseline interval for each individual separately. Then, the pupil change is estimated by contrasting information from the baseline and testing intervals. In the baseline period, users fixate on a predefined screen before the stimulus is presented. Baseline duration ranges from 400 milliseconds to 10 seconds [6,29-32]. In general, the variation in the baseline duration should play no substantial role in reporting pupil dilation [33]. Unsworth et al [32] suggest that better results can be obtained by using a longer duration; hence, they use 5 seconds to estimate the baseline.

A common practice is to use a neutral image, either black, gray, or white [31,34]; a gray image is more effective to reduce screen luminescence [35]. Using a neutral image is good enough for controlled tests that use luminance-controlled images, but there are significant changes in pupil size due to luminescence when participants play video games [36,37]. Studying the pupil dilation induced by mental activity when participants are exposed to environmental illumination changes is a challenge. For instance, several authors have reported that pupillary response features are directly correlated to cognitive load. Other authors, however, do not observe such correlations, and they suggest that this effect could be caused by luminescence changes [38,39].

Obtaining a baseline for each trial rather than for a whole test session is a common practice [33]; this is an applicable solution for settings where the screen luminescence remains stable for certain periods (eg, for a video game stage that is mainly dominated by the background). For these cases, the baseline is usually calculated from data generated by observing a scrambled image (ie, one image obtained by applying a scrambling scheme to a representative image in the period test).

Image scrambling [40] has two objectives: to transform a meaningful image into a meaningless or disordered image and to have the same mean intensity for the scrambled and original images.

The nonlinear relationship between luminescence changes and pupil size is one of the main difficulties when studying cognitive load in real conditions. Wong et al [41] study four approaches (ignoring, excluding, compensating, or using pupillary light reflex features) to mitigate the luminance change in cognitive load measurements. They found that ignoring the luminance change is the worst option. This paper proposes an initial
solution for studying cognitive load in real scenarios that is complementary to the approaches in the aforementioned study [41].

We hypothesize that a better baseline can be estimated from an image that maintains both the mean and local intensity. We tested grid scrambled images for obtaining the baseline. A grid scrambled image is generated by selecting a representative image within the measurement period, splitting it into a \( n \times m \) grid (\( n \) columns and \( m \) rows), and finally, scrambling each region to conform the image.

The contribution of this paper is twofold: we propose a grid scramble filter to reduce the effect of screen luminescence, and we test the hypothesis that using pupillary response data improves the classification of easy (or hard) difficulty levels.

The rest of this paper is organized as follows: the Methods section describes the experimental setup, including materials, participants, metrics, and procedure; the Results section discusses the results of each experiment; and finally, the Conclusions and Further Work section concludes this paper.

Methods

The goal of this study is to analyze the pupillary response and gamer data for different difficulty levels in a math EVG to evaluate the significant differences in perceived difficulty for participants with intermediate math skills. Selected relevant features are used to classify difficulty.

Materials

An eye-tracking device, the “EyeTribe” model ET1000 with 60 Hz sampling frequency, was used in a screen (24” extended monitor) with a resolution of 1440 \( \times \) 960 pixels, and both were connected to a laptop.

The eye tracker was located 50-60 cm from the participant’s face. A calibration was done before each test/play session by using the EyeTribe software development kit (twelve points). To remove atypical values, a Hampel filter was used in the preprocessing stage.

To avoid pupil dilation caused by sunlight, the windows in the testing room were covered with blackout curtains, which have a high light-blocking effect. We used the same brightness and settings of the screen throughout. In addition, no sounds and visitors were allowed in the experimentation area.

The educational Refraction video game [42,43] was used in the experiments, as shown in Figure 2. For research, “Refraction” is of particular interest because it is open-access, it provides a natural context for students to create fractions through splitting, and the log data for the game allows the use of learning analytics methods to examine the splitting process in detail [43,44]. Moreover, the design of the game allows us to modify mathematical and game difficulty semi-independently [42].

This game focuses on teaching fractions and discovering optimal learning pathways for math learning. It let gamers bend, split, and redirect lasers to power spaceships filled with lost animals. The general integrated instruction is “Help free as many animals as you can by expanding your knowledge of fractions.” As shown in Figure 2, game elements in Refraction are origins, which generate laser beams; targets, which receive the laser beams and contain spaceships with lost animals waiting to be released; pipe bends that change the laser direction; 2- or 3-way splitters that split the laser into two or three equal parts (eg, the operation of a 3-way splitter over half of a laser is \( \frac{1}{2} \div 3 = \frac{1}{6} \)); and obstacles that prevent the passage of any laser beams.
Figure 2. The Refraction EVG developed by the research group of the Center for Game Science [42,43]. The game mechanic is to use the pieces on the right to split lasers into fractional pieces and redirect them to the target spaceships.

Four levels of the Refraction game were selected for experiments and organized into two worlds: world A (levels $L_{1a}$ and $L_{2a}$), and world B (levels $L_{3b}$ and $L_{4b}$). As shown in Table 1, levels that almost have the same number of game elements were grouped into the same world (ie, $L_{1a}$ and $L_{2a}$ have about the same difficulty level).

Table 1. Number of game elements in the selected levels.

<table>
<thead>
<tr>
<th>Element</th>
<th>World A</th>
<th>World B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$L_{1a}$</td>
<td>$L_{2a}$</td>
</tr>
<tr>
<td>Origins</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Targets</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Two-way splitter (orange)</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Three-way splitter (orange)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pipe bends (blue)</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Obstacles</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total elements</td>
<td>18</td>
<td>19</td>
</tr>
</tbody>
</table>

Experiment 1

The objective of this experiment was to select the best baseline image (ie, a baseline image without semantic information that results in a smaller pupil-size change after the transition from the baseline image to the in-test image). Instances of tested baseline images are shown in Figure 3; they included the widely used white, black, and scramble backgrounds, but also grid scramble images of different sizes: 8×6, 10×10, and 20×20.
Figure 3. Baseline images tested. (Left) Baseline images can be uniform such as (a) black and (b) white, or can depend on the initial image like (c) scramble, (d) 8x6 grid scramble, (e) 10x10 grid scramble, and (f) 20x20 grid scramble. (Right) The in-test image.

Participants
All participants were asked about their general health and were excluded if they wore contact lenses or glasses with more than one power, had eye surgery or abnormalities (e.g., lazy eye, strabismus, nystagmus), or used medication or drugs. All participants were Hispanic and brown-eyed. Participants were not asked for personal information to preserve anonymity. A total of 14 volunteers (4 female, 10 male) between 16 and 37 years old (mean 21.81, SD 7.2) participated in this experiment.

Procedure
As illustrated in Figure 4, participants observed a randomly selected baseline image (an image from Figure 3) for 8 seconds (pupillary response data collected in the last 2 seconds are used as the baseline interval), and then they observed the in-test image for 8 seconds (pupillary data from the last 2 seconds are used as the testing interval). The MPDC is used to select the best baseline image (the MPDC definition is shown in Table 2). This procedure was repeated until all the baseline images were shown to participants.

Figure 4. The procedure used to generate pupillary response data for evaluating baseline images. First, the baseline image was shown on the screen for 8 seconds, and then the in-test image was shown.
Table 2. Pupillary and gamer features studied in this experiment.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>Total errors (TE) is the number of events performed in the wrong way (eg, the laser beam value does not match with the input value) on a level.</td>
</tr>
<tr>
<td>TC</td>
<td>Time to complete a stage (TC) is the time required to complete a given level.</td>
</tr>
<tr>
<td>CP</td>
<td>Number of changes of position (CP). A change of position is defined as the movement of a game element once it has been introduced in the gameplay—the area where the video game elements are dragged and dropped.</td>
</tr>
<tr>
<td>A</td>
<td>Attempts (A) is the number of attempts used by the gamer to complete a given level.</td>
</tr>
<tr>
<td>MPDC</td>
<td>The mean pupil diameter change is obtained by averaging the relevant data points in the measurement interval (time of the stage) and subtracting the mean diameter obtained in the baseline period [24-26].</td>
</tr>
<tr>
<td>PD</td>
<td>Peak dilation (PD) is defined as the maximal dilation obtained in the measurement interval time of the level [13]. First, mean baseline is established, then the single maximum value from the set of data points in the measurement interval time of level is selected.</td>
</tr>
<tr>
<td>LP</td>
<td>Latency to peak (LP) reflects the amount of time elapsed between the beginning of the measurement interval and emergence of peak dilation [13].</td>
</tr>
<tr>
<td>APCPS</td>
<td>Percentage change in pupil size (PCPS) is calculated as the difference between the measured pupil size and a baseline pupil size divided by the baseline pupil size [22,31,45]. The average PCPS (APCPS) is the average of PCPS in the measurement interval time of the selected level.</td>
</tr>
</tbody>
</table>

Statistical Analysis

After Mauchly’s test of sphericity, repeated-measures analysis of variance was performed on the normally distributed variables among MPDC values to explore the difference between the black, white, scramble, scramble 8x6, scramble 10x10, and scramble 20x20 baseline images. The Bonferroni test was used to make post hoc pairwise comparisons.

Experiment 2

The objective of this experiment was twofold: to evaluate which features are more related to the difficulty level, and to test the classification accuracy obtained by using different subsets of features. Studied features (both pupillary and gamer) of the video game levels (L1_a, L2_a, L3_b, and L4_b) are defined in Table 2.

Participants

A total of 20 volunteers (9 female, 11 male) between 23 and 31 years old (mean 27.16, SD 2.6) participated in experiment 2.

Procedure

As in the first experiment, we did not include volunteers with some characteristics that would make pupil-size estimation difficult. None of the subjects who participated in experiment 2 also participated in experiment 1.

As shown in Figure 5, the procedure consists of four phases: (1) participants observed the baseline image of world A for 8 seconds; (2) participants played the world A levels (L1_a and L2_a) without time restrictions; (3) participants observed the baseline image of world B for 8 seconds; and finally, (4) they played the world B levels (L3_b and L4_b) without time restrictions. The pupil baseline was estimated from the data of the last 2 seconds before playing a new world. Pupil size and gamer behavior data were collected along with each play session.
After obtaining features, all information was integrated into a data set \( \tau = \{(X_i, Y_i), i = 1,...,n\} \), where \( X_i \) corresponds to the uniform-length vector containing features \( X_i = (TE_i, TC_i, CP_i, A_i, MPDC_i, PD_i, LP_i, APCPS_i) \) and \( Y_i \) corresponds to the label associated to each level difficulty of the world A and world B. Each register of this data set is generated from a player and a single level. The following sets were defined: \( G = \{TE, TC, CP, A\} \), which includes all game behavior data features, and \( S = \{MPDC, PD, LP, APCPS\} \), which includes all pupillary features. Let \( G' \subseteq G \) and \( S' \subseteq S \) be the sets of features with a significant difference between worlds A and B.

From the 20 participants, 3 (15\%) were randomly selected, and their registers in \( \tau \) were used to train a random forest classifier [46] using different sets of features. Random forest classifier was selected because it is an ensemble meta-algorithm that improves accuracy and avoids overfitting by training on different random samples of the data. Registers in \( \tau \) associated with the rest of the participants were then used as the testing set.

### Statistical Analysis

Features were tested for normality; in this case, the Shapiro-Wilk test was used (because of the low size of the sample). Results show that the variables are not normally distributed. Then, the Wilcoxon signed rank test was used to detect significant differences in variables. Differences between values were considered significant when \( P<.05 \).

### Results

#### Experiment 1

Mauchly’s test of sphericity indicated that the assumption of sphericity had been violated (\( \chi^2_{14}=0.05; P<.01 \)); therefore, a Greenhouse-Geisser correction was used (\( \varepsilon=0.47 \)). The results show that there was a significant difference between MPDC estimated from different baseline images (\( F_{5,78}=30.965; P<.001 \)).

Table 3 shows the descriptive statistics for MPDC calculated for each baseline image. As expected, the 20×20 scrambled filter has the lowest average MPDC (0.32 pixels) as it more closely resembles the original image. Post hoc analyses using the Bonferroni post hoc criterion for significance indicated that there were no MPDC differences for different grid sizes, but there were significant MPDC differences between the group of images generated by the grid scrambled filter, and the group of conventional images used to estimate the baseline (white, black, and scrambled). We choose the 8×6 grid scramble operation for generating baseline images in experiment 2 because there are no differences in MPDC between grid scramble images, and it better obscures the meaning of the in-test image.

<table>
<thead>
<tr>
<th>Baseline image</th>
<th>MPDC(^a) (pixels), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White(^1)</td>
<td>3.356 (2.122)</td>
</tr>
<tr>
<td>Black(^2)</td>
<td>-1.754 (1.452)</td>
</tr>
<tr>
<td>Scramble(^3)</td>
<td>1.620 (0.746)</td>
</tr>
<tr>
<td>Grid scramble 8×6(^4)</td>
<td>0.471 (0.891)</td>
</tr>
<tr>
<td>Grid scramble 10×10(^4)</td>
<td>0.455 (1.392)</td>
</tr>
<tr>
<td>Grid scramble 20×20(^4)</td>
<td>0.320 (0.856)</td>
</tr>
</tbody>
</table>

\(^a\)MPDC: mean pupil diameter change.

#### Experiment 2

We did not find any feature with significant differences in measurements between levels of the same world, neither in the levels of world A (L\(_1\)\(_a\), L\(_2\)\(_a\)) nor in the levels of world B (L\(_3\)\(_b\), L\(_4\)\(_b\)). However, significant differences between worlds were found for the following features: \( TE \) between world A (median 0.00) and world B (median 2.90) (\( z=−2.159; P=0.03 \)); \( TC \) between world A (median 43.486) and world B (median 83,970) (\( z=−3.198; P=0.001 \)); MPDC between world A (median 2.25) and world B (median 2.90) (\( z=−2.159; P=0.03 \)); and PD between world A (median 5.1) and world B (median 18) (\( z=−3.587; P<.001 \)). Table 4 summarizes the statistics for pupillary and gamer features and the Wilcoxon signed rank results.

On the other hand, Table 5 summarizes the accuracy of the random tree classifier. As can be seen, the PD feature alone gives an accuracy of 62.5\%. The best accuracy was obtained by using the \( G' \cup P' \) features, with an accuracy of 87.5\%.
Table 4. Median values for pupillary and gamer measurements, and the Wilcoxon signed rank results.

<table>
<thead>
<tr>
<th>Feature</th>
<th>World A, median</th>
<th>World B, median</th>
<th>z</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TE</td>
<td>0.00</td>
<td>2.5</td>
<td>−2.900</td>
<td>.004</td>
</tr>
<tr>
<td>TC</td>
<td>43.486</td>
<td>83.970</td>
<td>−3.198</td>
<td>.001</td>
</tr>
<tr>
<td>CP</td>
<td>0.00</td>
<td>1.00</td>
<td>−0.382</td>
<td>.70</td>
</tr>
<tr>
<td>A</td>
<td>0.50</td>
<td>1.00</td>
<td>−0.282</td>
<td>.78</td>
</tr>
<tr>
<td>MPDC</td>
<td>2.25</td>
<td>2.90</td>
<td>−2.159</td>
<td>.03</td>
</tr>
<tr>
<td>PD</td>
<td>5.10</td>
<td>18.00</td>
<td>−3.587</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>LP</td>
<td>40.50</td>
<td>51.50</td>
<td>−0.973</td>
<td>.33</td>
</tr>
<tr>
<td>APCPS</td>
<td>0.135</td>
<td>0.136</td>
<td>−0.926</td>
<td>.36</td>
</tr>
</tbody>
</table>

Table 5. Results for a random forest classifier using different sets of features.

<table>
<thead>
<tr>
<th>Set</th>
<th>Features</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>TE, TC, CP, A</td>
<td>75.0</td>
</tr>
<tr>
<td>G'</td>
<td>TE, TC</td>
<td>75.0</td>
</tr>
<tr>
<td>P</td>
<td>MPDC, PD, LP, APCPS</td>
<td>50.0</td>
</tr>
<tr>
<td>P'</td>
<td>PD</td>
<td>62.5</td>
</tr>
<tr>
<td>G' ∪ P'</td>
<td>TE, TC, PD</td>
<td>87.5</td>
</tr>
</tbody>
</table>

Discussion

Experiment 1

Pupil-size changes at the beginning of the EVG (when going from the baseline image to the in-test image) can cause the participant's pupil to expand. A change caused by the screen luminescence would hide the change caused by the cognitive load produced by the reasoning task. This change was analyzed using the MPDC in experiment 1; it was found that baseline images with uniform colors (white and black) result in larger changes in pupil size (Table 3). The sign values of the MPDC are aligned with the optics of the human eye, as it is posited that pupil size increases when the intensity of environmental light decreases (in the case of black or white images); these changes occur even if baseline images resembles the general illumination conditions of the testing scenario such as the scrambled operation.

One could expect that a grayscale image, with the same average intensities as the in-test images, gives a good baseline estimator. Results of experiment 1 show that the conventional scrambled image (which has about the same intensities) just gives a rough estimation of the baseline. Alternatively, the proposed grid scrambled operation better estimates the baseline in comparison to the conventional scramble image. A possible explanation is that retinal ganglion cells (the output neurons of the retina) adapt to both image contrast (the range of image intensities) and to spatial correlations within the scene, even at constant mean intensity [47]. Hence, predicting the pupil size of an individual in different image scenes is challenging. John et al [48] propose a calibration protocol where the participant sees uniform slides of varying grayscale intensities in the range 0-255. We state that a better model could be found by using local and global information from the images.

Experiment 2

Many studies have shown that splitting objects is a promising way to teach fractions [43,49]. In any context, splitting items into halves is much more common than dividing into thirds; this could explain why the students prefer halving and struggle with creating thirds [43]. The Refraction game uses the process of splitting to teach fractions. As shown in Table 1, levels of world A (easy) have fewer 3-way splitters than levels of world B (hard). This means that participants must solve more operations that involve thirds in world B. The difficulty of the Refraction game not only depends on the mathematical operations but on the spatial difficulty. The spatial difficulty is directly correlated to the number of sources and targets; the number of source/target elements is smaller in the world A than in the world B. Results also evidence this change of difficulty, as we observed statistical differences in features G'—including TE and TC.

A random tree classifier that only uses the best game features, G’, only gives an accuracy of 75.0%. This accuracy was improved to 87.5% by using the peak dilation. The maximal dilation obtained in the measurement interval is a natural feature of many factors that dilate the pupil, including the cognitive load.

Pupillary features can be classified into subtractive (those that eliminate individual differences by subtracting the baseline value from the measurement interval, such as MPDC, PD, and LP) and divisive (those that calculate a ratio of a measurement value to baseline, such as APCPS). Subtractive features can be categorized into size-related, such as MPDC and PD, or time-related, such as LP. Results show that the subtractive size-related features, MPDC and PD, better describe the difficulty level.
Hunicke [50] states that difficulty adjustments must be implemented in a way such that users do not perceive difficulty changes. However, gamer data are recorded after human perception of difficulty; that is, a control that uses gamer data collected after the player finished each level could not completely fulfill the requirement of being imperceptible.

The proposed approach improves the accuracy of classification of the perceived difficulty to 87.5%, in contrast to 62% with only pupillometry. These results are aligned to other studies that suggest the relationship between pupil change and the level of a game; for instance, by using the Akaike Information Criterion, Strauch et al [51] propose that the pupil change is a quadratic function of the levels of Pong.

Video game difficulty adjustment is game data–dependent (ie, different games require different features). We argue that a generic framework for dynamic difficulty adjustment could be designed by fusing generic game features (such as score, elapsed time, etc) with the information provided by pupillometry. In this way, we can take advantage of ocular data as a general, noninvasive, near real-time option to sense the user perception of difficulty.

In a traditional pupillometry experiment, the researcher maintains tight control over luminance while manipulating a specific cognitive variable. Reilly et al [52] conducted the reverse approach (ie, holding cognitive task demands constant while manipulating luminance). We believe that the reverse approach must be used to obtain a model of the participants’ pupil size in the initial calibration stage by using the grid scrambled images, and then a subtractive approach should be used during the gameplay stage.

Conclusions and Further Work
This paper proposes a grid scramble filter to obtain a baseline image that reduces the effect of the screen light reflex on a participant’s pupil size. This filter simulates both the local and the mean luminance of a given image. To hide the meaning of an image, the 8x6 grid scramble filter can be used for tests that reasonably keep the same background in each interval. We consider that a more general baseline can be obtained by modeling luminescence factors that affect pupil size. Such a model could be used to estimate cognitive factors that affect the pupils in any setting (eg, a commercial video game).

Gamer data are a valuable resource for estimating the difficulty of EVGs, but adding cognitive load data measured by pupillary response data improves the accuracy of classifying the difficulty of game levels.

Using the human perception features from ocular data such as blinks, eye-fixations, and eye-saccade to measure the cognitive load may improve the classification accuracy of difficulty levels and gather imperceptible changes that gamer data can omit [53,54].

A key issue with approaches that estimate a baseline, like the proposed one, is that indoor light conditions and monitor brightness must be the same during the game time. Playing a game in specific conditions is restrictive; to address this, we are working on a model that relates luminescence to different screen configurations (instead of a baseline) This approach can be used in virtual reality headsets. The proposed approach can be included in a more elaborated calibration stage that tests different models of pupil change due to luminance, as in a previous study by Lara-Alvarez and Gonzalez-Herrera [55].

Acknowledgments
We thank the support given through the FORDECYT 296737 project "Consortio en Inteligencia Artificial" for the publication of this work.

Conflicts of Interest
None declared.

References


17. Abokyi S, Owusu-Mensah J, Osei KA. Caffeine intake is associated with pupil dilation and enhanced accommodation. Eye (Lond) 2017 Apr;31(4):615-619 [FREE Full text] [doi: 10.1038/eye.2016.288] [Medline: 27983733]


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>attempts</td>
</tr>
<tr>
<td>APCS</td>
<td>average percentage change in pupil size</td>
</tr>
<tr>
<td>CP</td>
<td>number of changes of position</td>
</tr>
<tr>
<td>EVG</td>
<td>educational video game</td>
</tr>
<tr>
<td>LP</td>
<td>latency to peak</td>
</tr>
<tr>
<td>MPDC</td>
<td>mean pupil diameter change</td>
</tr>
<tr>
<td>PD</td>
<td>peak dilation</td>
</tr>
<tr>
<td>TC</td>
<td>time to complete a stage</td>
</tr>
</tbody>
</table>
**TE:** total errors  
**TEPR:** task-evoked pupillary response

---

© Hugo Mitre-Hernandez, Roberto Covarrubias Carrillo, Carlos Lara-Alvarez. Originally published in JMIR Serious Games (http://games.jmir.org), 11.01.2021. 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 work on http://games.jmir.org, as well as this copyright and license information must be included.
Managing Game-Related Conflict With Parents of Young Adults With Internet Gaming Disorder: Development and Feasibility Study of a Virtual Reality App

Yu-Bin Shin1, PhD; Jae-Jin Kim1,2, PhD, MD; Hyunji Kim1, MA; Soo-Jeong Kim1, MD; Hyojung Eom1, MA; Young Hoon Jung1, MA; Eunjoo Kim1,2, PhD, MD

1Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
2Department of Psychiatry, Gangnam Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea

Abstract

Background: Individuals with internet gaming disorder (IGD) report facing family conflicts repeatedly because of their excessive internet gaming. With recent advancements in virtual reality (VR) technology, VR therapy has emerged as a promising method for the management of various psychiatric disorders, including IGD. Given that several risk and protective factors for young people with addiction can be influenced by their interpersonal context, the potential utility of VR-based apps for managing family conflicts needs to be examined with reference to IGD management. However, few studies have evaluated potential treatment modules related to interpersonal conflict management, such as emotion regulation and taking the perspective of others.

Objective: This preliminary study aims to examine the potential use of a VR-based app in the management of game-related conflicts with parents of young adults with IGD and matched controls.

Methods: In total, 50 young male adults (24 with IGD and 26 controls) were recruited to participate in the study. We developed a virtual room where game-related family conflicts arise. Using this room, participants completed 2 VR tasks that required them to express anger and then implement coping skills (ie, risk/benefit assessment of stopping a game and taking parents’ perspective) to deal with negative emotions in interpersonal conflict situations and to decrease one’s gaming behavior.

Results: The results showed that immersion in our VR app tended to provoke negative emotions in individuals with IGD. In addition, after a risk/benefit assessment of stopping a game, the response of stopping a game immediately increased significantly in the IGD group, suggesting that patients’ gaming behavior could be changed using our VR program. Furthermore, in individuals with IGD, longer gaming hours were associated with a lower level of perceived usefulness of the coping skills training.

Conclusions: The findings of this study indicate that our VR app may be useful for implementing more desirable behaviors and managing gaming-related family conflicts in individuals with IGD. Our VR app may offer an alternative for individuals with IGD to learn how a vicious cycle of conflicts is developed and to easily and safely assess their dysfunctional thoughts behind the conflicts (ie, perceived unreasonable risks of stopping a game and thoughts acting as a barrier to taking the perspective of others).

(JMIR Serious Games 2021;9(1):e22494) doi:10.2196/22494

KEYWORDS
internet gaming disorder; family conflict; coping behavior; virtual reality
**Introduction**

**Background**

Internet gaming disorder (IGD), defined as the persistent and recurrent use of the internet to engage in games, can lead to significant psychological distress and impairment in daily social functioning [1]. Given that social, behavioral, and neurobiological development continues until young adulthood, addiction in these periods could have a more striking effect on young populations [2,3]. Therefore, intervention and prevention targeting groups at risk for developing gaming-related problems may reduce the burden of the disease and provide lifelong benefits.

With global efforts to inform effective practice, appropriate therapeutic recommendations for IGD are being made [4]. The most commonly reported approach is a combination of various interventions, including cognitive behavioral therapy (CBT) [5,6], motivational enhancement therapy [7], and pharmacotherapy [4,8]. In addition, with recent advancements in virtual reality (VR) technology, the use of VR coupled with CBT has become a viable treatment option for IGD [9]. However, current VR treatments for patients with IGD have been developed in the context of substance abuse treatment, with a focus on aversive conditioning [9,10] and cue exposure [11]. Thus, few studies have evaluated other potential treatment modules, such as emotion regulation and interpersonal conflict management, which are critical for the management of IGD symptoms. Therefore, it is essential to build an evidence base for other specialized modules for individuals with IGD, which can be implemented using VR technology.

In some East Asian countries, as parental influences may continue even into young adulthood [12,13], young adults as well as adolescents are likely to experience game-related family conflicts. Youth with IGD could experience more difficulties related to handling such conflicts because they often exhibit avoidance coping styles, characterized by the denial of problems or their impact on them [14-16]. Thus, when confronted by their parents who ban gaming, they would deny and refuse to discuss their gaming problems with their parents, which can, in turn, contribute to a harmful cycle of problematic gaming [17]. Therefore, there is a need for specific therapeutic recommendations that not only take into account a repetitive pattern of conflict escalation in the family context but also seek to improve problem-solving skills to resolve these conflicts.

Interpersonal conflict management is one of the clinical domains of psychiatric treatment in which VR-based CBT has been applied increasingly [18,19]. In this context, although some researchers have developed anger-provoking VR [20], an intervention for anger management is still scarce. Studies on VR-based CBT have demonstrated the usefulness of VR in learning social skills via role-playing scenarios [21,22]. Such studies introduced exercises for challenging social contexts (eg, a virtual school) in a stepwise manner, thereby requiring patients to master a series of social skills (eg, starting a conversation and recognizing peers’ emotions) within a VR program [23]. Interactive VR, defined as a dyadic interaction that allows feedback from an avatar, is also known to benefit VR-based CBT [24]. Combining interactive VR with role-playing interpersonal scenarios can help individuals with IGD learn coping skills to handle family conflicts.

The VR scenario we chose for training coping skills in family conflict management is based on the risk/benefit assessment of addictive behavior. From the perspective of risk/benefit analysis, it can be assumed that if the perceived benefits of addictive behavior are weighted more heavily than are perceived risks, then people will continue to engage in such behavior. In accordance with this conjecture, previous studies demonstrated that the tendency to weigh risks more than benefits was predictive of risky and gambling behavior [25]. Moreover, recently, a behavioral intervention for cigarette smokers focusing on a personalized analysis of the risk/benefit of quitting (eg, the risks are the increase in negative affect and reduced ability to concentrate, and the benefits are better health, well-being, and finances) was found to be effective in improving smoking cessation outcomes [26]. Most of the relevant studies seem to focus primarily on perception of the general aspects of risks and benefits of addictive behavior, suggesting that this kind of intervention to treat addictive behavior is feasible in clinical settings [27]. As it is more helpful for adolescents/young adults to set a goal that is shortly and proximally framed, focusing on the risks and benefits of harmful behavior within a short time frame could be more effective. As such, our VR content was implemented to create a perception of stopping gaming immediately, focusing on risk-benefit analysis in a short-term framework.

The second scenario for coping skills training dealt with how individuals cope with negative emotions in interpersonal conflict situations. As disputes on gaming time in the family of individuals with IGD could lead to severe conflict or even violence among family members, emotion regulation in such situations seems to be particularly important for individuals with IGD. Previous research has suggested that one type of emotion regulation involves understanding the current situation from the perspective of others [28]. Thus, in this study, conflict situations escalating into violence were reproduced, and participants were asked to apply taking the perspective of their virtual parents.

**Objectives**

Given this background, the objective of this study is to examine the potential of a VR-based app designed to help young adults with IGD to manage game-related conflicts with their parents in a sample of individuals with IGD and matched controls. A virtual room, presumed to be the participants’ own, was developed, in which family conflicts were simulated. To determine whether our VR contents were capable of eliciting negative emotions (ie, anger), the participants engaged in a family conflict situation and were then encouraged to express their anger to their virtual parents who banned gaming. As a part of problem solving–focused training, we implemented the following 2 coping skills: evaluating the perceived benefits and risks (or disadvantages) of stopping the game immediately and taking the perspective of parents on their gaming behavior. As preliminary data on the potential role of VR as a promising approach in the management of IGD, we investigated whether...
patients with IGD expressed greater anger compared with matched controls and whether their response could be changed after experiencing the coping skill training within our VR program. Finally, we tested whether behaviors in response to our VR program were associated with the severity of IGD and other relevant clinical measures (eg, gaming frequency and motivation for changing gaming behavior).

**Methods**

**Sample Size (Power)**

Power analysis for an independent *t* test was conducted using the G*Power [29] program to determine the necessary sample size with an *α* of .05, a power of 0.80, a large effect size (f=0.8), and two tails. The results suggested a desired sample size of 26 participants in each group, totaling 52 subjects, to detect the group difference in our dependent measures.

**Recruitment**

Young Asian male adults (24 with IGD and 26 controls) were recruited from the local community using web-based advertisements. As men are known to have a higher prevalence of IGD than women and to avoid gender-specific confounding factors related to emotion expression and coping skills affecting the results, only male participants were recruited [30-32]. All participants were interviewed by a clinical psychologist for IGD diagnosis according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* [1]. All enrolled participants with IGD were at the time addicted to at least one of the popular internet games in South Korea (eg, Battle Grounds, League of Legends, and Overwatch). For the controls, reviews of history were performed to confirm whether the participants had ever been diagnosed with IGD. In addition, the controls were screened for a family history of psychiatric disorders; no one had a history of any psychiatric illness in first-degree relatives. The exclusion criteria for all participants were current use of psychotropic medication and any history of substance use disorder, neurological or neurodevelopmental disorder, major depressive disorder, bipolar 1 disorder, and psychotic disorder, as determined using the Mini-International Neuropsychiatric Interview [33]. However, 2 participants with depression were included in the IGD group. Ethical approval was obtained from the institutional review board of Yonsei University Gangnam Severance Hospital. Informed consent was obtained from all participants.

**Measures**

Estimated full intellectual functioning (IQ) was obtained using the short form of the Wechsler Adult Intelligence Scale-Revised [34]. Internet gaming behavior was measured in terms of frequency of use (eg, hours of use per week). Participants were asked to rate their current gaming craving on a visual analog scale (VAS), which represents the severity of craving on a 100-mm horizontal line, ranging from no craving at all (left side) to extreme craving (right side). The result is a score on a continuous scale, ranging from 0 to 100.

Readiness to change internet gaming behavior was assessed using a modified version of the readiness to change questionnaire [35]. The modified questionnaire contained 12 items on a 5-point scale, ranging from *strongly disagree* (−2) to *strongly agree* (2), with higher scores representing a higher readiness to change. The total score serves as an index of the motivation to stop playing internet games.

The presence questionnaire (PQ) [36] and the simulator sickness questionnaire (SSQ) [37] were used to measure the users’ experience with our system. PQ is a measure of users’ presence in VR, which is a psychological state of *being there*. Presence can enhance users’ active engagement in content, involving their senses and capturing their attention. SSQ assesses simulator sickness resulting from the discrepancy between simulated visual motion and the sense of movement coming from the vestibular system. Simulator sickness is negatively correlated with users’ enjoyment of VR programs.

**Procedure**

After each participant was informed of the purpose of the study and they consented to participate, the IQ assessment was conducted, and data on demographic and clinical characteristics were collected. Subsequently, the participants began the VR session. This study was conducted for approximately 20 to 30 min, with each of the 2 scenarios taking about 10 to 15 min. Their time to completion differed depending on how quickly each participant responded to the tasks in each scenario and which responses were chosen.

As shown in Figure 1, all participants began with the anger expression task, but tasks with the virtual mother or father were randomized and counterbalanced among participants to control for order effects.
**Figure 1.** Experimental procedure. The order of tasks on conflicts with the virtual mother and father was counterbalanced. VR: virtual reality.

**Virtual Reality**

We developed our own VR scenarios. Our VR scenarios emphasized the development of motivation for change based on the implementation of CBT principles by enabling the participants to access dysfunctional thoughts behind the conflicts (i.e., perceived unreasonable risks of stopping a game and thoughts acting as a barrier to perspective-taking) easily.

Participants were guided through the following 2 sets of VR tasks in a virtual room, presumed to be the participants’ own room: (1) an anger expression task and (2) a coping skill training task (see Figure 2 for screenshots of the virtual environment). Figure 3 presents a flowchart of the scenario. At the beginning of all tasks, participants found themselves seated in front of a monitor displaying internet game icons (e.g., StarCraft, Overwatch, League of Legends, and FIFA Online: Figure 2). The participant watched a video clip of their own choice played on the monitor in this VR program. The scenario continued depending on the task content. Every task ended by rating the usefulness of the strategy in resolving the conflict on a VAS built within the VR program (Figure 2).

**Figure 2.** Screenshots of the virtual environment. VR: virtual reality.

a. Desktop setting  
b. User interface for anger expression  
c. Score presentation  
d. Conflict scene with the virtual mother  
e. Conflict scene with the virtual father  
f. Visual analogue scale built in VR
Anger Expression Task

In the anger expression task, virtual parents would enter the room and begin compelling the participants to stop the game immediately. Next, the participants were offered a choice of whether to continue or stop the game. Regardless of which behavior they selected, the scenario would continue with another round of nagging from the parents to stop the game. Participants were then instructed to express their perceived anger about the 2 instances in which they were being forced to end the game. They were offered only a choice-based response format to express their initial anger (ie, suppression: ignore in silence or expression: express anger/irritation; Figure 2). If they selected suppression, the scenario continued to the next scene after 5 seconds of silence. If they selected expression, corresponding audio clips of a male actor’s voice were played. Following the completion of the initial expression, parents’ feedback was delivered, the content of which matched the participants’ anger expression response (see Multimedia Appendix 1 for the scripts of the clips). When expressing anger for the second time, participants were allowed to choose the anger expression style and were allowed to express their anger verbally to the parents’ avatars. At the end of each expression, the participants rated the extent of their anger experience on a VAS built within the VR program (0=not at all, 100=extremely high; Figure 2).

Coping Skill Training Task: Risk/Benefit Assessment and Perspective-Taking Task

In the coping skill training task, the virtual parents returned to the room, asking the participants why they had not stopped the game yet and attempting to limit their gaming time again. Participants were again offered a choice to either continue or stop the game immediately. Subsequently, they were instructed to employ the 2 coping skills of risk/benefit assessment and perspective-taking. Specifically, they were asked to provide an example of benefits and disadvantages of stopping the game immediately (ie, “What are the benefits/disadvantages of stopping the game immediately? Please describe an example”) and to assess the extent of the effect of each example on them (ie, “How much effect will the example you described have on you?”) using a VAS built within the VR program. They were asked to report as many examples as possible. These items were listed one at a time until they wished to move to the next question. After viewing the mean scores of their effect assessment (Figure 2), they were finally offered the choice of whether to continue or stop the game immediately. As before, regardless of their choice, the scenario continued with the parent’s pushing the participant to stop the game again. However, this time, the participants experienced the most negative social interaction in the VR situation. For example, the virtual mother turned off the game by pulling the computer’s power plug (Figure 2) or the virtual father used physical punishment by spanking the participant (Figure 2).
For the second coping skill, the participants were asked to talk about the parents’ behavior from their own perspective, followed by talking about it from the parents’ perspective. Subsequently, they rated the extent to which they were able to understand their parents’ perspective using a VAS built within the VR program.

The VR system consisted of a desktop computer (with the Microsoft Windows 10 operating system, an NVIDIA GeForce GTX 970 graphics card, and a 16 GB RAM graphics memory) and an Oculus Rift head-mounted display (Oculus VR: HD resolution of 1080×1200 per eye with a 51.6 diagonal field of view, a 3 degrees of freedom tracker for head rotation, and built-in headphones). The Oculus Touch controller was used for interactions with executable objects and avatars during the VR experience. The microphone built into the VR headset gathered verbal data of users’ self-speech in real time. Three-dimensional virtual environments included a virtual house and appliances using Autodesk 3D max and were integrated with Unity software. User data such as selected responses and selection time, head movement, hand gestures, speech contents, and speaking time were recorded and stored in the main server computer. This system allowed the therapist to track patients’ performance and analyze behavioral information.

### Statistical Analysis

We tested differences between groups on demographic and clinical variables using independent t tests. In addition, we conducted repeated measures analysis of variance (ANOVA) to test whether the 2 groups differed in the variables of interest in each VR scenario. To further examine the significant main effects and interactions, we conducted post hoc analyses, with Bonferroni correction at $P<.05$. Finally, Pearson correlation coefficients were used to investigate the relationships between the behavioral responses of the virtual environment and clinical variables. Statistical differences were considered significant at $P<.05$. All analyses were conducted using SPSS 23.0 (IBM Corporation).

### Results

#### Comparison of Clinical Symptom Measures Between IGD and Control Groups

Table 1 shows the results of a two-tailed independent t test between the IGD and control groups in demographic and clinical variables. There were no differences in age, education, and IQ between groups. As expected, the IGD group had higher scores on gaming craving, IGD symptoms, and internet gaming hours per week.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>IGD (n=24), mean (SD)</th>
<th>Healthy controls (n=26), mean (SD)</th>
<th><em>t</em> test (df)</th>
<th><em>P</em> value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>21.78 (2.33)</td>
<td>21.65 (1.83)</td>
<td>−0.22 (48)</td>
<td>.83</td>
</tr>
<tr>
<td>Education (years)</td>
<td>14.12 (1.19)</td>
<td>14.61 (0.94)</td>
<td>1.62 (48)</td>
<td>.11</td>
</tr>
<tr>
<td>IQ</td>
<td>119.54 (10.11)</td>
<td>122.04 (9.72)</td>
<td>0.92 (48)</td>
<td>.36</td>
</tr>
<tr>
<td>Gaming craving (visual analog scale)</td>
<td>6.87 (1.46)</td>
<td>0.85 (1.05)</td>
<td>0.41 (48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Severity of IGDb</td>
<td>7.29 (1.08)</td>
<td>N/Ac</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Internet gaming hours per week</td>
<td>50.12 (13.27)</td>
<td>1.99 (1.30)</td>
<td>−18.41 (48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Presence Questionnaire</td>
<td>102.29 (16.83)</td>
<td>114.92 (12.93)</td>
<td>2.99 (48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Simulator sickness questionnaire</td>
<td>7.45 (7.68)</td>
<td>4.53 (5.73)</td>
<td>−1.53 (48)</td>
<td>.13</td>
</tr>
</tbody>
</table>

aIGD: internet gaming disorder.
bSeverity of IGD was assessed using the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, criteria for IGD, providing a total score between 0 and 9.
cN/A: not applicable.

For scores on the PQ, there was a significant difference between the IGD and control groups ($t_{48}=2.99$; $P<.001$). As Internet games played among participants with IGD have high-quality graphics and animations, the VR environment and interactions might be perceived as less realistic and natural in the IGD group. Regardless of the group difference, the overall mean score was above the midpoint of the scale, indicating that participants experienced a high level of presence in the respective virtual environment.

The average score on the SSQ corresponded to lower levels of simulator sickness than the established norms for the SSQ [37]. An independent t test reported no significant difference between the 2 groups in their SSQ scores ($t_{48}=-1.530; P=.13$). This shows that both the IGD and control groups had similar VR experiences.

#### Anger Expression Style and Intensity of Anger Experience

Figure 4 shows the results of a three-way ANOVA of the 2 dependent variables in the anger expression task. The intensity of anger experience revealed a significant main effect of group ($F_{1,48}=8.154; P=.006$), target ($F_{1,48}=4.278; P=.04$), and expression order ($F_{1,48}=5.221; P=.03$; Figure 4). However, no main or interaction effects were observed for the anger expression style. These results indicate that the IGD group experienced higher levels of anger than the control group, the virtual father produced significantly higher anger than the virtual
mother, and the second anger expression produced significantly higher anger than the first expression.

**Figure 4.** Results of analyses of 3 dependent variables. (a) Mean ratings for the intensity of the anger experience; (b) mean percentage of response choice for “continue the game”; (c) mean ratings for usefulness of each strategy. Error bars represent SE of the mean. *: group or condition effect, \(P<.05\), **: \(P<.01\), ***: \(P<.001\); ¥: significant two-way interaction, ¥¥: \(P<.01\); ¥¥¥: \(P<.001\).

### a. Main effect of group and expression order

![Intensity of anger experience](image)

**b. Main effect of group and interaction between group and time point**

![Percentage of choice for "continue the game"](image)

**c. Main effect of group and strategy, and interaction between group and strategy**

![Scores on usefulness](image)

---

**Risk/Benefit Assessment of Stopping the Game Immediately**

A 3-way ANOVA on the number of perceived risks (disadvantages) and benefits of stopping the game immediately revealed a significant interaction between group and evaluation type \((F_{1,45}=10.157; P=.003)\), indicating the tendency of the control participants to report significantly more benefits than risks \((t_{25}=-3.277; P=.003)\). Difference scores (the number of benefit items minus the number of risk items) were used in the correlation analyses. Among individuals with IGD, higher
scores, indicating more benefits than risks, were significantly associated with enhanced motivation to change their gaming behavior ($r = 0.488; P = .01$), as measured by the scores on the readiness to change the questionnaire.

A 3-way ANOVA of the quantified scores on the extent of the effect of the risk-benefit assessment revealed a significant main effect of evaluation type ($F_{1,48} = 73.456; P < .001$) and interaction between group and evaluation type ($F_{1,48} = 23.445; P < .001$), indicating that benefits had a stronger effect on participants than did risks, and the control group reported a stronger effect of benefits than risks ($t_{25} = -7.245; P < .001$).

**Response of Stopping the Game After Each Time Point**

A 3-way ANOVA on deciding to stop or continue the game at different time points revealed a significant main effect for group ($F_{1,48} = 73.646; P < .001$), with the IGD group exhibiting a higher tendency to continue the game than the control group, and the main effect for time point ($F_{1,48} = 8.490; P = .001$; Figure 4), indicating a higher tendency to stop the game after the risk/benefit assessment, than that at baseline (mean difference 0.205, SE 0.058; $P = .003$) and after anger expression (mean difference 0.165, SE 0.057; $P = .02$). A significant group and time point interaction was also noted ($F_{1,48} = 3.886; P = .03$), which was caused by the higher tendency of the IGD group to stop the game after the risk/benefit assessment than at baseline ($t_{23} = 3.761; P = .001$) and after anger expression ($t_{23} = 2.933; P = .007$).

**Degree of Understanding the Virtual Parents’ Perspective**

A 2-way ANOVA revealed a significant main effect for group ($F_{1,48} = 17.656; P < .001$) and target ($F_{1,48} = 4.337; P = .04$). In other words, the control participants exhibited a better understanding of their parents’ perspective than those with IGD, and the participants had a better understanding of the virtual mother’s perspective than that of the virtual father.

**Usefulness of the Conflict-Resolution Strategy**

A 3-way ANOVA on ratings for the usefulness of each strategy in resolving conflicts showed a significant main effect for group ($F_{1,48} = 5.328; P = .03$) and strategy ($F_{1,48} = 52.667; P < .001$); the control group evaluated the VR contents as more useful in resolving conflicts (Figure 4), and participants rated the coping skill as more useful in resolving conflicts than anger expression. There was a significant interaction between group and strategy ($F_{1,48} = 9.201; P = .004$), indicating that the control group rated the coping skill as more useful than the IGD group ($t_{48} = 3.697; P = .001$), but there was no significant difference in anger expression ($t_{48} = 0.372; P = .71$).

As presented in Figure 5, in the control group, a higher number of gaming hours was associated with a higher tendency to perceive coping skills as more useful than anger expression ($r = 0.556; P = .003$). In contrast, in the IGD group, fewer gaming hours were associated with a higher tendency to perceive coping skills as more useful than anger expression ($r = -0.468; P = .02$).

**Discussion**

**Principal Findings**

In this study, we investigated the feasibility of a VR app for managing game-related conflicts with parents in young adults with IGD. The results showed that immersion in our VR program could provoke negative emotions (ie, anger) in individuals with IGD by placing them in interpersonal situations that would require them to manage these conflicts. Moreover, the clinical potential of our VR program was evidenced by the results of the (1) change in participants’ gaming behavior within the program and (2) relationship between patients’ usefulness rating for the implemented coping skills in resolving conflicts and the number of hours spent on gaming.

According to parents of adolescents with IGD, when they attempt to enforce time limits on the game, their children...
become angry, irrational, and even violent [12]. Thus, anger and violence seem to be common ways to manage such conflicts in families of individuals with IGD. In this context, our VR contents were developed to reproduce such anger-provoking situations in a hierarchical and interactive manner (eg, implementing staged manipulation to increase anger induction, matching the content of virtual parents’ feedback to participants’ anger expression type to simulate dyadic interactions, and allowing participants to express anger via a choice-based response format and verbal response) to increase the resemblance of our VR scenarios to reality. Data from the anger expression task showed that the intensity of the anger experience was higher in individuals with IGD than in control participants, suggesting that family conflict and its emotional response (ie, anger) can be simulated in VR scenarios. In line with this result, the perceived anger after the second expression was stronger than that after the first expression, indicating that the severity of the conflict increased. This experience may closely reflect the nature of conflict regarding time limits on games in families of youth with IGD. Therefore, the use of hierarchical and interactive scenarios for anger induction in VR apps could aid the closer-to-reality simulation of the progression of game-related arguments between family members.

In this study, the VR scenarios allowed participants to consider the benefits and risks of stopping the game immediately as a coping skill for managing family conflicts. The results showed that control participants reported more benefit items than risks. However, no difference was seen in the number of examples of risks and benefits expressed by participants in the IGD group. In line with this result, control participants reported that perceived benefits had a greater influence on them than perceived risks, whereas individuals with IGD did not assign more weight to the effect of benefits than risks. This perception could contribute to their continuous engagement in addictive behavior, as observed in other addictive disorders [25,27]. Hence, this study demonstrates the potential applicability of VR in treating IGD by helping participants to learn the use of benefit assessment of stopping a game.

Our most important finding is that after engaging in assessing the risk/benefit of stopping gaming immediately during repeated exposure to parents’ attempts to limit game time, the IGD group exhibited a higher tendency to choose to stop gaming. As one of the putative mechanisms of CBT to elicit adaptive behaviors is the acquisition of coping and problem-solving skills [38], the results of this study suggest that VR-CBT using risk/benefit assessment to stop gaming could potentially help individuals with IGD to exhibit more desirable behaviors and manage gaming-related family conflicts more effectively.

Taking another person’s perspective is one of the important processes involved in social cognition [39]. Deficits in social cognition may increase social withdrawal and aggression, which might lead to a vicious cycle of substance use [40]. In addition, consistent with the role of empathy in addiction, a previous study reported an association between empathy and IGD symptoms [41]. Similarly, in this study, although participants had the opportunity to talk about the virtual parents’ feelings or thoughts from their perspective, individuals with IGD exhibited a lower understanding of the perspective of the virtual parents compared with control participants. Thus, it can be inferred that impairment of the perspective-taking ability in individuals with IGD may hinder the resolution of game-related conflicts with significant others in real life, which, in turn, may lead them to continue gaming for excessive periods. Together, the findings of this study suggest that our VR content has the potential to be used as a medium for anger management by offering individuals with IGD opportunities to practice taking the perspective of significant others in real-life scenarios and to discuss their thoughts and feelings about difficulties in perspective-taking with clinicians.

After implementing each strategy to manage conflicts with the virtual parents, the participants were instructed to evaluate the overall usefulness of the conflict-resolution strategy employed. Overall, compared with control participants, individuals with IGD evaluated the 2 coping skills as less useful. Moreover, patients with longer gaming hours perceived coping skills as less useful than anger expression in resolving conflicts. This result suggests that IGD individuals with longer gaming hours may not feel the need to engage in more adaptive coping strategies to decrease their gaming hours. Moreover, the association between the usefulness rating of coping skill training in our VR and actual gaming hours indicates that VR parameters may reflect real-life gaming behavior of IGD youths or having a greater tendency to use adaptive coping skills are correlated with less severe symptoms of IGD measured by the gaming hours. Thus, the VR program may benefit young adults with IGD by aiding the assessment of IGD symptomatology. Moreover, given that adaptive coping acts as a buffer against psychological problems, a prior study suggested that a greater tendency to use coping skills such as risk/benefit assessment and perspective-taking of others could mitigate the development of more severe symptoms of IGD [42].

Limitations

This study has some limitations that warrant future research. The small sample size, sample composition of young Asian male adults, and relatively mild level of IGD severity among our participants limit the generalizability of our findings to the general population. In addition, 2 patients with mild depression were included; the presence of a comorbid condition may be a confounding factor for them. However, the main findings of our study generally remained the same as when excluding participants with comorbidities. Our VR contents still have to be tested in a larger, more heterogeneous sample before its potential role in IGD management can be discussed. Moreover, the present brief and time-limited application of our VR does not suggest that the behavior change achieved could also be applied to real-life situations. Furthermore, behavioral or questionnaire measures were not used to assess the empathic or perspective-taking ability of the participants. In addition, many feedback options were predesigned by the authors, such as the expressions of anger being read out by a male voice in our VR scenario. If participants had been able to verbalize their emotions in their own words, the results of our experiment might have differed. Prior studies have reported that self-speech is helpful for self-regulating cognition and behavior [43]. As computer technologies related to voice recognition and acoustic meaning-based speech analysis have developed, speech has also
become a viable interaction modality in VR environments. Therefore, a scenario with a self-speech component should be designed to explore this issue in future studies.

In addition, the levels of symptom severity or other behavioral measures in the VR program were based only on subjective reports. Objective physiological markers, such as measurement of psychophysiological reactivity or reports from caregivers regarding participants’ behavioral/emotional problems, would provide additional information. Finally, our preliminary study design did not allow for the validation of the ability of our program to reduce IGD symptoms or increase one’s perspective-taking ability. To further validate the effects of this program, a follow-up study should include a control group comprising individuals with IGD who are not exposed to this VR app.

Conclusions

We developed a VR program that simulated gaming-related conflicts between young adults and parents and tested its effectiveness in managing problems between family members in an adaptive manner. Considering that the traditional therapeutic approach is dependent on in vivo (real-life) exposure or imagination capabilities of individuals, our VR program may offer an alternative for individuals with IGD to learn how a chronic cycle of conflicts is developed. In addition, it enabled them to access dysfunctional thoughts behind the conflicts (ie, perceived unreasonable risks of stopping a game and thoughts acting as a barrier to perspective-taking) easily and safely. Given that a number of risk and protective factors for young individuals with addiction could be influenced by the family context, our findings suggest the potential use of VR-based apps in the management of family conflicts experienced by individuals with IGD. A clinical trial using VR combined with CBT may shed more light on the effects of learning in VR over time.

Acknowledgments

This study was supported by the Brain Research Program through the National Research Foundation of Korea, funded by the Ministry of Science, ICT and Future Planning (NRF-2015M3C7A1065053).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Content of participants’ anger expression and virtual parents’ feedback.
[DOCX File, 14 KB - games_v91e22494_app1.docx ]

References


Promoting Physical Activity in Japanese Older Adults Using a Social Pervasive Game: Randomized Controlled Trial

Luciano Henrique De Oliveira Santos\textsuperscript{1,2}, PhD; Kazuya Okamoto\textsuperscript{1,2,3}, PhD; Ryo Otsuki\textsuperscript{1}, MSc; Shusuke Hiragi\textsuperscript{1,2,3}, MD, PhD; Goshiro Yamamoto\textsuperscript{2,3}, PhD; Osamu Sugiyama\textsuperscript{4}, PhD; Tomoki Aoyama\textsuperscript{2}, MD, PhD; Tomohiro Kuroda\textsuperscript{1,2,3}, PhD

\textsuperscript{1}Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto, Japan
\textsuperscript{2}Graduate School of Medicine, Kyoto University, Kyoto, Japan
\textsuperscript{3}Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan
\textsuperscript{4}Department of Real World Data Research and Development, Graduate School of Medicine, Kyoto University, Kyoto, Japan

Corresponding Author:
Luciano Henrique De Oliveira Santos, PhD
Department of Social Informatics
Graduate School of Informatics
Kyoto University
54 Shogoin Kawahara-cho, Sakyo-ku
Kyoto, 606-8507
Japan
Phone: 81 75 366 7701
Email: lhsantos@kuhp.kyoto-u.ac.jp

Abstract

Background: Pervasive games aim to create more fun and engaging experiences by mixing elements from the real world into the game world. Because they intermingle with players’ lives and naturally promote more casual gameplay, they could be a powerful strategy to stimulate physical activity among older adults. However, to use these games more effectively, it is necessary to understand how design elements of the game affect player behavior.

Objective: The aim of this study was to evaluate how the presence of a specific design element, namely social interaction, would affect levels of physical activity.

Methods: Participants were recruited offline and randomly assigned to control and intervention groups in a single-blind design. Over 4 weeks, two variations of the same pervasive game were compared: with social interaction (intervention group) and with no social interaction (control group). In both versions, players had to walk to physical locations and collect virtual cards, but the social interaction version allowed people to collaborate to obtain more cards. Changes in the weekly step counts were used to evaluate the effect on each group, and the number of places visited was used as an indicator of play activity.

Results: A total of 20 participants were recruited (no social interaction group, n=10; social interaction group, n=10); 18 participants remained active until the end of the study (no social interaction group, n=9; social interaction group, n=9). Step counts during the first week were used as the baseline level of physical activity (no social interaction group: mean 46,697.2, SE 7905.4; social interaction group: mean 45,967.3, SE 8260.7). For the subsequent weeks, changes to individual baseline values (absolute/proportional) for the no social interaction group were as follows: 1583.3 (SE 3108.3)/4.6% (SE 7.2%) (week 2), 591.5 (SE 2414.5)/2.4% (SE 4.7%) (week 3), and −1041.8 (SE 1992.7)/0.6% (SE 4.4%) (week 4). For the social interaction group, changes to individual baseline values were as follows: 11520.0 (SE 3941.5)/28.0% (SE 8.7%) (week 2), 9567.3 (SE 2631.5)/23.0% (SE 5.1%) (week 3), and 7648.7 (SE 3900.9)/13.9% (SE 8.0%) (week 4). The result of the analysis of the group effect was significant (absolute change: $\eta^2=0.31$, $P=.04$; proportional change: $\eta^2=0.30$, $P=.03$). Correlations between both absolute and proportional change and the play activity were significant (absolute change: $r=0.59$, 95% CI 0.32 to 0.77; proportional change: $r=0.39$, 95% CI 0.08 to 0.64).

Conclusions: The presence of social interaction design elements in pervasive games appears to have a positive effect on levels of physical activity.

Trial Registration: Japan Medical Association Clinical Trial Registration Number JMA-IIA00314; https://tinyurl.com/y5nh6ylr
(Archived by WebCite at http://www.webcitation.org/761a6MVAy)
Introduction

The proportion of elderly people is increasing in populations worldwide as a natural consequence of the progression in health care and technology, as well as decreased birth rates [1]. This phenomenon intensifies the need for effective strategies to promote the well-being of elderly people. Previous work addressed that goal using different approaches, each with different levels of success [2]. Electronic games became a major topic of interest, being used in varied fields such as the rehabilitation of psychomotor functions [3,4], prevention of age-related diseases [5], and promotion of active lifestyles [6]. In that context, pervasive games are a relatively recent field of research [7].

A pervasive game is an electronic game that blends the real and virtual worlds, proposing interactions that incorporate elements from players’ daily lives into the game rules [8]. They can be especially beneficial for elderly people for three main reasons. First, such games often invite players to walk to real-world locations [9] or to perform different exercises with their bodies [10], and both types of interaction promote higher levels of physical activity, a practice strongly correlated with well-being among the elderly [11-13]. Secondly, these games can encourage social connection, which is an important factor in the maintenance of both physical and mental health among senior citizens [14]. Finally, pervasive gaming experiences also focus on casual gameplay by nature, allowing players to learn on their own pace and giving them a lot of freedom to choose how much they want to engage with the game, an aspect of particular interest for elderly players [15].

In this work, we explore pervasive games for elderly individuals from a design perspective, focusing on a specific design aspect—social interaction—and targeting a specific goal, which is to promote physical activity. To do that, we used a previously developed pervasive mobile game [16-18] and performed a randomized controlled trial with Japanese elderly people, evaluating whether the presence of social interaction elements has any effect on levels of physical activity.

Methods

Design

A single-blind randomized controlled trial was conducted to compare two groups. In the intervention group, participants played a version of a pervasive game that included social interaction elements. In the control group, participants played a single-player version of the same game. A 4-week protocol was adopted. During the first week, participants did not play the game, but their level of physical activity was measured to serve as the baseline. During the remaining weeks, participants could freely play the game, while their levels of physical activity and game actions were monitored. Participants were blinded to group assignment, but researchers were not.

Participants

Participants were recruited from community-dwelling senior adults living in Kyoto City, Japan, using flyers. Because this research is contextualized as a preventive health intervention and it is expected that experience with games will become increasingly common among older adults in the future, we adopted a broader age range of 50 years and older into the study’s inclusion criteria, aiming for middle-aged and older adults. Additional criteria included healthy people with independent ambulation and no cognitive or physical impairment that prevented them from understanding the instructions of the game or taking short walks.

All participants signed an informed consent term, and the research protocol was approved by the Kyoto University Hospital’s Ethical Committee, which reports compliance with the Declaration of Helsinki. Participants were offered a shopping voucher worth 10,000 Japanese yen (US $96.51) as compensation for their participation in the study. The only requirement to receive the voucher was answering the final questionnaire at the end of the study.

Game

Participants played a pervasive location-based mobile game called Shinpo, which, in free translation, means “steps of the gods” in Japanese. This game was first evaluated for its feasibility in Kyoto [19] and later used in a similar study in Brasília, Brazil [20].

In Shinpo, players must collect virtual cards (Figure 1) by visiting shrines and temples in Kyoto city. Each card features an animal and a color, which indicates its level. There are four levels, and the goal of the game is to collect the card of highest level for all animals.
On the main screen of the game (Figure 2A), players could see a map showing their current location and nearby hotspots, which were indicated by red icons. Hotspots were defined using Google Maps information about shrines and temples in Kyoto City. After visualizing nearby hotspots, players had to walk physically toward them, and when they were within 50 m of their locations, a “check-in” option appeared, allowing the player to register a visit. For safety reasons, players were not expected to keep the game screen open while walking and should have accessed the game only when they arrived at a destination. If the game was open and the player’s speed exceeded a certain threshold, the game warned the player not to walk while looking at the smartphone.

A player received a random level 1 card for every unique hotspot they visited within a day, or for every 1000 steps they walked. The step count was measured using a background service that operated whenever the smartphone was turned on. This software used Google Android’s Sensor application programming interface, which is the same as that of Google Fit, an application previously shown to have an accuracy comparable to that of wearable devices [21]. At any moment, players could trade 5 equal cards of the same level for 1 card of the next level (Figure 2B).

In the version of the game with social interaction, additional rules were added to stimulate players to interact and collaborate to obtain more cards:

- Players could leave one copy of any of their cards, up to level 2, on each hotspot they visited each day, once in the morning and once in the afternoon (Figure 2C). If this happened, that hotspot would be highlighted on other players’ devices, with an exclamation mark added to its icon. When a different player visited the highlighted hotspot, they received a copy of the card left there and the original visitor received additional random cards.
- Every day, players were randomly assigned to a challenge group, and anytime a person in the group collected a card, all other members also received a card. The members of the group and their contribution to the challenge were shown to all other members to promote a sense of unity (Figure 2D).
- All players could choose a public avatar and nickname and could make a short self-introduction; when players received cards as the result of other players’ actions, they had a chance to give them a “like” (Figure 2D).

The feasibility study and follow-up evaluations [17,22] suggested that these mechanics allowed players to feel more engaged in playing the game by working together with other people. We hypothesized that this setup would result in a greater positive effect on levels of physical activity.
Outcome Measures

The main observed outcome was the level of physical activity, measured by the average number of steps per week, in a 4-week period. During the first week, participants did not play the game, but their step count was still monitored. This monitoring was performed to assess their baseline level of physical activity. After that, they played the game for 3 weeks. Weekly cycles were chosen to consider different lifestyle behavior patterns during different days of the week. Also, to account for individual differences among participants, the proportional change in relation to the individual’s baseline level of physical activity was also analyzed.

To evaluate how much participants played the game, the weekly average number of visits to hotspots was also observed. Within a single day, this observation represented the number of unique hotspots visited by the player, while within a week, it was the sum of the visits in each day of the week (ie, the same hotspot was not counted twice for the same day, but it could be counted twice in a week). This measurement was used because players were directed to not keep the game open while walking, so play time was not a good measurement of how much a person played.

As a secondary evaluation, participants were also asked to answer a final survey composed of the Game Experience Questionnaire (GEQ) [23] (including the “Social presence” section) and the System Usability Scale (SUS) [24], since these questionnaires have been widely used in previous works. The items of these questionnaires were translated and cross-verified by 2 Japanese native speakers. All items consisted of 5-point Likert scales. In the case of the GEQ, the scale was 0=didn’t feel it at all to 4= felt it extremely; for the SUS, the scale was 0=don’t agree at all to 4=completely agree. Besides these two metrics, participants were also asked to respond to the statement “I remembered to carry the smartphone with me when I left home” according to the following scale: 0=every day, 1=often, 2=sometimes, 3=almost never, and 4=never. Finally, at the end of the survey, participants could freely write comments, criticisms, or suggestions.

Procedures

At the beginning of the study, participants signed an informed consent form and completed the first questionnaire (ie, regarding their previous experience with games and technology). Their smartphones were checked for compatibility, and the game was then installed. Compatible systems included Android-based smartphones with Android operating system version 6.0 or above with a global positioning systems (GPS) sensor and an internet connection. Participants who did not have a compatible smartphone or who could not or did not want to use their personal devices were lent a previously prepared one by the researchers. Participants were given an instruction manual and an oral presentation about basic smartphone usage and the rules and interface of the game.

Participants were randomly assigned to either the social interaction group (intervention group) or no social interaction group (control group), and remained blind to their group assignment. During the first week, no participant could play the game, and the application only showed a message indicating that their step count was being monitored and providing the date on which the game would become available. From the second week on, the game became available and participants were able to play freely. There was a follow-up meeting on the same weekday every week, in which researchers were available to clarify any doubts or solve technical problems.

Participants were instructed about the importance of keeping their smartphones turned on and carrying them with them whenever they left their homes, throughout the duration of the study, since they would be used to measure the participants’ step counts. In the final survey, they were also asked about their compliance with these instructions.

After the end of the trial, participants answered the final questionnaire to evaluate their experience while playing and received the voucher for 10,000 Japanese yen (US $96.51).

All questionnaires were administered by researchers, who were available to answer any questions about the items.

Data Analysis

Questionnaire data were consolidated to report percentages for each item, while means and standard errors were calculated for demographic data. Participants were considered active in a given day if they performed at least one check-in at a hotspot during that day. If a participant spent 3 or more days being inactive in a week, they were considered an inactive participant for that week. No participant explicitly asked to leave the study. Data were processed using Python (mainly the Pandas and Matplotlib packages) and exported into a suitable format for R, which was used for statistical analysis.

To analyze the effects, we calculated the change in the number of steps for each week from the baseline values in week 1. These measurements were made for each participant in relation to their own individual baseline value, and the proportional change was also calculated (ie, the absolute change divided by the baseline value). This was done to consider the effect on each participant in relation to their own initial baseline value.

In the statistical model, the change for each week after baseline was considered to be a repeated measure, and an analysis of variance (ANOVA) was performed—with group and week as factors—for participants who remained active until the end of the experiment. The relationship between the change in the step count and the number of hotspot visits was evaluated using Pearson’s correlation coefficient ($r$).

For the final questionnaires, the value of negatively phrased items was inverted, and the mean and standard error was calculated for each category of the GEQ and for the SUS as a whole, resulting in scores that ranged from 0 to 4.

Results

Participants

A total of 20 participants (16 females, 4 males) were recruited, and 18 participants (14 females, 4 males) remained active until the end of the study. Table 1 presents basic information about the participants, including their previous experience with games and technology.
### Table 1. Basic data of study participants (N=20).

<table>
<thead>
<tr>
<th>Participant data</th>
<th>Baseline</th>
<th>Social interaction</th>
<th>Active until the end</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No social interaction</td>
<td>Social interaction</td>
<td>No social interaction</td>
</tr>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants, n</td>
<td>10</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Sex (female), n (%)</td>
<td>9 (90)</td>
<td>9 (90)</td>
<td>7 (78)</td>
</tr>
<tr>
<td>Age (years), mean (SE)</td>
<td>66.4 (10.2)</td>
<td>63.2 (8.5)</td>
<td>64.6 (8.9)</td>
</tr>
<tr>
<td><strong>PC(^a) usage frequency, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>3 (30)</td>
<td>2 (20)</td>
<td>3 (33)</td>
</tr>
<tr>
<td>Approximately 1 time/week</td>
<td>1 (10)</td>
<td>3 (30)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>≥2 times/week</td>
<td>2 (20)</td>
<td>2 (20)</td>
<td>1 (11)</td>
</tr>
<tr>
<td>Daily</td>
<td>4 (40)</td>
<td>3 (30)</td>
<td>4 (44)</td>
</tr>
<tr>
<td><strong>PC activity(^b), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use email</td>
<td>7 (70)</td>
<td>5 (50)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Browse the web</td>
<td>4 (40)</td>
<td>6 (60)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Search the web</td>
<td>4 (40)</td>
<td>6 (60)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Read the news</td>
<td>5 (50)</td>
<td>5 (50)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Use social networks(^c)</td>
<td>3 (30)</td>
<td>3 (30)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Manipulate photos</td>
<td>5 (50)</td>
<td>5 (50)</td>
<td>4 (44)</td>
</tr>
<tr>
<td>Use office applications</td>
<td>6 (60)</td>
<td>6 (60)</td>
<td>5 (56)</td>
</tr>
<tr>
<td><strong>Mobile phone activity(^b), n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make calls</td>
<td>9 (90)</td>
<td>10 (100)</td>
<td>8 (89)</td>
</tr>
<tr>
<td>Use email</td>
<td>9 (90)</td>
<td>10 (100)</td>
<td>8 (89)</td>
</tr>
<tr>
<td>Browse the web</td>
<td>7 (70)</td>
<td>9 (90)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Use social networks</td>
<td>7 (70)</td>
<td>10 (100)</td>
<td>6 (67)</td>
</tr>
<tr>
<td>Install and use apps</td>
<td>7 (70)</td>
<td>10 (100)</td>
<td>6 (67)</td>
</tr>
<tr>
<td><strong>Play frequency of nonelectronic games, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>0 (0)</td>
<td>5 (50)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Very rarely</td>
<td>7 (70)</td>
<td>4 (40)</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Approximately 1 time/week</td>
<td>2 (20)</td>
<td>0 (0)</td>
<td>2 (20)</td>
</tr>
<tr>
<td>≥2 times/week</td>
<td>1 (10)</td>
<td>1 (10)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Daily</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Play frequency of electronic games, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1 (10)</td>
<td>3 (30)</td>
<td>1 (10)</td>
</tr>
<tr>
<td>Very rarely</td>
<td>6 (60)</td>
<td>2 (20)</td>
<td>5 (50)</td>
</tr>
<tr>
<td>Approximately 1 time/week</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥2 times/week</td>
<td>0 (0)</td>
<td>3 (30)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Daily</td>
<td>3 (30)</td>
<td>2 (20)</td>
<td>3 (30)</td>
</tr>
<tr>
<td><strong>Devices used to play(^b,(^d,(^e)) n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>3 (33)</td>
<td>2 (29)</td>
<td>3 (38)</td>
</tr>
<tr>
<td>Portable console(^f)</td>
<td>1 (11)</td>
<td>1 (14)</td>
<td>1 (13)</td>
</tr>
<tr>
<td>Conventional console(^g)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Mobile phone or tablet</td>
<td>5 (56)</td>
<td>5 (71)</td>
<td>5 (63)</td>
</tr>
</tbody>
</table>

\(^a\) PC usage frequency
\(^b\) PC activity
\(^c\) Use social networks
\(^d\) Mobile phone activity
\(^e\) Play frequency of nonelectronic games
\(^f\) Portable console
\(^g\) Conventional console
Participants in both groups had similar backgrounds of experience using a personal computer (PC), but there were some differences in other areas. For mobile phone usage, participants in the social interaction group had a higher level of self-reported skill, although both groups reported a high level. As for previous experience with games, more participants in the no social interaction group reported playing nonelectronic games than participants in the social interaction group, while the opposite was true for electronic games. Only participants in the no social interaction group reported ever playing games together with other people.

Participants could freely write down titles of games that they often played. For nonelectronic games, cited titles included the following (stated as number of citations by participants in the no social interaction group and in the social interaction group, respectively): crosswords (2 and 0), card games (1 and 1), sudoku (1 and 1), Go (1 and 1), puzzle games (1 and 0), Shogi (0 and 1), Reversi (0 and 1), and Concentration (0 and 1). For electronic games, cited titles included the following: Disney Tsum Tsum (1 and 4), Pokémon (1 and 0), Pokémon Let’s GO (1 and 1), Pokémon GO (1 and 2), Animal Crossing (several versions) (2 and 1), Solitaire (1 and 1), Splatoon (0 and 1), Homescapes (1 and 0), the Legend of Zelda (1 and 0), Monster Hunter (1 and 0), puzzle games (1 and 0), Everybody’s Golf (0 and 1), Mahjong (0 and 1), and LINE POP (0 and 1).

Except for 1 participant in the social interaction group, all participants used borrowed devices. Of the 20 total participants, 9 had an iPhone (Apple Inc) and could not install the application, while the remaining 10 participants that borrowed a mobile phone said that they did not want to install the application on their own device.

Main Outcome

Step count data are shown in Table 2 for participants who remained active until the end of the experiment. The baseline column shows the step counts recorded during the first week for each group. The remaining columns report the change for the subsequent weeks in relation to the participants’ own baseline values. If \( w_{p,i} \) is the number of steps on week “i” for participant \( p \), and \( b_p \) is the baseline number of steps for that participant, then the absolute change is calculated as \( \Delta_p,i = w_{p,i} - b_p \) and the proportional change is calculated as \( q_{p,i} = \frac{\Delta_p,i}{b_p} \).

For the absolute change, the effect of group as a factor in the ANOVA was significant \( (P=.04; \eta^2=0.31) \). No relevant relationship was found with week as a factor \( (P=.19) \). For proportional change, using group as a factor resulted in a \( P \) value of .03 \( (\eta^2=0.30) \), while using week as a factor resulted in a \( P \) value of .17.

For number of hotspot visits, the no social interaction group had a mean of 42.8 (SE 15.8) visits in week 2, 89.7 (SE 24.5) visits in week 3, and 96.2 (SE 35.4) visits in week 4. By comparison, the social interaction group had a mean of 169.0 (SE 63.9) visits in week 2, 115.0 (SE 40.0) visits in week 3, and 140.0 (SE 57.6) visits in week 4.

The correlation analysis between the absolute change in the number of steps and the number of hotspot visits resulted in a correlation factor of \( r=0.59 \) (95% CI 0.32 to 0.77). When proportional change was considered, the correlation factor was \( r=0.39 \) (95% CI 0.08 to 0.64).
Table 2. Number of steps at baseline for each group and individual variations in subsequent weeks.

<table>
<thead>
<tr>
<th>Group and change from baseline</th>
<th>Baseline, median (min; max)</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No social interaction group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change(^a)</td>
<td>42,305 (27,687; 71,028)</td>
<td>921 (–7986; 11,454)</td>
<td>437 (–7999; 8587)</td>
<td>–1143 (–6611; 6476)</td>
</tr>
<tr>
<td>Proportional change (%)(^b)</td>
<td>NA</td>
<td>4.8 (–16.0; 28.1)</td>
<td>1.8 (–11.3; 16.3)</td>
<td>–1.6 (–9.7; 18.7)</td>
</tr>
<tr>
<td><strong>Social interaction group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute change</td>
<td>51,254 (20,356; 65,594)</td>
<td>10,494 (–508; 27,167)</td>
<td>5954 (4944; 20,711)</td>
<td>3722 (–2691; 22,891)</td>
</tr>
<tr>
<td>Proportional change (%)</td>
<td>NA</td>
<td>24.4 (–1.0; 54.5)</td>
<td>24.2 (7.6; 41.5)</td>
<td>10.0 (–12.2; 45.9)</td>
</tr>
</tbody>
</table>

\(^a\)Absolute values indicate the change in the weekly number of steps when compared with the user’s own baseline value.

\(^b\)Proportional values indicate the absolute value divided by the user’s own baseline value (ie, if \(\Delta p,i\) is the change from baseline to week “i” for participant \(p\), and \(b_p\) is the baseline number of steps for that participant, then the proportional change \(q_{p,i} = \Delta p,i / b_p\).

\(^c\)NA: not applicable.

Game Experience

The scores for the usability and game experience questionnaires are summarized in Table 3. The scores of component items ranged from 0 to 4; therefore, a value of 2 or greater indicates a high valuation. The data included are only those from participants who stayed active until the end of the experiment. Since these data were used only as complementary information for future interventions, no further statistical analysis was performed.

For the statement “I remembered to carry the smartphone with me when I left home,” 8 participants in the no social interaction group responded “everyday,” and 1 participant in that group responded “often.” In the social interaction group, all participants responded “everyday.”

Participants’ comments about the game were summarized by one of the researchers who is a native Japanese speaker and briefly interacted with the participants at the last meeting, in which they completed the final questionnaire. Comments made during the last meeting and also written on the questionnaires included the following: “The accuracy of the GPS could be improved,” “I enjoyed visiting new places,” “I want to continue playing,” “I want Shinpo to notify me when I am near the temple,” “It is difficult to understand the communication function,” “I think it would have been more interesting if I was used to using smartphones,” and, “There were too many notifications.”

Table 3. Mean scores for Game Experience Questionnaire (GEQ) and System Usability Scale (N=18).

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean score (SE)(^a)</th>
<th>No social interaction group (n=9)</th>
<th>Social interaction group (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GEQ: Core</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>1.6 (0.3)</td>
<td>2.1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Sensory and imaginative immersion</td>
<td>1.7 (0.4)</td>
<td>2.3 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Flow</td>
<td>1.4 (0.4)</td>
<td>1.6 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Tension</td>
<td>0.5 (0.2)</td>
<td>1.1 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Challenge</td>
<td>1.3 (0.3)</td>
<td>1.5 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Negative affect</td>
<td>0.7 (0.3)</td>
<td>1.0 (0.2)</td>
<td></td>
</tr>
<tr>
<td>Positive affect</td>
<td>2.6 (0.4)</td>
<td>3.0 (0.3)</td>
<td></td>
</tr>
<tr>
<td><strong>GEQ: Social presence</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychological involvement: empathy</td>
<td>0.8 (0.2)</td>
<td>2.0 (0.4)</td>
<td></td>
</tr>
<tr>
<td>Psychological involvement: negative feelings</td>
<td>0.5 (0.2)</td>
<td>1.0 (0.3)</td>
<td></td>
</tr>
<tr>
<td>Behavioral involvement</td>
<td>0.5 (0.3)</td>
<td>1.5 (0.5)</td>
<td></td>
</tr>
<tr>
<td>System Usability Scale</td>
<td>2.6 (0.2)</td>
<td>2.8 (0.1)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Scores are normalized in the 0 to 4 interval; values equal or greater than 2 indicate a high score for the item.
Discussion

Principal Results

Due to the limited number of participants, there were a few differences between the groups after randomization; however, both groups had similar previous experience with games and technology, and the proportions remained similar when only participants who remained active until the end were considered. For PC usage, most participants had at least some experience, and for smartphones, all participants knew at least the basic operations. Unsurprisingly, when previous experience with games was considered, most participants in both groups reported never playing games or playing them very rarely. Participants in both groups reported using PCs, handheld consoles, and smartphones to play, and no participant reported using a traditional gaming console. Most people reported playing alone, with only a few people in the no social interaction group and no one from the social interaction group reporting playing with other people.

For the main outcome, a greater positive effect was observed in the social interaction group than in the no social interaction group. The statistical analysis with respect to the absolute change indicated a large effect size (η²=0.31), and P=.04 indicates a statistically significant difference. For proportional change, similar results were observed, with P=.03 (η²=0.30). The correlation between the number of hotspot visits and the effect was also statistically significant, with a positive value in both cases (absolute change: r=0.59, 95% CI 0.32 to 0.77; proportional change: r=0.39, 95% CI 0.08 to 0.64), which suggests that the gameplay was one of the main factors that generated the effect. These results lead us to believe that the social interaction elements had a relevant positive effect on the levels of physical activity of the players.

The evaluations of game experience and system usability were used as complementary information only, and statistical analysis was not performed. The mean scores were higher for all items in the social interaction group. This suggests that this version of the game had a greater effect on players in “positive” aspects, such as “competence” and “sensory and imaginative immersion,” but also in “negative” ones, such as “tension” and “psychological involvement—negative feelings.” A possible explanation is that this version of the game provided a more intense experience in general. Nonetheless, if we consider as “positive” the categories of “competence,” “sensory and imaginative immersion,” “flow,” “positive affect,” “psychological involvement—empathy,” and “behavioral involvement,” the social interaction version was highly evaluated in 4 of them, and in none of the “neutral” or “negative” categories. For usability, both versions were highly evaluated.

Limitations

The limitations of this study are as follows. The main outcome was measured using smartphone software. The methodology has been evaluated in previous studies, and the authors of those studies concluded that it is adequate; however, the comparative studies considered indoor environments, and outdoor measurements may render different results. In addition, there might be differences in accuracy for different age groups. In this study, we considered weekly cycles and also analyzed the proportional values to evaluate the effects for each individual participant.

Future interventions might test similar settings with a different device, such as an external pedometer, and compare the results. In both cases, because the data are collected not in a controlled environment but rather in a user-dependent context and participants’ adherence to carrying the device is self-reported, there might be inaccuracies in the measurements. We specifically instructed participants about the importance of carrying the smartphones with them so that their step counts could be measured; with the exception of 1 participant who responded “often” to the statement “I remembered to carry the smartphone with me when I left home,” all other participants reported following this guideline every day. Nonetheless, future interventions might use more objective ways of measuring compliance.

Step counts were observed in a continuous state, considering any daily activity of participants, and the number of visits to hotspots was used as a proxy measurement of the amount of game play, since participants were encouraged to only open the game to check in at hotspots and close it between visits. Because step counts for the baseline week were also measured continuously and the analysis considered the observed change, the results are still relevant. Nonetheless, further interventions might also separate in-game counts explicitly and analyze if there is any difference.

The sample population was mostly composed of women, due in part to the fact that the majority of the Japanese elderly population is female, but mainly because of recruitment difficulties caused by the cultural context, such as the fact that most elderly people in Japan engage primarily in activities in gender-specific groups. A more gender-balanced sample would better represent the actual adult elderly population.

Although this study is included in the more general field of interventions to improve the quality of life of older adults, it focused specifically on increasing physical activity based on previous results that show a strong correlation between these variables. Future interventions could measure these two variables explicitly and evaluate their relationship in the context of pervasive games. Also, the questionnaires used to assess usability and game experience were not statistically validated, and were used only to provide complementary information rather than to draw conclusions about the effect. The use of validated metrics would allow us to further understand the data and test more hypotheses.

The proposed social interaction mechanics focused mainly on collaboration and virtual interaction. More types of social interaction and different variables can be tested, such as competition, direct (in-person) interaction, group dynamics, and interaction with family and friends, among others.

The final purpose of promoting physical activity is to increase the well-being of the elderly population. In this study, we evaluated specifically whether social elements in pervasive
games would have any effect on the level of physical activity, but to evaluate the effect on well-being in general would require many additional metrics and a much longer period of time.

**Comparison With Prior Work**

This work is a follow-up of a similar study in Brazil [20], which was performed with a version of the game that was adapted to Brazilian culture but was otherwise identical. The results found in this evaluation were stronger than the results in the Brazil study, with similar values for proportional effects (the absolute effect was also larger, but the baselines were very different) and larger effect sizes for both absolute ($\eta^2=0.30$ versus $\eta^2=0.19$) and proportional ($\eta^2=0.31$ versus $\eta^2=0.27$) measurements. Additionally, while the Brazil study was inconclusive for the correlation between hotspot visits and proportional effect, the results in this study were statistically significant in all cases.

A few studies have already employed pervasive games or gamified apps targeting older adults, usually focusing on specific goals, such as cognitive training [25] and the promotion of physical activity using social incentives [26,27]. A successful commercial example that does not target elderly adults in particular but that became extremely popular among people of all ages is Pokémon GO [28]. Different studies have analyzed its effects on levels of physical activity, finding overall positive results, especially in the first weeks of use [29-31].

Recently, new research has emerged [32,33] analyzing in greater depth the experience of elderly people in play, based on the principle that games—even serious games—should primarily be fun because health benefits come later as a natural consequence of play [34]. In that respect, a few studies have attempted to clarify elderly players’ needs and motivations and investigated possible challenges in designing games for older audiences, listing common physical and cognitive limitations that should be taken into consideration [35-39]. Other studies have attempted to identify the preferences of elderly people regarding the content and/or genre of the games [15,40-42]. In this study, we evaluated social interaction as a design element in the context of pervasive games, a new kind of game that is only now being explored. This study was limited and focused on a specific metric, namely physical activity, which was taken as a proxy, but the results suggest that this topic should be further investigated, with the consideration of additional variables related to the game experience.

**Conclusions**

In this work, we investigated whether the new genre of pervasive games could be used to increase physical activity among older adults by focusing on a specific design element, namely social interaction. Our results indicated that a pervasive game using social interaction had a greater positive effect on levels of physical activity than the same game without social interaction. This study had limitations, but the results are promising, and corroborated a previous study using the same game. In future interventions, other types of social interaction and/or design elements could be evaluated, and additional variables might be considered, such as indicators of physical and psychological health, among others.

**Acknowledgments**

This study was partially supported by the Hayao Nakayama Foundation for Science and Technology and Culture, by Kyoto University’s Collaborative Graduate Program in Design, and by the Japan Society for Promotion of Science’s KAKENHI, grant number 19K12820. The authors thank the participants for volunteering and the students and staff for helping conduct the experiment.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1000 KB - games_v9i1e16458_app1.pdf ]

**References**


38. Brox E, Konstantinidis ST, Evertsen G. User-Centered Design of Serious Games for Older Adults Following 3 Years of Experience With Exergames for Seniors: A Study Design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]


Abbreviations

ANOVA: analysis of variance
GEQ: Game Experience Questionnaire
GPS: global positioning systems
PC: personal computer
SUS: System Usability Scale
Promoting Physical Activity in Japanese Older Adults Using a Social Pervasive Game: Randomized Controlled Trial

Please cite as:

Promoting Physical Activity in Japanese Older Adults Using a Social Pervasive Game: Randomized Controlled Trial
JMIR Serious Games 2021;9(1):e16458
URL: https://games.jmir.org/2021/1/e16458
doi:10.2196/16458
PMID:33404507

©Luciano Henrique De Oliveira Santos, Kazuya Okamoto, Ryo Otsuki, Shusuke Hiragi, Goshiro Yamamoto, Osamu Sugiyama, Tomoki Aoyama, Tomohiro Kuroda. Originally published in JMIR Serious Games (http://games.jmir.org), 06.01.2021. 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 http://games.jmir.org, as well as this copyright and license information must be included.

https://games.jmir.org/2021/1/e16458
Abstract

Background: People with chronic obstructive pulmonary disease (COPD) who are less active have lower quality of life, greater risk of exacerbations, and greater mortality than those who are more active. The effectiveness of physical activity interventions may facilitate the addition of game elements to improve engagement. The use of a co-design approach with people with COPD and clinicians as co-designers may also improve the effectiveness of the intervention.

Objective: The primary aim of this study is to evaluate the feasibility of a co-designed mobile game by examining the usage of the game, subjective measures of game engagement, and adherence to wearing activity trackers. The secondary aim of this study is to estimate the effect of the game on daily steps and daily moderate-to-vigorous physical activity (MVPA).

Methods: Participants with COPD who were taking part in the co-design of the active video game (n=9) acted as the experiment group, spending 3 weeks testing the game they helped to develop. Daily steps and MVPA were compared with a control group (n=9) of participants who did not co-design or test the game.

Results: Most participants (8/9, 89%) engaged with the game after downloading it. Participants used the game to record physical activity on 58.6% (82/141) of the days the game was available. The highest scores on the Intrinsic Motivation Inventory were seen for the value and usefulness subscale, with a mean of 6.38 (SD 0.6). Adherence to wearing Fitbit was high, with participants in both groups recording steps on >80% of days. Usage of the game was positively correlated with changes in daily steps but not with MVPA.

Conclusions: The co-designed mobile app shows promise as an intervention and should be evaluated in a larger-scale trial in this population.

(KEYWORDS

fitness trackers; chronic obstructive pulmonary disease; physical activity; video games; smartphone; mobile phone

Introduction

Background

Chronic obstructive pulmonary disease (COPD) is a leading cause of death worldwide [1] and is associated with persistent respiratory symptoms, reduced exercise capacity, and poorer quality of life [2,3]. People with COPD are generally less active than people without COPD [4,5], and physical activity levels generally decline as the disease progresses [6]. Inactivity in people with COPD is associated with poorer health-related outcomes, including a lower quality of life, greater risk of exacerbations, and greater mortality [6-10]. International guidelines promote regular physical activity in people with...
COPD, generally targeting 30 min of moderate physical activity on most days [11].

Interventions that are effective in achieving this targeted level of physical activity in people with COPD are limited [12,13]. Pulmonary rehabilitation, involving supervised exercise training, is strongly recommended for people with COPD [14,15], as it is very effective at increasing exercise capacity [16,17]. Despite this, many people with COPD struggle to maintain their physical activity levels in the months after pulmonary rehabilitation [18], and some do not become more active at all [19]. A longer period of pulmonary rehabilitation may be more effective at improving activity levels, with one systematic review finding that all studies showing no impact of exercise training on physical activity had durations of less than 12 weeks, whereas all interventions lasting longer than 12 weeks improved physical activity levels [12]. However, lengthening the duration of pulmonary rehabilitation beyond the 12 weeks may lead to reduced availability of places in pulmonary rehabilitation programs, which are already limited in many countries [20].

Consumer-grade electronic pedometers, such as those developed by Fitbit (Fitbit Inc), have been shown to be valid devices for measuring physical activity in people with COPD [21] and may enable people with COPD to be more conscious of their physical activity levels. Behavioral interventions that use technology such as wearable pedometers to facilitate self-monitoring of physical activity have shown some short-term effectiveness in improving physical activity levels in people with COPD [22,23]. However, the benefits of physical activity may be short lived, possibly because of poor long-term engagement with these interventions [24]. A recent Cochrane review of technology-based COPD self-management interventions concluded that “researchers also must take into consideration strategies that will promote long-term engagement with smart technology” [22].

Gamification is an emerging strategy to improve engagement with digital technology, including within the context of health care [25]. Gamification is the use of game design elements in nongame contexts [26] and is a common feature in health and fitness apps, including fitness tracker apps [27,28]. Most commonly, the apps include game features such as digital rewards for goal attainment, avatars (visual representations of players), social or peer pressure (including leaderboards), and the provision of feedback on performance. However, little research exists on game interventions paired with wearable activity trackers in people with COPD, and trials of gamified interventions in other populations have shown conflicting results. For example, a trial in healthy adolescents of an activity tracking website known as Zamzee (Zamzee Co) demonstrated a 54% increase in moderate-to-vigorous physical activity (MVPA) over 6 weeks [29], but a trial Active Team (Portal Australia), a gamified smartphone app for healthy adults, had no effect on objectively measured MVPA over 3 months [30].

Although both interventions in these studies were gamified, they differed substantially in the game elements that were used [29,31], underscoring the impact that different designs can have on the effectiveness of gamification.

Active video games (AVGs), defined as video games that require physical activity to play [32], are another approach that has been used in an attempt to increase physical activity. A number of studies have investigated the use of AVGs in COPD and other chronic respiratory conditions, showing that they can evoke a similar physiological response to more traditional exercises (eg, stationary bicycle) while being more enjoyable [33]. However, the effect that AVGs have on habitual physical activity in an unsupervised setting has not been extensively studied in respiratory disease populations [34] or older adults [35]. In addition, the studies to date have generally used commercially available AVGs that are designed for the general population rather than for older adults [36] or to address the preferences of people with chronic diseases. Trials of AVGs in older adults with chronic diseases, such as COPD, are required, and such trials might be expected to demonstrate greater adherence or effectiveness if those AVGs are designed to take into account the needs and preferences of the patient population involved in the trial.

Aims

Using a co-design process with people with COPD and clinicians, we developed an AVG called Grow Stronger to promote physical activity in people with COPD after pulmonary rehabilitation. A co-design methodology known as participatory design was used. Participatory design is a research and design practice where the users of a particular system participate as co-designers throughout the design process rather than merely as testers providing feedback to designers [37]. A participatory design process can, at least in some circumstances, improve the effectiveness of serious games for health [38].

The primary aim of this study is to evaluate the feasibility of the Grow Stronger AVG intervention in people with COPD by assessing the usage of and engagement with the AVG, along with adherence to wearing the Fitbit activity tracker. The secondary aim of this study is to assess the effect of the Grow Stronger AVG, when combined with a Fitbit activity tracker and Fitbit app, on physical activity in comparison to the Fitbit activity tracker and Fitbit app alone. Primary outcomes included usage of the AVG in the experiment arm of this pilot trial (how often the AVG was used, what types and difficulties of activity goals were chosen, and what breathlessness values were reported), subjective measures of engagement with the AVG in the experiment group, and adherence to wearing Fitbit activity trackers in both the experiment and control arms. Secondary outcomes were daily steps and daily physical activity levels in both the experiment group and control group, as assessed by the Fitbit activity monitor.

Methods

Overview

This study is a pilot trial nested within an iterative co-design process to develop an AVG. This co-design process comprised a series of focus groups with people with COPD (n=10) and clinicians (n=18), aiming to outline, design, and develop an AVG. For the trial, 9 of the 10 people with COPD who were taking part in the co-design process comprised the experiment group, who received the AVG app in addition to a Fitbit activity tracker.
The control group comprised individuals with COPD who did not take part in the co-design process, who received only the Fitbit activity tracker and the Fitbit app.

The study was approved by the Prince Charles Hospital Human Research Ethics Committee and ratified by the University of Queensland Human Ethics Research Office.

The co-design process took place between June 2019 and November 2019, with the pilot trial being conducted for 3 weeks at the end of this process, from October 4, 2019 to October 25, 2019.

**Participants**

People who reported they had been clinically diagnosed with COPD were recruited. A letter containing information about the study was sent to recent (previous 12 months) attendees of pulmonary rehabilitation programs across 4 sites operated by Queensland Health in the Moreton Bay Region of Queensland, Australia. Interested potential participants were screened to ensure they met the inclusion and exclusion criteria. Participants were included if they had attended pulmonary rehabilitation in the past 12 months, were able to read and speak English, and were able to exercise independently (with or without the use of mobility aids and supplemental oxygen). Participants were excluded if they did not have access to a smartphone, were unable to exercise because of medical or physical limitations, required 24-hour supplemental oxygen, or lacked the visual acuity to view the text displayed on typical mobile devices.

**Procedures**

After all participants were recruited, a randomized sequence of participants was generated. Participants were alternately allocated into 2 groups: (1) an active co-design group (hereafter the experiment group), which took part in the focus groups and received an activity monitor and the AVG intervention and (2) a control group, which received an activity monitor but did not take part in focus groups or received access to the AVG. For the 19-week duration of the co-design process, participants in both the experiment and control groups were provided with a consumer-grade wearable activity monitor, namely, a Fitbit Alta HR or Fitbit Charge HR 2 (Fitbit Inc.). This activity monitor was paired to the participant’s smartphone and was capable of tracking steps, physical activity, and heart rate. Participants in both groups were provided with instructions on how to use the Fitbit app, and participants in both groups were set up as friends with other participants within their group, allowing participants to see the weekly step total of other participants and access other social features. It was not possible to blind the participants to their group allocation.

The control group did not participate in the focus groups and only had in-person contact with the research team during a group enrollment session and study conclusion session. As per Figure 1, the control group received regular telephone check-ins across the trial duration to set appropriate step goals on the Fitbit app and to provide the same opportunity to raise any device-related issues as was afforded the experiment group before and after the focus groups. Participants in the experiment group were able to trial the test version of the AVG during the final 3 weeks of the development process. However, not all participants were able to download and use the app the day it became available, resulting in some participants having a shorter period to experience the AVG than others. At the conclusion of the trial, participants relinquished their wearable fitness trackers, but those using the app continued to have access to it for at least a month after the conclusion of the trial.

**Figure 1.** Contact times, types of contact, and interventions received for each group. Group sessions for the experiment group comprised focus groups in the co-design process along with an install session in week 16, whereas the 2 group sessions for the control group were a group enrolment session in the first week and a study conclusion session in the final week. AVG: active video game.

http://games.jmir.org/2021/1/e23069/
Game Intervention

The Grow Stronger game and the co-design process undertaken to develop it will be described more fully elsewhere. In brief, Grow Stronger is a smartphone app that functions as both a game and a physical activity diary. Progress in the game requires the player to report the completion of upper body and lower body physical activities commonly used in the physical rehabilitation of people with COPD. The game features a simple stick figure image of each activity, and players are provided with an additional handout with more complete instructions for each activity. Each day, players choose an upper body and lower body activity and set at what difficulty or intensity they wish to perform these activities. At the completion of each activity, users must report their perceived Borg breathlessness value using a slider present in the app to receive their reward for that activity.

The game features 2 parallel game modes, which can be used together or separately. The first mode functions as a single player mode and uses the theme of growing a garden, where players are rewarded with water in a watering can that can be used to grow a potted plant. The second game mode functions as a cooperative multiplayer game mode and has the theme of a caravan trip around Australia, visiting multiple well-known Australian destinations. As a team, players are rewarded with progress on the trip, determined by the average number of activities completed by the team. All data from the use of the game are reported to a web interface that allows clinicians to monitor the progress of all players and sends encouraging messages. A more complete description of the game, along with representative screenshots and a full list of all available activities, is available in Multimedia Appendix 1.

Outcome Measures

Several primary outcome measures were collected by the AVG in the experiment group, namely, the usage of the app, type of activities completed, difficulty level selected by participants for each activity, and reported Borg breathlessness ratings for each physical activity. Adherence to wearing the Fitbit activity tracker was assessed using step data collected from the Fitbit devices, with nonwear defined as zero steps recorded for an entire day.

Additional primary outcome measures of subjective game engagement were collected at the conclusion of the study by asking the experiment group to complete 3 questionnaires. First, the Game Engagement Questionnaire (GEQ) was used [39], which was measured on a three-level scale (1=no, 2=maybe or sort of, 3=yes). Second, 5 subscales of the intrinsic motivation inventory (IMI) were employed: interest or enjoyment, perceived competence, effort or importance, value or usefulness, and relatedness [40]. Each of these subscales was measured on a seven-point Likert scale (1=completely disagree to 7=completely agree). Finally, a cognitive processing and cognitive activation (CPCA) questionnaire was developed for use in this study, adapted from Hollebeek et al [41] and measured on a 7-point Likert scale as for the IMI. All questions were given using either paper-based or web-based forms immediately after the final focus group.

Secondary outcome measures collected from all participants included total steps and duration and the intensity of physical activity. These measures were automatically collected for the entire 19-week duration of the study by the Fitbit activity trackers provided to participants in both the experiment group and the control group. Devices such as these are considered to be valid low-cost devices to measure physical activity in people with COPD [21]. MVPA was assumed to be the sum of the 2 highest Fitbit categories for active minutes (fairly active and very active categories). This approach has been previously used when comparing consumer-level activity monitors to research-grade accelerometers, demonstrating moderate-to-strong validity for MVPA measured by Fitbit devices in healthy adults in free-living conditions [42].

Before the first focus group, participants also filled in a prestudy survey, providing information on their gender, age, employment status, confidence in technology (on a 0-10 scale), and degree of self-perceived functional limitation because of breathlessness, as assessed using the Medical Research Council (MRC) dyspnea scale [43].

Data Analysis

Data were analyzed and visualized using Python (Python 3.7; Python Software Foundation). Step counts and minutes of activity were collated to a daily figure for each participant, which was used to compute each participant’s average for steps per day and MVPA per day over the period before and after the AVG was downloaded. Days where no step data were recorded were ignored when calculating each participant’s average steps per day and MVPA per day, effectively interpolating these missing days with the participant’s own average for that period. One participant in the control group did not wear the Fitbit during the final 3 weeks of the study and so was excluded from the pre-post comparisons of steps per day and MVPA per day.

Owing to the small sample size and nonnormality evident in some outcomes, the Spearman rank correlations were used to examine relationships between outcome measures (ie, pre-post change in daily steps, change in MVPA, game engagement on GEQ and IMI scales, and game usage). The Spearman rank correlation is not affected by skewness and generally copes better with light-tailed distributions than the Pearson correlation [44].

As this was a pilot randomized controlled trial, with a sample size determined by optimum focus group size during the co-design process rather than being adequately powered to detect differences in primary outcome measures, no statistical tests were performed, and data are presented as mean and SD only.

Results

Participants

Figure 2 shows the progression of participants throughout the study in a Consolidated Standards of Reporting Trials diagram. Of the 89 participants invited to participate in the study, 37 responded and were screened against the inclusion and exclusion criteria. The 25 eligible and consented to participate were randomized into the 2 arms of the study. Of these, 10 from each group attended the first session and completed the prestudy
Overall, 2 participants discontinued and withdrew from the study, both during week 4. One participant withdrew for personal reasons, whereas the other withdrew because of reported skin irritation from the Fitbit device. Two other participants reported some skin irritation, resulting in low Fitbit adherence, but did not withdraw from the trial.

**Figure 2.** Consolidated Standards of Reporting Trials diagram showing the flow of participants through the study.

One Fitbit device had to be replaced during the trial because of issues with synchronization between the activity tracker and phone, but data were not lost. A number of participants also experienced issues with Bluetooth synchronization, but these issues were resolved after troubleshooting discussions with the research team, and this did not appear to result in a loss of data for any full day for any participant (although loss of part of the data for that day may have occurred).

The results from the prestudy survey for the control and experiment groups are shown in **Table 1**. Aside from the gender balance, there were no obvious differences between the groups. Both groups had an MRC dyspnea score between grade 2 and grade 3, indicating moderate functional limitation because of breathlessness.

**Table 1.** Details of participants who received the full intervention in each arm of the study (n=9).

<table>
<thead>
<tr>
<th>Group</th>
<th>Control</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>6 (67)</td>
<td>5 (56)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>65 (7)</td>
<td>70 (6)</td>
</tr>
<tr>
<td>Retired, n (%)</td>
<td>8 (89)</td>
<td>9 (100)</td>
</tr>
<tr>
<td>Confidence with technology (0-10 scale), mean (SD)</td>
<td>5.2 (2)</td>
<td>5.3 (2.3)</td>
</tr>
<tr>
<td>Medical Research Council dyspnea, mean (SD)</td>
<td>2.4 (1.1)</td>
<td>2.4 (1.2)</td>
</tr>
</tbody>
</table>
Game Usage Statistics

Game Usage Frequency

The number of activities logged per day by each participant in the experiment group is shown in Figure 3. The game allowed a maximum of 2 activities to be recorded each day. Most participants (8/9, 89%) engaged in the game after downloading it. Excluding participants who did not use the game at all, the remaining participants logged at least one activity on 58.6% (82/141; SD 21%) of the days when they had access to the game during the test period. Note that not all participants downloaded and installed the game on their smartphone on the same date. Although the test period concluded on October 25, some participants continued to use the game after this date.

Figure 3. Number of activities recorded per day for each participant in the experiment group. The period where each participant had downloaded the game, but before the test period had concluded, is indicated by the shaded background. One participant was not shown, as they did not use the game after downloading it on their phone.

Types of Activities Recorded

Figure 4 shows the frequency of activities that were recorded using the app as well as which participants recorded which activity. Outdoor walking was by far the most recorded activity, recorded 41 times. Walking either indoors or outdoors represented 34.5% (57/165) of all recorded activities.
Borg Breathlessness Values

The frequencies of Borg breathlessness values, reported on the 0 to 10 modified Borg breathlessness scale, after activities when using the app are shown in Table 2. The mean Borg breathlessness score was 3.8 (SD 1.3).

### Table 2. Frequency for Borg breathlessness values recorded by the experiment group.

<table>
<thead>
<tr>
<th>Breathlessness value (modified Borg scale)</th>
<th>Number of events recorded in the app</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td>4</td>
<td>65</td>
</tr>
<tr>
<td>5</td>
<td>34</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

Difficulty Values Selected

The frequency at which participants selected the various difficulty or intensity options in using the app is shown in Table 3. Higher numbers represented greater difficulty for a given physical activity task. All difficulties were roughly equally represented, with no evident skew to higher or lower difficulties.
Table 3. Frequency for difficulty values selected by the experiment group.

<table>
<thead>
<tr>
<th>Selected difficulty</th>
<th>Number of events recorded in the app</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
</tr>
<tr>
<td>3</td>
<td>36</td>
</tr>
<tr>
<td>4</td>
<td>46</td>
</tr>
</tbody>
</table>

Game Engagement

People with COPD in the experiment group completed 3 measures of subjective game engagement: IMI, GEQ, and a series of CPCA questions developed for this study.

The scores on the IMI subscales are shown in Table 4. The highest scores were seen for the value and usefulness subscale, with a mean of 6.4 (SD 0.6). All other subscales had lower scores, ranging from approximately 5.1 to 5.7.

Table 4. Intrinsic Motivation Inventory subscale scores (7-point Likert scale).

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest and enjoyment</td>
<td>5.4 (0.5)</td>
</tr>
<tr>
<td>Perceived competence</td>
<td>5.7 (1.0)</td>
</tr>
<tr>
<td>Effort and importance</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>Value and usefulness</td>
<td>6.4 (0.6)</td>
</tr>
<tr>
<td>Relatedness</td>
<td>5.1 (1.1)</td>
</tr>
</tbody>
</table>

The total score for the 19 items of the GEQ for each participant is presented in Table 5. The GEQ score totals ranged from 19 to 39, with a mean GEQ total score of 30.4 (SD 6.9). As this scale has a minimum possible score of 19 and a maximum possible score of 57, a score of 30.4 represents 30% (11.4/38) of the distance between these extremes.

Table 5. Total scores for the Game Engagement Questionnaire.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Game Engagement Questionnaire score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>36</td>
</tr>
<tr>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>7</td>
<td>26</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>10</td>
<td>39</td>
</tr>
<tr>
<td>12</td>
<td>19</td>
</tr>
</tbody>
</table>

The results for each individual question in the CPCA questionnaire are presented in Table 6. Participants generally had a moderate to high degree of agreement across all questions, ranging from 5.2 to 6.5 on a 7-point Likert scale. Mean scores were higher for items relating to their health goals (items 1-3), all of which had a mean of 6.5 (SD 0.8).

Table 6. Cognitive processing and cognitive engagement individual item results.

<table>
<thead>
<tr>
<th>Question</th>
<th>Score, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPCA1: Using the game gets me to think about my health goals (n=8)</td>
<td>6.5 (0.8)</td>
</tr>
<tr>
<td>CPCA2: I think about my health goals a lot when I'm using the game (n=8)</td>
<td>6.5 (0.8)</td>
</tr>
<tr>
<td>CPCA3: Using the game stimulates my interest to learn more about achieving my health goals (n=8)</td>
<td>6.5 (0.8)</td>
</tr>
<tr>
<td>CPCA4: I spend a lot of time using the game, compared to other ways of being physically active (n=8)</td>
<td>5.4 (1.4)</td>
</tr>
<tr>
<td>CPCA5: Whenever I'm trying to be more active, I usually use the game (n=5)</td>
<td>5.6 (1.5)</td>
</tr>
<tr>
<td>CPCA6: The game is what I usually play when I think about being physically active (n=5)</td>
<td>5.2 (2.4)</td>
</tr>
</tbody>
</table>

*CPCA: cognitive processing and cognitive activation.*
Fitbit Wear Adherence

Figure 5 shows the weekly average adherence to Fitbit activity trackers for participants across the study period, as calculated by days with a nonzero step count divided by total days and expressed as a percentage. Overall, participants in the control group had a slightly higher average adherence than the experiment group, wearing the Fitbit on 94.5% (1069/1131) of days compared with 84.3% (975/1157) of days in the experiment group. This was especially evident during the middle of the study period when adherence in the experiment group decreased for several weeks.

Figure 5. Average weekly adherence to wearing Fitbits in each group before and after the Grow Stronger app was downloaded by the experiment group. The period where each participant had downloaded the game is indicated by the shaded background. The control group, which did not download the game, were aligned with the majority of the experiment group for ease of comparison. As weeks are assumed to start on Mondays, but most participants downloaded the game on a Friday (day 0), the value for the final week before the game was downloaded (the last week in the unshaded area) included values from days 1 and 2 after the game was downloaded. All participants are included in the data presented in this figure. Ctrl: control; Exp: experiment.

Steps

Across all weeks before the game intervention was downloaded, the experiment group averaged 4730 (SD 1959, range 1493-7522) steps per day, as shown in Figure 6. Steps in the experiment group in the weeks after downloading the Grow Stronger AVG averaged 4649 (SD 2357, range 1853-8130) per day, representing a decrease of 81 steps per day or a 2% decrease. In the period before the experiment group downloaded Grow Stronger, the control group was averaging 6394 (SD 4306, range 2700-15,000) steps per day, which then decreased by 800 steps per day (800/6394, 12.5%) to 5593 (SD 4277; range 1924-14,367) steps per day.

Figure 6. Average steps per day and MVPA per day in each group before and after the game was downloaded. Ctrl: control; Exp: experiment; MVPA: moderate-to-vigorous physical activity.
Individual step count charts for each participant are shown in Multimedia Appendix 2 for the experiment and control groups.

**MVPA**

As shown in Figure 6, before the game intervention was downloaded, the experiment group was averaged 33 (SD 30; range 3-76) min of MVPA per day, and in this period, the control group had an average daily MVPA of 34 (SD 41; range 3-120) min. During the game intervention, the experiment group was averaging 42 (SD 48; range 2-122) min of MVPA, and the control group was averaging 33 (SD 62; range 1-182) min of MVPA each day. This represented an increase of approximately 9 min per day or a 26% increase for the experiment group and an approximately 1 min or 2% decrease for the control group per day.

**Correlations Between Outcome Measures**

To explore the relationships within and between the secondary outcome measures (pre-post change in MVPA and steps) and the primary outcome measures (game adherence, game engagement on GEQ, and game engagement on IMI), the Spearman correlation coefficients were calculated. Table 7 shows the results in the form of a correlation matrix (a scatter matrix for these comparisons can be found in Multimedia Appendix 3). There appeared to be a moderately high positive correlation, with a Spearman rank correlation coefficient of 0.62 between the pre-post change in daily step and the usage of the Grow Stronger app (as assessed by percentage of days during the test period with at least one activity logged). Physical activity was weakly correlated with game usage. The total score on the GEQ correlated moderately strongly and positively with the mean score on the IMI, with correlation coefficients of 0.61. The pre-post change in daily steps appeared to be strongly negatively correlated with both the subjective measures of game engagement, with correlation coefficients of −0.79 for the GEQ and −0.71 for the IMI. The subjective measures of game engagement (GEQ or IMI) did not appear to correlate with changes in daily MVPA or to game adherence.

**Table 7.** Spearman rank correlation matrix between primary and secondary outcomes in the experiment group.

<table>
<thead>
<tr>
<th>Outcome measures</th>
<th>ΔModerate-to-vigorous physical activity per day</th>
<th>ΔSteps per day</th>
<th>Game usage</th>
<th>Game Engagement Questionnaire</th>
<th>Intrinsic Motivation Inventory</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔModerate-to-vigorous physical activity per day (pre-post)</td>
<td>1.00</td>
<td><em>b</em></td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>ΔSteps per day (pre-post)</td>
<td>0.21</td>
<td>1.00</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Game usage (percentage of days with activities logged on Grow Stronger)</td>
<td>0.45</td>
<td>0.62</td>
<td>1.00</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Game Engagement Questionnaire</td>
<td>0.21</td>
<td>−0.79</td>
<td>0.07</td>
<td>1.00</td>
<td>_</td>
</tr>
<tr>
<td>Intrinsic Motivation Inventory</td>
<td>0.19</td>
<td>−0.71</td>
<td>−0.19</td>
<td>0.61</td>
<td>1.00</td>
</tr>
</tbody>
</table>

\( \Delta \): change in.

\( _b_ \): These cells are deliberately left blank. The table is symmetrical about the diagonal, so the data are not repeated.

**Discussion**

**Summary of Primary Findings**

This study assessed the feasibility of the AVG intervention by measuring usage of the AVG in the experiment arm of this pilot trial and adherence to wearing Fitbit activity trackers in both the experiment and control arms. The usage of the game in this study was moderate, with participants who used the game recording at least one activity on 58.6% (82/141) of the days. This is slightly lower than the usage rates reported in the first few weeks of similar studies of apps or websites for physical activity promotion. For example, 68% of users of the gamified Active Team app were accessing the gamified app each day by day 7, although this fell to 31% by day 91 [45]. However, usage of Grow Stronger was measured by days with logged activities; therefore, it may be expected to be lower than a usage measurement based on days where the intervention was merely accessed. The difference between these 2 methods of measuring usage was illustrated by the pilot trial of ActivityCoach (Roessingh Research and Development), which found that participants visited the website an average of 23 times over 4 weeks but filled in the daily symptom diary just 16 times on average [46]. Similarly, in the pilot trial of Condition Coach (Roessingh Research and Development), which included a website with a symptom diary and exercise module, the website was visited on 86% of the days available, but the exercise module was only completed on 21% of the days [47]. Although the pilot trial lasted 9 months compared with just 3 weeks in this study, adherence to the Condition Coach exercise portal was clearly low throughout: only 5 of the 12 participants completed any exercises in the exercise portal. In comparison, 8 of the 9 participants completed at least one activity in Grow Stronger. However, these other apps were able to send notifications to participants during these trials, whereas our Grow Stronger app lacked the ability to provide notifications, which may have improved usage if it was available.

The Fitbit was worn on 84% of the days by participants in the experiment group and on 94% of the days by participants in the control group. These values are comparable with those reported by Wan et al [48], where pedometer wear adherence ranged between 80% and 90% in people with COPD. It is worth noting, however, that this pedometer was worn on the hip, pocket, or with a lanyard, in contrast to the wrist-worn Fitbits provided in...
this study. Adherence is thought to be higher for wrist-worn activity trackers than hip-worn in adolescents [49], although several participants in our study reported skin irritation, which may have decreased wear adherence.

Difficulty options selected in the app by participants were roughly equally distributed, indicating that the difficulty options presented were adequate for participants with COPD. Reported Borg breathlessness values were clustered around 3 to 5 on the modified Borg scale, indicating moderate shortness of breath and representing an appropriate target for exercise in people with COPD [50].

Engagement with the game in the experiment group as assessed on the GEQ averaged 30.4, out of a minimum possible score of 19 and a maximum possible score of 57. This is comparable with the GEQ results originally reported during the development of the GEQ, which were generally around 30 to 32 for adolescents and undergraduates playing common video games [39]. However, the result in this study is lower than the score given by healthy adults for AVGs [51], which would have corresponded to a score of 42 if the adapted GEQ with a seven-point scale used in that study was transformed into the three-point scale used in this study and in the original development study of the GEQ. The scores on the GEQ found in this study may be lower than that seen by other games because several items on the GEQ assess the subjective experiences of immersion and flow that may not be elicited by a game, such as Grow Stronger, which does not provide real-time feedback to players. GEQ results for games similar to Grow Stronger are unfortunately not available for comparison.

Participants in the experiment group appeared to have given a higher average rating for the value and usefulness subscale of the IMI than other subscales such as interest or enjoyment, relatedness, and effort or importance. This would suggest that participants were primarily motivated to play Grow Stronger because they saw value in it rather than because they enjoyed the game, felt a degree of relatedness to other players, or were motivated to put effort into it. It is possible that by being involved in the design of the game from its inception, participants in this trial were especially aware of the value or potential usefulness of the game. In addition, it has been shown that multiplayer AVGs offer greater relatedness than single player ones [52], so designing Grow Stronger to provide the option of a single player experience in addition to multiplayer may have diluted ratings on the relatedness dimension.

The CPCA scale showed somewhat higher mean scores for health goal questions than physical activity questions, indicating that participants primarily associated the game with achieving health goals rather than performing physical activity. This could be because of the design of the game, which allowed players to perform physical activity away from their smartphone and then return to their phone to play Grow Stronger. This temporal asynchrony between being physically active and playing the game may have decreased the association between the game and physical activity, whereas participation in the design process may have increased the association between playing the game and reflecting on health goals. As this scale was significantly modified for the purposes of this trial, no comparisons to other studies are available.

It is also noteworthy that a response rate of around 41% (37/89) was observed among participants invited to participate in this study. Although we cannot determine the reason for nonresponse, some may not have responded because of lack of access to, or reservations about, technological interventions. A recent survey of people with COPD attending pulmonary rehabilitation in metropolitan areas of Australia found that 48% had personal access to a smartphone and 57% felt that their technology skills were adequate or better [53]. Nonetheless, the results of that survey and the engagement of participants in our study suggest that a substantial portion of people with COPD are willing to use technology in their rehabilitation.

**Summary of Secondary Findings**

In addition to the primary findings mentioned previously, this study also examined the effect on daily steps and daily physical activity levels of the combination of the Grow Stronger AVG and Fitbit activity tracker with the Fitbit app alone, compared with the Fitbit activity tracker with the Fitbit app alone, in patients with COPD. A noteworthy finding of this study was that the experiment group performed an average of 9 min more MVPA after downloading the game, whereas the MVPA in the control group remained roughly similar. Although these results must be interpreted cautiously because of the low sample size of this pilot trial and the large sample SD in the results, this nonetheless suggests that the AVG may have a positive effect on physical activity in this population. To the best of our knowledge, no published research has examined the effect on physical activity of a mobile game intervention combined with a wearable activity tracker versus an activity tracker alone in patients with COPD.

Another secondary finding of this study was that the average steps per day decreased by 13% in the control group but only by 2% in the experiment group. Despite the large variability in this small sample, this finding is consistent with the hypothesis that AVG could ameliorate the decline in physical activity often experienced after pulmonary rehabilitation [18]. It is also possible that the activity tracker itself caused an initial increase in daily steps, which slowly waned over the duration of the study before being partly counteracted by the effect of the AVG in the experiment group. Other studies of people with COPD have, however, examined the combination of activity trackers with mobile apps to encourage daily steps [23]. These mobile apps were not considered games and lacked clear game features but did incorporate behavior change techniques (such as self-monitoring, goal setting, and social support) that could have an effect similar to game design elements. For example, Moy et al [54] and Wan et al [48] compared the effect of a pedometer alone to the same pedometer combined with a website intervention, which encouraged incremental goal setting, allowed social communication between users, and provided educational and motivational messages. These studies, respectively, showed steps per day increased 13% in the intervention group, with no significant change in the control group at 17 weeks [54], and an increase of 19% in steps per day in the intervention group versus a decrease of 5% in steps per day in the control group after 13 weeks [48]. These 2 studies...
demonstrated an increase in the steps for the intervention with little change to the control group, whereas we found little change in the experiment group but a decrease in the control group. A decrease in daily steps in the control group was also observed in a trial of a mobile app that enabled clinician feedback combined with activity tracker compared with activity tracker alone at 3 or 6 months [55]. However, in contrast to this study where daily steps decreased only in the control group, in that study, both the control group and the experiment group decreased steps per day by approximately 14%. Vorrink et al [55] suggested that either the use of a smartphone app rather than a website or the involvement of health professionals could have contributed to the lack of difference between groups, but the Grow Stronger intervention used in this pilot trial was also a mobile app and also involved clinicians, yet a difference between groups was observed. Furthermore, although both Moy et al [54] and Vorrink et al [55] used activity tracking websites or apps that were specifically designed for users with chronic diseases, only the intervention used in a study by Vorrink et al [55] consulted people with COPD during the design phase and only after the initial design had been developed [56].

The observed change in daily steps was positively correlated with game usage, indicating that those who used the game more often also had a greater increase (or smaller decrease) in daily steps. This is consistent with walking, which is the most commonly recorded activity in the game. A correlation between game usage and MVPA was also present but weaker than that between steps and game usage. A stronger correlation between game usage and physical activity may have been expected, given that a recent study found that those in the top quartile for the usage of a gamified smartphone app, Active Team, increased their daily physical activity by 18 min (around 17% of baseline), whereas users in the lowest 3 quartiles of app usage decreased their daily physical activity by 8 min (around 8% of baseline) [45]. Physical activity in that study was assessed by research-grade accelerometry rather than the wearable activity tracker provided to participants, as was used in this study. It is possible that Fitbit activity trackers did not count toward the MVPA measurement of the short bouts of strength training that the Grow Stronger app encouraged participants to do.

Surprisingly, subjective game engagement, as measured on the GEQ and IMI, appeared to be negatively correlated to steps but not correlated to MVPA. This is contrary to the implicit hypothesis that the engaging nature of AVGs would encourage both more steps and more physical activity. This also conflicts with the results of other AVGs in other populations such as children, where an increased level of intrinsic motivation and enjoyment were correlated with increased physical activity [57]. It is possible that those who found Grow Stronger most enjoyable were more likely to have performed other forms of physical activity encouraged by the app, such as upper limb strength exercises, in place of their usual walking. If strength training could not be readily detected as MVPA by the Fitbit, this would not appear as a correlation between game engagement and MVPA. Possible reasons for the negative relationship between subjective game engagement and physical activity remains speculative, as no other studies have examined this relationship in the context of either smartphone AVGs or AVGs for older adults.

Subjective measures of game engagement did not appear to be correlated with game usage, which is contrary to a previous study that demonstrated a correlation between the IMI rating of a smartphone game to improve physical activity and the usage of said game [58]. However, the correlation in that study was assessed before and after a 24-week intervention in a group of 18 people with diabetes. At just 3 weeks with only 9 people with COPD, this study may not have been long enough for a correlation between game usage and subjective game enjoyment to arise or be large enough to detect whether such a correlation did exist. In addition, that study [58] employed a linear regression rather than the Spearman correlation as used in this study, precluding direct comparison between the 2 studies.

Limitations

This study is limited in several ways. As a pilot trial with a small number of participants and short duration, with the sample size, and trial duration oriented around the co-design process, this trial was underpowered to detect differences in steps and physical activity, which could be expected from such a short intervention. A larger and longer duration trial would be required in the future to gain a meaningful estimate of the effect of the AVG on daily steps and physical activity. In addition, participants in the experiment group were co-designers of the intervention, so they may have been more invested in the game that they helped to create. Therefore, this study’s results may not be generalizable to a population who are naïve to the game.

Furthermore, the trial period formed part of the design process, with feedback from participants used to develop a final version of the game intervention beyond the test version trialed in this study. Therefore, the results of this study do not account for the effect that these revisions may have had on the effectiveness of or adherence to the final version of the game, which will be tested in future studies. For instance, the version of the Grow Stronger app used in this study was not able to record usage data regarding when participants accessed the app for purposes other than to record the completion of an activity. As such, no data were available on how often participants used the pause button feature or used the app to interact with one another or check their individual or team progress. Similarly, as mentioned, the app was unable to send notifications, and app usage was likely lower than it would be with reminder notifications enabled. The future version of the app will address these limitations.

As this pilot trial was embedded within a co-design process, the control condition was selected without full knowledge of the eventual design of Grow Stronger and so may not have been an ideal comparison. The control group received Fitbit activity trackers and the Fitbit app under the assumption that the AVG provided to the experiment group would most closely resemble the Fitbit app, albeit in the form of a game. However, as a result of the co-design process, Grow Stronger more closely resembled a mobile exercise diary. As such, future studies aiming to explore the effect of the game elements of Grow Stronger on physical activity should compare this game with a mobile exercise diary app that functions similar to Grow Stronger but...
lacks game elements. In addition, the use of any digital technologies in COPD may not represent the usual standard of care for people with COPD. Therefore, future research may also compare Grow Stronger with usual care of people with COPD after pulmonary rehabilitation, namely, providing a control group with an entirely unsupervised home exercise program.

In this study, activity outcome measures were recorded by commercial-grade Fitbit activity monitors, rather than a research-grade physical activity monitor, which presents 2 main limitations. First, although older hip-worn Fitbit devices have been shown to have a good correlation with research-grade activity monitors [21] in people with COPD, this study was conducted with newer wrist-worn Fitbit activity monitors for which such validity data are not known in COPD. The wrist-worn Fitbit Charge 2 devices used in this study have shown high validity for step counts but only moderate validity for MVPA when compared with research-grade accelerometry in older adults [59]. Second, the Fitbit devices were part of the intervention given to both groups, and data from such devices were visible to participants, which may have caused participants to alter their physical activity in response. Future studies may therefore benefit from employing accelerometers with established validity for MVPA and concealing such measurements from participants.

It is also worth noting that the CPCA questions used herein, although they were based on previously validated scales, were modified significantly to fit the purposes of this study and so may no longer retain their validity. Future research should seek to validate these measures or use alternative validated measures.

Finally, no objective tests of lung function or functional status were performed, nor were results from such tests available for this trial. Although it is very likely that all participants had COPD, as they all had attended a pulmonary rehabilitation class that requires referral by a physician, future research may benefit from confirmation of a clinical COPD diagnosis. In addition, the results of spirometry or exercise tolerance tests could be used to appropriately stratify participants in a future trial.

Conclusions

To our knowledge, this is the first study that trials an AVG designed by people with COPD and clinicians to maintain or enhance physical activity levels. Although the results are limited because of the small sample size, this study is an initial demonstration of the potential value of an app that facilitates physical activities for people with COPD. Future work is required to further improve adherence and to investigate the long-term effects of this intervention. Despite this, the Grow Stronger app shows promise as an intervention worthy of a larger-scale trial in this population.

Acknowledgments

The authors would like to thank Bitlink for their work in developing the software used in this study. Additional thanks go to Anita Keightley for her assistance with recruitment of participants in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots and description of the Grow Stronger smartphone app. [DOCX File, 698 KB - games_v9i1e23069_app1.docx ]

Multimedia Appendix 2

Daily step counts for individual participants in the experiment group and control group. [DOCX File, 307 KB - games_v9i1e23069_app2.docx ]

Multimedia Appendix 3

A scatter matrix for correlations between the primary and secondary outcome measures in the experiment group. [DOCX File, 73 KB - games_v9i1e23069_app3.docx ]

References


Abbreviations

AVG: active video game
COPD: chronic obstructive pulmonary disease
CPCA: cognitive processing and cognitive activation
GEQ: Game Engagement Questionnaire
IMI: Intrinsic Motivation Inventory
MRC: Medical Research Council
MVPA: moderate-to-vigorous physical activity

http://games.jmir.org/2021/1/e23069/
A Bowling Exergame to Improve Functional Capacity in Older Adults: Co-Design, Development, and Testing to Compare the Progress of Playing Alone Versus Playing With Peers

Jorge Luiz Andrade Da Silva Júnior1*, BPEd, MSc; Daiana Biduski2*, BSc, MSc; Ericles Andrei Bellei2*, BTech, MSc; Osvaldo Henrique Cemin Becker2; Luciane Daroit2, BMath, MSc; Adriano Pasqualotti1,2, BMath, MSc, PhD; Hugo Tourinho Filho1, BPEd, MSc, PhD; Ana Carolina Bertoletti De Marchi1,2, BSc, MSc, PhD

1School of Physical Education and Physiotherapy, University of Passo Fundo, Passo Fundo, Brazil
2Institute of Exact Sciences and Geosciences, University of Passo Fundo, Passo Fundo, Brazil

*these authors contributed equally

Corresponding Author:
Ericles Andrei Bellei, BTech, MSc
Institute of Exact Sciences and Geosciences
University of Passo Fundo
B5 Building, São José, Highway BR285, Km 292
Passo Fundo, 99052900
Brazil
Phone: 55 5433168354
Email: 168729@upf.br

Abstract

Background: Older people often do not meet the recommended levels of exercise required to reduce functional decline. Social interaction is mentioned by this cohort as a reason for joining group-based exercises, which does not occur when exercising alone. This perspective shows that exergames can be used as motivational resources. However, most available exergames are generic, obtained from commercial sources, and usually not specifically designed or adapted for older people.

Objective: In this study, we aim to co-design and develop a new exergame alongside older participants to (1) tailor the game mechanics and optimize participants’ adherence to and enjoyment of exercise; (2) test the participants’ functional capacity, motivation, and adherence to the exergaming program; and (3) compare these scores between those who played alone and those who played with peers.

Methods: We conducted a co-design process to develop a new exergame adapted to older people. For user testing, 23 participants were divided into 2 groups to play individually (alone group) or to compete in pairs (with peers group). They played the game twice a week, resulting in 21 exergaming sessions. We assessed the participants’ General Physical Fitness Index (GPFI) before and after the user testing. We also administered questionnaires about the gaming experience and exercise adherence with its motivators and barriers.

Results: We introduced a new bowling exergame for Xbox with a Kinect motion sensor that can be played in single or multiplayer mode. For the GPFI measurements, the sample was homogeneous in the pretest (with peers group: mean 40.5 [SD 9.6], alone group: mean 33.9 [SD 7.8]; P=.11). After the exergame testing sessions, both groups had significant gains (with peers group: mean 57.5 [SD 8.7], P=.005; alone group: mean 44.7 [SD 10.6]; P=.02). Comparing the posttest between groups, it was found that the group in which participants played with peers had better outcomes than the group in which participants played alone (P=.02). Regarding the gaming experience and exercise adherence, both groups recognized the benefits and expressed enthusiasm toward the exergame.

Conclusions: The findings suggest that the developed exergame helps in improving the functional capacity and adherence to physical exercise among older people, with even better results for those who played with peers. In addition to leading to more appropriate products, a co-design approach may positively influence the motivation and adherence of participants.

(JMIR Serious Games 2021;9(1):e23423) doi:10.2196/23423
KEYWORDS
functional status; elderly; virtual reality therapy; user-centered design; software design; video games

Introduction

Functional capacity expresses an individual’s capacity to perform submaximal activities [1]. For older people, decreased functional capacity may negatively influence the performance of daily activities and everyday life, interfering with the person’s independence, which is directly associated with impairments in functional performance [2,3]. To reduce functional decline, health professionals recommend physical activity, as it positively influences physical aptitude, cognition, and overall health conditions [4]. Frequent and regular physical exercise is an appropriate recommendation, as it can prevent numerous age-related declines, improve functional capacity, and improve quality of life [5,6]. However, despite the advantages of exercise routines, physical inactivity is highly prevalent among older people, and they might find it difficult to meet the minimum recommended activity levels needed to maintain health benefits [7-9].

The search for appropriate and attractive ways to engage older people in physical exercise is a challenging process [10]. In this perspective, exergames are presented as safe and beneficial resources for the practice of physical exercises by older people [11,12]. This type of game uses motion sensors to capture the player’s actual movements, stimulating physical interaction [13]. Exercise routines based on digital games are promising to keep players physically and socially motivated [14-16]. In addition, they present motivational factors such as performance evaluation and feedback, which may reduce dropout rates [17,18]. However, a considerable part of the available exergames is from commercial sources, generally not designed exclusively for older people or based on specific protocols [19]. Furthermore, a generic approach is not the best solution to engage older people; thus, exergames need to be personalized to match their goals and performance levels [11].

Studies indicate that the practice of physical activity when performed alone does not show good adherence by older people because of the lack of motivation [20,21]. In contrast, social interaction is the central aspect mentioned by older people as a reason for joining group exercises [22-24]. Practices that combine the advantages of group-based exercise instructed by a professional can increase effectiveness through a personalized program with specific exercise routines, keeping older people motivated to exercise [10,25]. Socializing, either through cooperation or competition, has been investigated as a motivational factor mentioned by players who prefer exergaming with peers [26,27].

Although several studies report that exercising with other people influences the motivation of older people, further research is needed to verify whether the practice of exergaming in the same conditions as the conventional exercise may increase adherence. In addition, studies should verify whether exergames can positively affect the health outcomes of participants. We assumed that an exergame designed specifically for older people could produce better outcomes and engagement. Before implementing an exergame, the needs, wishes, and motivation of the intended players should be assessed, for example, by applying user-centered design methods [11]. Moreover, different processes and activities should be considered to increase engagement in the exergame [11]. A co-design process can significantly contribute to product development, reducing social stigma, empowering users [28], and giving voice to patient groups in the creation process [29]. This study has 3 objectives. First, we intended to co-design and develop a new exergame alongside older participants to tailor the game mechanics and optimize their adherence and enjoyment for exercise. Thereafter, we tested the participants’ functional capacity, motivation, and adherence to the exergaming program and compared these scores between those who played alone and those who played with peers.

Methods

Recruitment and Participants
We conducted a co-design process and tests with 23 older participants who attended Caixeiral Campestre Club, based in the city of Passo Fundo, Rio Grande do Sul, Brazil. The inclusion criteria were as follows: participants who reported practicing some type of physical activity at least once a week before this study and who agreed to pause it during the testing period. All participants also had to have prior knowledge about the use of any type of technology, such as smartphones, computers, or video games. Men and women older than 60 years were accepted if they had cognitive aptitude attested by the Mini-Mental State Exam [30], which we administered before the intervention. As exclusion criteria, we excluded participants who reported having been discouraged by a physician to practice physical activity, had been hospitalized in the past 20 days, or would not have been able to attend all gaming sessions. Written informed consent was obtained from all participants. The local ethics committee of the University of Passo Fundo approved all procedures involving human participants under the opinion number 2.784.343.

For the test procedure, participants were separated into 2 groups: the alone group, initially with 11 participants who played the exergame individually, and the with peers group, initially with 12 participants who played the exergame, competing in pairs. Participants were allocated to the groups on an alternate basis by the order of agreement to participate. Throughout the tests, the pairs could watch each other play. As they were colleagues at the club, they also frequently talked about the game even outside the test sessions.

Game Co-Design and Development
The development team consisted of 3 fitness trainers, 2 software developers, and 3 human-computer interaction specialists with prior experience in game development for older people. A multidisciplinary team was fundamental for project development, encouraging communication between professionals and participants. The human-computer interaction
specialists assisted in enrolling participants for the co-design practice and specifying aspects of game interaction such as aesthetics, gameplay, levels, scoring, competition, and other features. Fitness trainers explored the possibilities of training exercises, assisted with gameplay orientations, and assisted with the levels of difficulty, according to the body movements required to complete the actions in the game.

Our study was guided by 3 main aspects, which examined the interface design, game design, and player experience [31]. For game design and development, we applied a co-design process using experience-based design theory, based on the method described by Askenäs et al [32]. We also followed the guidelines formulated by Planinc et al [33], which consider the development of exergames controlled by body movements designed especially for older players. The guidelines include the use of appropriate gestures, minding the physical condition of the participants and avoiding small objects. Planinc et al [33] also mentioned the encouragement of social interaction and that the exergame should be adjusted to the interests of older people, based on activities that are familiar to them.

Following the co-design actions, in the first stage of the process, the development team had 6 meetings to discuss aspects related to technology, game design, testing, purpose of the study, and expected results. The meetings were held once a week for approximately 50 min at the club where the tests would be performed. In the meetings, other exergames developed and practiced specifically by older people were investigated to ascertain what aspects and characteristics should be incorporated in the development of a game for this target audience.

In the second stage, we set up the co-design group, which included the development team and members of the club who were willing to participate in the design process. The contributions of the other team members helped encourage the user participants to join the group. All participants met at the same location and were introduced to each other and the objectives of the study. The pace of the meetings was controlled by the researchers to ensure that the process remained on track and that the procedures were fully understood by all participants. During 2 meetings, the group brainstormed ideas to discuss the types of games that older people would like to play. The team of professionals provided insights on the choice of gaming, each according to their expertise. The user participants suggested that they would like to play bowling, because of its similarity to the real-world game, that is, the way of playing was already familiar to them. They also stated that they would enjoy playing virtual bowling because it reminded them of the activity they practiced with family and friends.

The software developers believed that they could implement socializing and competition features in the bowling game. The fitness trainers affirmed that bowling was a suitable choice as the movements performed by the player are the same in the real and virtual world. Therefore, all members of the group agreed that bowling was the most appropriate choice of game. Bowling is a closed skill sport that requires fine-tuning and training of self-paced predefined movements as well as a high level of concentration and balance to completely knock down the pins [34-36]. This kind of game may also help prevent an inactive life, promote physical and cognitive functions, and result in entertainment and motivation with social presence [6]. Furthermore, bowling allows slow movements that correspond to typical physical activities that are recommended for older people [37]. In this sport, they can practice by performing body movements such as real sports where flexibility, abductions, and extensions of the lower and upper limbs are required.

In the third stage, the group described the main tasks, activities, requirements, and objectives that they believed should be a part of the game. The tasks were designed to be simple in structure to facilitate user understanding and attention span. Some exergaming concepts were analyzed, and a few design ideas were tested. After this stage, possible design solutions were elaborated, building ideas on how the game could be presented to users, from sketches following the overall concept and the information collected in the previous stages. Participants could freely explore and express their thoughts. No detailed models were expected as the data collection method was new to the user participants.

The game was implemented using a Kinect v2 motion sensor (Microsoft Corporation) and motion capture technology to guide and immerse the avatar in a 3-dimensional (3D) environment. The game is compatible with devices running the Windows operating system (Microsoft Corporation). We used Unity 3D (Unity Technologies) as the game engine, Blender (Blender Foundation) for modeling and texturing the game objects, and Adobe Fuse (Adobe Inc) for the creation and modeling of the 3D characters. When the first stable version of the game was concluded, we began the tests. Throughout the testing period, the group reported feelings and impressions about the game, including difficulties, problems, and improvements that could be implemented. The technique used to gather information at this stage was observation and the think-aloud method to collect the participants’ immediate feedback [38]. All collected information was implemented during the testing period. As new characteristics were implemented, the game was updated for the sessions with the user participants. This process is characterized as iterative, in which activities and assessments must be repeated until a satisfactory solution is found [39]. The iterative development process helped researchers reliably collect tangible data. At the end of this study, everyone who was a part of the research met in a focus group to ensure that the results met the genuine needs of older people and to clarify previous experiences and impressions, considering each member’s vision and level of satisfaction with the game developed.

**Outcome Measures**

We used the following measurements to assess participants and their interaction with the game:

- The Senior Fitness Test [40] was used to assess functional capacity and obtain the General Physical Fitness Index (GPFI). It consists of 6 tests designed to assess physiological parameters associated with independent functioning: lower and upper body strength, aerobic endurance, lower and upper body flexibility, and agility and dynamic balance. Numerous studies have shown that functional capacity can be assessed with several instruments.
to provide important diagnostic and prognostic information in a wide variety of clinical and research settings [1].

- With the Physical Exercise Adherence Questionnaire [41], we used the 13 yes or no statements of motivators and 12 yes or no statements of barriers for exercising, in this case, related to the exergame. Picorelli et al [41] developed and validated this questionnaire in Portuguese, which was adapted for the Brazilian older adult population, considering Brazil’s cultural, economic, and social contexts.

- Game Experience Questionnaire (GEQ) [42] was used to characterize players’ gaming experience during and after the gaming session. We used its in-game and postgame modules, which contain 14 statements and 17 statements, respectively, with a semantic differential scale response ranging from 0 to 4. Each component score is computed as the average value of its specific items. GEQ has been widely used in traditional game studies [43] and studies dealing with exergames including older participants [44].

Testing Procedure

In the pretest, participants were interviewed in a predetermined setting on the club premises. On that occasion, a fitness trainer administered the Senior Fitness Test to all participants. In the test period, participants from both groups played the exergame in sessions twice a week, resulting in 21 exergaming sessions. This twice-weekly design and the 10-week duration were defined based on an average of similar studies [3,10]. Each participant had 10 rounds per session, with 2 throws each, regardless of whether the player scored points. There were 10 throws with the right arm and 10 throws with the left arm. Therefore, each participant had 20 throws, resulting in approximately 10 min of gameplay per session. A fitness trainer followed all sessions and instructed participants on the gameplay rules. For the with peers group, the throws were interspersed between the 2 participants of the pairs, who played competing and seeing the opponent’s score on the game screen. During the testing period, participants were not allowed to perform any other type of physical exercise. After the gaming test period, all participants had their functional capacity reevaluated using the Senior Fitness Test. The participants also answered the GEQ and Exercise Adherence Questionnaire.

Statistical Analysis

Quantitative data were analyzed using the SPSS 22.3 statistical package (IBM Corp). We performed Wilcoxon signed intergroup comparisons based on negative posts to obtain the GPFI through the Senior Fitness Test for pre and posttest in both groups. We also performed Mann-Whitney tests with signed intragroup comparisons. We performed Mann-Whitney U test with signed intragroup comparisons to analyze the GEQ scores. All analyses considered a 95% CI (significance for P ≤ 0.05).

Results

Game Evolution

We intended to design a playful experience in which the physical movements themselves would contribute to the meaning of the game. The first contact participants had with the game was with a simpler first version, which had only one character and no sound effects. In this version, the game consisted of 2 distinct scenarios, one with the main menu and one with the bowling alley and surrounding bleachers, both with simple and homogeneous textures. The menus and the scoreboard of the rounds had only basic features.

During the testing period, configurations and improvements were implemented gradually, according to the need observed in the interactions through the participants’ reports. The game scenario was set with realistic texturing. Background songs were added to the home menu and for the game rounds. Rolling ball and falling pins sound effects were also added. Later, we included the option to return to the menu during the game without closing it. In that version, a female avatar was included, as the participants questioned the absence of such a character. In the menu implementation, we added the input of the participant’s name, gender, and age as well as a ranking with all registered players’ scores. Various colors and textures of bowling balls were also added and randomly chosen when the participant had to catch the ball to throw. In that version, the ball was set at a variable speed, according to the player’s throwing strength.

For the intermediate version, the characters were replaced by 6 new ones, 3 from each gender. The characters resembled older people. The bleachers were composed of a cheering audience to motivate the participants. We updated the instance system for balls, pins, scores, and match history. The pin counting and replacement system were simplified, and the balls on the rack were no longer static as before, as they had to be instantiated and rolled through the rack to be released. The participant could choose the character and start the round using an arm gesture.

For the final version, the interface was redesigned with new colors and harmonic proportions, including the option of gesture interaction to go through the menus. Four more character options were added, resulting in 10 different characters. Strike and spares counts were added. At the end of the game, it reported the accuracy and performance of each player based on their score. A complete match history was implemented. New camera animations and unified scenes were added. When there was a strike, confetti effects appeared in the scenario, and the audience cheered. In the menu, we added options for sound volume, sensor distance, and activation of gesture interaction. The lighting system and scripts were optimized, and new background sounds were added. Figure 1 illustrates the evolution of the exergame with the 3 versions used over the testing period.

http://games.jmir.org/2021/1/e23423/
The final developed exergame, called *Boliche Virtual*, has 2 game modes, one for single players and another to be played in pairs (multiplayer). The Brazilian National Institute of Industrial Property granted us the software copyright registration code BR5120190019121. The game works similar to the real-world 10-pin bowling game, which consists of 10 frames. In each frame, the player has 2 chances to knock down as many pins as possible with the bowling ball. When playing with peers, 2 players are interspersed in each round. Each dropped pin equals 1 point; therefore, the maximum score a player can obtain is 200 points per session. The game’s sequence and scoring system is not entirely similar to the actual bowling system, as it was designed in a more straightforward way to facilitate understanding and gameplay for older players. The total scores for each round, strikes, and spares were stored. At the end of the gaming session, the game announces the winning player and shows the accuracy of each player, ranging from 0% to 100%. Multimedia Appendix 1 provides the screenshots and details of the bowling exergame’s final version.

### Participant Characteristics

At the end of the gaming sessions, 19 participants completed the tests (age: mean 68.9, SD 5.3 years). In the with peers group, there were 10 participants (7 women and 3 men) with a mean (SD) age of 68.8 (5.2) years. However, in the alone group, there were 9 participants (6 women and 3 men) with a mean age of 69.1 (SD 5.8) years. The loss of testing was because of the participants’ loss of interest. All participants were retired, living in the city where the study was performed, and had a higher education level. Although all participants had already used some type of technology, all said that they had never interacted with an exergame before this study. No statistically significant differences were observed between the groups across the baseline demographic variables of gender (\(P=0.89\)), age (\(P=0.90\)), education level (\(P\geq.99\)), or prior knowledge about exergames (\(P\geq.99\)).

### User Testing

Table 1 shows the statistical results of the GPFI score pretest and posttest. Initially, considering the pretest, the sample was homogeneous (\(P=0.11, Z=-1.603\)). Both groups showed significant gains from the exergaming sessions (\(P=0.005\) and \(Z=-2.807\) for the with peers group; \(P=0.02\) and \(Z=-2.273\) for the alone group). Comparing the posttest between groups, the with peers group had significantly better results than the alone group (\(P=0.02, Z=-2.416\)).

<table>
<thead>
<tr>
<th>Test</th>
<th>Played with peers (n=10)</th>
<th>Played alone (n=9)</th>
<th>(P) value[^{a}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest, mean (SD)</td>
<td>40.5 (9.6)</td>
<td>33.9 (7.8)</td>
<td>0.11</td>
</tr>
<tr>
<td>Posttest, mean (SD)</td>
<td>57.5 (8.7)</td>
<td>44.7 (10.6)</td>
<td>0.02</td>
</tr>
<tr>
<td>(P) value[^{b}]</td>
<td>.005</td>
<td>.02</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\[^{a}\]Intergroup comparison using Mann-Whitney \(U\) test.

\[^{b}\]Intragroup comparison using Wilcoxon test.

\[^{c}\]N/A: not applicable.

Figures 2 and 3 show the results of the Physical Exercise Adherence Questionnaire, with statements of barriers and motivators encountered by participants when playing the exergame. As for the motivator statements, the game was motivational for both groups. There was a significant difference between the groups for the statements “Groupmates would help me deal with my problems” (\(P=0.04\)) and “I’d prefer group games to individual games” (\(P=0.04\)), which had a greater agreement in the with peers group. With regard to the barriers faced by the participants, the with peers group showed higher percentages than the alone group in the questions “It is difficult to play when I am in pain” and “It is difficult to play when I am sad.” In turn,
several participants in the alone group reported that “Bad weather hinders me from playing” and “Transport difficulty hinders me from playing.” Finally, all participants in both groups agreed that “If my health was better, I would be more active.” In either group, no participant answered affirmatively to statements about tiredness, difficulty in use, and fear of falling when interacting with the game.

Table 2 shows the results of the GEQ scores and the comparisons with the Mann-Whitney U test, signed intragroup comparisons, with groups as the grouping variable. The only component with a statistically significant difference between groups was sensory and imaginative immersion, which had a higher average score for participants who played with peers. Although nonsignificant, the with peers group obtained a slightly higher average than the alone group in the competence category. In the categories challenge and flow, the values were higher but nonsignificant for the alone group. In the evaluation of participants’ experiences after using the game, the tiredness category had no answers. For both groups, the tension category obtained a lower average of results compared with the other categories. Similarly, there were no answers linked to some aspects of the negative affect category. In contrast, both groups obtained a high average in the positive affect category.

Figure 2. Responses to the adherence statements about barriers when playing the bowling exergame.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Played with peers (n=10)</th>
<th>Played alone (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I have difficulty performing all stages of the game</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>I'm afraid of falling when I play</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>I feel very tired when I play</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Bad weather hinders me from playing</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Transport difficulty hinders me from playing</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>It's hard to play when I'm sad</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>It's hard to play when I'm in pain</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>I'm not interested in games</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>If my health was better, I would be more active</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>I feel like I don't have the strength to play</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>I'm afraid of getting hurt playing</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>I feel the same whether I'm playing or not</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>

Answer: [ ] Yes [ ] No
Figure 3. Responses to the adherence statements about motivators when playing the bowling exergame. The asterisk indicates \( P = .04 \) intergroup.

Table 2. Mean (SD) scores for the components of the Game Experience Questionnaire.

<table>
<thead>
<tr>
<th>Module and component(^a)</th>
<th>Played with peers (n=10), mean (SD)</th>
<th>Played alone (n=9), mean (SD)</th>
<th>( P ) value(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-game</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>3.80 (0.26)</td>
<td>3.72 (0.67)</td>
<td>.55</td>
</tr>
<tr>
<td>Sensory and imaginative immersion</td>
<td>3.90 (0.21)</td>
<td>3.67 (0.25)</td>
<td>.05</td>
</tr>
<tr>
<td>Flow</td>
<td>3.85 (0.24)</td>
<td>3.94 (0.17)</td>
<td>.33</td>
</tr>
<tr>
<td>Tension</td>
<td>0.00 (0.00)</td>
<td>0.11 (0.22)</td>
<td>.13</td>
</tr>
<tr>
<td>Challenge</td>
<td>3.05 (0.90)</td>
<td>3.56 (0.68)</td>
<td>.18</td>
</tr>
<tr>
<td>Negative affect</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Positive affect</td>
<td>4.00 (0.00)</td>
<td>3.89 (0.33)</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Postgame</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive experience</td>
<td>3.28 (0.08)</td>
<td>3.22 (0.25)</td>
<td>.92</td>
</tr>
<tr>
<td>Negative experience</td>
<td>0.67 (0.00)</td>
<td>0.74 (0.22)</td>
<td>.29</td>
</tr>
<tr>
<td>Tiredness</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Returning to reality</td>
<td>1.07 (0.41)</td>
<td>1.04 (0.26)</td>
<td>.45</td>
</tr>
</tbody>
</table>

\( ^a \)Scores can range from 0 to 4.

\( ^b \)Mann–Whitney \( U \) test with signed intragroup comparisons.

---

Table 2. Mean (SD) scores for the components of the Game Experience Questionnaire.

<table>
<thead>
<tr>
<th>Module and component</th>
<th>Played with peers (n=10), mean (SD)</th>
<th>Played alone (n=9), mean (SD)</th>
<th>( P ) value(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-game</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competence</td>
<td>3.80 (0.26)</td>
<td>3.72 (0.67)</td>
<td>.55</td>
</tr>
<tr>
<td>Sensory and imaginative immersion</td>
<td>3.90 (0.21)</td>
<td>3.67 (0.25)</td>
<td>.05</td>
</tr>
<tr>
<td>Flow</td>
<td>3.85 (0.24)</td>
<td>3.94 (0.17)</td>
<td>.33</td>
</tr>
<tr>
<td>Tension</td>
<td>0.00 (0.00)</td>
<td>0.11 (0.22)</td>
<td>.13</td>
</tr>
<tr>
<td>Challenge</td>
<td>3.05 (0.90)</td>
<td>3.56 (0.68)</td>
<td>.18</td>
</tr>
<tr>
<td>Negative affect</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Positive affect</td>
<td>4.00 (0.00)</td>
<td>3.89 (0.33)</td>
<td>.29</td>
</tr>
<tr>
<td><strong>Postgame</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive experience</td>
<td>3.28 (0.08)</td>
<td>3.22 (0.25)</td>
<td>.92</td>
</tr>
<tr>
<td>Negative experience</td>
<td>0.67 (0.00)</td>
<td>0.74 (0.22)</td>
<td>.29</td>
</tr>
<tr>
<td>Tiredness</td>
<td>0.00 (0.00)</td>
<td>0.00 (0.00)</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>Returning to reality</td>
<td>1.07 (0.41)</td>
<td>1.04 (0.26)</td>
<td>.45</td>
</tr>
</tbody>
</table>

\( ^a \)Scores can range from 0 to 4.

\( ^b \)Mann–Whitney \( U \) test with signed intragroup comparisons.
Discussion

Principal Findings

In this study, either playing alone or with peers, older participants showed significant improvement from pretest to posttest, reaffirming that the practice of exergame is a therapeutic instrument to improve the functional capacity of older people [45]. Nevertheless, the results also suggest that even better outcomes in individuals’ functional capacity and adherence can arise from group-based exergaming, which is similar to the outcomes of group-based conventional exercise [46-48]. Technology is an innovative method to assist changes associated with aging. Supporting older people to participate in physical activity through evaluations offers the added benefit of tracking user activity by monitoring health events and behaviors in real time [49].

Other studies [23,25] suggest that, when comparing traditional exercise with exergaming, both modalities showed more successful results when performed in groups, which can significantly promote social interaction [37]. However, similar experiments [50] have revealed differing outcomes regarding exergaming alone or with peers. Thus, this type of finding deserves further investigation. In this study, we assumed that the participatory co-design process was a key determinant. Notwithstanding, participation in exergames can improve cognitive and physical skills that are directly involved in the functional abilities needed by older people in daily life [51].

Understanding the opportunities, challenges, and training processes that optimize the benefits of exergames has important implications for improving the quality of life and longer-term independence for older people [52]. According to Meekes and Stanmore [27], comprehending the motivation of older people to use exergames can help in the game development process. Thus, throughout the co-design and evaluation period of this study, users were part of the creation process. The contribution that the participant brings to the development of a particular product is crucial to its success, as, by participating in this process, users can see their suggestions being implemented before the product is completed, feeling valued during this process [53,54].

According to Zhang et al [55], the study of familiarity issues among older people is necessary to guide the interaction design of new technologies. Thus, they can be built upon the prior knowledge of older people on real-world interactions, making it easier to apply their existing knowledge and skills to a new domain. Zhang et al [55] also explained that 3 specific design aspects should be followed: representation as similar as possible to the real world to help older people understand the game; manipulation, which indicates how older people manipulate the objects in the exergame; and meaningful design, which includes the stimulation of the older people’s memories and emotions. From a game design perspective, to increase the preference of older people’s attitude toward exergames, it is necessary to consider games with more attractive and comprehensible elements so that the older people can enjoy playing [50,56]. In this sense, a co-design creation process can lead to tailormade challenges that are optimized to the participant’s own preferred level of physical exertion.

In this study, older people were motivated to cooperate directly and indirectly in the game development, similar to what was reported by Chen et al [57] and Kappen et al [58]. We observed that all participants answered no to all statements about barriers that were directly related to the game itself. In addition, virtually all participants responded positively to all statements about motivators (241 of 247 individual responses). Furthermore, the participants’ positive experience and interest in their co-designed exergames were evidenced by high scores for complementary components and low scores for derogatory components of the GEQ results.

Other studies highlighted the importance of including the user in the process of development, especially when considering older people [59]. Brox et al [60] reported their results from 3 years of research in the field, providing a user-centered protocol for exergames adapted to the needs of older people. According to the authors, the time devoted to the older people was the main element essential to establish communication and to earn the trust of each one of them. In addition, all types of game elements should be considered during the exergame project, even if they seem too obvious, and particular needs of this target audience must be met during the design process to make exergames more effective [61]. Considering the development of exergames for older people, Chen et al [57] state that playfulness and perceived utility are 2 of the main aspects that affect behavior and the intent to use. According to their study, older people’s acceptance of exergames depends not so much on the fun when playing but on the perceived usefulness of physical and cognitive skills employed on it.

According to Bird et al [62] and Tobaigy et al [63], older people often lack knowledge and perception about how to interact with exergames and how their use can improve health care, until they are exposed to them. In this study, users began to feel comfortable with the game as the sessions progressed, gradually becoming familiar with the way of playing and identifying the benefits of exergaming. The results showed that the interaction with the exergames provided a friendly environment for social interaction, fun, and rehabilitation of the older people in both the groups. When assessing adherence to physical exercise from the motivators, we observed 2 statistically more positive answers in the with peers group to the statements about playing alongside other people. Social aspect was one of the main motivators that led older people to interact with exergames, as mentioned in other studies [10,20,25]. In this context, socializing refers to the interaction between participants, that is, the exchange of experiences facing the same challenge together and competition [64]. Older people can regularly play exergames not only to enhance their health but also to take advantage of the opportunities to socialize with others, which further increases their emotional well-being [50].

Most participants’ experience with the exergames was considered positive, however, the sensory and imaginative immersion achieved a significantly better result for the with peers group, suggesting that the participants in this group became more interested in the game’s story and found it...
impressive. The fact that playing with a peer is very similar to a real bowling game where participants compete with each other may have contributed to this result, as the exergame may have an effect similar to the conventional bowling game [65]. Usually, exergames use an avatar to represent a player in the virtual world. Studies suggest that the avatar’s appearance may be a key factor that influences player behavior [66]. When choosing an avatar reflecting themselves, players have higher perceived game interactivity. This can be a powerful motivator, leading to higher engagement during gameplay. In this study, female participants reported that they felt more attached to the game when a female character was available to play. In addition, participants felt more captivated by the game during the final sessions when improved characters and new elements were added.

**Limitations**

Some limitations of this study include the fact that we only explored competition instead of collaboration among the user participants and the fact that the study had a small sample size because of dropouts, which makes it difficult to generalize the results. With regard to the study design, nonblinding may have negatively influenced the findings, although the groups were homogeneous in age and functional capacity. We had to rely on the participants’ claims about not performing other physical exercises throughout the study, thus avoiding confounders. In addition, the bowling game has specific movements; other types of games may use different ones and therefore may not elicit the same outcomes. During the sessions, participants kept interacting with the game even when it was not completely ready in its final version. If participants had only interacted with the final version of the game, the results might have been different. In addition, participants attended the same club and became even more familiar with their peers throughout the sessions. In this sense, playing with stranger peers or in a smaller number of sessions would not necessarily lead to the same findings. Only older people who already practiced some physical activity participated in the experiment; therefore, the results do not necessarily apply in an equivalent way to older people with sedentary habits.

**Conclusions**

Exergames as motivators of physical activity have been increasingly seen as promising and encouraging tools. The findings of this study suggest that exergaming, either alone or with peers, leads to a statistically significant improvement in functional capacity. However, evidence suggests that an exergame is more attractive and successful when participants play it with peers. With regard to adherence to physical exercise, both groups showed interest in playing alongside other people. Another contribution of this study was the creation of the exergame Boliche Virtual. Our approach offered a collective experience with the use of technology, placing the user at the center of the game design process. This approach may have influenced the motivation and adherence of participants and, consequently, elicited better health outcomes. Therefore, co-design and testing using a participatory and iterative process might foster positive results.

More research applying this design approach is needed to understand the opportunities, challenges, and training processes that can optimize exergames to improve the quality of life and long-term independence of older people. In future works, we intend to perform new tests with a larger number of participants. Other aspects related to player experience can be assessed, such as decision making. We also suggest the long-term monitoring of the participants to verify the maintenance of results and to further investigate the differences in exergaming with peers.

**Acknowledgments**

The authors are grateful to the Caixeiral Campestre Club of Passo Fundo for their support. The authors acknowledge the Brazilian National Council for Scientific and Technological Development for the fellowships. The authors are solely responsible for designing and performing this study and all its analyses, for drafting and editing the manuscript, and for the final contents.

**Authors’ Contributions**

JJ, DB, and EB share first authorship, conducted major parts of the methods and experiment design, and contributed to the majority of the writing and reviewing of the manuscript. OB contributed to developing the software and drafting part of the manuscript. LD and AP provided methodological support and conducted the statistical analysis. HF provided methodological suggestions and critical review for the text. AM supervised all stages of this study, including conceptualization, methodology, validation, and text review.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1

Screenshots and details of the bowling exergame’s final version.

[PDF File (Adobe PDF File), 1349 KB - games_v9i1e23423_app1.pdf ]

**References**

http://games.jmir.org/2021/1/e23423/


40. Rikli RE, Jones CJ. Development and Validation of a Functional Fitness Test for Community-Residing Older Adults. Journal of Aging and Physical Activity 1999 Apr;7(2):129-161 [FREE Full text] [doi: 10.1123/japa.7.2.129]

41. Picorelli AMA, Pereira DS, Feliciano DC, Gomes DA, Dias RC, Pereira LSM. Adhesion of older women to a home exercise program post ambulatory training. Fisioterapia e Pesquisa 2015;22(3):291-308 [FREE Full text]


Abbreviations

3D: 3-dimensional
GEQ: Game Experience Questionnaire
GPFI: General Physical Fitness Index
Utilizing Theories and Evaluation in Digital Gaming Interventions to Increase Human Papillomavirus Vaccination Among Young Males: Qualitative Study

Gabrielle Darville1*, MPH, PhD, CHES; Jade Burns2*, PhD; Tanaka Chavunduka2*, MPH; Charkarra Anderson-Lewis3*, MPH, PhD

1Department of Public Health, College of Health Professions, Mercer University, Macon, GA, United States
2Department of Health Behavior and Biological Sciences, Center for Sexuality & Health Disparities, University of Michigan School of Nursing, Ann Arbor, MI, United States
3School of Health Professions, College of Nursing and Health Professions, The University of Southern Mississippi, Hattiesburg, MS, United States

*all authors contributed equally

Corresponding Author:
Gabrielle Darville, MPH, PhD, CHES
Department of Public Health
College of Health Professions
Mercer University
1501 Mercer University Drive
Macon, GA, 31207
United States
Phone: 1 478 301 5129
Email: darville.gc@mercer.edu

Abstract

Background: Human papillomavirus (HPV) is the most common sexually transmitted infection in the United States. HPV attributes to most cancers including anal, oral, cervical, and penile. Despite infection rates in the United States, recommendations and communication campaigns have traditionally focused on females. Because of this, males lack knowledge about reasons for vaccination, the benefits of being vaccinated, and their HPV risk, overall. Gaming as a health education strategy can be beneficial as mechanism that can promote behavior change for this key demographic because of the popularity of gaming.

Objective: We sought to explore the relationship between gamification and HPV vaccine uptake.

Methods: Interviews were conducted with experts (n=22) in the fields of cancer prevention, sexual and reproductive health, public health, game design, technology, and health communication on how a game should be developed to increase HPV vaccination rates among males.

Results: Overwhelmingly, theoretical models such as the health belief model were identified with key constructs such as self-efficacy and risk perception. Experts also suggested using intervention mapping and logic models as planning tools for health promotion interventions utilizing a digital game as a medium. In game and out of game measures were discussed as assessments for quality and impact by our expert panel.

Conclusions: This study shows that interventions should focus on whether greater utilization of serious games, and the incorporation of theory and standardized methods, can encourage young men to get vaccinated and to complete the series of HPV vaccinations.

(JMIR Serious Games 2021;9(1):e21303) doi:10.2196/21303

KEYWORDS
digital games; behavior change; theory; evaluation; game design; health care providers
Introduction

According to the Centers for Disease Control and Prevention, human papillomavirus (HPV) is one of the most common STIs among adolescents and young adults [1]. Approximately 14 million new cases of HPV occur in the United States each year, with nearly 80 million people estimated to be currently infected [1,2]. An equal opportunity virus affecting both men and women, HPV can be spread easily during skin-to-skin contact during anal, oral, or vaginal sex [2,3]. Routine vaccination starting in early adolescence (aged 11 or 12 years) can prevent common HPV infections and their associated diseases, such as oropharyngeal and cervical cancers [1,3,4]. In comparison to females, males are less likely to be partially or completely vaccinated [3,5,6]. While there are 3 approved vaccines for use in the United States that are covered by insurance, vaccination rates remain low among young men [1,6].

The HPV vaccine has been available to males since 2009, and in 2019, the Advisory Committee on Immunization Practices recommended catch-up vaccinations among this population, with certain special populations being immunized up to age 26 years [1,5]. Yet, recent statistics have shown that because of the lack of or delayed symptoms with HPV-associated diseases, adolescent and young adult males are at a higher risk for certain HPV strains than adolescent and young adult females [3,4,7]. To date, there is no approved HPV test for men. Furthermore, routine screening is not recommended by the Centers for Disease Control and Prevention for HPV-related diseases that may be associated with anal, penile, or oropharyngeal cancers [3]. According to one major study [5], males lacked knowledge about HPV and perceived no benefit to getting the vaccine. In this same study, they were also unaware that they could be given the vaccine and did not receive any HPV-related recommendations by their health care providers [5]. Additionally, socioeconomic and behavioral factors, including substance abuse, racial and ethnic disparities, sexuality, distrust of the health care system, stigma, and lack of access to health care have increased the spread of the STI [5,8]. With that being said, innovative strategies need to be employed to facilitate the uptake of this vaccine.

Access to digital materials has been shown to be a highly effective method of receiving, retaining, understanding, and actionizing health information [9]. The use of interactive games (also referred to as serious games) has increased in popularity among health professionals as a viable educational tool and behavior intervention strategy for addressing health disparities. Gamification is the use of digital game mechanisms in a nongaming context to engage users, motivate activities, and solve problems. Gamification can prove to be better than traditional programs and interventions because digital games provide immediate feedback when the player takes an action or selects an option within the game. Because of this feedback (negative or positive), players are able to correct their decisions or make changes to their strategy. This feedback is beneficial to the user because it allows for increased cognitive processing and experiential learning [10,11]. When designing a serious game, gaming components should be based on behavioral theories that are capable of measuring its impact and effectiveness on individual-level determinants [12]. Interactive digital interventions including health promotion games that are based in theory, such as the self-determination theory [13,14], and dual-processing models of cognition, were effective in educating and reducing the risk of STIs including HPV [15].

While gaming has been utilized in many domains, very few have explored gaming for vaccination uptake within public health. Furthermore, of the games developed around the influenza and general vaccinations, they have been widely researched as preventive and care engagement tools. In a systematic review [16], published in 2016, that highlighted these games, none utilized behavioral change theory. Additionally, they lacked specific evaluation methodologies and long-term assessments of the games [16]. To our current knowledge, there is only one interactive game that exclusively explores improving HPV vaccination for boys and girls aged 11 to 12 years [13]. In a 2019 study [17] conducted in Norway, researchers developed and released a mobile HPV game [17]. However, this game was not theory-based and is only meant to increase awareness and knowledge of HPV. A game that is developed as a health intervention must have the ability to be appropriately evaluated and engaging [18]. In the absence of standardized evaluation methods for serious games, design principles (ie, usability, playability, and visualization) are incorporated into designing a gameplay experience that does not distract the user from the intended outcomes. This must be simultaneously balanced with meeting user expectations of a gaming experience that combines challenge, fantasy, and curiosity [19]. Recent work has begun to explore digital gaming to improve HPV and HPV-related vaccine outcomes among young men [20], but to date, little attention has been focused on designing and evaluating an evidence-based serious game targeting HPV vaccination tailored for young men.

This study seeks to explore the relationship between gamification and HPV vaccine uptake. With this approach, we will examine (1) theoretical models that would be most useful in game design to increase HPV vaccination rates in young males aged 18 to 26 years and (2) explore evaluation study designs for digital-based gaming. As an area with limited research, digital games that incorporate theory and integrate evaluation may have the potential to normalize perception of sexual health and STI prevention and to improve vaccination rates among this population [8].

Methods

Participants

This study was approved by the University of Florida’s institutional review board. Experts were recruited from the fields of sexual and reproductive health, cancer prevention, public health, game design, and technology and health communication to collect qualitative data. Because the topics of HPV and gaming are interdisciplinarily in nature, we aimed to get at least 2 experts from each of the fields mentioned above to participate in the interviews. Besides the disciplines within which experts were employed, sampling was not done based on any other variables. A nonrandom purpose sampling approach was implemented to ensure heterogeneity with respect to expert...
backgrounds. Experts were initially identified based upon their publication record, word of mouth recommendations, and body of work or experience. Initially, a total of 18 experts were contacted and invited to participate in the interviews. All 18 experts accepted the invitation to participate in the study. A snowball method of interviewee identification and selection was also employed to recruit additional participants. At the end of each interview, experts were asked whom else the research team should contact to participate in the study. An additional 6 experts were referred and invited to participate based on their relevant areas of expertise. One of the referred experts declined to participate.

**Procedure**

The purpose of the interviews was to elicit in-depth input and recommendations to explore the current state of games for health research and practice and strategies for using games as a means to promote and improve HPV vaccination rates among college-age men. In-depth interviews were carried out with the identified experts who were contacted via email and invited to participate in the study. If the response was positive, a telephone or in-person interview was scheduled by research staff.

The interviewee was provided with a copy of the interview guide (see Textbox 1) and informed consent form prior to their interview in an effort to help them prepare. Interviews were conducted using open-ended questions. All interviews were audiotaped, and each participant was made aware of the audiotaping when they gave their verbal or written consent to participate in the study. Interviews lasted an average of 80 minutes and were conducted by the same member of the research team in a location conducive to private conversation. Interviewees were not compensated for their time but were very willing to participate due to their interest in the purpose of the interview. Because the experts had participated in interviews before, combined with the training of interviewers, all interviews were conducted in one round and sufficient detail was provided. Interviews were carried out by the first and fourth authors who were trained in qualitative research inquiry.

**Textbox 1. Expert interview questions guide.**

1. What’s your experience in HPV research or just HPV?
2. Have you had any experience in game development design and/or testing?
3. Would you recommend someone using a digital game to improve sexual health outcomes? And why or why not?
4. What do you think are some of the benefits of using digital games for sexual health?
5. What do you think are some of the challenges of using digital games for sexual health?
6. Do you think that college men would be receptive to using a digital game about the virus and about the vaccine? And why or why not?
7. When you think of a game, a digital game aimed at increasing in particular risk perception of the HPV virus and vaccine uptake, what health messages do you think would be most beneficial for college aged males and why?
8. What features or characters should be included in the game for college age men to increase their risk perception for the virus?
9. What features or characteristics do you think would be more instrumental for increasing risk perception?
   - What game mechanics would be beneficial for increasing risk perception?
10. What features or characteristics do you think would be more instrumental for increasing HPV vaccine uptake?
    - What game mechanics would be beneficial for increasing HPV vaccine uptake?
11. If you were to design a game on the vaccine itself how would you foresee that being done? Specifically the development and designing of the game?
12. What are the most relevant theories, models, or evaluation strategies for HPV vaccine programs targeting college age students?
13. How should we evaluate the game for quality and impact?
14. How would you define success in a game for health behavior change related to the HPV vaccine?
15. Are there any other advice or recommendation would you give to someone developing a game for HPV for college age men?

**Data Analysis**

Interviews were transcribed verbatim by trained transcriptionists. Transcripts were organized and coded across all 22 interviews. Using grounded theory, emergent theory were identified and reported [21]. To establish reliability, 2 independent researchers individually coded each transcript. Coding was conducted in rounds and researchers coded 5% of each transcripts then met to discuss the identified codes. By the second round all differences were discussed, and a final coding scheme was established. This approach was continued until all of the transcripts were completed. After coding, the data were sorted by themes and subthemes with accompanying quotes for clarity and organization. A matrix was used to identify patterns within the data to be explored.

**Results**

**Overview**

The results reported in this study reflect only a portion of the data collected from the study. Specifically, results discuss the
use of theory, program planning models or approaches, and evaluation strategies that should be employed in game design and development pertaining to the HPV prevention and uptake of the HPV vaccine for college-age men.

Theories and Theoretical Constructs
The experts who were interviewed emphasized that any gaming intervention designed for males should incorporate theory. Some theories that were mentioned in the interviews included social progress theory, social cognitive theory, normative theory, psychosocial theories, elaboration likelihood model, theory of reaction, theory of reasoned action/planned behavior, lifestyle risk deduction model, theory of gender and power, transtheoretical model (stages of change), health belief model, extended parallel process model, theory of triadic influence, integrative behavioral model, diffusion of innovation, and Anderson health care utilization model. Although interviewees either worked in research, government, or academia, collectively, they all encouraged the use of behavioral theories. In particular theories that focused on increasing a male’s perception of benefits and addressing those key barriers that could prohibit HPV vaccine uptake were highly regarded. This resulted in the health belief model being highlighted frequently.

Of the behavioral theories, constructs most mentioned being instrumental to the intervention design included perceived susceptibility, risk perception, and self-efficacy (see Table 1). In fact, 17 of the 22 interviewees mentioned self-efficacy and highlighted self-efficacy as one of the predictors of behavior change. In addition to HPV vaccine self-efficacy, consideration of gaming self-efficacy was encouraged. While men of all ages are the highest gamers, gaming self-efficacy is predicated on making sure that whatever is designed is easy enough that people with that confidence voluntarily engage in gameplay and continue this behavior [22]. Apart from self-efficacy, risk perception was the second-most mentioned theoretical construct in the qualitative interviews. While risk perception is instrumental to the uptake of the game (ie, feeling though you are at risk which encourages you to play the game to learn how to reduce your risk), interviewees felt as though highlighting that risk within the game and engaging within game risk reduction activities would be most influential to end game and out of game behavior. Overall, the two most mentioned theoretical constructs (self-efficacy and risk perception), were both components of the most frequently cited theory (health belief model) in this particular study. Thus, strengthening the health belief model’s appeal as the theoretical foundation that should be used when developing a digital gaming intervention around HPV and the HPV vaccine.
Program Planning Models or Approaches

In addition to theory, interviewees mentioned the role and impact of using program planning tools, approaches, and models to develop a gaming intervention for males. Items that emerged during the one-on-one interviews included intervention mapping and logic models. One interviewee preferred the intervention mapping approach because this model goes step-by-step thorough how to design an intervention program, which can also be applied to the game design and development process. Intervention mapping can also be of great benefit because it can help developers think of the determinants within the target population that should be considered and help the game be more effective. Social marketing was also mentioned as a possible dissemination tool to encourage game play and promote health among males (see Table 2).

Logic models are also an instrumental tool to program and intervention development in the fields of public health, health education, and health promotion. While it is an approach that is applied more in the traditional setting, it can be easily adapted to a virtual or technological setting. As discussed by the experts, logic models can provide a snapshot view of the entire intervention. This can include the processes that should be done beforehand (prior to game design) such as formative research and usability or beta testing, the strategies (game mechanics) that should be integrated into the game, the activities that should encourage game play and then the short-, intermediate-, and long-term outcomes that we would like to see to indicate success after the intervention is done.

Table 1. Theory or theoretical construct quotations.

<table>
<thead>
<tr>
<th>Frames</th>
<th>Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health belief model</td>
<td>&quot;In this age group, I think—and I’ve not done work with this age group, its speculation—but based on different behavioral change theories, you know, breaking down some of those barriers, 'Where can they get the vaccine?'; 'Is there a cost attached to it?'; 'What are the benefits to them, specifically?'; 'What are the benefits to their relationships, that they care about?'; 'What are the potential side effects to the vaccine' and being able to sort of weigh those against the majority of the benefits that they would experience&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;To implement your program more than anything. But if you're talking, that's not, that's not really a theory so if you're really talking about actual theories, I mean I would say more the Health Belief Model. Cause I know it was developed to...to explain vaccinations&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;Well, health belief model is the one that’s used a lot in sexual health because it looks at perceived susceptibility of a disease or other threat and then kind of cues to action and perceived benefits&quot;</td>
</tr>
<tr>
<td></td>
<td>&quot;I mean cause like for example, you know, you could say well people may not know where to go to get the HPV vaccination. They may think that they may have to go to the health department, they may not have health insurance... it could be a situation they could get it for free like those sorts of things. So, applying each individual... concept of the Health Belief Model, I think would be, would be really good. I think there is a very tangible, applicable, applicable way to do it”</td>
</tr>
<tr>
<td>Self-efficacy</td>
<td>&quot;But you know the game while it can make them behave in a certain way, which is cool, because that is one of the advantages of the game, but it can translate to reality, we only really are targeting the cognitive things, so even when I do it I’m not only playing the game but I’m potentially changing my cognition, my behavioral capability, my self-efficacy, my knowledge and maybe that’s going to change my intention of behavior”</td>
</tr>
<tr>
<td></td>
<td>&quot;That’s a lot more interesting, and from a theory perspective, I can bide with that that’s building self-efficacy because you’re engaging in the behavior, even if it’s simulated.”</td>
</tr>
<tr>
<td></td>
<td>&quot;So, you’ve, you know you’re going to have to apply methods that will boost that self-efficacy. So, you are probably going to be using things like persuasion, modelling, um successive approximation to a goal maybe. So that makes me think about simulations, when you’re thinking about the group because they could take an avatar and take them through what they think is a risk reduction behavior and see the results or not.”</td>
</tr>
<tr>
<td>Risk perception</td>
<td>&quot;You know so I think, things about genital warts, probably would be much more realistic, um, you know, improving risk perception and severity, you know, that’s where a game I think will have um, a big impact, because its um, especially if it was the simulation where um, you know like that whatever plague is that they spread in World of Warcraft, similar to that, like how easy it is to be unaffected, that content takes only one time, you know colleges full of all kinds of risk behaviors as people figure out their identity and those coupled with drinking and things like that, and the fact that this is an STD does not necessarily get protection from using condoms”</td>
</tr>
<tr>
<td></td>
<td>&quot;Well I mean if it’s risk perception is one. Low levels of risk perception is one of the determinants that are negatively influencing vaccine uptake. Then I think messages and of things like you know, 'Anyone can get it,' and maybe characters that have similar attitude of, 'Yeah I’m not—that’s not going to happen to me,’ and you know it spreads the misconceptions that then are addressed and I think that those kinds of things can help. Also, you know increasing information about risk, about the common-ness of the infection”</td>
</tr>
</tbody>
</table>
Table 2. Program planning models quotations.

<table>
<thead>
<tr>
<th>Frames</th>
<th>Quotations identified</th>
</tr>
</thead>
</table>
| Intervention mapping | “So again I’m off topic, but I think intervention mapping is my recommendation because that’s what we used here and I think it covers the basis and you can use it to the degree you want to use it in terms of the depth. But it basically says find out about the problem, come up with a concept of the solution, build a prototype to that, fix it and design, and figure out how you are going to disseminate it as well.”  
  
  “We usually, we always use an intervention map which is a—which is an approach to develop things any kind of intervention but it really helps you think through the, you know, the-the desire to change in whatever determinants and then make decisions about what are the message and the practical applications that could influence change in the determinant. So, so a systematic process for making decisions about the approach, I think, is critical.”  
  
  “Yeah I think that from a generic perspective I would probably recommend a process such as intervention or mapping. You’ve heard about intervention mapping......Yeah, so that’s really fairly well known. It’s a pretty standard approach to develop an intervention. What it offers to this particular behavioral science and health education is it helps with determining health behavior. So, if your game is designed to encourage risk reduction behavior um, let’s say, um then you will, you will expect to delineate the specific behaviors to, delineate the specific determinants like a social determinant of those behaviors. So you are talking about self-adequacy expectations, norms, perceived norms, whatever is important for those in your particular target population and it will sort of force you to really understand the nature of the problem and, the nature of the best delivery system”  
  
| Logic models     | “So, I would do that first and then once we have that, what we do in this process that I mentioned is that we developed, you know, a logical model of change. And in that logical model of change you have the behavior that you want which in this case is probably vaccine uptake and completion of the, of the theory.”                                                                                                                                                                                                                          |

**Evaluation for Quality and Impact**

As with any health promotion program or intervention, it is important to evaluate it to determine its effectiveness. However, successful evaluation relies heavily on well thought out data collection and assessment activities. Because game design can be an expensive and lengthy process, our experts advised conducting alpha or usability testing to ensure that it captures what it needs to prior to final data collection and assessment. Apart from testing the gaming mechanics and features prior to the final development stage, they also encouraged thinking about the measures that are to be collected. More importantly developers should determine if the final measure will be collected as a pre- and post gaming intervention variable external to the game; or whether there a way to use built-in features within the game to collect data points. Measures for quality and impact discussed in the interviews included feedback from the players, message reception or acceptance, dissemination to others, social impact, and increased engagement or retention within the game. However, the greatest predictors of success, highlighted by experts, for a male gaming intervention included changes in knowledge and awareness, changes in behavior, and gaming to vaccine correlation (see Table 3). Ultimately, all interviewees who participated in our study highlighted the need to conduct process, impact, and outcome evaluations.
“So I think, yeah I would tell people about this game or I would play this with my friends or, I believe this game is as good as other games that are related to health or how does this game compare to the games you play? Is it better or worse? More fun less fun? These are questions that we typically ask in usability but you can ask those type of questions. One of the questions I really like is ‘how could we make this game better?’ And the reason I like it is that it is positively oriented, its um, it’s got an altruistic style to it and it really find out what sucks in the game without them having to say it, so it’s really acknowledging the short fall of the game. So, it’s like they can say I would really change this up and blah, instead of them having to endure the sort of social desirability problem of telling you this thing sucks”

“Usability is giving it your user, a new player and asking the same questions. You can do usability by doing that. You could do it with a survey. You can also do talk aloud usability test. You get them to play it and tell you what is going through their mind so um well, I’m going to pick the pink button because they think that if you know that’s going to lead to a better result, I’m going to choose the middle door on my quest you know, why, I’m going to choose this avatar, why? We did our Asthma simulation, we found that the young males all gravitated to this older Hispanic woman who had this lighter color hair because she was just perceived as more attractive. So, you learn a bit about how you are presenting everything with that approach. So, then you may tweak it or you may think it’s ready for a trial and you can apply it. You can get enough samples, that would give you the power to check the outcome and you can write a pre-post that randomize the laboratory conditions”

Let’s say it’s a simple card game and so now we have all these cards and all these rules, we sort of think how are we going to play it and we play it amongst ourselves and we play it till we thin kits cool and we thin kits a good one, so let’s now take it to some of our target population and see it too and check their responses. We haven’t programmed something yet, we are still working with the concept and the rules versus a table game. So that is proof of the test of the concept and it’s an acceptability test and a usability test. But, what we’re asking, what we have the opportunity to look at is, questions that are looking for evidence like, ease of use, ease of play, um, understandability, acceptability, of this a s a game, um, credibility, I think the information here is trustworthy. The perceived value um so I think the information I am getting here will help me make better decisions in the future.”

“Um, so I think that part of it is super helpful, I mean, I like love self-measurement projects where its like already built in um, I mean I’m so, but, here’s something, okay. It’s actually something we are doing with google right now. It’s really cool. Um, we have all these digital ads that are like measuring behavior physicians. So we know predominantly what websites these physicians frequent, you have our ads on there, they are exposed to various kind of ads banner ads, PSAs, maybe a 6 second pre-roll that they can continue watching or mandatory 30 second yadda yadda. And we’re actually able to use IP address chasing to determine if they have taken an action after seeing this. And um, those who have watched our 30 second video are 300 times more likely to then search around HPV vaccines and vaccinations”

“I don’t know how technology works but I think there is ways to do that relatively easily. Um you could say on some level just, you know one measure of success is that engaging and do they play it. And the other is does it actually have the desired effects. Maybe it has undesired effects like you wouldn’t want to measure but behavior changes are often what the ultimate end goal but sometimes were harder to evaluate and practice”

“if there’s a point in the game on where to get the vaccine you can track to see how many people get to that point. If they get to the location maybe there’s something on that end that they could verify, sort of prize or coupon or something like that. Obviously if you build in social media then you can track those analytics as well. I think you’d have a lot of data now. You can quantify that to success”

“It just depending on what you want your audience to do. Do you want people to remember messages? So maybe you could build in a measure for recall. Send them an email in a month or something and give them a short quiz and see if they remember that stuff.”

“Pre and post the game playing. You could do behavioral stuff and then even long-term behavioral of your same cohort”

“And I would think, that doing the same kind of measurement, that this would have to be a planned evaluation project from the outset where we are going to enroll 200 young men or whatever the number is and we are going to have them play this game, and we are going to naturally allow them to accept or not accept the HPV vaccine, we’re going to find out what their rates were before and after that. We are going to pick HPV vaccine naive men, so that we know how many get vaccinated or there could easily be a survey component in it too, because I think it’s more than just one person being vaccinated”

“Like...this is not the best evaluation but like pre-your game what percent of the students or whatever vaccinated and then after. Did you change it? You could even do like demand at the clinic”

“So, if you’re going to do this on college campus, maybe look at the pre vaccinations levels the university implements freshman orientation than game and then you look at the end of the year you look at the vaccination levels. Or the game can be more specific, questions for user have you gotten this pre vaccination? Maybe you can throw away a pop up that they can click on to get the vaccine”
Discussion

General

In this study, we sought to examine the theoretical models that would be the most useful in game design for increasing HPV vaccination among adolescent and young adult males. We also explored an evaluation of study design topics (for creating digital games). Our major findings included the use of the health belief model and normative theory for game design. The major constructs that we found included perceived susceptibility and risk perception. Specifically, experts in sexual health, cancer prevention, health and communication, college health and wellness, and game development all found self-efficacy to be the primary construct for predicting behavior change. Other significant findings included the use of logic models and intervention mapping as program-planning tools and for evaluating effectiveness, measuring quality and impact, and conducting usability tests, all of which were essential in the front end of game design for this population.

Digital games that incorporate theories and integrate evaluation into their design can improve HPV vaccination among young males. Similar games for health that have used theoretical foundations have shown to be successful with changing behavioral outcomes. In a systematic review [23] conducted in 2015 on the development of digital games, of the 54 games were designed for health promotion, 7 focused on sexual health outcomes. Of the 7 games, 5 incorporated theoretical constructs in the design of the game and as outcome measures, and 2 specifically measured the effects of the games on behavior. While the sample was small, effect sizes were positive and significant for knowledge, self-efficacy and behavioral intention; which has great implications for overall behavior change [23].

While the focus of this study was college-age males who were not previously vaccination at 11 to 13 years old due to varying and outdated vaccine recommendations, the implications of this study can have greater impact upon younger adolescent age males. Benefits are especially prominent if implemented and used prior to sexual activity initiation to offer full protection against HPV-related cancers. Previous research has shown that theoretical models such as the health belief model show promise as theoretical frameworks for understanding for vaccine uptake among young adult males living in urban settings [24]. Our study clarifies the broader relationships between multiple theories and constructs (including the health belief model) that can be used to measure the impact of gaming on individuals, particularly for educating this population on the importance of HPV [18].

While there are several models that can be employed, the health belief model and its varying constructs have shown to be the best choice because of its proven application. Developed in the 1950s to explain why people did or did not participate in programs to prevent disease and illness, the health belief model is one of the most known and widely applied theories [25]. Specifically, in public health, it has been proven effective in behavioral interventions focused in areas such as health screenings, immunizations, vaccines, sexually transmitted infections, and cancer prevention [25].

In order to predict behavior change, the health belief model takes into account the perceived severity of the disease or illness, susceptibility to the disease or illness, benefits and barriers to performing a new behavior, and cues to action (motivators) to changing behavior. Often, perceived severity and perceived susceptibility are collectively referred to as perceived threat or perceived risk. Self-efficacy was later added to the theory, because the importance of one’s confidence in their ability to change their individual behavior was identified as an important component to behavior change as well [26]. Therefore, when developing a digital game to increase vaccine uptake, one must take into consideration the self-efficacy of the players. Self-efficacy, perceived susceptibility to HPV, and extensive exposure to HPV-related information have been shown in previous studies to predict vaccine acceptance and uptake by young males [24,27]. Self-efficacy is an essential construct, as it regulates several cognitive processes for developing and maintaining behaviors [28]. These include confidence, control, and success with specific tasks [28]. In this case, the goal is to design a game that educates and creates an experience for an individual involving HPV vaccination and sexual health promotion using the identified theories and constructs. The aim is to develop behavioral skills that will promote consistent vaccination among this population.

This study also showed the importance of evaluating game design for quality, impact, and effectiveness. Logic models are essential because they provide an organized schematic depiction of the way theory and intervention intersect, and they can be useful for understanding how outcomes are produced for a given project [29]. Intervention mapping may also be useful in designing a game to promote HPV vaccination, as it is a systematic way of looking at a health problem, design materials, and protocols based on theory and practice [30]. Historically, beta testing a game for playability, usability, visualization, effectiveness, and quality has been essential to the evaluation of any serious game [19]. In all, these data suggest that understanding the purpose of theory and evaluation may have a profound impact on the quality of the game design and on what individuals learn about HPV, sexual health, and STI prevention.

The limitations of this study include the interviewees personal biases as qualitative researchers. Qualitative research can depend heavily on the skill of the individual researcher and is thus easily influenced by researchers’ values and opinions. [31] Another bias could have emerged through the use of purposive and snowball sampling to recruit study participants. This type of nonprobability sampling often relies on the judgement and discretion of the researcher making it subjective in nature. It is also argued that this type of sampling technique is not truly representative of the entire population [32]. However because of the interdisciplinary nature of this project and specificity of the results targeting to a specific population that the methodology employed is not as much of a limitation overall. In fact, reviewing the data with peers and experts in the field can help one maintain objectivity in the data-analysis process. Another limitation is that the qualitative data gathered here are a subset of a broader data set. The overall volume of qualitative
data can affect the analysis and the interpretation of the findings [33,34].

Future Implications

Although a large majority of interventions and programs are implemented in traditional health education settings, health care providers and nurse practitioners are becoming increasingly vital to increasing vaccination rates among adolescent and young adults [35,36]. Most often, these programs or approaches are implemented in health care settings and school-based programs. In clinical settings recommendations by health care providers are cited as the primary motivator for HPV vaccination of children by parents [37]. Therefore, there has been great discussion and emphasis on the nurse’s role as a champion in promoting the HPV vaccine especially since they interact more with families during in clinic visits [35]. Alternatively, in a 2014 study [36], it was identified that public health nurses used a variety of strategies to increase the rate of HPV vaccination in schools. These included but were not limited to providing HPV health education sessions alongside an informative package, question and answer sessions, flexible appointments, the inclusion of males in health promotion activities, and communication through a variety of media formats such as the school newsletter and website [36].

Since the COVID-19 pandemic has affected medical visits and school-based interactions (because of social distancing, stay at home orders, and governmental bans), we have seen the number of vaccinations, including HPV vaccinations, decline dramatically over a short period [38]. Coupled with the already lagging vaccination rates, this is cause for concern [39]. With the increase of telemedicine and telehealth to respond to the need of health service provision, health care providers, public health practitioners, and nurses should consider gamification activities and digital games as a way provide accurate information and initiate behaviors that would lead to HPV vaccine uptake [40]. Consistently using data reduction methods can make the concepts and relationships in the interviews more visible [34]. The limitations notwithstanding, our findings contribute to the games for health literature from an evidence-based perspective. Future research should consider using focus groups with young men to aid in identifying salient content for a serious game intervention. Creating a safe space to educate young men around safe sex practices and vaccination can help to promote discussions around use, health benefits, and concerns about the virus and vaccine.

Conclusion

Research in this area should focus on whether greater utilization of serious games and the incorporation of theory and standardized methods can encourage young men to get vaccinated and complete the series of the HPV vaccinations. In analyzing the effectiveness of a digital game, the focus should be on change in uptake measures that include not only vaccination initiation but 3-dose completion. Moreover, individuals designing health interventions for practice should consider tailoring the design to role models and or opinion leaders within the community. By providing information and focusing on the rewards to vaccination, gaming can create a constructive learning model and make learning personalized, interactive, and fun.

Authors’ Contributions

All authors wrote, reviewed, and approved the final manuscript.

Conflicts of Interest

None declared.

References


Abbreviations

HPV: human papillomavirus
STI: sexually transmitted infection
Perception of Game-Based Rehabilitation in Upper Limb Prosthetic Training: Survey of Users and Researchers

Christian Alexander Garske¹, MSc; Matthew Dyson¹, PhD; Sigrid Dupan², PhD; Kianoush Nazarpour², PhD

¹Intelligent Sensing Laboratory, School of Engineering, Newcastle University, Newcastle upon Tyne, United Kingdom
²Edinburgh Neuroprosthetics Laboratory, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom

Corresponding Author:
Christian Alexander Garske, MSc
Intelligent Sensing Laboratory
School of Engineering
Newcastle University
Merz Court
Newcastle upon Tyne, NE1 7RU
United Kingdom
Phone: 44 191 20 86682
Email: c.a.garske2@ncl.ac.uk

Abstract

Background: Serious games have been investigated for their use in multiple forms of rehabilitation for decades. The rising trend to use games for physical fitness in more recent years has also provided more options and garnered more interest for their use in physical rehabilitation and motor learning. In this study, we report the results of an opinion survey of serious games in upper limb prosthetic training.

Objective: This study investigates and contrasts the expectations and preferences for game-based prosthetic rehabilitation of people with limb difference and researchers.

Methods: Both participant groups answered open and closed questions as well as a questionnaire to assess their user types. The distribution of the user types was compared with a Pearson chi-square test against a sample population. The data were analyzed using the thematic framework method; answers fell within the themes of usability, training, and game design. Researchers shared their views on current challenges and what could be done to tackle these.

Results: A total of 14 people with limb difference and 12 researchers participated in this survey. The open questions resulted in an overview of the different views on prosthetic training games between the groups. The user types of people with limb difference and researchers were both significantly different from the sample population, with $\chi^2=12.3$ and $\chi^2=26.5$, respectively.

Conclusions: We found that the respondents not only showed a general willingness and tentative optimism toward the topic but also acknowledged hurdles limiting the adoption of these games by both clinics and users. The results indicate a noteworthy difference between researchers and people with limb difference in their game preferences, which could lead to design choices that do not represent the target audience. Furthermore, focus on long-term in-home experiments is expected to shed more light on the validity of games in upper limb prosthetic rehabilitation.

(JMIR Serious Games 2021;9(1):e23710) doi:10.2196/23710

KEYWORDS

upper limb; rehabilitation; arm prosthesis; serious games; engagement; transfer

Introduction

Background

Serious games have been shown to enhance the outcome of movement rehabilitation after stroke or cerebral palsy [1,2]. Computer games have also been researched for upper limb prosthetic control training since the early 1990s [3-14]. Although the field has received considerable attention over recent years, it has yet to succeed in finding general proof for successful transfer of training myoelectric control in the games to an increase in prosthetic control ability [15]. For instance, transfer has only been found in training a strongly activities of daily living (ADL)–relevant task in a virtual environment [15].

http://games.jmir.org/2021/1/e23710/
The skills required to play a game can be learned with practice. A serious game incorporating training exercises consistent with a rehabilitation regimen allows the numerous repetitions of exercises necessary during rehabilitation therapy to take place. Nonetheless, it is hypothesized that the training needs to be engaging to ensure regular and consistent adherence to the exercises [12,16]. Current prosthetic training games revolve around 2 core themes: the engagement of the player [7,8,12,14,16-20] and the skill transfer from the game to prosthetic use [15,20,21]. These 2 objectives are pursued with 2 types of video games that aim to improve myoelectric control (myo-games) [10] in prosthetic training: (1) games that are based on existing game platforms with input adjusted to the electromyogram (EMG) signals [4,6] and (2) completely novel games around specific training goals [9,20].

Few researchers specifically acknowledge the importance of considering the diversity of the target audience when developing games in general [16,17,22,23] and, therefore, myo-games specifically. Even fewer researchers have attempted to address this difference in preferences in their developed games and take the views of people with limb difference into account before starting the development process. A notable exception is the work of Tabor et al [16], who collected qualitative feedback from a small testing group and incorporated the suggestions into their game design. Owing to the diversity in the population, where people have varying definitions of fun aspects and expectations for games, multiple attempts have been made to categorize people into user types [24-26]. Professional game development is based on the psychology behind those types of gamers and the choice of the appropriate game design elements fitting for the target audience. However, academic game development is typically carried out over a relatively short period and by a small, nonspecialist team, which is in stark contrast to the years of development time often carried out by a large and highly specialized team that goes into modern games. This means that the decision making in these games is potentially subject to the preferences of a small team that does not generally reflect the preferences of the target audience.

Objectives

We hypothesized that there are considerable differences between the views of people with limb difference and researchers with regard to engaging aspects of a game. To test this hypothesis, we created a survey and sought to determine the focus points of each of these groups. In addition, we included a user type questionnaire to ascertain the distribution of each participant group for comparison. Using such a user type distribution can deliver useful information to lead general design choices in game development. It could also be used for presets that emphasize certain game design elements over others for increased user engagement. Furthermore, this survey aimed to identify other challenges than potential disparities in game preferences that the community of researchers could have to face on the path of game-based upper limb prosthetic rehabilitation. This study adds the opinions of researchers and people with limb difference about games in upper limb rehabilitation to the research that has been conducted on the opinions of clinicians [27].

Methods

Study Design

The study was approved by the University Ethics Committee of Newcastle University under the reference number 905/2020. The survey was conducted from February 2020 to May 2020. Participants were either people with upper limb difference or researchers who were active in the research of games for prosthetic training or in prosthetic research in general. All participants gave their consent by filling out and submitting the survey as stipulated on the first page of the survey form.

The recruitment of this study was conducted predominantly on the web via personal contacts. Additional outreach was done via social media and by contacting charities in the United Kingdom that are involved with people with limb difference. Specifically, the survey was sent out to the main and local branches of 13 different charities as well as 40 researchers involved in upper limb prosthetic research. The inclusion criterion for people with limb difference was the absence of the upper limb, irrespective of level, side (unilateral or bilateral), or use of a prosthesis. A total of 14 people with limb difference and 12 researchers filled out the survey. An overview of the demographic data of the participants is presented in Tables 1 and 2. We chose a web-based survey and expected that the web-based nature would increase the number of people willing to participate because of ease of access. In addition, the web-based survey offered the participants time to think about their answers without the pressure of coming up with an answer on the spot. The survey was developed in English and was not altered over the course of the study. The participants were given the option to contact the authors if they did not want or were not able to fill out the survey on the web. No participant used this option.

The survey first introduced the general aim of the study and the contact information of the first and the last author and the Data Protection Officer of Newcastle University. The survey asked for general demographics and, in case of people with limb difference, for anamnesis with regard to their limb. This was followed by a user type questionnaire originally developed by Tondello et al [26]. The survey concluded with open questions about the preferences and opinions of participants with regard to games in general and games in prosthetic training specifically. The researchers were asked to answer additional questions concerning the challenges in this field of research.
Table 1. Participants’ demographics.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age group (years)</th>
<th>Type of limb difference</th>
<th>Side of limb difference</th>
<th>Level of limb difference</th>
<th>Prosthesis use</th>
<th>Former participation in research</th>
</tr>
</thead>
<tbody>
<tr>
<td>F^a</td>
<td>31-40</td>
<td>A^b</td>
<td>Dominant</td>
<td>Above elbow</td>
<td>No, but interested</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>31-40</td>
<td>C^c</td>
<td>Dominant</td>
<td>Below elbow</td>
<td>Former</td>
<td>No</td>
</tr>
<tr>
<td>F</td>
<td>41-50</td>
<td>A</td>
<td>Both</td>
<td>Below elbow</td>
<td>Tried</td>
<td>No</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>C</td>
<td>Nondominant</td>
<td>At shoulder</td>
<td>No, but interested</td>
<td>No</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>A</td>
<td>Dominant</td>
<td>Below elbow</td>
<td>Active</td>
<td>Yes</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>C</td>
<td>Nondominant</td>
<td>Below elbow</td>
<td>Active</td>
<td>No</td>
</tr>
<tr>
<td>F</td>
<td>41-50</td>
<td>A</td>
<td>Dominant</td>
<td>Above elbow</td>
<td>Tried</td>
<td>Yes</td>
</tr>
<tr>
<td>M</td>
<td>51-60</td>
<td>A</td>
<td>Dominant</td>
<td>Below elbow</td>
<td>Active</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>51-60</td>
<td>A</td>
<td>Dominant</td>
<td>Below elbow</td>
<td>No, but interested</td>
<td>Yes</td>
</tr>
<tr>
<td>F</td>
<td>51-60</td>
<td>C</td>
<td>Nondominant</td>
<td>At wrist</td>
<td>No, no interest</td>
<td>Yes</td>
</tr>
<tr>
<td>M</td>
<td>51-60</td>
<td>C</td>
<td>Nondominant</td>
<td>Below elbow</td>
<td>Active</td>
<td>Yes</td>
</tr>
<tr>
<td>M</td>
<td>51-60</td>
<td>A</td>
<td>Dominant</td>
<td>Below elbow</td>
<td>No, but interested</td>
<td>No</td>
</tr>
<tr>
<td>M</td>
<td>61-70</td>
<td>A</td>
<td>Nondominant</td>
<td>Below elbow</td>
<td>Active</td>
<td>No</td>
</tr>
</tbody>
</table>

^aF: female.  
^bA: amputation.  
^cM: male.  
^dC: congenital.

Table 2. Researchers’ demographics.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age group (years)</th>
<th>Profession</th>
</tr>
</thead>
<tbody>
<tr>
<td>M^a</td>
<td>20-30</td>
<td>Researcher</td>
</tr>
<tr>
<td>M</td>
<td>31-40</td>
<td>PI^b</td>
</tr>
<tr>
<td>M</td>
<td>31-40</td>
<td>Medical doctor</td>
</tr>
<tr>
<td>M</td>
<td>31-40</td>
<td>PI</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>PI</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>PI</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>PI</td>
</tr>
<tr>
<td>M</td>
<td>41-50</td>
<td>PI</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>Undisclosed</td>
<td>PI</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>Undisclosed</td>
<td>Undisclosed</td>
</tr>
</tbody>
</table>

^aM: male.  
^bPI: principal investigator.

Data Analysis
The results of the user type questionnaire were processed using MATLAB (The MathWorks, Inc). A goodness-of-fit test using the Pearson chi-square test with a significance level of α=.05 was conducted for both participant groups, compared with the distribution published by Tondello et al [26]. This test was chosen to identify potential differences between the distribution of user types of the participant groups and the distribution of a larger sample population.

For the analysis of the resulting data for the open questions of the survey, we applied the thematic framework approach [28]. This approach consists of 5 steps:
1. Familiarization: All authors familiarized themselves with the collected data. The first author created an initial theme set, which was discussed and agreed upon by all authors.

2. Identifying a thematic framework: The first author created a set of subthemes for the data set. This was approved by the last author for use in the next steps. The full set of themes can be found in Table 3.

3. Indexing: All authors coded the interview data independently. These were discussed between all authors until a consensus was reached.

4. Charting: The data were sorted by themes and subthemes by the first author.

5. Mapping and interpretation: The first author summarized and interpreted the charted data according to the themes.

Table 3. Thematic framework.

<table>
<thead>
<tr>
<th>Themes</th>
<th>People with limb difference (n=39), n (%)</th>
<th>Researchers (n=108), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accessibility</td>
<td>4 (10.3)</td>
<td>4 (3.7)</td>
</tr>
<tr>
<td>Data management</td>
<td>0 (0)</td>
<td>5 (4.6)</td>
</tr>
<tr>
<td>Hardware</td>
<td>2 (5.1)</td>
<td>11 (10.2)</td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Muscle development and control</td>
<td>4 (10.3)</td>
<td>4 (3.7)</td>
</tr>
<tr>
<td>Prosthetic ability</td>
<td>7 (17.9)</td>
<td>15 (13.9)</td>
</tr>
<tr>
<td>Additional benefits for users</td>
<td>2 (5.1)</td>
<td>8 (7.4)</td>
</tr>
<tr>
<td>Clinical and research benefits</td>
<td>1 (2.6)</td>
<td>4 (3.7)</td>
</tr>
<tr>
<td>Education</td>
<td>0 (0)</td>
<td>4 (3.7)</td>
</tr>
<tr>
<td>Feedback</td>
<td>0 (0)</td>
<td>3 (2.8)</td>
</tr>
<tr>
<td>Training</td>
<td>14 (35.9)</td>
<td>38 (35.2)</td>
</tr>
<tr>
<td>Game</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affect</td>
<td>5 (12.8)</td>
<td>17 (15.7)</td>
</tr>
<tr>
<td>Personalization</td>
<td>3 (7.7)</td>
<td>7 (6.5)</td>
</tr>
<tr>
<td>Social aspects</td>
<td>0 (0)</td>
<td>1 (0.9)</td>
</tr>
<tr>
<td>Mechanics</td>
<td>11 (28.2)</td>
<td>7 (6.5)</td>
</tr>
<tr>
<td>Challenges</td>
<td>—b</td>
<td>27 (25.0)</td>
</tr>
<tr>
<td>Justification and reasoning</td>
<td>—</td>
<td>11 (10.2)</td>
</tr>
<tr>
<td>Design and development</td>
<td>—</td>
<td>10 (9.3)</td>
</tr>
<tr>
<td>Involvement</td>
<td>—</td>
<td>2 (1.8)</td>
</tr>
<tr>
<td>Recognition</td>
<td>—</td>
<td>4 (3.7)</td>
</tr>
</tbody>
</table>

*aItalicized values denote the subtotals for the respective main themes.

b—: Not available. People with limb difference were not asked about research-specific challenges.

Results

General

A total of 14 people with limb difference and 12 researchers in the prosthetic field participated in the survey (Tables 1 and 2, respectively). Of the 14 people with limb differences, 7 had experience with studies pertaining to rehabilitation with a computer-based device before the survey.

An overview of the themes identified from the survey data is presented in Table 3. In the following sections, the responses of the survey participants are presented. The texts in italic font are direct quotes from the participants. The results are reported thematically: usability, training, game design, and challenges. The raw survey results are provided in Multimedia Appendix 1.

Usability

The majority of the participants, both people with limb difference and researchers, indicated that they believed that people in prosthetic training would use game-based training at home (Figure 1). A total of 3 of the 14 people with limb difference and 5 of the 12 researchers cautioned that factors predominately concerning usability and game design can affect adoption, as usability is one of the issues as to why former gamers have stopped playing virtual games:
Figure 1. Participant response proportions for (A) their willingness to use a game-based prosthetic training tool at their own home and (B) their preferred peripheral or device for such a game.

I used to play an Atari, as it had joystick controls, but I struggle with handsets, so [I have] not played for some time.

In addition, participants pointed out that such a tool should prove easy and robust to use both in setup as well as in the actual use of the tool.

An additional concern was raised with regard to the compatibility of the software with potentially existing hardware at home. One participant said:

Requirements for game-based prosthetic training include providing a setup that is easy to use and robust; ensuring a game framework that allows people to use their own devices.

Creating a game framework that would work cross-platform with potentially outdated hardware is a significant challenge for the myoelectric control research community. A step in this direction could be the development of web-based gaming platforms, as mentioned by one of the researchers. Furthermore, researchers have indicated the importance of the exchange of code and knowledge within the community.

The preferred choice of platforms and peripherals, the input and output devices with which the game is played, was more distinct in the case of the participants with limb difference than the researchers. Figure 1 shows that the majority of people with limb difference would prefer a computer or television screen. However, researchers have shown no clear preference for the mobile over fixed screen options.

Training

The topics connected to the main theme of training were grouped into muscle development and control; prosthetic ability; additional benefits for users, clinicians, and research; education; and feedback.

The answers concerning muscle development and control mostly centered on the participants’ hopes and expectations. People with limb difference expect the game to help develop the musculature in their remaining limb and improve prosthetic dexterity with these muscles. For instance, a participant with limb difference wished for a game that highlighted the following:

my ability to use advanced prosthetics.

Another participant with limb difference would like to see a game made that also helps to reduce the phenomenon of phantom limb pain.

Researchers additionally specified their hopes in the direction of accelerated learning speeds, the training of ADL-related tasks, and the transfer of the skills learned in the game to prosthesis use. One researcher stated that any tool created should be based on the needs of the patients as well as clinicians. Outside of the physical benefits, researchers pointed out that a game-based tool could allow the user to share data with other users, if they wanted to, and potentially lead to a reduced need for input through in-clinic appointments.

In addition, a training game could support therapists in assessing patients’ abilities. One researcher indicated:

In the end the games should also relieve the therapists from time spending on rehabilitation. Moreover, the games should be helpful as an assessment tool for the appropriateness of a prosthesis for a given patient.

The need of personalized feedback for users was raised by a couple of researchers. They envisioned the benefits of regular
and individualized training progress feedback by which users as well as therapists can track their skill progression properly.

**Game Design**

The answers of the participants directly relating to design choices for the game were grouped under the main theme of game design. The subtheme of affect included all answers that aimed at an emotional response of the user; personalization contains in-game control of design for the player; social aspects revolve around the possibility of interaction with other players; and answers in the mechanics subtheme pertain to ideas for the mechanical features of the game.

With regard to the affective aspect of the design, participants of both groups indicated that the game had to be engaging for in-home adaption. For instance, one researcher points out that rehabilitation games should not only be interesting in the short term but also be able to secure the long-term engagement of the user. However, engaging the user to a sufficient level was recognized as a challenge.

Furthermore, researchers say that the game should be attractive to users, although it could be hard to appeal to a wide range of potential users. A range of games or in-game options could be beneficial because of the differing motivating factors. The degree of immersion was wished for by a participant with limb difference, and another participant pointed out the potential of an empowering portrayal of life postamputation in the game, for example:

*That the game would be so immersive that the participant would not realize they are training their remnant muscles for optimal EMG based control.*

Interestingly, one participant with limb difference suggested an *end of world* setting. As this might not feel appealing for all users, the setting should be customizable by the player. Some participants from the limb difference group wanted to see a relatable character using a prosthesis in the game, tying in with the empowering effect mentioned earlier. This affords a viable option for personalization:

*Perhaps the protagonist could be a prosthetic wearer [...]*

The variety of preferences in the thematic setting of the game can be seen in Multimedia Appendix 2. These are results of a multiple-choice question and therefore do not add up to 100%.

The mechanics of a game-based prosthetic training tool are predominately addressed with respect to the game genre. Participants of the limb difference group identified a variety of different genres as desirable, including quiz and puzzle games, but also adventure, shooter and fighting games, and horror games. One participant mentioned more specific activities, such as camping, fishing, and shooting. A researcher argued against the use of war and fighting mechanics in clinical settings. The different genre preferences of the participants can be found in Multimedia Appendix 2. As mentioned earlier, these results do not add up to 100%.

An additional influence on the type of game and the game mechanics involved can be the type of gameplay the user prefers. As it cannot be assumed that a user is an active or former gamer and therefore knows what they look for in a game, an assessment of the user types using the Hexad Scale [26] was conducted in this survey. The outcome of this assessment for both people with limb difference and researchers can be seen in Multimedia Appendix 2. Among the people with limb difference, the philanthropist and the achiever user type are tied as clearly the most common types, whereas the player and the disruptor are the least commonly occurring. None of the researchers were grouped into players and disruptors, which is similar to them being the least common type in the other group. We observed a notable difference in the overall distribution of user types to people with limb difference. A Pearson chi-square test was conducted to test the statistical significance of this difference. The results showed that both groups were significantly different from the control group in a study by Tondello et al [26] at a significance level of $\alpha=.05$. However, the results for the people with limb difference at $\chi^2=12.3$ are noticeably closer to the critical value of $\chi^2=11.1$ than the result of the researcher participant group of $\chi^2=26.5$. This indicates that the group of participants with limb difference showed a higher similarity to the sample population presented in a study by Tondello et al [26] than the group of researchers showed to the same sample population.

Apart from the genre, people with limb difference indicated that they would like a progressive and appropriate increase in difficulty, would like to use both hands to play the game, and would like for the game to motivate them to make enough repetitions of the trained arm to form habits. A researcher pointed out that the abstraction of the signals to rewarding or menacing game elements could be beneficial.

**Challenges**

Only researchers participated in this part of the survey. They were asked to formulate their opinions on the challenges that the field of games in upper limb prosthetic rehabilitation faces. In addition, they were invited to propose potential actions that could be undertaken by the community to address these challenges.

Part of the challenges mentioned by the researchers was the justification for the use or the development of serious games in prosthetic rehabilitation. The meaningful impact in terms of skill transfer to prosthetic use by myo-games must be investigated. This was stated not only for short-term effects but also for long-term benefits, when compared with other rehabilitation methods. The recognition of the difference between in-game improvement and actual benefit for prosthetic use has not yet been widely acknowledged:

*However, most game studies focus only on in-game improvement. Now in-game improvement is a requirement for transfer to daily life performance. However, in-game improvement is not a sufficient requirement for transfer.*

Therefore, researchers have called for longitudinal and large-cohort studies in the field to show the appropriateness of the medium used and the transfer capabilities of—potentially only certain types of—games.
The development of these myo-games faces its own problems and challenges. They should make the benefits clearly visible for the user but at the same time make the training imperceptible by shifting the focus of the user away from the underlying reason for the training onto the task-specific in-game goals. Researchers have indicated that the formation of bad habits to win the game by potentially compromising the training efficiency should be avoided. Therefore, it was recommended to involve game developers in the process and to parallelize game development and transfer testing procedures to avoid losing sight of either aspect. One researcher proposed the development of a knowledge and information sharing platform for myo-games. He also indicated that such a platform could lead to a wider and easier access to developed games for users at home with their existing hardware.

Finally, the recognition of the value of myo-games is another key obstacle to tackle. Some people, especially certain age groups, might dismiss games as frivolous and a waste of time. Both clinicians and patients might need to be convinced that a serious game for prosthetic rehabilitation could benefit them as well as potentially benefit research. For example, a researcher suggested:

_Educating participants about serious games and why the time they spent playing is well spent_ [...]

Reviewing the number of mentions of the main themes, a clear separation of themes becomes visible. In Figure 2, it can be seen that the main expectation of the participants for the prosthetic training lies within the training benefits that it should provide. The fewer mentions of game design topics indicate that the design is of concern but that the training aspect takes priority and should be the base minimum of any game-based training. The themes of training and game design almost have the same amount of mentions overall in this survey, with 52 and 51 responses, respectively. In contrast to the clear focus on training in the expectations, the game design–related themes were more split between preferences expectations. This suggests that many game design traits are considered desirable but not a necessary component, which is reflected in some research in the field of game-based prosthetic training.

_Figure 2. Number of coded responses per question and main theme over both participant groups._

![Number of main themes mentioned per question](image)

**Discussion**

**Myo-Games: Opportunities and Challenges**

The aim of this study is to determine the preferences, expectations, and views of both researchers and people with limb difference on game-based prosthetic training. The outcome of this survey indicates a general willingness and tentative optimism toward the topic. However, for the wide adoption of myo-games, several scientific and engineering challenges should be addressed.
In the following sections, we will discuss all identified challenges in the context of usability, training, and game design themes.

Usability
With respect to hardware usability, the discontinuation of the commonly used Thalmic Myo Gesture Control armband [9,13,20] means there is now a lack of low-cost, easy-to-use EMG sensors for home use. Of the commercially available dry EMG options suitable for game-based training, almost all require sensors to be accurately positioned using adhesives. As such, researchers are increasingly developing custom-built EMG acquisition solutions for game-based systems [14,18,29]. This approach does not scale well and is likely to contribute to the slow translation of laboratory research to translational research with a larger number of participants.

The accessibility of in-home myo-game software is also a key factor for overall usability. Accessibility must be balanced against client privacy and patient confidentiality, key points that therapists have identified as important in game-based upper limb rehabilitation [27]. The small target population of upper limb amputees and the niche nature of game-based rehabilitation mean cross-platform software with widespread hardware compatibility is not likely to be financially viable. The low cost of Android mobile devices and the fact that they can be locked down to a restricted set of sandboxed apps make this platform the most likely candidate for in-home use.

Training
The topic of training with regard to goals was fairly unanimous between the participants with limb difference and the researchers. In current research in this field, it appears to have mostly been assumed that an improvement in game control would translate readily to an improvement in prosthetic control as only in-game improvement or abstract control has been measured in most studies [4,6,8,12-14,20,30,31]. The question of whether this transfer can happen and with which type of game this might happen has yet to be answered, as current research challenges the idea that a general myocontrol skill exists [10,32]. However, it could prove beneficial to not only consider direct transfer but to also investigate indirect improvements like increases in training speed in learning the use of a prosthesis after training with myogames. Therapists have especially expressed that gaming in therapy should be balanced with other forms of therapy [27], and as such, it might prove beneficial to examine them in conjunction.

The only testing of the effect on actual prosthetic skill in game-based prosthetic training research was conducted on the direct effect that the developed game would have on prosthetic skill [10,15]. We could not find any study that has conducted research if the training with a game before or alongside actual training with a prosthesis could be beneficial to the learning process as opposed to an immediate effect on prosthetic skill. Prior training could be sensible as post amputation, the site of surgery may not yet be ready for the fitting of a prosthetic socket. The British Society for Rehabilitation Medicine states that the fitting may be deferred from 4 to 6 weeks after the amputation [33]. In both the amputees and people with congenital limb difference, the muscle sites could need development before a prosthesis can be considered. Training alongside real-life prosthetic training could add to the fine control without being reliant on the other arm musculature when those muscles are tired from the weight of the prosthesis.

In contrast to the popular assumption that a prosthetic training game would reduce the time investment necessary for the therapist, it should be pointed out that this is highly dependent on the type of training tool and the way it is included in the exercise regime. Almeida and Nunes [34] point out that therapists might need to be involved in the setup and fulfill a supervisory role in the exercises performed by users. Optimally, this would not be the case after the system was tested and approved apart from a first introductory session with the therapist to ensure that the user understands how to use the system properly. Therefore, ease of use is an important factor, both with respect to setting up equipment and its daily use.

Responses from the participants with limb difference group indicate that they would like a prosthetic training tool that makes rehabilitation feel less like rehabilitation. This was reflected in the call of the researchers for immersion in the game, which allowed the user to shift from their limb and muscles to the task at hand, allowing the motions to become intuitive. This can also be found as part of the Optimizing Performance Through Intrinsic Motivation and Attention for Learning theory of Wulf and Lewthwaite [35], which states that an external focus has beneficial effects on motor learning as well as the sense of accomplishment. According to that theory, the external focus as well as intrinsic motivation feed into a virtuous cycle of enhanced motor learning. However, the question of how to incorporate these aspects efficiently is yet to be answered.

Game-Design
One of the main points mentioned for the game design was the facilitation of both short- and long-term engagement. The potential users of a prosthetic training tool seem willing to take the leap to use it, which is supported by research in games for other conditions [36], but the appeal of such a novelty can quickly wane. It is the task of good game design to support the user by providing motivating gameplay for the entire training period. This again feeds into the aforementioned virtuous cycle, according to Wulf and Lewthwaite [35]. A shared experience among prosthetic users as well as between prosthetic users and able-bodied people could further enhance the motivation and potentially the training intensity [36-38].

The involvement of all stakeholders, including game developers, is a recommended path for research to take. This can be especially motivated, as game developers have more experience in catering toward a specific audience with their games and which game design elements work best to keep the users motivated to play the game. Academic teams are likely to have different views on fun themes and activities to the general population based on the differences in the user types in this study. These differences further support the need for co-design between researchers and potential end users [17,29,39]. The inclusion of the design preferences of the users as well as the input by their families can provide valuable insights for the researchers developing the games [40]. In addition, an
understanding of the practical activities of the therapists and their involvement in game-based rehabilitation is necessary for the effective development of a training tool that benefits all parties [34]. As of yet, cocreation in the wider field of prosthetics remains difficult to integrate with current academic methods but is now becoming a focal point for translational research [Jones et al, forthcoming].

Nonetheless, this gives rise to the question of how to achieve effective targeting of the game in a highly varied target audience. Working toward increasing the engagement of a game before knowing whether transfer will happen for this game could lead to fruitless efforts. However, it is possible that engagement is a contributing factor in the transfer process and therefore worthy of further exploration.

As the current state of myo-game development is very diverse, it was expected that the opinions on the matters to focus on and the potential resolutions are just as varied.

The Survey: Strengths and Limitations
It is the first time a survey like this has been conducted in this field with these participant groups. We did not include clinicians in this survey, as a previous study covered therapists’ views on the use of video games in general upper limb rehabilitation [27]. The format of the study allowed the easy spread of the survey in the field. This led to the recruitment of 12 researchers and 14 participants with limb difference. However, the sample size of the survey was smaller than would be desirable for a good cross-section of the population with limb difference as well as researchers. In an attempt to acquire many participants, this survey has been spread in various ways, but it is possible that survey fatigue has stopped people from participating. This could indicate that other means of interacting with people with limb difference might be advisable for future research. If a similar study was to be repeated in this format, we recommend collaborating on it with several research groups. This would provide access to a larger group of researchers as well as people with limb difference.

In addition, regarding the data set of the researchers, the clear majority of participants identifying as male could influence the outcome of the answers. Female researchers were included in the list, and the survey was distributed among them. However, because of the general gender disequilibrium in the field, this list already contained a higher percentage of male researchers. This could have influenced the opinions and preference distribution of the researchers, as differences in gaming preferences have been identified in both gamer and nongamer populations [41].

The Future of Myo-Games
In this study, the results indicate a higher level of similarity of the participants with limb difference group to the sample population in a study by Tondello et al [26] than of the researcher participant group compared with the same sample population. Although the low number of participants does not provide conclusive evidence in this matter, it is still worth considering the implications of this dissimilarity. The influence of personal preferences and assumptions made by the researchers could have a significant impact on the engaging and motivating aspects of the game they are developing. This could be mitigated by bringing in professional support from the game development sector or by increasing collaboration with game experts within academia, who have more practical knowledge in catering a game experience toward a diverse target audience. Moreover, this would likely benefit the efficient incorporation of features facilitating external focus and intrinsic motivation. However, as the market for serious games for upper limb prosthetic rehabilitation is fairly small and therefore the potential profit margin is small, including professional game developers in this research could prove challenging. In addition, involvement of the users, their families, and other stakeholders can provide additional benefits to the development process and should be considered.

Furthermore, the effect of myo-games has mostly been investigated in short trials of up to a week in experimental scenarios at a university or at the home of the participant. A much-improved assessment of the effect of these games in both the short term and the long term as well as whether a significant level of transfer occurs could be achieved by conducting longer-term home trials with people with limb difference. This would also provide more information about the engaging aspects of the game in the long term, that is, if the feelings of the participants change because of the waning novelty of the game and if it becomes a chore. Conversely, if the attrition over a long-term study would prove to be significantly lower than comparable experiments, this could be an indicator for a positive effect of games on exercise and training adherence. As discussed, for long-term experiments in a home environment, Android mobile devices appear to be a promising choice for hardware. In addition, the effect of myo-game training in conjunction with conventional prosthetic training could prove beneficial to assess.

The current state of isolation because of the 2020 coronavirus pandemic poses many challenges for academic research, some of which may persist in the long term. Thus, a shift in experimental procedures to the home environment of the participants could be a beneficial route to follow. Many of the technical challenges surrounding home-based experimentation, such as precision timing, low latency networking, and data security, have already been widely addressed in gaming. More widespread adoption of gaming technology to facilitate the shift of experimentation to participants’ homes may provide an alternative route to bridge the gap between academic research and viable prosthesis training solutions.
Acknowledgments

This work was supported by the Leverhulme Doctoral Scholarship Programme in Behaviour Informatics (DS-2017-015), the National Institute of Health Research via Devices 4 Dignity Starworks STWK-006 and the Engineering and Physical Sciences Research Council via grants EP/R004242/1 and EP/M025594/1.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questions and results.

[PDF File (Adobe PDF File), 121 KB - games_v91e23710_app1.pdf]

Multimedia Appendix 2
Preferences and user type distributions.

[PNG File, 221 KB - games_v91e23710_app2.png]

References


Communications Workshops (PerCom Workshops). 2019 Presented at: 2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); 11-15 March 2019; Kyoto, Japan, Japan p. 151. [doi: 10.1109/percomw.2019.8730824]


34. Almeida J, Nunes F. The Practical Work of Ensuring Effective Use of Serious Games in a Rehabilitation Clinic: A Qualitative Study. JMIR Rehabil Assist Technol 2020 Feb 21;7(1):e15428 [FREE Full text] [doi: 10.2196/15428] [Medline: 32130177]


Video Game–Based Rehabilitation Approach for Individuals Who Have Undergone Upper Limb Amputation: Case-Control Study

N A Hashim1*, MSc; N A Abd Razak1*, PhD; H Gholizadeh2*, PhD; N A Abu Osman1,3*, PhD

1Department of Biomedical Engineering, Faculty of Engineering, Kuala Lumpur, Malaysia
2Ottawa Hospital Research Institute, Ottawa, ON, Canada
3The Chancellery, University of Malaysia, Terengganu, Malaysia
*all authors contributed equally

Corresponding Author:
N A Abu Osman, PhD
Department of Biomedical Engineering
Faculty of Engineering
Universiti Malaya
Kuala Lumpur, 50603
Malaysia
Phone: 60 0379675341
Email: azuan@umt.edu.my

Abstract

Background: Brain plasticity is an important factor in prosthesis usage. This plasticity helps with brain adaptation to learn new movement and coordination patterns needed to control a prosthetic hand. It can be achieved through repetitive muscle training that is usually very exhausting and often results in considerable reduction in patient motivation. Previous studies have shown that a playful concept in rehabilitation can increase patient engagement and perseverance.

Objective: This study investigated whether the inclusion of video games in the upper limb amputee rehabilitation protocol could have a beneficial impact for muscle preparation, coordination, and patient motivation among individuals who have undergone transradial upper limb amputation.

Methods: Ten participants, including five amputee participants and five able-bodied participants, were enrolled in 10 1-hour sessions within a 4-week rehabilitation program. In order to investigate the effects of the rehabilitation protocol used in this study, virtual reality box and block tests and electromyography (EMG) assessments were performed. Maximum voluntary contraction was measured before, immediately after, and 2 days after interacting with four different EMG-controlled video games. Participant motivation was assessed with the Intrinsic Motivation Inventory (IMI) questionnaire and user evaluation survey.

Results: Survey analysis showed that muscle strength and coordination increased at the end of training for all the participants. The results of Pearson correlation analysis indicated that there was a significant positive association between the training period and the box and block test score ($r_p=0.95$, $P<.001$). The maximum voluntary contraction increment was high before training (6.8%) and in the follow-up session (7.1%), but was very small (2.1%) shortly after the training was conducted. The IMI assessment showed high scores for the subscales of interest, perceived competence, choice, and usefulness, but low scores for pressure and tension.

Conclusions: This study demonstrated that video games enhance motivation and adherence in an upper limb amputee rehabilitation program. The use of video games could be seen as a complementary approach for physical training in upper limb amputee rehabilitation.

(JMIR Serious Games 2021;9(1):e17017) doi:10.2196/17017

KEYWORDS box and block test; Intrinsic Motivation Inventory; maximum voluntary contraction; motor rehabilitation; upper limb amputee; video games

https://games.jmir.org/2021/1/e17017
Introduction

A myoelectric prosthesis is a popular choice among upper limb amputees [1], although most users stop using it after sometime owing to control difficulties, muscle fatigue, and lack of motivation to practice before getting used to the control mechanism [2]. Myoelectric prosthesis task training is required before and after prosthesis fitting [3], and it takes several months of practice under the guidance of physicians and therapists to regulate these prostheses naturally. This process is exhausting and tedious [4].

The use of interactive technology can foster intrinsic motivation and thus the effort invested in the training of neuromuscular rehabilitation [5]. Various computer-based systems, including video games, have been suggested [6-9] to support motor training. Physiological data suggest that gaming can cause neuroplastic reorganizing, leading to the long-term retention and transfer of skills; however, further clinical research is needed in this field [10]. Besides, it has been shown that virtual reality (VR) platforms, in the form of video games, provide amputees with an interactive and immersive technique for enhanced muscle coordination and overall control, which has been found to be very helpful for a patient to start training before getting an actual prosthetic hand [11].

Previous studies have shown that rehabilitation tasks based on a fun and playful concept provide better outcomes compared to traditional physiotherapy exercises [12,13]. Video games have been shown to offer some positive effects on behavior and physiology [10]. Participants who learn motor skills through games, especially in the VR setting, can have better skills in the real environment [14]. Therapy needs to have high repetition, supervision, clear rewards, and a long duration over time for maximum effectiveness [15,16]. In this case, patient motivation needs to be maintained in order to increase patient engagement and the dosage of therapy-relevant movements by incorporating these movements into an interactive environment that can enhance cognitive, motor, and affective measures [10]. Video game–based rehabilitation is popular in geriatric populations [17], people with stroke [18,19], and people with Parkinson disease [20,21]. In 2018, a similar study on upper extremity amputees and able-bodied participants showed an overall improvement in electromyography (EMG) control, fine muscle activation, and electrode separation [4].

Patient motivation was usually evaluated using Intrinsic Motivation Inventory (IMI) questionnaires and observations by physiotherapists [22-24]. This study investigated whether the inclusion of video games in the upper limb amputee rehabilitation protocol could have a beneficial impact on the course of treatment and patient motivation, with the hypothesis that the functional outcome of motor rehabilitation is directly proportional to the duration of the therapeutic session [25-27] and patient interest to take part in rehabilitation [28]. Other study outcomes measured in this work are maximum voluntary contraction (MVC) readings of the forearm muscle and the box and block test (BBT) score.

Methods

Participants

The participants of this study were five able-bodied and five transradial amputee subjects with a mean age of 26.3 years (SD 4.47), who were recruited from University Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia. Two out of the five transradial amputee participants had congenital amputation, while the other three had undergone amputation owing to traumatic injury at a mean of 3.7 years (SD 1.69) at study inclusion. One participant was a bilateral amputee; therefore, only one residual limb was considered in this study. The rest of the amputee participants had unilateral amputation (right side for all), and the amputation level ranged from long to short transradial amputation. All amputee participants had never used any prosthetic device since amputation and had no experience with an EMG system. The amputee participants were independent, could perform all activities of daily living on their own, and received no other treatment or rehabilitation throughout the study period. In this study, the amputee participants represented the test group and the able-bodied participants represented the control group.

The EMG signal generated by each participant’s forearm was evaluated before the training session. Four video games (Crate Whacker, Race the Sun, Fruit Ninja, and Kaiju Carnage) and their respective control variations were assessed for suitability to be used in the rehabilitation protocol. All participants had normal vision and were guided throughout the study. Two questionnaires were administered (modified version of the IMI and the user evaluation study). The IMI questionnaire was administered at the end of the session, whereas the user evaluation study was administered at the end of each game on the 10th training session. The test was approved by the Medical Research Ethics Committee and the Ministry of Health Malaysia (approval ID: NMRR-16-2106-32880). The Medical Research Ethics Committee considered that the data collection for this study would only involve physical evaluation. Participants were required to sign a written consent form prior to the tests.

Games

The following four different games were used in this study: Crate Whacker, Race the Sun, Fruit Ninja, and Kaiju Carnage (Figure 1). The games were selected because of the suitability of the main player control to prosthesis usage and the compatibility with the Myo armband (Thalamic Lab). The main control used for Crate Whacker and Fruit Ninja is “make a fist,” which can improve the overall strength of the forearm muscle. The amputee participant’s residual muscle must be in good condition [29,30], and precise EMG control ability is required to control the prosthesis naturally [31]. Meanwhile, Race the Sun and Kaiju Carnage were included because of their ease of control. The game control methods are summarized in Table 1.
Figure 1. Selected open-source games.

Table 1. Game control methods with the Myo armband.

<table>
<thead>
<tr>
<th>Game</th>
<th>Player control</th>
<th>Action</th>
<th>Muscle contractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crate Whacker</td>
<td>Make a fist</td>
<td>Equip mace</td>
<td>Quick but frequent contraction</td>
</tr>
<tr>
<td>Crate Whacker</td>
<td>Spread the fingers</td>
<td>Equip grenades</td>
<td>Quick but frequent contraction</td>
</tr>
<tr>
<td>Crate Whacker</td>
<td>Wave left</td>
<td>Centre your arm</td>
<td>Quick but frequent contraction</td>
</tr>
<tr>
<td>Crate Whacker</td>
<td>Wave right</td>
<td>Reset player</td>
<td>Quick but frequent contraction</td>
</tr>
<tr>
<td>Crate Whacker</td>
<td>Pan</td>
<td>Smash crate</td>
<td>Quick but frequent contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Make a fist</td>
<td>Center position</td>
<td>No contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Spread the fingers</td>
<td>Open menu</td>
<td>No contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Wave left</td>
<td>Mouse click</td>
<td>No contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Wave right</td>
<td>Enable/disable mouse control</td>
<td>No contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Pan</td>
<td>Fly side to side</td>
<td>No contraction</td>
</tr>
<tr>
<td>Race the Sun</td>
<td>Rotate</td>
<td>Jump up or down</td>
<td>No contraction</td>
</tr>
<tr>
<td>Fruit Ninja</td>
<td>Make a fist</td>
<td>Cut the fruits (hold)</td>
<td>Prolonged contraction</td>
</tr>
<tr>
<td>Fruit Ninja</td>
<td>Double tap</td>
<td>Move mouse</td>
<td>Prolonged contraction</td>
</tr>
<tr>
<td>Fruit Ninja</td>
<td>Pan</td>
<td>Unlock (hold) or lock</td>
<td>Prolonged contraction</td>
</tr>
<tr>
<td>Kaiju Carnage</td>
<td>Make a fist</td>
<td>Charge/release kaiju shockwave</td>
<td>Combination of gestures</td>
</tr>
<tr>
<td>Kaiju Carnage</td>
<td>Wave left</td>
<td>Recalibrate Myo armband</td>
<td>Combination of gestures</td>
</tr>
<tr>
<td>Kaiju Carnage</td>
<td>Pan</td>
<td>Quickly swipe with the hand for kaiju claw</td>
<td>Combination of gestures</td>
</tr>
<tr>
<td>Kaiju Carnage</td>
<td>Rotate</td>
<td>Charge/unleash kaiju beam</td>
<td>Combination of gestures</td>
</tr>
<tr>
<td></td>
<td>Hold arm up</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The keyboard mapper utility available in Myo Connect was used to assign keyboard and mouse commands to any of five gestures to control the game. The gestures used to control the games included fist, wave in, wave out, finger spread, and double tap. Pan motion from highly sensitive motion sensors embedded in the Myo armband is important as it allows the player to move in the game setting. The custom configuration for each game is exported as connector scripts. In Crate Whacker, the player repeatedly smashes the crate with a mace and destroys them with grenades. The player controls the mace position with the acceleration and gyroscopic data of the armband. In Race the Sun, the player controls a solar-powered spacecraft, dodging various objects on the way, such as laser beams, other ships, and other stationary obstacles, while collecting pick-ups that can be used in the game, with the sun slowly setting on the horizon. The game ends either when the player hits an object and the ship gets destroyed or when the sun sets. Meanwhile in Fruit Ninja, the player needs to slice the fruit into half to collect point. The Kaiju Carnage game enables the player to claw, smash, and shoot nuclear lasers on buildings.

Experimental Protocol

The set up for this study includes a monitor that displays the game that the participant plays with the Myo armband connected via Bluetooth. The system utilized commercially available equipment of the Myo gesture control armband. It consists of eight plastic pods held together by a rubber lining that measures 7.5 inches (19.05 cm) in circumference and can be extended to 13 inches (33.02 cm). There are three medical steel EMG sensors on the bottom of each pod, which are responsible for reading the muscles’ electrical activity and arm movement. With the use of a machine learning process, the armband can recognize the gesture perform based on the electrical activity of the arm in real time. Amputee participants wear the Myo armband on the affected limb. Able-bodied participants wear the armband on the dominant hand. Ten 1-hour sessions were carried out within a 4-week period. Participants were initially
instructed to perform the VR BBT designed by the authors, and provisional MVC levels were recorded before the four Myo-controlled computer games were introduced to each participant in a randomized order. Randomization of the games was performed by the study facilitator. At the end of the training session, the MVC value was recorded again, and a modified IMI questionnaire and a simple user evaluation survey regarding the gaming experience, similar to the approach in the study by Prahm et al [32], were administered to the participants. Follow-up MVC values were taken 2 days after the last intervention session.

Assessment
In order to investigate the effect of the video game rehabilitation protocol used in this study, VR BBTs and basic EMG assessments evaluating approximate coordination and muscle strength were performed. The participants were asked to perform the BBT in a virtual environment (Figure 2). The test was made up of the following two phases: an initial 15-second phase, where the user gets a trial run, followed by a scored 60-second phase. Participants were told to move as many blocks as possible in a span of 60 seconds (one by one) from one compartment to another. The participant’s hand must cross over the partition in order for a point to be given, and blocks that drop or bounce out of the second compartment onto the floor are still rewarded with a point. The instructions are scripted, and they were read to each user according to the manual for the test. The experiments were performed according to the World Medical Association Declaration of Helsinki [33]. The player needs to move the virtual hand near the cube, contract the forearm muscle to pick the cube up, and then stop contracting the muscle to release the cube into another compartment. The score generated on the top left of the screen refers to the amount of successfully transferred blocks, and the red line on the top right corner of the screen refers to the linear timer (Figure 2).

*Figure 2.* A transradial amputee participant performing the virtual box and block test.

The MVC was recorded for each participant before, immediately after, and two days after playing the game. Participants were instructed to maximally contract their forearm muscle and to hold this contraction for 10 seconds, and the highest value was recorded as the MVC value. The assessment was carried out with the Myo visualization tool (Figure 3) on a laptop with a discrete graphics processing unit (NVIDIA GeForce GTX 1060) for signal acquisition, filtering, calibration, training, and prediction. The Myo visualization tool is an open-source Myo armband data reader created using Myo SDK, Qt Creator, and QCustomPlot that streams and displays the EMG signal and accelerometer, gyroscope, and orientation data of the Myo armband.
**Questionnaires**

Two questionnaires were provided for each participant to be completed at the end of the training session, which included a modified IMI questionnaire [3] and a user evaluation survey [32]. The participants’ experiences with each game were evaluated using the IMI questionnaire, which has previously been used with a virtual environment for motor rehabilitation [22-24]. A modified version of the IMI questionnaire made up of five selected subscales was utilized to evaluate participants’ experiences with the protocol. The subscales included were interest/enjoyment, perceived competence, perceived choice, pressure/tension, and value/usefulness. The questionnaire included statements such as “I found this activity very interesting,” and they were rated on a 7-point Likert rating scale from 1 (not true) to 7 (true).

A brief questionnaire about the gaming experience was presented after the completion of each game. This short survey included questions about the gameplay, fun factor, motivation, and input and control methods, and it included the following questions: (1) Did you like this game in general? (2) Did you have fun playing the game? (3) How would you rate the input mechanism? (4) How would you rate the game control? (5) Did this game motivate you? Question could be answered on a 5-point Likert scale ranging from 1 (“do not agree”) to 5 (“agree”). The participants were asked to focus on the differences between each game and to avoid giving the same answer to one statement for all the games if possible.

**Results**

In this study, we showed that incorporating video games in an upper limb amputee rehabilitation protocol would provide benefits to the rehabilitation output. The platform consists of four video games controlled by an off-the-shelf Myo gesture control armband. The games were played by five able-bodied and five transradial amputee participants. The use of the four video games was widely explored in this context. No technical difficulties were encountered with the game settings.

The average BBT score in every training session gradually increased, with the mean score reported in the 10th training session of this study being 23.2 (SD 4.62) in the control group and 18.6 (SD 3.93) in the test group (Figure 4). The results of the Pearson correlation assessment indicated that there was a significant positive association between the training period and the BBT score ($r = 0.95$, $P < .001$). The highest score achieved was 29 blocks transferred over 60 seconds by one able-bodied participant during the final session.
From sessions 1 to 10, the MVC value for able-bodied participants increased during follow-up and shortly after the training was conducted, but not immediately after the training session. This may be caused by muscle fatigue. Meanwhile, for amputee participants, the MVC value increased before and shortly after the training was conducted, and during the follow-up session (Figure 5). This proves that the muscle strength of all participants increased with the training duration. Shortly after the training session, six out of 10 participants showed an improvement in the MVC value from the initial to the final session, two participants showed no improvement, and two participants showed a reduction in the MVC value at the end of the training session. In the follow-up session, nine out of 10 participants showed an improvement in the MVC value compared to the initial training session. The MVC increment was greater before training and in the follow-up session but was very small (approximately 2.1%) shortly after the training was conducted.

The results obtained from the IMI questionnaire are shown in Table 2. Participants had high interest in playing the games, perceived playing the games as their own choice, and felt competent and at ease while playing the games. Participants also agreed that this activity provided benefits that included strengthening their muscles, motivating them, and keeping them active, resulting in higher motivation to perform the activity for a longer period of time. Pressure or tension was the only subscale with a low reported average score. The difference in the scores for both groups was insignificant for all subscales, except for the perceived choice subscale ($d=0.58$).
Table 2. Intrinsic Motivation Inventory questionnaire results based on five subscales.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Likert scale score, mean (SD)(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Able-bodied participants</td>
</tr>
<tr>
<td>Interest/enjoyment</td>
<td>6.8 (0.40)</td>
</tr>
<tr>
<td>Perceived competence</td>
<td>6.2 (0.75)</td>
</tr>
<tr>
<td>Perceived choice</td>
<td>5.8 (0.75)</td>
</tr>
<tr>
<td>Pressure/tension</td>
<td>1.8 (0.75)</td>
</tr>
<tr>
<td>Value/usefulness</td>
<td>6.4 (0.80)</td>
</tr>
</tbody>
</table>

\(^a\)Score range: 1 (low) to 7 (high).

According to the results of the user evaluation survey (Table 3), the favorite game (derived from the score of question 1 and question 2) was Race the Sun, followed by Fruit Ninja, Crate Whacker, and Kaiju Carnage. There was no difference in the distribution of the games for both general and fun categories, but a significant difference was observed in Crate Whacker, where amputee participants appeared to less enjoy Crate Whacker compared with able-bodied participants (\(P<.001\)).

Table 3. Results of the user evaluation survey.

<table>
<thead>
<tr>
<th>Category</th>
<th>Score</th>
<th>Category</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Able-bodied participants</td>
<td>Amputee participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crate Whacker</td>
<td>Race the Sun</td>
<td>Fruit Ninja</td>
</tr>
<tr>
<td>General</td>
<td>3.6</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Fun</td>
<td>3.6</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Input</td>
<td>3.6</td>
<td>3.8</td>
<td>4.0</td>
</tr>
<tr>
<td>Game control</td>
<td>4.0</td>
<td>3.8</td>
<td>4.2</td>
</tr>
<tr>
<td>Motivation</td>
<td>3.8</td>
<td>3.8</td>
<td>4.4</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

An important finding from this study lies in the usage of the platform in 10 1-hour sessions within a 4-week period. The BBT is a standard test to assess hand functionality [34,35]. In this study, the test was used in a virtual setup in order to check the participants’ ability to control and operate a myoelectric hand, as the mechanism utilized to pick and place an object in the virtual setting is similar to that for a myoelectric prosthesis. Compared to the mean conventional BBT score for a healthy person as reported by Mathiowetz [34], which is in the range of 78.9 (SD 15.3) to 75.1 (SD 11.1) per 60 seconds, the score for the virtual setup was very low (approximately 74% lower score). However, the mean score increased from session 1 to 10.

The MVC test is a standardized method for measurement of muscle strength [36]. The results obtained demonstrate improvements in muscle strength from pregaming to follow-up assessments. A high increment was observed among amputee participants after playing the game, which is consistent with the finding in the study by Prahm et al [32]. This result may be associated with the electrode resistance that may be affected by a change in body temperature or sweat, which, at the same time, proves that the use of video games in upper limb amputee rehabilitation is not fatiguing for the participants. It can be safely said that certain improvements in EMG control could be detected by the MVC test. Besides, from the results, MVC values recorded for able-bodied participants were greater than those for transradial amputee participants, with the lowest MVC value recorded for an amputee participant with congenital amputation, which may have been associated with muscle atrophy. Atrophy of the remaining muscle occurs when the muscle is not used for long periods of time, and the EMG signal becomes very weak after amputation [37]. Male participants recorded greater MVC values compared to female participants.

Based on the IMI questionnaire and user evaluation survey conducted, the motivational aspects of training gamification were traceable. Participants agreed that the activity can induce desirable physiological changes that are consistent with the findings in the study by Lohse et al [10]. Pressure (psychologically) could be felt by amputee participants while playing the Race the Sun game because of its difficulty level. Nevertheless, the participants were still motivated to play the game according to the survey conducted. It has been reported that enjoyment, control mechanisms, music, direct feedback on scores, and ability to receive upgrades influence a player’s motivation [32], and choice, rewards, and goals lead to increased engagement [10].

Aside from game suitability and main player control, the four games were included in this study because each of the games involved a different engaging element. Crate Whacker is continuous without a specific time limit, scoring system, and...
level upgrade. Race the Sun and Fruit Ninja have both a scoring system and specific time to complete the game, but the player can receive upgrades only in Race the Sun. Meanwhile, in Kaiju Carnage, a specific time limit is available, but the player receives no score or upgrade. All games, except for Kaiju Carnage, have no background music. The majority of the participants reported that they enjoyed and wanted to continue playing Race the Sun and Fruit Ninja after the training period, which indicated increased participant motivation as reported in previous studies [25-27]. Both of these games provide direct feedback for the player interaction, and the player can see the timer and score obtained clearly while playing, which is supported by the findings of Liepert et al [15]. According to question 3 and question 4, participants preferred the control in Fruit Ninja, as they felt that they were slicing the fruit in the game with their own hands. A significant difference in the score distribution was observed in Crate Whacker and Kaiju Carnage under game control input (P<.001), indicating that amputee participants had control difficulty in these games. In terms of motivation (question 5), all of the participants only felt less motivated while playing Kaiju Carnage owing to lagging in EMG control and long loading time. Participants also reported that this game showed no score while playing, had a low level of difficulty, and was not interesting. In addition, all amputee participants appeared to be struggling to imitate the clawing gesture, but this was not the case with able-bodied participants. This part of the study showed that engaging elements contribute to increasing a player’s motivation during rehabilitation.

Conclusions
This study demonstrated that a video game–based rehabilitation protocol can be used as a complementary method for upper limb amputee rehabilitation. Participants showed greatly improved muscle strength, coordination, and control. A good muscle condition and induced neuroplasticity enabled better control in the BBT, which is related to readiness to use a myoelectric prosthesis. Participants agreed that this activity is beneficial to them both behaviorally and psychologically. According to the questionnaire responses, it could be shown that the use of video games maintains patient interest during training, therefore improving patient motivation and exercise intensity and enhancing adherence throughout the rehabilitation period. The engaging elements in the game could be identified with the questionnaire. Video games with high intensity and difficulty create pressure for the player, but not all Myo-controlled video games are suitable for transradial amputee rehabilitation.

Acknowledgments
This work was supported by University of Malaya (RU013-2017).

Conflicts of Interest
None declared.

References


Abbreviations

BBT: box and block test
EMG: electromyography
IMI: Intrinsic Motivation Inventory
MVC: maximum voluntary contraction
VR: virtual reality

© N A Hashim, N A Abd Razak, H Gholizadeh, N A Abu Osman. Originally published in JMIR Serious Games (http://games.jmir.org), 04.02.2021. 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 http://games.jmir.org, as well as this copyright and license information must be included.
Serious Games for Improving Technical Skills in Medicine: Scoping Review

Tycho Joan Olgers\textsuperscript{1*}, MD; Anne Akke bij de Weg\textsuperscript{1*}, BSc; Jan Cornelis ter Maaten\textsuperscript{1*}, MD, PhD
Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
\*all authors contributed equally

Corresponding Author:
Tycho Joan Olgers, MD
Department of Internal Medicine
University of Groningen
University Medical Center Groningen
Hanzeplein 1
Groningen, 9700RB
Netherlands
Phone: 31 0503616161
Email: t.j.olgers@umcg.nl

Abstract

Background: Serious games are being used to train specific technical skills in medicine, and most research has been done for surgical skills. It is not known if these games improve technical skills in real life as most games have not been completely validated.

Objective: This scoping review aimed to evaluate the current use of serious games for improving technical skills in medicine and to determine their current validation state using a validation framework specifically designed for serious games.

Methods: We used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines. A multidatabase search strategy was adopted, after which a total of 17 publications were included in this review.

Results: These 17 publications described five different serious games for improving technical skills. We discuss these games in detail and report about their current validation status. Only one game was almost fully validated. We also discuss the different frameworks that can be used for validation of serious games.

Conclusions: Serious games are not extensively used for improving technical skills in medicine, although they may represent an attractive alternative way of learning. The validation of these games is mostly incomplete. Additionally, several frameworks for validation exist, but it is unknown which one is the best. This review may assist game developers or educators in validating serious games.

(JMIR Serious Games 2021;9(1):e24093) doi:10.2196/24093

KEYWORDS
serious games; technical skills; ultrasound skills; validity of serious games

Introduction

Point-of-care ultrasound is an important bedside diagnostic tool for various specialties. For internal medicine, it is a relatively new tool, and educational programs have been created in The Netherlands for residents and internists to become competent in ultrasound [1]. Learning to make the right probe movements and constructing a 3D mental image from a 2D screen image may cost time. To assist in training eye-hand coordination with an ultrasound probe, a serious game involving a 3D-printed probe and an underwater game is under development in The Netherlands [2]. However, it is not known if this game will actually improve ultrasound skills (technical skills of probe movements and thereby image optimization) in real practice. To the best of our knowledge, there is no serious game available at this moment for learning ultrasound skills. A review in 2012 showed that some games were available to train other technical skills like laparoscopic psychomotor skills, but none of the serious games had completed a full validation process [3]. In this review, we aimed to explore the current use of serious games for training technical skills in medicine, including...
personal factors of influence while playing these games, and we determined their validation status using a framework for assessing the validity of serious games [4,5]. Knowledge about validation and the current use of serious games for technical skills may provide useful information to develop games for training ultrasound skills.

We have determined the following research questions: (1) Which games exist for training technical skills in medical education or practice? (2) What is known about the validity of these games? (3) Which personal factors influence the performance in these games?

**Methods**

**Identification of Relevant Studies**

We conducted a scoping review using the recommended items from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extensions for Scoping Reviews (PRISMA-ScR) guidelines [6]. We included original studies investigating serious games for technical skills in health care. Studies were excluded if they (1) evaluated nontechnical skills, such as cognitive skills; (2) included a game that was designed as a therapy for patients or to teach anything other than a medical intervention; (3) did not have full text available; (4) were written in a non-English language; and (5) only described a simulator instead of a serious game.

The databases PubMed, Cochrane Library, EMBASE, and CINAHL were searched on April 9, 2020, using the following terms or abstracts of these terms without limitation of published date: serious game, video game, computer game, education, teaching, training, and skill. This search resulted in 2006 articles (Figure 1). One Author (AbdW) screened all articles and removed duplicates (n=832). After reading the title and/or abstract, another 764 articles were excluded, as they did not concern medical skills, and 282 were excluded for other previously defined reasons (only simulators, describing only cognitive skills, or no game at all). The inclusion and exclusion criteria are presented in Textbox 1. From the 128 remaining articles, the full text was obtained, after which another 101 were excluded for the previously mentioned reasons that could not be determined by the title/abstract. Additionally, we excluded one article because the abstract and full text were not available, four articles because they had non-English text, and 10 articles because they concerned only conference abstracts. The remaining 11 articles were critically assessed, after which six additional articles were found that had not been included in our original search (May 29, 2020); one article was published after our query, and two additional articles were found with a specific google search for “arthroscopy VR Tetris game.” The article describing a game for arthroscopy, included in our primary search, refers to it as the “arthroscopy VR Tetris game.” A specific search on Google for this term and VirtaMed, the operating platform, revealed that this game appears to be part of the ArthroS FAST simulator. A search on PubMed for “ArthroS FAST” and “arthroscopy” produced one article, which, in turn, cited another relevant article about this simulator. However, it was described as a simulator and not as a game, and therefore, was not found in the primary search. Finally, three articles were only found with a specific google search based on two conference abstracts related to the primary search. Full text was not published but could be found on the internet separately. Our strategy described above identified 17 articles to be included in this review.
Figure 1. Flow chart of study inclusion. KTS: Kheiron training system.

Selected articles (2006) from PubMed (594), EMBASE (762), Cochrane Library (262), CINAHL (388)

Duplicates (832)

Title/abstract screening after removing duplicates (1174)

764 articles excluded for
- Patients (371)
- Public Health and other populations (393)

282 articles excluded for
- Simulators (117)
- Knowledge or cognitive skills (85)
- Other/no game (80)

Full text screening of remaining articles (128)

101 articles excluded for
- Other populations (10)
- Simulation (13)
- Knowledge or cognitive skills (36)
- Other/no game (42)

15 articles excluded for
- Conference abstract Underground (4)
- Conference abstract KTS (2)
- Conference abstract other (4)
- Not in English (4)
- No abstract and full text (1)

11 articles included
- Underground (8)
- Arthroscopic VR Tetris game (1)
- Relive (1)
- Orthopedic blood management game (1)

6 articles included after search on Pubmed and Archivo Digital UPM on May 29, 2020
- Underground (1)
- Kheiron Training System (3)
- ArtroS™ FAST (2)
Textbox 1. Inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The game was designed to teach a medical intervention usually performed by health care personnel.</td>
</tr>
<tr>
<td>- The game was designed to teach a technical skill.</td>
</tr>
<tr>
<td>- The technical skill was performed by the player and simulated the real-life technical skill.</td>
</tr>
<tr>
<td>- The game was described as a serious game in at least one article.</td>
</tr>
<tr>
<td>- The article was available in full text and in the English language.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>- The game was designed as a therapy for patients or to teach anything other than a medical intervention.</td>
</tr>
<tr>
<td>- The game was designed to teach non-technical skills (not involving hands-on activity), for example, only knowledge, cognitive skills, and attitudes.</td>
</tr>
<tr>
<td>- The technical skills used in the game do not resemble real-life technical skills (training of cognitive skills).</td>
</tr>
<tr>
<td>- The approach was described solely as a simulator or other medium (eg, courses, online modules, and mannequin).</td>
</tr>
<tr>
<td>- The conference abstract or article was available in a non-English language.</td>
</tr>
</tbody>
</table>

Validity Types for Assessment of Games

Validity in game design for technical skills means that playing the game will actually improve the specific skill in real life. There are several frameworks used for assessment of game validity. The classical framework consists of five different types of validity, and more integrative models are exploring different sources of validity [7]. However, as most reported studies on game validity for technical skills use the classical framework, we have chosen to use that framework for this review. This classical framework consists of the following five different phases of validity: content validity, face validity, construct validity, concurrent validity, and predictive validity [4,5]. Content validity concerns the content of the game to be legitimate (eg, its specifications: Is the game complete and correct, and has nothing but the intended construct [no additional content other than what it was designed for]?). Face validity means that the game appears to be similar to the construct it attempts to represent and is essentially the concept of the game (Do educators or trainees view it as a valid way of instruction?). Construct validity means the game actually measures (or trains) what it intends to measure (Is the game able to measure different skills?). It can be determined by testing prototypes and comparing scores of experts in real life to those of novices. The last stage integrates the construction phase with performance in real life. Concurrent validity reflects the correlation between the performance in the serious game and the performance with the actual instrument. Predictive validity relates the performance in the game to outcomes in reality or predicts skills in real life. In theory, this may require a randomized controlled trial. If the type of validity was not explicitly mentioned, we interpreted the experiment that was conducted and scored the applicable validity.

Results

Search Strategy

The abovementioned search strategy resulted in 17 articles describing five different serious games to train technical skills in health care. We will discuss these five games and their current validity state.

Underground Game

The “Underground” game is the most extensively studied and described. A total of nine articles discussed this Nintendo Wii-U–based game for training basic laparoscopic skills [8-16]. The game was released in 2015 and uses two Wii remote controllers in a custom-made laparoscopic tool shell. The aim is to save robots in a fictional mining world by demolishing and rebuilding the environment. The learning objectives include learning inverted movements, eye-hand coordination, depth perception, and ambidexterity. The face validity, construct validity, and concurrent validity have already been demonstrated and published. It has been shown that playing the game in advance to laparoscopic simulated tasks increased skills [9,16]. Additionally, a study using the game as a preoperative warm-up for 15 minutes showed improvement in task performance [12]. However, the final stage of predictive validity has not been completed yet. This may require comparing surgical skills in the operating theatre between surgeons playing and those not playing the game before the surgery. Several participant characteristics were assessed during these studies. Men outperformed women, and prior video game experience was correlated with “Underground” game scores, although independence of these two variables could not be established as women had less video game experience [10].

Arthroscopic VR Tetris Game

The setup integrates the well-known game “Tetris” into a virtual reality platform for arthroscopic training. The platform consists of a dome with several entry portals for the camera and graspers, and a video screen. The participant can manipulate the Tetris blocks to the preferred position before putting them down. As in the real Tetris game, a line is cleared if it is completely filled with blocks. The aim is to clear 10 lines. The learning objective in this game is to train motor skills, such as opening and closing of graspers and eye-hand coordination. A construct validity study used this setup but with a different assignment, consisting of three activities, and compared the following three groups of users: postgraduate students, fellows, and faculty [17]. Strikingly, the combined scores of the three activities did correlate with year of training but not with prior total...
arthroscopies performed. It would be expected that a higher year of training relates to a higher number of arthroscopies performed and therefore higher scores. Unfortunately, an explanation of this finding was not provided. It is possible that these associations were not significant because the sample size was small or because the game design itself was unable to discriminate the three groups. It is important to emphasize that two validation studies used the arthroscopic simulator setup but not the serious game Tetris. This means that the serious game itself was not validated and that the setup was in fact an arthroscopy simulator. One study with the Tetris game showed that residents performed better with their dominant hand, but this difference disappeared in experienced surgeons [18]. Unfortunately, scores between residents and surgeons were not compared. Interestingly, the second study using the arthroscopic setup showed a gender difference in performance unrelated to previous experience [19].

Kheiron Training System
The Kheiron training system is a serious game for minimally invasive surgery training. The setup includes a box trainer with a box and a camera inside it, real laparoscopic instruments (in contrast to the game “Underwater” that uses the Wii console), a computer, and a monitor. The game is about a young alchemist who has to find the Philosopher’s stone by completing different recipes. Two articles provided technical details of the setup and machine learning, but without any kind of validation of the game itself. Only one additional article described the start of the content validity of this game [20]. Other validation studies were announced but have not been published yet.

Relive Game
The setup of Relive consists of a motion detection device (Kinect version 1; Microsoft Corp), a Resusci Anne mannequin, and a laptop. The game is staged on the planet Mars where chest compressions have to be performed on a person with real-time feedback. It can be played in tournament mode. The game was evaluated by a small study with 65 students who played the game at three different time intervals [21]. After a few months, chest compression depths were better than at baseline. However, there was no comparison with students who did not play the game, and no validation was performed in terms of concurrent and predictive validity.

Orthopedic Blood Management Game
This game consists of a computer, a screen, and a haptic device to manipulate surgical instruments and has been developed to train eye-hand coordination by manipulating instruments to stop bleeding on surfaces and in a virtual patient. The game was tested with students, and a subsequent questionnaire indicated that they found the game realistic and helpful, which is the first step in determining content and face validity [22]. Other validation studies have not been published.

An overview of the included articles, the games they discuss, and the validity types is provided in Table 1 [8-24].
Table 1. Articles included in the review.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Article title</th>
<th>Game</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jalink et al [8]</td>
<td>Construct and concurrent validity of a Nintendo Wii video game made for training basic laparoscopic skills</td>
<td>Underground</td>
<td>Construct and concurrent validity</td>
</tr>
<tr>
<td>IJgosse et al [9]</td>
<td>Saving robots improves laparoscopic performance: transfer of skills from a serious game to a virtual reality simulator</td>
<td>Underground</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>IJgosse et al [10]</td>
<td>Construct validity of a serious game for laparoscopic skills training: validation study</td>
<td>Underground</td>
<td>Construct validity</td>
</tr>
<tr>
<td>Rosser Jr et al [14]</td>
<td>Impact of Super Monkey Ball and Underground video games on basic and advanced laparoscopic skill training</td>
<td>Underground</td>
<td>Concurrent validity</td>
</tr>
<tr>
<td>Harrington et al [16]</td>
<td>Playing to your skills: a randomised controlled trial evaluating a dedicated video game for minimally invasive surgery</td>
<td>Underground</td>
<td>Construct and concurrent validity</td>
</tr>
<tr>
<td>Tofte et al [17]</td>
<td>Knee, shoulder, and fundamentals of arthroscopic surgery training: validation of a virtual arthroscopy simulator</td>
<td>Arthroscopic VR (Tetris) game</td>
<td>Construct validity</td>
</tr>
<tr>
<td>Pedowitz et al [18]</td>
<td>Asymmetry in dominant/non-dominant hand performance differentiates novices from experts on an arthroscopy virtual reality serious game</td>
<td>Arthroscopic VR (Tetris) game</td>
<td>Construct validity</td>
</tr>
<tr>
<td>Walbron et al [19]</td>
<td>Evaluation of arthroscopic skills with a virtual reality simulator in first-year orthopaedic residents</td>
<td>Arthroscopic VR (Tetris) game</td>
<td>Construct validity</td>
</tr>
<tr>
<td>Sanchez Peralta et al [20]</td>
<td>E-learning serious game for surgical skills training: Kheiron training system</td>
<td>Kheiron Training System</td>
<td>Content validity</td>
</tr>
<tr>
<td>Sanchez Peralta et al [23]</td>
<td>Serious game for psychomotor skills training in minimally invasive surgery: Kheiron Training System</td>
<td>Kheiron Training System</td>
<td>Only technical description of the setup, no validation of the game</td>
</tr>
<tr>
<td>Semeraro et al [21]</td>
<td>Kids (learn how to) save lives in the school with the serious game Relive</td>
<td>Relive</td>
<td>Construct validity</td>
</tr>
<tr>
<td>Qin et al [22]</td>
<td>Learning blood management in orthopedic surgery through gameplay</td>
<td>Orthopedic blood management</td>
<td>Content and face validity</td>
</tr>
</tbody>
</table>

Discussion

Principal Results

This review provides an overview of the currently used serious games for improving technical skills in health care and the subsequent validation process. We reviewed 17 articles describing five serious games available for improving technical skills. The game “Underground” has been the most extensively validated, including content, face, construct, and concurrent validity. Most other games only had a description of the initial steps in the process of validation. This means that we are not sure if playing the game will lead to better performance of that specific skill in clinical practice. Most serious games for technical skills need additional validation studies. Guidelines on how to perform validation for serious games may assist game designers and educational experts to develop these games. It is necessary for serious games to be well constructed and evaluated and to impact trainees’ performance in real life, especially if expensive software or equipment is needed to play the game.

Frameworks for Validation

Validation of serious games is the process of collecting and interpreting validity evidence, which, in this case, is used to evaluate the appropriateness of a game for improving technical skills in real life [25]. The classical validation framework identified at least three different “types” of validity (content, construct, and criterion). Criterion validity includes correlational,
concurrent, and predictive validity and denotes the correlation between actual test scores and the “true” (criterion) scores, for example, the correlation with a gold standard. The specifically designed framework for serious games, suggested by Graafland and Warmelink, also uses content and construct validity but adds concurrent and predictive validity instead of criterion validity [4,5]. A more contemporary framework was proposed in 1989 and finally adopted as standard in this field in 1999 and 2014 [26]. Many elements of the classical validation framework are recognizable in this framework, including the construction of the game and its effect on task performance in real practice. It consists of the following five sources of evidence: content, internal structure, relationship with other variables, response process, and consequences [14,15]. Finally, the most recent validation framework was proposed by Kane [27]. The model of Kane is based on inferences and consists of scoring, generalization, extrapolation, and implication. If we apply this framework to serious games, it would start with a player who has a specific score (performance of technical skills) in the game. We assume that the score reflects the overall level of performance, but this score is very dependent on the scoring system/game itself. Multiple scores (or game levels) are combined to generate a total score, assuming this better reflects the performance (technical skills in our case) across the whole test domain (internal consistency). The generalization of the score still deals with performance in the test world and reflects how well the selected test items (scores) represent all of the theoretically possible items. Next, this test world performance is extrapolated to the real world, assuming that this test performance also reflects the skills in real life. Evidence to support extrapolation can be collected by comparing test results with a conceptually related real-world assessment. The final stage is the impact/consequence of this assessment (eg, performance in the game) on the real world (eg, clinical performance, patient safety, length of training, and pass/fail standard). Important questions are as follows: Will playing the game improve or predict technical skills in real life and what are the potential consequences for the trainee?

Although different frameworks for serious games may be used, the validation has the following two key elements: evidence must be collected about the construct of the game itself and its effects on performance in real life. The type of evidence may vary across different games and stages of validation.

Serious Game, Gamification, and Simulation

There is considerable overlap between a serious game, gamification, and a simulator. A serious game is an interactive computer application, with or without specific hardware, that has a challenging goal, is engaging, incorporates some scoring system, and increases the skills, knowledge, or attitudes of its user [3]. These games are designed for specific objectives and therefore differ from commercial video games. A serious game differs from a simulator or gamification in that it uses another context than the actual performance in real life. The game “Underground” is an example of a serious game. In this game, no surgical task is performed or simulated, but the goal is to improve surgical skills. In gamification, there is addition of a game or gaming elements to a nongame context. The game “Relive” is an example of gamification. It uses a normal mannequin to train chest compression but with a scoring element, and the mannequin is “lying on Mars.” A simulator is a device that enables the operator to reproduce or represent under test conditions those phenomena that are likely to occur in actual performance. In health care, simulators can be high fidelity, which means they have a high resemblance to the actual context, for example, a fully equipped operating theatre with a mannequin as a patient instead of a real patient. In this review, the “Arthroscopic VR Tetris game” is actually a simulator with realistic instruments, but instead of a virtual patient, it uses the game Tetris, although the same setup is also used for more realistic simulations.

We have attempted to identify factors that influence performance within a game. Unfortunately, the results were inconsistent and differed between games and simulators. It seems that previous gaming experience is at least an independent predictor of game performance. It is unknown if there is also gender inequality. These are interesting topics for future research.

The idea for this review originated from our interest in multiple learning modalities for learning ultrasound skills. However, there are no specific serious games for learning ultrasound skills available at this moment. It is noteworthy that one article describes a serious game for ultrasound-guided needle placement [28]. Although the authors use an interesting setup for learning needle placement, we excluded this article from our review, because in our opinion, it is an ultrasound simulation setup with some gaming elements (gamification) but not a serious game.

Limitations

The literature selection process was primarily done by one reviewer, which may have caused selection bias. However, we used stringent criteria formulated in advance of our search and we received assistance from our experienced librarian. Additionally, the first author cross-checked PubMed for missed publications, although an extensive second search was not performed. Nevertheless, we are confident that no relevant publications have been missed. The search strategy included the word “skill” to eliminate serious games concerning cognitive skills. However, relevant articles describing serious games for technical skills using different words for skills may have been missed, although in the references, we did not find any relevant articles without the word “skill.” Moreover, an additional search on PubMed with “psychomotor skills” or “psychomotor performance” did not result in additional articles. It is possible that serious games are being used for teaching technical skills but without any publication in the medical literature. We were not able to address the specific quality of each study as there are no specific quality criteria for validation studies of serious games. We excluded four non-English articles and 10 conference abstracts. We translated and reviewed them, but they were not of additional value. Thus, the exclusion did not cause relevant selection bias.

Conclusions

To date, only a few serious games exist for the training of technical skills in medical education, and serious games for learning ultrasound skills are lacking. Factors predicting performance in serious games are only briefly known. The
majority of games still need full validation. This is especially true if they require expensive software and/or hardware. Serious games can be evaluated with the classical concept of validation consisting of content validity, face validity, construct validity, concurrent validity, and predictive validity, although more integrative frameworks are advocated. This review may help serious game developers in the validation process of their games. Despite the specific process of validation, the ultimate goal of serious games is to improve technical skills in real life in a fun way.

Conflicts of Interest
None declared.

References


Edited by N Zary; submitted 03.09.20; peer-reviewed by W IJgosse, X Wang; comments to author 04.10.20; revised version received 30.10.20; accepted 14.12.20; published 25.01.21.

Please cite as:
Olgers TJ, bij de Weg AA, ter Maaten JC
Serious Games for Improving Technical Skills in Medicine: Scoping Review
JMIR Serious Games 2021;9(1):e24093
URL: http://games.jmir.org/2021/1/e24093/
doi:10.2196/24093
PMID:33492234

©Tycho Joan Olgers, Anne Akke bij de Weg, Jan Cornelis ter Maaten. Originally published in JMIR Serious Games (http://games.jmir.org), 25.01.2021. 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 http://games.jmir.org, as well as this copyright and license information must be included.
A Web-Based Game for Young Adolescents to Improve Parental Communication and Prevent Unintended Pregnancy and Sexually Transmitted Infections (The Secret of Seven Stones): Development and Feasibility Study

Ross Shegog1, BSc, DipND, DipBMC, MPH, PhD; Laura Armistead2, MPH; Christine Markham1, PhD; Sara Dube3, MPH, CHES; Hsing-Yi Song4, MPH; Pooja Chaudhary1, MD, MPH; Angela Spencer5, MPH; Melissa Peskin1, PhD; Diane Santa Maria6, RN, MSN, DrPH; J Michael Wilkerson1, PhD; Robert Addy1, PhD; Susan Tortolero Emery1, PhD; Jeffery McLaughlin7, MA

1Department of Health Promotion and Behavioral Sciences, School of Public Health, University of Texas Health Science Center Houston, Houston, TX, United States
2Mathematica Policy Research, Princeton, NJ, United States
3The Widen Lab, University of Texas at Austin, Austin, TX, United States
4School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States
5Special Supplemental Nutrition Program for Women, Infants and Children (WIC), Washington, DC, United States
6School of Nursing, University of Texas Health Science Center Houston, Houston, TX, United States
7Radiant Digital, LLC, Vienna, VA, United States

Corresponding Author:
Ross Shegog, BSc, DipND, DipBMC, MPH, PhD
Department of Health Promotion and Behavioral Sciences
School of Public Health
University of Texas Health Science Center Houston
7000 Fannin St, Suite 2668
Houston, TX, 77006
United States
Phone: 1 17135009677
Email: Ross.Shegog@uth.tmc.edu

Abstract

Background: Early adolescent unintended pregnancy and sexually transmitted infection prevention are significant public health challenges in the United States. Parental influence can help adolescents make responsible and informed sexual health decisions toward delayed sexual debut; yet parents often feel ill equipped to communicate about sex-related topics. Intergenerational games offer a potential strategy to provide life skills training to young adolescents (aged 11-14 years) while engaging them and their parents in communication about sexual health.

Objective: This study aims to describe the development of a web-based online sexual health intergenerational adventure game, the Secret of Seven Stones (SSS), using an intervention mapping (IM) approach for developing theory- and evidence-based interventions.

Methods: We followed the IM development steps to describe a theoretical and empirical model for young adolescent sexual health behavior, define target behaviors and change objectives, identify theory-based methods and practical applications to inform design and function, develop and test a prototype of 2 game levels to assess feasibility before developing the complete 18-level game, draft an implementation plan that includes a commercial dissemination strategy, and draft an evaluation plan including a study design for a randomized controlled trial efficacy trial of SSS.

Results: SSS comprised an adventure game for young adolescent skills training delivered via a desktop computer, a text-based notification system to provide progress updates for parents and cues to initiate dialogue with their 11- to 14-year-old child, and a website for parent skills training and progress monitoring. Formative prototype testing demonstrated feasibility for in-home use and positive usability ratings.
Conclusions: The SSS intergenerational game provides a unique addition to the limited cadre of home-based programs that facilitate parent involvement in influencing young adolescent behaviors and reducing adolescent sexual risk taking. The IM framework provided a logical and thorough approach to development and testing, attentive to the need for theoretical and empirical foundations in serious games for health.

(JMIR Serious Games 2021;9(1):e23088) doi:10.2196/23088

KEYWORDS

serious game; intervention mapping; sexual health; adolescents; sexually transmitted infections; teenage pregnancy; parent, communication; intergenerational; mobile phone

Introduction

Background

Early sexual debut in adolescents is a pervasive public health challenge in the United States. Nearly half (46.8%) of high school students reported having engaged in sexual intercourse, whereas 5.6% reported having engaged in sex before the age of 13 years [1]. Furthermore, teenagers and young adults (aged 10-24 years) represent approximately 90% of single parents and half of all new cases of sexually transmitted infections (STIs) 
[2,3]. Sexuality and gender identity typically emerge during early adolescence around 10 to 14 years of age, a period corresponding with early experimentation with sexual behaviors
[4]. However, it is also a time of receptivity to health messages and therefore opportune for interventions to positively influence future sexual health decision making
[4].

School- and clinic-based prevention programs often achieve broad support and success in reducing sexual risk behaviors in young adolescents
[5-14]. Conversely, these programs often face implementation barriers such as perceptions of sex education as controversial, limited time and resources, and lack of fidelity
[14-17]. Furthermore, school- and clinic-based programs face challenges in effectively involving parents as an unintended pregnancy and STI and HIV prevention mediator. A growing body of research supports the importance of parents’ influence on young adolescent risk behaviors; however, many parents express reluctance to discuss sex with young adolescents because of the belief that they lack time, knowledge, or appropriate skills
[18-20]. Young adolescents also report a desire to discuss sexual health-related topics with their parents but feel that their parents need training on how to communicate about these topics
[21]. Web-based and mobile technology may provide utility in reaching young adolescents and parents with novel sexual health skills training programs
[22-24]. Serious gaming offers promise as an innovative and efficacious approach to sexual health education, operating as a forum to promote a common intergenerational experience, and as a catalyst for increased communication
[25-36]. Recent research on parent-based adolescent sexual health education suggests promise for the use of intergenerational games in enabling collaborative engagement across age groups
[37].

Purpose

The purpose of this study is to describe the development of a novel in-home web-based intergenerational game for parent and young adolescent (11-14 years) dyads, the Secret of the Seven Stones (SSS), designed to provide sexual health skills training to young adolescents and to positively impact dyadic sexual health communication. SSS is played on internet-accessible devices through the Adobe Air framework and comprises

18 game levels (each of 45 min of game play) with 50 interactive skills training clusters, 54 card battle sequences, and 7 game-mediated parent-young adolescent partner-engage-plan

(PEP) talks. Players adopt an avatar to negotiate the town of Seven Stones and assume a hero’s quest to liberate the inhabitants when their personal life rules are challenged in contexts of maintaining healthy friendships, understanding puberty and reproduction, having healthy dating relationships, refusing sex, and negotiating safe sexual practices. The player must power-up their wisdom, skill, and support capabilities in these domains (represented in the game as battle cards) in a dojo using skill training and rehearsal mini-games, animations, puzzles, quizzes, role modeling, and peer video. The player can then liberate an inhabitant by winning a card battle, releasing them from misperceptions, poor judgment, and bad decision making. At 7 milestones in the game, the player is cued to have a

PEP talk with their collaborating parent who is a gatekeeper, conferring a code that enables continuation of the quest. PEP talks focus on the concepts and strategies covered in the game and also introduces each of the 7 character traits that are important in navigating one’s life decisions (eg, respect, vision, persistence, caring, responsibility, courage, and integrity). Throughout the game, SMS text prompts notify the parent of the player’s progress, cue them to an upcoming PEP talk, and link them to parent website resources that can assist with the PEP talk, providing progress tracking, supporting videos, and downloadable fact sheets. Theoretical methods and practical applications guide behavioral skills training that draws from 135 performance behaviors and over 1300 learning objectives within 15 sexual health domains encompassing responsible decision making about communication, friendships, dating relationships, sex, and social support. SSS was developed using the intervention mapping (IM) framework
[38]. The development process is described in the context of each of the six IM steps.

Methods

Overview

We developed SSS through a National Institutes of Health (NIH) Small Business Technology Transfer Research (STTR) project collaborative between UTHealth (The University of Texas Health Science Center Houston) and Radiant Creative Group, LLC. Our development team comprised specialists in adolescent sexual health, computer-based interventions, parent-child
communication, and digital media development. The Parent-Youth Advisory Group (P-YAG) provided conceptual guidance and formative evaluation. Parents (n=20) and young adolescents (n=19, aged 11-14 years) were recruited through flyers, targeted Facebook advertisements, and word of mouth. Young adolescents were mainly female (13/19, 68%), mean 12 (SD 0.28) years old, African American (9/19, 47%), and White (8/19, 42%). Parents were mainly mothers (17/20, 85%), African American (8/20, 40%), White (9/20, 45%), and Hispanic (3/20, 15%). IM, a 6-step framework for developing evidence- and theory-based intervention programs, guided our development process (Table 1) [38]. Our study was approved by the University of Texas UTHHealth Institutional Review Board. At the initial P-YAG meeting, parents and young adolescents signed consent and assent forms, respectively.
**Table 1.** Intervention mapping steps with associated tasks and intermediate development products.

<table>
<thead>
<tr>
<th>IM steps</th>
<th>IM tasks</th>
<th>Intermediate development products</th>
</tr>
</thead>
</table>
| Step 1: Assess need & develop a logic model of the problem | • Establish and work with a planning group.  
• Describe the context for the intervention, including the population, setting, and community.  
• Conduct a needs assessment to create a logic model of the problem. | P-YAG\(^c\)  
• Literature review—evidence table  
• PRECEDE\(^d\) model |
| Step 2: Develop matrices of change objectives | • State expected outcomes for behavior and environment.  
• Specify performance objectives for behavioral and environmental outcomes.  
• Select determinants for behavioral and environmental outcomes.  
• Construct matrices of change objectives. | Matrices for parent (n=6), youth (n=8), and dyadic (n=1) outcome behaviors comprising performance objectives for parent (n=65), youth (n=70), and dyad (n=8) and learning objectives for parent (n=869), youth (n=781), and dyad (n=72).  
• Conceptual model for SSS\(^e\) game flow (model of change). |
| Step 3: Identify theory-based methods and practical applications for program design | • Choose theory- and evidence-based methods to create change.  
• Select or design practical applications to deliver change methods.  
• Generate program themes, channels, components, scope, and sequence. | Table of content domains (n=9).  
• SSS design document comprising specifications including functional inventory, game flow, screen map design, game mechanics, scripts, character descriptions, and interactive activities. |
| Step 4: Produce program components and materials | • Refine program structure and organization.  
• Prepare plans for program materials.  
• Draft messages, materials, and protocols.  
• Pretest, refine, and produce materials. | SSS game consisting of 18 levels of content.  
• SSS parent website including parent training videos (n=7) and tip sheets (n=10).  
• Pilot test protocols and results:  
  • Manual of procedures.  
  • Usability rating results table (parent & youth) with ratings on ease of use, acceptability, credibility, motivational appeal, and applicability for 2 prototype levels.  
• Qualitative data (parent and youth) on acceptability for in-home use. |
| Step 5: Plan for program adoption, implementation, and sustainability | • Identify potential program implementers.  
• State outcomes and performance objectives for implementation.  
• Construct matrices of change objectives for implementation.  
• Design implementation interventions. | Marketing and commercialization plan for future implementers.  
• Written University of Texas Tech Transfer agreement.  
• SSS game and website revisions for future implementation. |
| Step 6: Plan for evaluation | • Write effect and process evaluation questions.  
• Develop indicators and measures for assessment.  
• Specify evaluation design. | Efficacy study design Manual of Procedures comprising:  
• Study hypotheses and protocols.  
• Baseline and first and second follow-up Questionnaire Development System (QDS) software and paper-based surveys.  
• Qualitative exit interview prompts. |

\(^{a}\)IM: intervention mapping.  
\(^{b}\)Youth indicates young adolescents (11-14 years).  
\(^{c}\)P-YAG: Parent-Youth Advisory Group.  
\(^{d}\)PRECEDE: predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation.  
\(^{e}\)SSS: Secret of Seven Stones.

**Intervention Development**

IM is a stepped framework to guide the development of behavioral interventions, providing a process by which program developers can apply social and behavioral science theories within the practice of health behavior change [38]. It comprises 6 primary steps: (1) assess needs and develop a logic model of the health problem; (2) develop matrices of behavioral change objectives; (3) identify theory-based methods and practical applications to design the program; (4) produce program components and materials; (5) plan for program adoption, implementation, and sustainability; and (6) plan for evaluation [38]. IM is widely used to design disease prevention and disease management interventions worldwide. A recent systematic
review has demonstrated a significant increase in the uptake of disease prevention behaviors associated with IM-based interventions and placebo control groups [39]. IM has been successfully applied in the domain of sexual health, including interventions to promote increased communication between parents and young adolescents on relationships and sex [39,40]. However, few applications of IM have been reported in the context of serious games for health, and to our knowledge, none in the context of intergenerational video games for health [41,42].

**Results**

**Step 1: Assess Need and Develop a Logic Model of the Problem**

In step 1, we conducted a needs assessment to understand the health problem and priority population and to describe a theoretically- and empirically-based model for sexual health behavior (Table 1) [38]. PRECEDE (predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation) provided a framework for developing a logic model of the problem [38]. The model prescribes an analysis of causation for a health promotion problem that accounts for multiple ecological levels as well as the multiple determinants of a health-related behavior and environment. National statistics, data from our previous empirical studies on sexual health, qualitative data from 6 P-YAG focus groups, and a literature review of behavior change theories (principally Social Cognitive Theory, social influence models, and the theory of triadic influence) informed our development of a PRECEDE logic model describing quality of life issues and behavioral, social, and physical influences related to adolescent sexual risk behaviors (Figure 1) [28,43-49]. National statistics indicated that early sexual debut is correlated with increased risk of STIs and unintended pregnancy, increased number of sexual partners, more frequent sex, use of alcohol or drugs before sex, and reduced condom use [50-56]. Possible quality of life consequences includes increased high school dropout, welfare dependence, and negative health outcomes for children of teens [57,58]. Environmental factors include limited parent-child communication about sexual health and parental monitoring (Figure 1) [59,60].

**Figure 1.** Predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation “logic model of the problem” of young adolescent sexual behavior for the Secret of Seven Stones. STI: sexually transmitted infections.

Our needs assessment determined a priority population of young adolescents (aged 11-14 years) and their parents and program goals to increase young adolescent intentions to delay initiation sex until they are older and increase parent-child sexual health communication to delay sexual initiation. P-YAG recommendations for sexual health topics included puberty, sexual behavior, and STIs and skill building on parent-young adolescent communication, negotiation, and decision making [46]. Parents wanted to be a credible and focal resource in their child’s sexual health education, and both parents and young adolescents wanted to be more comfortable and effective in initiating and maintaining the conversation around sexual health [46]. Our needs assessment confirmed that parents and young adolescents enjoyed playing a variety of digital games, often on smartphones, and acknowledged the bonding experience of games which, when played together, were played most frequently at home [46].
Step 2: Develop Matrices of Change Objectives

In step 2, we described the behavioral outcomes, delineated these behaviors into their component parts (performance objectives), specified behavioral determinants, and developed change (learning) objectives (Table 1).

Behavioral Outcomes

Drawing from our needs assessment findings and our previous studies, we identified 15 outcome behaviors that were important for the current program (Table 2) [30,31,46,61]. Outcome behaviors for young adolescents comprised delayed initiation, condom use, contraceptive use, human papillomavirus (HPV) vaccination, healthy peer and dating relationships, HIV and STI testing, parental monitoring, and young adolescent to parent communication. Decision making for sexual risk reduction can follow a self-regulation framework of self-monitoring to determine if behaviors are in accordance with one’s values and consistent with one’s goals, self-judgment to identify the presence of threats to these values, such as risky situations, self-reaction to take appropriate action (e.g., avoidance or using refusal skills), and self-evaluation to assess the success of the chosen action [62]. Young adolescents have demonstrated capability in engaging in self-regulation that can provide a cognitive framework for navigating life challenges and reducing health risks [63-65]. Furthermore, young adolescents have demonstrated achievement in processing extensive sexual health curricula content and effectively translating this into behavioral outcomes [28,29,31]. Outcome behaviors for parents comprised parental monitoring; general parent to young adolescent communication; and parent communication about condom use, contraceptive use, HIV and STI testing, and HPV vaccination. The outcome behavior at the level of the dyad (parent and young adolescent) was dyadic communication.
Table 2. Outcome behaviors for young adolescents, parent, and dyad with performance objectives for the dyadic (parent-young adolescent) communication outcome behavior.

<table>
<thead>
<tr>
<th>Learner–domain</th>
<th>Outcome behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Youth</strong></td>
<td></td>
</tr>
<tr>
<td>Healthy peer and dating relationships</td>
<td>Youth will have healthy peer and dating relationships</td>
</tr>
<tr>
<td>Abstinence</td>
<td>Youth will not have sex</td>
</tr>
<tr>
<td>Condom use</td>
<td>Youth who are sexually active or considering having sex will use condoms correctly and consistently when having sex</td>
</tr>
<tr>
<td>Contraceptive use</td>
<td>Youth who are sexually active will use effective method of birth control along with condoms</td>
</tr>
<tr>
<td>HIV and STD&lt;sup&gt;b&lt;/sup&gt; Testing</td>
<td>Youth who are sexually active will get tested and counseled for HIV and STD and unintended pregnancy</td>
</tr>
<tr>
<td>HPV&lt;sup&gt;c&lt;/sup&gt; vaccination</td>
<td>Youth will complete the 3-dose HPV vaccination series</td>
</tr>
<tr>
<td>Parental monitoring</td>
<td>Youth will establish common rules with parents about supervision and monitoring</td>
</tr>
<tr>
<td>Youth to parent communication</td>
<td>Youth will communicate with their parents about dating, intimate or healthy relationships, and sexual behaviors</td>
</tr>
<tr>
<td><strong>Parent</strong></td>
<td></td>
</tr>
<tr>
<td>Parental monitoring</td>
<td>Parents will monitor their youth’s adherence to personal rules</td>
</tr>
<tr>
<td>Parent to youth communication</td>
<td>Parents will communicate with their youth about dating, healthy intimate relationships, and sexual behaviors</td>
</tr>
<tr>
<td>Condom use</td>
<td>Parents will talk to their youth about condom use when having sex</td>
</tr>
<tr>
<td>Contraceptive use</td>
<td>Parents will talk to their youth about contraceptive methods</td>
</tr>
<tr>
<td>HIV and STD testing</td>
<td>Parents will talk to their youth about getting tested and counseled for HIV and STD and unintended pregnancy</td>
</tr>
<tr>
<td>HPV vaccination</td>
<td>Parents will talk to their youth about completing the 3-dose HPV vaccination series</td>
</tr>
<tr>
<td><strong>Dyad</strong></td>
<td></td>
</tr>
<tr>
<td>Dyadic communication</td>
<td>Parents and youth will interact in a mutually engaging and responsive communication process to achieve shared goals</td>
</tr>
</tbody>
</table>
| Performance objectives for dyadic communication | • PO<sub>d</sub>1: Parent and youth will pick the right time and place (T&P) to talk.  
• PO<sub>d</sub>2: Parent and youth will converse with respect.  
• PO<sub>d</sub>3: Parent and youth will assess the youth’s motivation to engage in the behavior under discussion.  
• PO<sub>d</sub>4: Parent and youth will assess alternative actions to the behavior under discussion and their benefits and consequences.  
• PO<sub>d</sub>5: Parent will share their values and expectations regarding possible behaviors (the behavior under discussion and alternate actions).  
• PO<sub>d</sub>6: Parent and youth will develop the best plan of action together.  
• PO<sub>d</sub>7: Parent and youth will encourage each other to keep communicating openly.  
• PO<sub>d</sub>8: Parent and youth will reflect on what to do the same or differently next time. |

<sup>a</sup>Youth refers to young adolescent (11-14 years).  
<sup>b</sup>STD: sexually transmitted disease.  
<sup>c</sup>HPV: human papillomavirus.  
<sup>d</sup>PO: performance objective.

**Performance Objectives**

We identified 143 performance objectives (sub-behaviors) that are necessary to complete the outcome behaviors. Table 2 shows the performance objectives for dyadic communication.

**Behavioral Determinants of Sexual Behavior**

Once the target behaviors were defined, we used theory and empirical applications of theory, including our literature review and previous studies to guide the identification of determinants that likely influence successful performance. Determinants that have been described as impacting sexual behavior in young adolescents include constructs derived from the Social Cognitive Theory, social influence models, and the theory of triadic influence [28-31,47-49,61]. Programs grounded in these theories demonstrated success [28,29,31,44,45]. The determinants included behavioral capability (declarative and procedural knowledge of risk reduction and communication behaviors),
skills and self-efficacy (capability and confidence to perform risk reduction and communication behaviors), outcome expectations (belief that risk reduction behaviors and communication will lead to important results), perceived norms (belief that significant others believe in and use risk reduction behaviors), and social support (recognition of social others who can assist in risk reduction; Figure 1) [59,60].

Behavioral Determinants of Game Play

We also attended to determinants of game play using motivational theory to optimize learner attention [66-71]. Hypothesized determinants included challenge (defined and personally meaningful goals and uncertain, difficult, and yet attainable game outcomes that are predicated on personal effort), curiosity (through novel and surprising game environments, including novel sensory stimuli, and results that can only be confirmed through play), control (through game environments and consequences that are subject to learner action), and self-efficacy (through successive positive reinforcement provided through play).

Matrices of Change Objectives

We defined the program learning objectives by creating separate matrices of performance objectives (row headings) and determinants (column headings) for each outcome behavior. Table 3 provides an example of a partial matrix for dyadic communication behavior (from the example in Table 2). In each matrix cell, we described the learning objective, related to a particular determinant, which contributes to achieving the performance objective. For example, in Table 3, the dyad needs to demonstrate the capability (skill determinant S1.1) to pick the right time and place to converse (performance objective). For interested readers, matrices are available from the corresponding author.

Table 3. Parent matrix for the dyadic parent-youth communication outcome objective that parents and children will interact in a mutually engaging and responsive communication process to achieve shared goals.

<table>
<thead>
<tr>
<th>POa</th>
<th>Determinants of behavior</th>
</tr>
</thead>
</table>
| PO.1 Parents and child will pick the right T&Pb to talk | • K1.1. State that the right T&P is one where both parent and child are focused and calm.  
• K1.2. Describe the influence of emotions, preconceived thoughts, and distractions on communication.  
• K1.3. State the importance of being aware of these influences and setting them aside before initiating communication.  

| PO.1 Parents and child will pick the right T&Pb to talk | • S1.1. Demonstrate the ability to pick the right time and place to converse.  
• S1.2. Demonstrate the ability to set aside emotional or cognitive predispositions before conversing.  

| PO.1 Parents and child will pick the right T&Pb to talk | • SE1.1. Demonstrate the confidence to pick the right time and place to converse.  
• SE1.2. Demonstrate the confidence in ability to set aside emotional or cognitive predispositions before conversing.  

| PO.1 Parents and child will pick the right T&Pb to talk | • OE1.1. State that picking the right T&P will lead to a more focused and calm discussion.  
• OE1.2. State that reflecting on one’s emotions, preconceived thoughts, and distractions before communication will facilitate open and respectful communication.  

| PO.1 Parents and child will pick the right T&Pb to talk | • PN1.1. State that other parents and children have great communication success when they pick the right T&P.  
• PB1.1. State ways to overcome barriers to selecting a right T&P to communicate (schedule or environment).  

| PO.1 Parents and child will pick the right T&Pb to talk | • PB1.2. Recognize barriers to being aware of and setting aside one’s emotions or cognitions before conversing.  

| PO.1 Parents and child will pick the right T&Pb to talk | • SS1.1. Identify others who can help in arranging a right T&P to converse.  

aPO: performance objective.  
bT&P: time and place.

Step 3: Identify Theory-Based Methods and Practical Applications for Program Design

In step 3, we identified theoretical methods and practical applications to inform program design (Table 1).

Theory-Based Methods

A theoretical method is a general technique that influences the determinants of behaviors. If a young adolescent theoretically requires the knowledge, skills, self-efficacy, positive outcome expectations, positive perceived norms, and engagement of social support to perform sexual risk reduction and communication behaviors, then an effective sexual health
education program needs to elicit positive change in these determinants. We drew from empirical literature and our previous research on sexual health and web-based curricula [38]. Table 4 provides an example of the dyadic performance objective to pick the right time and place to converse. We derived methods to increase skills and self-efficacy that comprised informing and consciousness raising, goal setting, chunking, verbal persuasion, modeling, enactive mastery, and public commitment (Table 4).

Table 4. Partial (example) matrix of methods and applications.

<table>
<thead>
<tr>
<th>Method (and theory)</th>
<th>Practical application</th>
<th>For parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information and consciousness raising</td>
<td>Communication activity describing PEP steps and importance of respectful communication; SDP activity teaching how to use this tool to protect personal rules; RRR activity to manage emotions; and “What kind of friend are you” quiz and activity.</td>
<td>PEP Talks 101 and SDP tip sheets with communication tips and “Ask the Expert” advice on communicating about specific topics and information on planning ahead to protect rules; and parent/youth video testimonials illustrating benefits of talking.</td>
</tr>
<tr>
<td>Goal setting (theories of self-regulation and Social Cognitive Theory)</td>
<td>Prompt to set personal rules before each PEP Talk and develop strategies with parent.</td>
<td>Personal rule to orient communication and expectation.</td>
</tr>
<tr>
<td>Chunking (information processing)</td>
<td>PEP to teach steps of PEP Talk; RRR to manage emotions; and SDP tool to help youth maintain personal rules.</td>
<td>PEP to teach steps of healthy communication and SDP tool to help youth maintain personal rules.</td>
</tr>
<tr>
<td>Verbal persuasion (Social Cognitive Theory)</td>
<td>SDP activity training and practice by presenting youth with pressure lines in various situations and asking youth to select appropriate response and RRR activity to manage emotions.</td>
<td>PEP Talk videos; parent and youth video testimonials illustrating how other parents talk to their children and describing how to talk about family values to set rules; Virtues tip sheet reviewing virtues and how to talk about them; and PEP Talk question prompts.</td>
</tr>
<tr>
<td>Modeling (Social Cognitive Theory)</td>
<td>Parent and youth video testimonials on communication and discussing values.</td>
<td>Parent and youth video testimonials on communication and discussing values.</td>
</tr>
<tr>
<td>Enactive mastery (Social Cognitive Theory)</td>
<td>Communication activity reviewing PEP steps and how to choose right time and place to talk; SDP activity training and practice by presenting youth with pressure lines in various situations and asking youth to select appropriate response; RRR activity to manage emotions; and prompt for youth to enter personal rules and strategies. Content progresses in terms of maturity.</td>
<td>PEP Talk videos; Virtues tip sheet reviewing virtues and how to talk about them; and PEP Talk question prompts.</td>
</tr>
<tr>
<td>Public commitment (trans-theoretical model)</td>
<td>Discussing rules and strategies with parent during PEP Talk and then entering rules into game where they can be viewed by youth and parent throughout the game.</td>
<td>Rules and strategies inform social support.</td>
</tr>
</tbody>
</table>

Table 4. Partial (example) matrix of methods and applicationsa.

aOutcome behavior: Parents and children will interact in a mutually engaging and responsive communication process to achieve shared goals. Performance objective #1: Parents and children will pick the right time and place to talk. Determinant and change objective: Skills (S1.1) and self-efficacy (SE1.1) to pick the right time and place to converse. Youth refers to young adolescent (ages 11-14 years).
bPEP: partner-engage-plan.
cSDP: select, detect, protect.
dRRR: relax, rewind, replay.

Given that we were designing a serious game, we also adopted methods to influence the determinants of young adolescent persistence in game play. To address learner challenge, we designed the game to include goals to accomplish and milestones to reach and provided game scenarios of moderate difficulty [66,67,72-75]. Progress was designed to be reinforced verbally by the dojo master and with performance-based rewards (eg, higher quiz scores provide stronger battle cards and virtue tokens that can assist in winning battles) to impact self-efficacy for game play [47]. To address learner curiosity, we embedded scenarios of real-life challenges in the fantasized content of a gaming motif, featuring novel locations and characters [66,68,69]. We appealed to learner sensory curiosity by providing multiple modalities to convey information (sound, graphics, video, and animation), and provided opportunities for learner control with flexibility in the selection of content exposure and battles [71,76-78].

Practical Applications

Practical applications refer to the mode and context of program delivery that fits with the priority population. These comprise channel, scope and sequence, and theme. We designed the program to operationalize the theory-based methods and to be responsive to needs assessment recommendations from the P-YAG [46].

Theme

We provided 2 underlying themes. The first theme was that young adolescents do not have to “go it alone,” that the parent is a social support, dyadic communication is important, and that the discomfort and lack of confidence to discuss sexual health
topics (by both young adolescent and parents) can be overcome with skills to initiate and maintain the sexual health dialogue. The second underlying theme was that young adolescents have control of their life decisions and that smart life decisions (based on self-regulation by selecting, detecting, and protecting their personal rules) have positive consequences. The game motif was a quest to liberate the citizens of the town of Seven Stones from the control of an evil villain, Frostbyte. The young adolescent is victorious if they can defeat Frostbyte in a final boss battle.

Channels

SSS comprised (1) an adventure game for young adolescent skills training delivered via a desktop computer, (2) a text-based notification system to provide progress updates for parents and cues to initiate dialogue with their young adolescent, and (3) a website for parent skills training and progress monitoring (Figure 2).

Figure 2. The Secret of Seven Stones program scope and sequence of play and content domains (inset). PEP: partner-engage-plan; SSS: the Secret of Seven Stones.

Scope and Sequence

We designed the SSS interactive adventure game to provide sexual health skills training primarily for young adolescents because, although dyads supported a gaming strategy, parent time constraints would not accommodate extensive parental gameplay. As such, we designed SSS to accommodate parents in an adjunct, supporting, and gatekeeper role. The game had 18 levels to accommodate the educational content. Each game level approximated 45 to 60 min of game play and contained multiple short duration (2-5 min) activities. The SSS game sequence required young adolescents to (1) visit a location in Seven Stones and encounter citizens who are facing a sexual health challenge (eg, a conflict between being true to yourself vs maintaining a friendship); (2) enter the Dojo to complete educational activities, which were adapted from pre-existing evidence-based curricula and to power-up on relevant knowledge and skills; (3) complete a challenge quiz to earn battle cards of wisdom, skill, and support; and (4) initiate a card-based battle to liberate the citizen opponent (Figures 2 and 3). Dojo activities comprised interactive and noninteractive animations, mini games, role model videos, and role-playing activities from our previous curricula [28,30,31,43,61]. These activities used methods of consciousness raising, chunking, goal setting, verbal persuasion, modeling, enactive mastery, and planning coping responses and have been demonstrated to be effective in promoting behavior change in previous studies (Table 4) [28,31,43,79,80].
The program title, SSS, derives from (1) the town of Seven Stones, (2) the 7 supportive parental interactions during the quest, and (3) the 7 character virtues acquired at each of these dialogues (Figure 2 inset). SSS was designed to (1) provide a common foundation for talking points drawn from the game experience; (2) train young adolescents (and provide resources to parents) on how to communicate; (3) take a life skills approach that introduces character traits, healthy friendships, and rule setting long before discussing sex; and (4) providing the opportunity for ongoing dialogue to normalize such discussions, allowing a more subtle transition to sexual health topics over time.

The parent is updated on the young adolescents’ progress through text messages (Figure 2). At each of the 7 game milestones the young adolescent is locked out of the game and the parent is cued (via text message and e-mail) to have a PEP talk with their young adolescent. In a PEP talk, parents and young adolescents decide a time and location to have a face-to-face discussion (partner), discuss personal and family rules (engage), and develop goals and strategies to maintain these rules (plan). Upon completion of a PEP talk, the parent provides an unlock code, enabling the young adolescent to receive cards and virtue tokens to use in future battles (Figures 2 and 3).

SSS provides increased challenge by sequencing content from topics with less mature content early in the game (eg, healthy friendships) to more mature content later in the game (eg, sexual risk reduction). Early PEP talks focus on friendships and decision making to normalize discussions and make the transition to later discussions of reproduction and sexual relationships less abrupt and awkward. The SSS is designed to accommodate family values. A control feature allows parents to delay the delivery of more mature content of condoms and contraceptives to a time when they perceived their young adolescent to be developmentally ready. The SSS parent website recommends that children be exposed to all content and that content is provided sequentially ensuring foundational material is mastered before exposure to more mature content.

SSS was designed to respond to game preferences emanating from the focus groups. A general preference for boys was the quest to defeat Frostbyte in a boss battle, with girls to resolve interpersonal relationship conflicts among the citizens of Seven Stones, and both girls and boys to use card games for battle. SSS was designed to be responsive to the needs of lesbian, gay, bisexual, transgender, queer, and intersex (LGBTQI) young adolescents with activities that are inclusive of sexual minority preference (eg, scenarios in which young adolescents can choose a partner of either gender and use gender-neutral names) and with materials (eg, fact sheets) that focus on sexual minority issues (eg, LGBTQI and sexuality defined, self-acceptance, the notion of what normal means, social support [“Who can I talk to?”], things to consider before coming out at home, and organization resources).

We designed an SSS website (Figures 2 and 3) to promote communication skills training for both mothers and fathers to enable parents to be a more credible resource in their young adolescents’ sexual health education. Resources comprised 15 PEP talk and communication role model videos featuring parents and young adolescents and 10 communication tip sheets. PEP talk videos were of 2-min duration, introduced the content of the PEP Talk, updated on the educational content covered by their young adolescent, and provided tips on preparing for their PEP Talk. Testimonial videos showed both mothers and fathers and young adolescent role models describing their real-world experiences.
interactions and positive communication experiences. Tip sheets provided summaries of game content, strategies for engaging their young adolescents in conversation, and exercises to increase communication skills and self-efficacy.

**Step 4: Produce Program Components and Materials**

In step 4, we produced and pilot tested an SSS prototype comprising the first 2 game levels (Table 1).

**Prototype Development**

We designed SSS for installation on desktop computers (both Windows and Mac) using the Adobe Interactive Runtime through a broadband connection. The back end mini Structured Query Language (MSQL) database and parent website were implemented using a web server running Hypertext Preprocessor (PHP) built on the Yii framework and accessible through standard browsers using a broadband connection. The back end database was designed to store game data allowing pause points that allow players to exit and re-enter SSS without loss of game progress [46].

**Prototype Feasibility Testing**

We conducted a 2-week pilot test of feasibility in the homes of 10 dyads to determine functional integrity, acceptability by parents and young adolescents, and to explore psychosocial impact. This sample size is sufficient for usability testing and comprises young adolescents (aged 11-14 years; mean 13.1, SD 1.20 years, predominantly males (7/10, 70%), and of White (5/10, 50%) and Hispanic (3/10, 30%) ethnicity [81,82]. Young adolescents were experienced with games; they reported playing for 5 to 8 hours a week (4/10, 40%), playing first-person shooter and multiplayer games (3/10, 33%), and playing on gaming consoles (6/10, 60%) and cell phones (5/10, 50%). Parents comprised mainly mothers (8/10, 80%) and reported either not having played in the last 3 months (4/10, 40%) or playing for less than 2 hours (3/10, 30%). Most played creative or casual games (10/10, 100% and 8/10, 80%, respectively) and all participants played on their cell phones (10/10, 100%).

**Measurement**

We collected parent and young adolescent self-report data using computer-assisted self-administered surveys on study laptops at baseline and at the 2-week follow-up. Our pilot study enabled the testing of protocols to be employed in subsequent efficacy testing. Parent consent and young adolescent assent were obtained before data collection. The feasibility process measures comprised system access logs and user reports (written and verbal) of program issues. Dyads rated SSS on likeability, ease of use, duration, understandability, credibility, perceived impact, and motivational appeal using previously described rating scales [30]. They also assessed the commercial potential of the SSS (discussed in step 5 below). An exploratory analysis was conducted on the impact of SSS on psychosocial determinants of parent-young adolescent communication regarding sex (perceived quality of communication, self-efficacy for parent-child communication, outcome expectations for communication, communication ability, communication openness, and parent-young adolescent connectedness) and attitudes toward using digital games for learning (using an adapted 12-item scale) [83-87]. Scales were provided as pretests before SSS use and at the 2-week follow-up. The Wilcoxon signed-rank test was used as a nonparametric alternative to a t test as the sample size (n=10 dyads) was too small to assume normal distribution.

**Results of Feasibility Prototype Testing**

The SSS prototype functioned according to specifications with players completing the 2 levels within the 2-week period. Most young adolescents rated SSS as likable and credible (6/10, 60%-10/10, 100%) and helpful in making healthy choices (9/10, 90%; Figure 4). The prototype was rated as more fun or as much fun as other sexual health lessons (5/9, 54%), but it was not rated more favorably than favorite computer games. Conversely, less young adolescents agreed that SSS was easy to use (5/10, 50%) and indicated that they needed help to play (5/10, 50%). These lower ratings were primarily associated with technical issues (reported bugs and system crashes), which were a source of frustration for participants and a barrier to completing the assigned activities and led to navigation difficulties where young adolescents would lose track of their next destination in the town of Seven Stones.

Parents rated the website as likable, easy to navigate, credible, and understandable and the game as useful in helping young adolescents make healthy choices (6/10, 60%-10/10, 100%; Figure 5). Conversely, ratings of website ease of use and acceptability were lower (4/10, 40% and 2/7, 29%, respectively) as was SSS game duration (2/7, 29%).
Figure 4. Young adolescent ratings for prototype levels 1 and 2.
Psychosocial Variables for Communication

Exploratory analysis demonstrated positive change in young adolescent attitudes toward using computer games for learning and parent communication outcome expectations ($P \leq 0.05$), driven principally by increased confidence and perceived ease of learning. Other psychosocial variables for communication were not significantly impacted within the 2-week test period (Table 5).
SSS Completion in Preparation for Implementation

Prototype testing informed our completion of the full 18-level prototype. SSS had demonstrated feasibility and compared favorably to other sexual health education programs. However, design modifications were indicated for both the adventure game and parent website. Program bugs and stalls were tracked and fixed, and the player navigation was modified. An open world architecture allowing the player freedom to move at will and explore the gaming environment, originally supported by the P-YAG, was redesigned to be a more directed sequence of destinations with more clearly articulated instructions. Parents reported website navigational problems when trying to locate the parent information associated with their young adolescent’s game level, largely due to forgetting how to use the site between visits. To create a more intuitive site, visual cues that highlighted the relevant information were provided upon log-in. Printed and electronic parent guides were developed, and a web-based tutorial was included to improve the understanding of navigation.

Step 5: Plan for Program Adoption, Implementation, and Sustainability

In step 5, we planned for SSS implementation and dissemination (Table 1). We developed a commercialization plan in accordance with the STTR funding mechanism, which comprises direct-to-consumer sales and bulk licensing to intermediary organizations.

Direct-to-Consumer Sales

Parents and young adolescents in the feasibility pilot provided information on purchase interest, cost points, barriers, and facilitators for purchase and expected marketing channels. Most young adolescents (6/10, 60%) expressed interest in purchasing SSS if it was for sale. They cited potential barriers to purchase of cost (56%), uncertainty about SSS efficacy (2/9, 22%), and long play duration (2/9, 22%). Most parents (6/10, 60%) were willing to pay at least US $20 for SSS. They cited barriers to purchase of duration (5/10, 50%) and possible misalignment with their values (4/10, 40%), potential facilitators to purchase as testimony from other parents (5/9, 56%), and evidence of effectiveness (4/9, 44%). Parents (4/10, 40%) and young adolescents (5/9, 56%) expected to hear about SSS mainly through school.

Bulk Licensing

Discussions with representatives from third-party distribution channels that promote family wellness (eg, WebMD, Aetna, Humana, and ActiveHealth) resulted in awareness of the SSS proof-of-concept and interest in ongoing discussion as the product matures out of prototype through efficacy testing. Additional market analysis will use the Strategyzer strategic management framework to further develop the business model.
increasing exposure to sexual health content [33]. Most young recruit and retain in sexual health education programs, was also population group that has traditionally been more difficult to (7/10, 70% of our feasibility sample) to play SSS, which is a utility of this strategy. The interest of young male adolescents toward the use of computer games for learning supports the health education program, parents positively rated SSS, reporting while providing an engaging and immersive experience. As a educational content and optimal exposure for behavioral impact. This introduces a tension between ensuring sufficient balance between strategies for behavior change and playability.

**Step 6: Plan for Evaluation**

In step 6, we planned to evaluate the SSS (Table 1). Our SSS evaluation plan comprised a randomized controlled efficacy trial with 85 parent-young adolescent (aged 11-14 years) dyads to test the full SSS game. Dyad psychosocial and communication data will be collected at baseline and at the third and sixth month. Hypotheses would be that dyads accessing SSS will report increased frequency and quality of communication about sexual health and young adolescents will demonstrate greater intentions to delay initiation of sexual behavior compared with those not receiving SSS.

**Discussion**

**Principal Findings**

The SSS represents a novel application of an intergenerational serious game for sexual health education, adding to the limited cadre of home-based programs that facilitate parent involvement in influencing young adolescent behaviors and reducing adolescent sexual risk taking [8,18,46].

SSS is an intergenerational game to the degree that it provides both parents and young adolescents’ roles in the gaming experience and encourages dialogue to accelerate game play. Parents and young adolescents could sit together to play SSS, but currently the game does not allow parents and young adolescents to synchronously play or compete in the game space. Parental time constraints necessitated a gatekeeper role for parents rather than a dual-player mode. This is consistent with what Voida et al [35] have described as the performance/audience role pair that was pervasive in their explorative studies of intergenerational gaming. In this instance, there is a more active performer (usually the young adolescent) and a spectator role (usually the older person). The analogy is that the SSS adventure game is experienced more actively by young adolescents and more vicariously by the parent gatekeeper.

Success in serious game design is predicated on achieving a balance between strategies for behavior change and playability. This introduces a tension between ensuring sufficient educational content and optimal exposure for behavioral impact while providing an engaging and immersive experience. As a health education program, parents positively rated SSS, reporting it to be valuable and credible. The positive impact on attitudes toward the use of computer games for learning supports the utility of this strategy. The interest of young male adolescents (7/10, 70% of our feasibility sample) to play SSS, which is a population group that has traditionally been more difficult to recruit and retain in sexual health education programs, was also encouraging [28,43]. The use of a serious gaming strategy may be intrinsically motivating to this population segment, thereby increasing exposure to sexual health content [33]. Most young adolescents reported that the duration of SSS was just right.

This is encouraging as approximately 13 hours of sexual health curricula exposure is optimal to see the impact on delayed sexual initiation [89]. Conversely, even within the limits of a 2-level feasibility test, most parents suggested that SSS was too long compared with too short or just right (both 29%) despite considerable design efforts to reduce parental burden. Rectifiable bugs in the prototype may have contributed to this perception. It is uncertain to what degree this perception translates to reduced use or attrition of young adolescents using the full SSS program.

Young adolescents rated SSS as fun as other computer games (56% agreement) but, perhaps predictably, none rated SSS as much fun as their favorite computer game. SSS exhibited common gaming features such as the quest motif, characters and bosses, virtues, power-ups, battles, and points. However, there were dissimilarities relating to the educational content, including power-up dojo activities, quizzes, parent updates, PEP talks, and skills training around life skills issues. Furthermore, SSS could not compete with the production value of high-end commercial games that feature richly textured graphics, epic musical scores, massive scope, greater user control, and sophisticated three-dimensional game mechanics. SSS, similar to other serious games for health, may best be marketed as a palatable way to consume health information and training rather than as a direct competitor to commercially available entertainment games. Health-oriented games occupy a commercial niche that offers social value and has the potential to operate in community settings with an aligned mission (eg, schools, work places, clinics). In the context of the home, where there is an array of competing demands, the use of serious games may be more tenuous and contingent on parent and child commitment.

**IM is one of a number of useful development frameworks [34,90,91]. It has been applied for game-based sexual health curricula for middle school students and interventions for gay men [28,30,38,61]. As a general development framework, it has demonstrated utility in designing, developing, implementing, evaluating, and disseminating theory-based serious games and in enabling the development of interventions that provide skills training for complex health behaviors. Commercial product development, under the NIH STTR grant mechanism, represents a novel application of the IM framework despite the burgeoning trend of innovation incubators and academic-corporate partnerships. The commercial dissemination of serious games holds the promise of generating a revenue stream to sustain them. Efforts in this arena have been ongoing for over a decade; however, research on the potential for serious games in sexual health is in its infancy and exemplary models of commercial success are yet to be reported.**

**Limitations**

The findings need to be interpreted in light of study limitations. The pilot study was formative in nature. A small sample size (n=10 dyads), abbreviated intervention dose (2 levels over 2 weeks), and the use of self-selected sampling, although appropriate in this setting for feasibility assessment, were insufficient to assess the efficacy of the game and impact on psychosocial and dyadic communication outcomes. The sample
was inherently biased, attracting parents predisposed to improving communication with their children, and predominantly of mothers, which may have affected the parental content and resources (and hence appeal and relatability) of SSS. Our development was not powered to provide a meaningful comparison of mother and father perceptions of SSS. Mothers and fathers provided similar responses regarding the usability and feasibility of SSS (eg, that SSS was useful in helping young adolescents make healthy choices and that most parents would tell their friends about SSS). Further investigation of varied parental perspectives would be useful. It is possible that the male (father) perspective was underrepresented in the SSS development process. Postponing commencement of the field testing in favor of more extensive alpha testing may have mitigated some program bugs and thus some usability ratings. This is difficult to guarantee, however, and the usability data gathered was valuable despite these issues. The formative pilot testing of 2 levels, although important and useful in product evolution, inherently limited the conclusions we could offer about SSS, as it limited exposure to content topics of communication, healthy friendships, decision making, and sexual health topics that are delivered later in the game. There is a precedent in previous school-based studies to indicate that the content scope and volume is appropriate for the young adolescents, but this remains to be further empirically tested within the home context [28,29,31]. Dyads were not challenged to start having conversations on topics that traditionally cause greater discomfort, a necessary focus for the efficacy study. Program-led discussions are not always well received if they are perceived as contrived or forced and not organic. Thus, naturally occurring conversations are preferred. Future studies will focus on how parent-young adolescent dyads perceive PEP talks and whether these can promote naturally occurring discussions beyond any need for program cues. Despite limitations, the exploratory results were consistent with the content covered and indicative of the potential of exposure to the full program. The criteria by which developers pilot and field test complex interventions can vary as a function of scientific, resource, and timeline constraints, but guides exist to assist in decision making and determining evaluation designs. In this regard, a useful adjunct to IM is the United Kingdom Medical Research Council guidance document on developing and evaluating complex interventions, providing consideration of study designs with case study examples [91].

Important empirical questions remain regarding SSS. SSS allows families to choose the match of parent-young adolescent dyadic combinations (mother or father with daughter or son) based on existing communication dynamics, desire, and logistics. Young adolescents may prefer same-sex caregiver support for sexual health discussions. Further studies pertinent to the field of intergenerational gaming could investigate the differential benefits of alternate dyadic combinations, triadic combinations, and inclusion of siblings in improving communication and family dynamics. Altering the game mechanics to allow for text messaging updates and prompts for 2 parents and adjusting the expected communication dynamics to be inclusive of parents and young adolescents are relatively easy adjustments. Future studies could also contribute to our understanding of the optimal exposure to achieve behavioral impact.

Conclusions

SSS provides a feasible strategy to overcome parent and young adolescents discomfort about discussing sexual health topics and enhancing the skills required to initiate and maintain this dialogue. IM is a useful framework for developing a theoretically and empirically based intergenerational, sexual health computer game (SSS) for in-home use. Further testing to assess the efficacy of the complete SSS program on parent-young adolescent communication is indicated.

Acknowledgments

This work was conducted through a collaborative between UTHealth and Radiant Creative LLC and was supported by grant 1R42HD074324-01 from the National Institute of Child Health and Human Development, NIH.

Conflicts of Interest

None declared.

References


http://games.jmir.org/2021/1/e23088/


45. Peskin MF, Hernandez B, Markham C, Johnson K, Tyrrell S. Sexual health education from the perspective of school staff: implications for adoption and implementation of effective programs in middle school. J Appl Res Child 2011;2(Article 9) [FREE Full text]


58. Shuger L. he National Campaign to Prevent Teen and Unplanned Pregnancy and America's Promise Alliance. URL: https://www.colorincolordo.org/research/teen-pregnancy-high-school-dropout-what-communities-can-do-address-these-issues [accessed 2021-01-12]


83. Miller KS, Kotchick BA, Dorsey S, Forehand R, Ham AY. Family communication about sex: what are parents saying and are their adolescents listening? Fam Plann Perspect 1998;30(5):218-22, 35 [FREE Full text] [Medline: 9782044]


89. Shegog R, Baumler E, Addy R, Peskin M, Thiel M, Tortolero S. Sexual health education for behavior change: how much is enough? Child At Risk 2017;8(1) [FREE Full text]


**Abbreviations**

- **HPV**: human papillomavirus
- **IM**: intervention mapping
- **LGBTQI**: lesbian, gay, bisexual, transgender, queer, and intersex
- **NIH**: National Institutes of Health
- **PEP**: partner-engage-plan
- **PRECEDE**: predisposing, reinforcing, and enabling constructs in educational diagnosis and evaluation
- **P-YAG**: Parent-Youth Advisory Group
- **SSS**: Secret of Seven Stones
- **STI**: sexually transmitted infection
- **STTR**: Small Business Technology Transfer Research
©Ross Shegog, Laura Armistead, Christine Markham, Sara Dube, Hsing-Yi Song, Pooja Chaudhary, Angela Spencer, Melissa Peskin, Diane Santa Maria, J Michael Wilkerson, Robert Addy, Susan Tortolero Emery, Jeffery McLaughlin. Originally published in JMIR Serious Games (http://games.jmir.org), 27.01.2021. 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 http://games.jmir.org, as well as this copyright and license information must be included.