A multidisciplinary journal on gaming and gamification for health education/promotion, teaching and social change

Volume 6 (2018), Issue 4. ISSN: 2291-9279

CONTENTS

Original Papers

Digital Gaming to Improve Adherence Among Adolescents and Young Adults Living With HIV: Mixed-Methods Study to Test Feasibility and Acceptability (e10213)
Amanda Castel, Saba Qasmieh, Daniel Greenberg, Nicole Ellenberger, Tyriesa Howell, Caleb Griffith, Brittany Wilbourn, Kavitha Ganesan, Nadia Hussein, Gabriel Ralte, Natella Rakhmanina. 3

Developing Digital Games to Address Airway Clearance Therapy in Children With Cystic Fibrosis: Participatory Design Process (e18)
Fabio Balli. 44

Emotion-in-Motion, a Novel Approach for the Modification of Attentional Bias: An Experimental Proof-of-Concept Study (e10993)
Lies Notebaert, Ben Grafton, Patrick Clarke, Daniel Rudaizky, Nigel Chen, Colin MacLeod. 61

Creating a Theoretically Grounded, Gamified Health App: Lessons From Developing the Cigbreak Smoking Cessation Mobile Phone Game (e10252)
Elizabeth Edwards, Hope Caton, Jim Lumsden, Carol Rivas, Liz Steed, Yutthana Pirunsarn, Sandra Jumbe, Chris Newby, Aditi Shenvi, Samaresh Mazumdar, Jim Smith, Darrel Greenhill, Chris Griffiths, Robert Walton. 73

A Web-Based Serious Game on Delirium as an Educational Intervention for Medical Students: Randomized Controlled Trial (e17)
Kiki Buijs-Spanjers, Harianne Hegge, Carolien Jansen, Evert Hoogendoorn, Sophia de Rooij. 93

Potential of an Interactive Drug Prevention Mobile Phone App (Once Upon a High): Questionnaire Study Among Students (e19)
Máté Kaptály-Fővény, Eszter Vagdalt, Zsófia Ruttkay, Róbert Urbán, Mara Richman, Zsolt Demetrovics. 102

A Mobile App Delivering a Gamified Battery of Cognitive Tests Designed for Repeated Play (OU Brainwave): App Design and Cohort Study (e10519)
Martin Thirkettle, Jennifer Lewis, Darren Langridge, Graham Pike. 120

Patterns Among 754 Gamification Cases: Content Analysis for Gamification Development (e11336)
Sungjin Park, Sangkyun Kim. 135

Viewpoint

Virtual Reality Clinical Research: Promises and Challenges (e10839)
Bernie Garrett, Tarnia Taverner, Diane Gromala, Gordon Tao, Elliott Cordingly, Crystal Sun. 21
Digital Gaming to Improve Adherence Among Adolescents and Young Adults Living With HIV: Mixed-Methods Study to Test Feasibility and Acceptability

Amanda D Castel¹, MPH, MD; Saba Qasmieh¹, MPH; Daniel Greenberg², BA; Nicole Ellenberger³, MPH; Tyriesa Howard Howell³,⁴, MSW, PhD; Caleb Griffith³, MPH; Brittany C Wilbourn¹, MPH; Kavitha Ganesan⁵,⁶, MPH; Nadia Hussein³, MPH; Gabriel Ralte², ASN; Natella Rakhmanina³,⁷,⁸, MD, PhD

¹Department of Epidemiology and Biostatistics, Milken Institute School of Public Health, George Washington University, Washington, DC, United States
²Media Rez, Inc, Washington, DC, United States
³Children’s National Health System, Washington, DC, United States
⁴Center for Prevention Science, School of Social Work, Rutgers University, New Brunswick, NJ, United States
⁵Henry M Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, United States
⁶Walter Reed Army Institute of Research, United States Military HIV Research Program, Silver Spring, MD, United States
⁷Department of Pediatrics, School of Medicine and Health Sciences, George Washington University, Washington, DC, United States
⁸Elizabeth Glaser Pediatric AIDS Foundation, Washington, DC, United States

Corresponding Author:
Amanda D Castel, MPH, MD
Department of Epidemiology and Biostatistics
Milken Institute School of Public Health
George Washington University
5th Floor
950 New Hampshire Avenue Northwest
Washington, DC,
United States
Phone: 1 202 994 8325
Fax: 1 202 994 0082
Email: acastel@gwu.edu

Abstract

Background: An estimated 50% of adolescents and young adults (AYA) living with HIV are failing to adhere to prescribed antiretroviral treatment (ART). Digital games are effective in chronic disease management; however, research on gaming to improve ART adherence among AYA is limited.

Objective: We assessed the feasibility and acceptability of video gaming to improve AYA ART adherence.

Methods: Focus group discussions and surveys were administered to health care providers and AYA aged 13 to 24 years living with HIV at a pediatric HIV program in Washington, DC. During focus group discussions, AYA viewed demonstrations of 3 game prototypes linked to portable Wisepill medication dispensers. Content analysis strategies and thematic coding were used to identify adherence themes and gaming acceptance and feasibility. Likert scale and descriptive statistics were used to summarize response frequencies.

Results: Providers (n=10) identified common adherence barriers and strategies, including use of gaming analogies to improve AYA ART adherence. Providers supported exploration of digital gaming as an adherence intervention. In 6 focus group discussions, 12 AYA participants identified disclosure of HIV status and irregular daily schedules as major barriers to ART and use of alarms and pillboxes as reminders. Most AYA were very or somewhat likely to use the demonstrated game prototypes to help with ART adherence and desired challenging, individually tailored, user-friendly games with in-game incentives. Game prototypes were modified accordingly.

Conclusions: AYA and their providers supported the use of digital games for ART adherence support. Individualization and in-game incentives were preferable and informed the design of an interactive technology-based adherence intervention among AYA living with HIV.

(JMIR Serious Games 2018;6(4):e10213) doi:10.2196/10213

KEYWORDS

youth; HIV; adherence; video games; Wisepill; adolescents; digital technology; serious games

Introduction

The goals of the US National HIV/AIDS Strategy include increased HIV testing, linkage to care, retention in care, and viral suppression for all persons living with HIV [1]. For an estimated 36,000 adolescents and young adults (AYA) aged 13 to 24 years living with HIV, these are particularly critical goals [2]. AYA accounted for 22% of all new HIV infections in 2015, and those AYA who were perinatally infected are aging into adulthood [3]. Furthermore, with treatment guidelines recommending early initiation of antiretroviral treatment (ART), AYA represent a significant population being prescribed daily ART medications [4]. Despite the high efficacy of new ART drugs and more available fixed-dose single tablet per day ART regimens, adherence to daily treatment remains suboptimal among AYA; less than half (47%) of youth aged 18 to 24 years living with HIV in 2016 were virologically suppressed [5].

As the chronic nature of HIV becomes apparent to youth, placing emphasis on adherence support is essential. As many as 50% of AYA or their caregivers report a failure to adhere to prescribed regimens [6-9], and a review of more than 50 studies on pediatric HIV infection confirmed that 42% to 80% of youth had suboptimal ART adherence [10]. AYA living with HIV are at particular risk of poor adherence due to specific challenges. They are experiencing a developmental period with increased peer pressure, risk-taking, and a desire to be similar to their HIV-negative peers [11]. Autonomy also differs between AYA and adults living with HIV, as younger persons must frequently rely on parents or caregivers for treatment [12,13]. Furthermore, AYA living with HIV may experience a higher frequency of cognitive deficits, depression, and substance abuse issues compared to general populations [14]. Therefore, the development of effective, youth-friendly interventions to improve adherence among AYA living with HIV is critical to improving adherence and treatment outcomes among this population.

Current approaches to ART adherence monitoring and support include text messaging, self-reporting, pill counts, pharmacy refill assessments, and electronic monitoring with devices such as Wisepill, a portable medication dispenser that emits a wireless signal when opened, used for monitoring real-time adherence [15-24]. While a few studies have provided evidence of the feasibility, acceptability, and performance of electronic monitoring of ART adherence, use of the Wisepill dispenser for real-time adherence monitoring is currently understudied; the dispenser appears to be a promising adherence strategy to explore. Digital game-based interventions also have the potential to be a promising approach to increasing treatment adherence among AYA. These approaches capitalize on the daily interests of AYA who are estimated to spend as many as 7 hours per day interacting with computers, cell phones, and video games [25,26].

Combining Wisepill technology and youths’ interest in digital gaming, we aimed to develop an intervention to link on-time Wisepill openings to in-game incentives to increase ART adherence among AYA living with HIV. As part of our gaming design process, we sought to better identify factors associated with nonadherence and identify digital gaming preferences among AYA and their health care providers. The objectives of this analysis were to use a formative research approach to describe barriers and facilitators to ART adherence among AYA living with HIV, identify their preferred digital game characteristics, and assess the feasibility and acceptability of the proposed intervention.

Methods

Study Setting and Approach

The study was conducted by a multidisciplinary research team comprising epidemiologists, qualitative researchers, HIV clinicians, and game development experts in Washington, DC. All study materials were reviewed and approved by the George Washington University and Children’s National Medical Center (CNMC) institutional review boards. Participants were recruited from CNMC Special Immunology Services (SIS), a clinic that provides care for HIV-exposed and infected infants, children, and adolescents in the DC metropolitan region. The SIS primarily serves perinatally and behaviorally infected patients from birth to 24 years old and provides a range of HIV specialty services including medical care, case management, subspecialty referrals, mental health services, and nutritional support. These services are provided by an experienced team of physicians, nurses, social workers, psychologists, and clinical and behavioral researchers.

The formative portion of this study included 3 phases: phase I, game prototype development and subsequent game revisions; phase II, a focus group discussion (FGD) with treatment providers of AYA; and phase III, a series of FGDs among AYA ages 13 to 24 years living with HIV. In phases II and III, each FGD was conducted by 2 moderators: (1) a clinician researcher to facilitate discussions regarding HIV knowledge and adherence and (2) the videogame developer to demonstrate the game prototypes and lead discussions on game design and development.
**Phase I: Game Prototype Development and Revisions**

**Theoretical Framework**

Previous games for behavior change have shown positive results by using the compelling nature of game play to make health education entertaining and shift attitudes and emotions about chronic illness and treatments [27,28]. In our gaming design, we sought to use Wisepill dispensers for real-time adherence measurement and as a mechanism to provide incentives (eg, power-ups) for game play through the granting of microtransaction points for on-time pill box openings. Guided by Bandura’s social cognitive theory [29], a widely used framework that can be used to assess HIV adherence [30-32], we hypothesized that our intervention would positively influence self-efficacy, risk perceptions, knowledge of HIV treatment goals, and social support of study participants. Using this framework, iterations of game prototypes were developed by Media Rez including a project code-named Virus Fighter/Body Voyage, designed to teach players about the effects of HIV on the body and immune system, including important concepts such as poor adherence and medication resistance (Figure 1).

This prototype demonstrated through gaming how HIV would get stronger if the player did not use ART as a weapon against the virus while also reinforcing how HIV is weakened if the player used ART. A second game, Adherence Warrior, was also developed with no explicit intent to be educational that included features that HIV care providers could potentially use as teachable moments to help educate AYA about the importance of ART adherence (Figure 2).

**Game Modifications**

Participant feedback from both health care providers and AYA was analyzed and informed revisions made to the initial game prototypes. This feedback was used to determine the optimal set of game features for the game prototypes used in a subsequent pilot intervention. As a result of the AYA feedback specifically, production on the Virus Fighter/Body Voyage game was suspended. In order to create sufficient game content to engage a variety of participants, prototypes of 3 distinctly different Wisepill-connected games were developed based on FGDs and survey feedback regarding desired game features. Participants would be able to play any of the 3 games at their discretion: a modified version of Adherence Warrior, the Berry Game, and Cat O’Polt, a new game that was developed in order to provide a game that was of moderate difficulty, potentially interesting to AYA, and adventurous. The games and pillbox interactions were also designed to meet important privacy, motivation, and information goals (Table 1).

**Adherence Warrior Game**

Adherence Warrior (Figure 2) features heroes in a fantasy world protecting their home village from invading monsters. The player manages a group of 3 brave adventurers who use weapons and magic to stop the invading monsters. To attack the monsters, 3 or more tiles are required to be matched in a row. Tiles can be moved 3 times for each of the 3 characters, which allows 9 attempts to make a match before the monster’s attack. Players can collect many new kinds of tiles resulting in new kinds of magic spells, skills, attacks, and healing. Players also need energy to progress in the game, and unlike in similar games, players cannot gain this energy by in-app purchases; they can earn it through adherence. Players can figure out unique vulnerabilities in different monsters and use the most effective tiles against those weaknesses.

![Virus Fighter images include HIV-like monsters that would invade CD4 cells.](http://games.jmir.org/2018/4/e10213/)
Figure 2. In the Adherence Warrior game, players act as heroes protecting their village from invading monsters. To attack the monsters, three or more tiles need to be matched in a row. Players need energy to progress in the game, and they can gain this energy by opening their pillbox on time.
Table 1. Features of final games.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy</td>
<td>Making games about patients fighting HIV conveys a risk of unintentionally revealing a patient’s HIV status when they would prefer privacy. To prevent this risk, we changed plans for a Body Voyage game fighting viruses and made 3 games designed to protect patient privacy and minimize the risk that using the intervention would disclose their status.</td>
<td>Increases app engagement by meeting the privacy needs of adolescent and young adult HIV patients.</td>
</tr>
<tr>
<td>Motivation</td>
<td>By making a direct connection between daily adherence and the kind of daily, in-game power-ups that players usually have to pay for with microtransactions, we attempted to motivate patients to take their medication on time. Opportunities to earn greater in-game power and access increases motivation for both engagement with the games and adherence with antiretroviral treatment (ART).</td>
<td>Direct motivation to increase adherence.</td>
</tr>
<tr>
<td>Motivation</td>
<td>The game incentivizes cooperation by letting the player build available healing tiles and defensive tiles by not using them on less-injured heroes but instead saving them up for bigger healing effects on heroes who are more injured. In this way, the game puts the player in a caregiver role.</td>
<td>Having player in caregiver role motivates patients in their current self-care role to adhere to their ART.</td>
</tr>
<tr>
<td>Information</td>
<td>The system knows when patients do not open their pillboxes on time and sends out a feedback reminder when this happens in the form of a text message (which does not disclose anything about HIV). Providing this timely information about adherence affirms adherence behavior and serves as a cue to action in a way that does not breach privacy.</td>
<td></td>
</tr>
</tbody>
</table>

Although the game contains no explicit HIV education (to avoid violating patient privacy), it was designed to have teachable moments that HIV care providers could potentially use to explain complex concepts associated with treatment. Provider FGDs reported that one of the most difficult concepts to teach is the potential for viral resistance to ART, which is commonly taught by metaphor. By creating the vulnerability system in the game, the provider can emphasize this similarity with HIV resistance. This provides a conceptual approach to understand complex adherence concepts and behaviors. In this way, the game not only avoids breaching the privacy of AYA living with HIV but also avoids coming across as didactic. The game itself does not try to lecture the patient through onscreen text but provides experiences designed to be used as teachable moments by caregivers. Enabling educational opportunities and points of connection for providers and patients through the game could be more effective for behavior change than simply having the game display health messages that are often ignored.

**Berry Match**

The objective of the Berry Match (Figure 3) game is to match as many identical berries as possible. If 3 or more berries are matched in a row, the berries will switch places and the 3 matching berries will disappear and new berries descend. Points are awarded for all berries “eaten.” Some levels have limited amounts of moves to win while others can be beaten by reaching the target score. Special blocking and trapping tiles add to the challenge. When the player takes his or her medication on time, they receive in-game currency that can be spent on special power-ups to solve especially tricky levels of the game. In similar commercial games, players must purchase these power-ups with cash through in-app purchases; in our gaming design they can earn the in-game currency by opening their Wisepill boxes on time.

**Cat O’Polt**

The goal of Cat O’Polt (Figure 4) is to direct a cat as it travels through the city collecting as many coins as possible while simultaneously avoiding obstacles. With the right timing, the cat can make huge jumps over obstacles and even vault over buildings. The objective is to achieve the highest score possible before losing game play lives. In our gaming design, players can gain new lives by opening their pillbox on time.

As the next phase of the study, the Wisepill dispenser was integrated within each game and was linked to in-game power-ups for daily Wisepill openings as an incentive to improve ART adherence.

**Phase II: Focus Group Discussion With HIV Providers**

A semistructured FGD using a single category design was conducted among the specialty HIV care provider team at CNMC SIS and was aimed at identifying key facilitators and barriers to ART adherence among AYA and eliciting feedback on the game prototype and use of the Wisepill dispenser. The semistructured format allowed FGD moderators to probe emergent topics and themes during the discussion. Prior to the FGD, providers were administered a brief survey to collect basic demographic and practice-related characteristics. Providers were not remunerated for their participation in the FGD.

**Phase III: Focus Group Discussions With Adolescents and Young Adults**

AYA ages 13 to 24 years living with HIV and disclosed as to their HIV status were recruited during routine clinic visits at CNMC SIS. A series of multiple-category FGDs were stratified by age (13 to 17 and 18 to 24 years) to account for developmental differences and potential variances in game characteristic preferences among participants. A total of 12 AYA participated in 6 FGDs; 3 FGDs among participants aged 13 to 17 years and 3 FGDs among participants aged 18 to 24 years. Participant feedback from the first FGD was analyzed, and revisions were made to the initial game prototypes. During the second stage of FGDs, participants from the initial FGDs as well as AYA who had not previously viewed the games were invited to participate.
Figure 3. In the Berry Match game, players try to match as many berries as possible. Points are awarded for all the berries eaten. Players receive in-game currency in the form of special power-ups by opening their pillbox on time.
In the Cat O’Polt game, players control a cat travelling through the city trying to collect coins while avoiding obstacles. The objective is to achieve the highest score possible before losing all of one’s gameplay lives. The player gains new lives by opening their pillbox on time.

Participants had the flexibility to explore multiple games and provide unrestricted input regarding user interface performance, acceptability of the art and animation, and response to gameplay. Additional input was also provided regarding the feasibility and acceptability of the games with respect to their potential to improve ART adherence.

Before the start of each FGD, participants completed a survey that collected information on their demographics, ART regimens, and current video game use. A separate survey using Likert scale ratings was administered to participants after demonstration of the video game prototypes to quantify their impressions of the games and their potential to improve adherence. All AYA participants were remunerated with gift cards and transportation vouchers for their participation in each FGD.

**Data Collection and Analysis**

All FGDs were audio-recorded and professionally transcribed verbatim for data analysis. A systematic analysis guided by a constant comparative analytic framework was used to employ content analysis strategies through the use of thematic coding to determine major findings. Constant comparative frameworks objectively identify patterns in data and discover relations between ideas and concepts [33]. This framework supported the use of summative content analysis, which involves comparing the content of data to interpret the underlying context [34]. Content analysis was achieved through thematic coding, which links the common themes and ideas discussed among study participants [35]. Qualitative data analysis was organized using both NVivo version 9 (QSR International) and ATLAS.ti version 7 (Scientific Software Development GmbH) qualitative data analysis software by 2 independent coders and included 2 cycles: (1) each interview transcript was objectively analyzed using the aforementioned analytic approach and (2) coding across FGDs was compared to produce overall themes that support the purpose of the study. Codes were reviewed, and intercoder agreement was achieved where there were discrepancies in the categorization of the codes. Descriptive statistics were used to summarize response frequencies from the surveys. Mann-Whitney $U$ tests were used to compare median responses to survey questions across 2 AYA age groups (13 to 17 years and 18 to 24 years).

**Results**

**Provider Focus Groups**

**Provider Demographics and Focus Group Themes**

The HIV provider FGD (n=10) included physicians (n=2), case managers (n=3), research coordinators (n=3), a clinic nurse (n=1), and a psychologist (n=1). The majority (6/10, 60%) of providers had at least 5 years’ experience prescribing ART and delivering treatment and adherence support services, conducting clinical research, and providing HIV counseling and testing to AYA. During the FGD, several recurrent themes emerged related to ART and the potential role of video games to improve adherence, including (1) patient understanding (or lack thereof) of the concept of adherence, (2) daily barriers to patient adherence, (3) facilitators and promotion strategies for ART adherence, (4) provider educational methods to educate and support HIV treatment, (5) perceptions of the role of video games in ART adherence support, and (6) desired characteristics of video games for the patients (Textbox 1).
Textbox 1. Summary themes identified from the HIV provider focus group discussions and representative quotes.

### Patient understanding of adherence and resistance
- “The concept of resistance...I don’t think people understand what it means to be resistant. We keep trying to explain [you know what] it means the drugs don’t work anymore...people do not see the consequence...”
- “…it’s associated with negativity as opposed to being something that’s empowering that I can take control of my medicines, I can take control of my life, I can have control over...”

### Barriers to adherence
- “A lot of the times the patient...may have, like, issues about disappointing the clinicians and the providers and so they’ll be less forthcoming about their actual adherence, and then when you confront them about their viral load and why what they’re saying isn’t matching then they’ll, you know, give the puppy dog eyes and kind of disclose that ‘actually I haven’t been taking my medicine.’”
- “They don’t feel sick, they don’t feel like they need to take medicines, it’s very abstract in terms of me talking about ‘this is your CD4 count, this is your viral load, you take your medicine so your viral load goes down and your CD4 comes up’...but they don’t feel the importance of it...”
- “And I think that taking their meds a lot of what I hear is ‘Oh when I take my meds that reminds me that I have HIV. If I am reminded about HIV I just don’t want to deal with it.’ And as they get older and older it becomes more of an issue and they don’t see the consequences. They don’t see anybody dying around them from HIV, they don’t see the ramifications of it.”

### Facilitators and promotion strategies
- “Actually our kids who are very concrete thinkers tend to do really well that they just take the medicine; they don’t think about all the ramifications of why I’m taking medicines and I’m going to have to do this for the rest of my life...”
- “I feel like those kids [who have been told their HIV-positive status sooner] do better than the ones who are 13 or 14 and it’s been a secret hidden and then they’re being told that [they have HIV]. Those are the ones that really seem to struggle [with medication adherence].”
- “…we do have a few kids though, very few, but we have a few kids where the opposite happens where they get old enough to take their own meds and they actually weren’t taking them and they all better and all of the sudden become undetectable. When it was the parents’ responsibility it didn’t happen.”

### Provider educational methods
- “So the way I’ve used it...gaming analogies, I’ve told ‘you need the grenade, you need the pistol and you need the AK-47, you need all 3 to kill off the virus. You can’t just do one or else it’s not going to work it’s going to get stronger...like you need all 3 you don’t need just one, it’s not going to work...’”
- “Red, green, and yellow light and it’s in regards to the immune system...I will tell patients frequently you’re now in the yellow zone that means that I’m seeing the decline in your immune system and I’m trying to urge you not to let it become red zoned. And then obviously there [are] some patients when I will say... we are in the red zone your immune system is weak and it’s not recovering. The longer we see it in the red zone the harder it will be to recover the immune system.”
- “Because I’ve actually tried the video game analogy many times in terms of ‘you know those little power packs you have to get so that your energy level goes really high that’s what you’re trying to do with your medicines, get that power pack...’”

### Initial perceptions of role of video games
- “Yes that incentive is there and yes that reminder is there but after a while it just becomes like ‘oh this is nothing I don’t really care, whatever.’”
- “I wonder, first of all I think about gender differences, I wonder if it’s going to be girls versus boys. And I think boys might be more interested in doing it.”
- “I don’t know how that’s going to help because I think some of our kids forget but I don’t think a lot of them forget. I think the thing rings and they go, ‘oh this thing’s ringing,’ and they continue with playing the video game...Maybe if the video game dispensed it...”

### Desired characteristics of a video game
- “One thing that I think is really important is that the game not peg them as being HIV positive”
- “And it might send you a message saying you’ve been doing great...and the second reward is missing and this is what you could do if you want.”
- “And also I just want to say, to me, like, with the teenager is a very change in milieu. So I think having the same system of awards is boring. So I think what needs to be in any game is you need to completely change the system once they get it.”
witness people dying from HIV and therefore did not perceive
the long-term consequences of not taking the medications as a
threat to their lives. Providers described characteristics of AYA
living with HIV that made adherence challenging, including
concrete thinking, being less organized, and lacking the maturity
to control their own ART regimens. Providers also believed that
learning one’s HIV-positive status at a young age was associated
with improved ART adherence. Providers also described their
efforts to use age-appropriate, culturally relevant analogies to
increase understanding of viral resistance and ART adherence
among AYA in care.

**Perspective on the Potential Role of Video Games**

With regard to the feasibility and acceptability of using video
games to improve adherence, many providers were initially
skeptical about the ability of gaming to improve adherence.
They suggested that once the novelty of the gaming wore off,
adherence levels might decrease. Providers also expressed
concern that gender differences could emerge with males
potentially being more engaged in gaming compared to females.
Despite providers’ initial skepticism, after seeing and
experiencing the game prototype, the majority of providers
expressed optimism and provided specific feedback for
improving game design. Overall, providers were supportive of
exploring the proposed gaming intervention and recognized its
potential positive role in improving adherence among AYA in
care.

**Focus Group Discussions With Adolescents and Young Adults**

**Demographics**

A total of 31 AYA were recruited for the study; however, only
12 were able to attend and participate in the FGDs. Among the
12 AYA who participated in the 2 rounds of FGDs, 7 were aged
13 to 17 years old and 5 were aged 18 to 24 years old (Figure
5).

![Figure 5. HIV-infected adolescent and young adult focus group participant recruitment. FGD: focus group discussion.](http://games.jmir.org/2018/4/e10213/)
Table 2. Adolescent and young adult demographic, disclosure status, adherence, and access to video games characteristics (N=12).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>13-17 years (n=7)</th>
<th>18-24 years (n=5)</th>
<th>Total (N=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (range)</td>
<td>14.7 (14-16)</td>
<td>19.2 (18-22)</td>
<td>16.6 (14-22)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4 (57)</td>
<td>3 (60)</td>
<td>7 (52)</td>
</tr>
<tr>
<td>Male</td>
<td>3 (43)</td>
<td>2 (40)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Race/ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Hispanic black</td>
<td>7 (100)</td>
<td>4 (80)</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Other</td>
<td>0</td>
<td>1 (20)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>State of residence, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>6 (86)</td>
<td>2 (40)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>District of Columbia</td>
<td>1 (14)</td>
<td>3 (60)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Education level, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grades 1-8</td>
<td>4 (57)</td>
<td>0</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Grades 9-11</td>
<td>3 (43)</td>
<td>1 (40)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>High school graduate/general equivalency diploma</td>
<td>0</td>
<td>4 (80)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Time since HIV disclosure prior to study enrollment, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;6 months</td>
<td>2 (29)</td>
<td>0</td>
<td>2 (17)</td>
</tr>
<tr>
<td>6 months to 5 years</td>
<td>2 (29)</td>
<td>0</td>
<td>2 (17)</td>
</tr>
<tr>
<td>&gt;5 years</td>
<td>3 (43)</td>
<td>5 (100)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>Storage of HIV medications, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In my bedroom</td>
<td>3 (43)</td>
<td>4 (80)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>In the kitchen</td>
<td>4 (57)</td>
<td>1 (20)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Pills per day, mean (range)</td>
<td>4.3 (1-8)</td>
<td>2.2 (1-4)</td>
<td>3.4 (1-8)</td>
</tr>
<tr>
<td>Medication regimen, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Once per day</td>
<td>4 (57)</td>
<td>5 (100)</td>
<td>9 (75)</td>
</tr>
<tr>
<td>Twice per day</td>
<td>2 (29)</td>
<td>0</td>
<td>2 (17)</td>
</tr>
<tr>
<td>≥3 times per day</td>
<td>1 (14)</td>
<td>0</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Pillbox/dispenser use, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In the past</td>
<td>1 (14)</td>
<td>1 (20)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Currently</td>
<td>4 (57)</td>
<td>3 (60)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Never</td>
<td>2 (29)</td>
<td>1 (20)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Adherence strategies, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I do it myself</td>
<td>3 (43)</td>
<td>2 (40)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>A relative helps me remember</td>
<td>5 (1)</td>
<td>2 (40)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>I set an alarm</td>
<td>0</td>
<td>2 (40)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>I receive a text reminder</td>
<td>0</td>
<td>1 (20)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>7-day adherence, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All the time (100%)</td>
<td>4 (57)</td>
<td>1 (20)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Most of the time (51-99%)</td>
<td>3 (43)</td>
<td>4 (80)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Half of the time to never (0-50%)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Video game device in household, n (%)</td>
<td>7 (100)</td>
<td>5 (100)</td>
<td>12 (100)</td>
</tr>
<tr>
<td>Gaming devices in household, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Total (N=12)  
18-24 years (n=5)  
13-17 years (n=7)  

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>13-17 years (n=7)</th>
<th>18-24 years (n=5)</th>
<th>Total (N=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wii</td>
<td>5 (71)</td>
<td>2 (40)</td>
<td>7 (58)</td>
</tr>
<tr>
<td>Xbox</td>
<td>4 (57)</td>
<td>2 (40)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Nintendo</td>
<td>0</td>
<td>3 (60)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>Computer</td>
<td>2 (29)</td>
<td>2 (40)</td>
<td>4 (33)</td>
</tr>
</tbody>
</table>

**Frequency of game play, n (%)**
- Daily: 0 (0) 1 (20) 1 (9)
- Only on weekends: 2 (29) 0 2 (18)
- Anytime: 2 (29) 4 (80) 6 (54)

*Totals may not sum to 100% due to missing data.
Participants were allowed to check all that applied.

The mean age of participants was 16.6 years, 92% (11/12) were black, and 58% (7/12) were female (Table 2). All participants had been told about their HIV status, and many participants (8/12, 67%) had known for more than 5 years. All participants were prescribed ART, and 75% (9/12) were on a once-a-day ART regimen. Younger adolescents (ages 13 to 17 years) had a higher pill burden compared to older AYA (4.3 vs 2.2 mean number of pills, respectively). Seven participants (58%) were currently using some type of pillbox at the time of the FGDs.

**Perspectives on Adherence**
The following themes were identified during the AYA FGD: general HIV and ART knowledge, knowledge of viral resistance and ART adherence, barriers and facilitators to adherence, interest in and use of video games, and acceptability of video games for improving adherence (Textbox 2).

**Knowledge of HIV, Adherence, and Barriers and Facilitators to Adherence**
Participants in the younger FGD had a conceptual understanding of HIV, whereas older FGD participants demonstrated a deeper understanding of HIV and adherence concepts. Participants across both age groups expressed a similar understanding of viral resistance and adherence. When discussing barriers associated with adherence to ART, participants revealed a broad array of factors such as not wanting to be different from their peers, being afraid of inadvertently disclosing their status by taking the medications, having difficulties with the taste and size of medications, and facing time management challenges. Participants in all AYA FGDs shared personal motivating factors and facilitators associated with improving adherence.

**Feasibility and Acceptability of Video Games for Adherence**
All AYA participants had a video game device in their house with the most commonly reported gaming platforms being Wii (7/12, 58%) and Xbox (6/12, 50%). Additionally, 80% (4/5) of older AYA reported unrestricted access to video games compared to 29% (2/7) of younger AYA (Table 2). Prior to introducing the initial video game prototype, participants were asked to describe enjoyable aspects of video games and motivating factors for playing games. Participants generally expressed that they liked games that were engaging, had intense adventure, and posed a challenge. Discussion about the use and frequency of video game playing indicated that length of time spent on video game playing was dependent on the challenging nature of the game. Participants in the younger age group reported enjoying games as a leisure activity, while participants in the older age group were more interested in challenging games.

During the first round of FGDs, participants viewed demonstrations of the Wisepill ART dispensers, artwork and images for the prototype of Virus Fighter/Body Voyage (Figure 1), a game about the immune system, and an initial version of the Adherence Warrior (Figure 2) game. All FGD participants were comfortable with the Wisepill dispensers and believed that their medications would fit and that they would be able to carry the dispenser around without any issues. When introduced to the concept of linking the Wisepill to in-game incentives to improve medication adherence, AYA demonstrated acceptance of the concept. Younger participants thought the Wisepill would serve as an additional reminder to help them remember to take their medications, whereas the older participants focused more on the game content and emphasized a desire for the games to be engaging and interactive.
Textbox 2. Summary themes identified from the HIV-infected youth and young adult focus group discussions and representative quotes.

General HIV/antiretroviral treatment knowledge
How antiretroviral therapy works and how it affects the body

- “I mean I heard it before, I forgot. We used to talk about it a lot but I wasn’t listening ’cause I didn’t understand. ‘Cause I didn’t want to take my medicine.” [13 to 17 years]
- “My viral load goes down by taking the meds...taking the meds every day.” [13 to 17 years]
- “I guess if you don’t take it [ART] for a long period of time then the virus can start, like, making your body feel weird. If you don’t start taking your medicine. You might, like, feel weaker, or...Medicines will stop being effective. ‘Cause the virus will, I guess it’ll get used to it. Or, something like that.” [13 to 17 years]

Antiretroviral therapy and HIV knowledge

- “I just know that you get born with it or you get it through like sex and stuff. And it’s like a disease that’s in your immune system that takes over your immune system.” [18 to 24 years]
- “You want the viral load to be really low because that’s how much the virus is in your body, you want it to be low.” [18 to 24 years]
- “They got some [pills] out here that’s one a day, they got like, probably like 2 or 3 different kind that’s one a day and then they got other kinds that’s not one a day but for the people who take the ones that’s not one a day it’s probably because when they were younger, like if they had to live with it all their life like me then it messed up their medicine at first then now we can’t take the one a day pill.” [18 to 24 years]

Knowledge of adherence and resistance
Impact of not taking meds daily

- “It’s like when you stop taking your medicine and it stops working or you want to take it and it doesn’t work.” [13 to 17 years]
- “It’s when your medicine don’t work, right, or don’t want to...like help your body...I mean ‘cause it doesn’t work anymore.” [13 to 17 years]
- “If you take it for a long time period that regardless. I don’t know...medicines stop being effective...” [13 to 17 years]
- “Resistance is if I’m supposed to take 4 pills a day, right, so I stop taking those 4 pills or something like that or maybe I might just take 3 of the 4 pills and that 1 pill or those 4 pills that I’m supposed to be taking that will eventually help me to, you know, like [new] medicine and stuff like that it’s not going to work because I didn’t take it and I was playing around with it.” [18 to 24 years]
- “I think I heard it before but now I really remember what they were saying...” [18 to 24 years]
- “Yea, I’ve heard of resistance...resistance to medication...so like if you take it [medicine] and then if you don’t take it for a long time and then take it again you could be resistant to it and then it won’t work no more.” [18 to 24 years]

Barriers/facilitators to adherence
Not wanting to be different, motivation from family, and forgetfulness

- “I always have been taking my medicine. But yes, I did have a time when I didn’t [take my medicine]. Nobody else was taking medicine and I felt like I was the only one who was and I didn’t want to be different and I didn’t want to be the only one who was taking medicine and then questions of why I am taking it...” [13 to 17 years]
- “I felt I need to be around to see my brother graduate and I want to still be around to talk to people, motivation...” [13 to 17 years]
- “I can remember taking it in the morning but sometimes after school I forget because I’m doing like my homework and stuff.” [13 to 17 years]

Difficulty swallowing, reminders, and forgetfulness

- “It’s, um, it was just hard taking it ‘cause the pill was nasty or too big to swallow and the liquid was just disgusting...” [18 to 24 years]
- “I set 3 alarms every hour and I put those 3 alarms on repeat 3 times.” [18 to 24 years]
- “When I turned 15 and 16 going to high school and all that then you start forgetting because it was like a whole bunch of stuff on my plate you know at the time.” [18 to 24 years]

Interest and use of video games
Appeal of games

- “Just a good way to pass time. Especially if I’m not doing anything, like on Sundays. Just to get passed time I guess.” [13 to 17 years]
- “I like a lot of adventure...” [13 to 17 years]
- “I play after homework...but in the summer, my mom doesn’t care.” [13 to 17 years]
- “Well it keeps my attention but I like a game that has an interesting story behind it or just keeps me involved ‘cause I get a game quick.” [18 to 24 years]
“Well I play all types of games but one game my head was stuck on and that’s because I can’t beat it because it keeps changing is War Frame.” [18 to 24 years]

...the harder the challenge the better...” [18 to 24 years]

Acceptability of video games for adherence

Perceptions regarding the potential of games to assist with adherence

- “Yeah...because I need something besides my alarm.” [13 to 17 years]
- “I think if you like playing video games you’d remember to take your medication.” [13 to 17 years]
- “I wish we could like have our own little profile.” [13 to 17 years]

Suggestions for improving the game prototype

- “I don’t know, it might work but for me there’d have to be a game where it’s like, kind of like how XX said, like a puzzle or something like that or like a challenge game or whatever...” [18 to 24 years]
- “Like I mean every person is different but for me it would be like a puzzle game. Like some of it makes you think and you have to like work for it at the end and then they’ll be like a message or something.” [18 to 24 years]
- “I like the power-ups because you get it to do different like variety of stuff so you get to choose different stuff...if you have different stuff you can earn a higher score.” [18 to 24 years]

In addition to the interactive FGDs, brief surveys using Likert scales were used to evaluate acceptability of the Wisepill and game prototypes among AYA (Table 3).

On a scale of 1 to 5, with 5 indicating strong agreement, most participants agreed the dispenser format was convenient (median 4.0) and linking it to the games would not be an issue (median 5.0). In assessing game design upon viewing the artwork for Virus Fighter, younger participants were less likely than older participants to like the idea of playing games about the immune system (median 4.0 vs 5.0, respectively; \( P = .01 \)) and significantly more likely to agree that they preferred playing games about topics other than the immune system (median 4.0 vs 5.0, respectively; \( P = .01 \)).

The feedback from the initial AYA FGDs led to suspension of production on the Virus Fighter/Body Voyage game, revisions of the Adherence Warrior game, and the development of a third game prototype, the Berry Game, both of which were shared with AYA during the second round of FGDs.

After playing the two intervention game prototypes, most participants were optimistic of the games’ abilities to help them improve their adherence and provided suggested revisions that were incorporated by the researchers. Among the suggested changes were allowing for the customization of characters including both male and female avatars, adding music, having an in-game notification or reminder that encourages players to take their medication, and increasing the diversity of the games. Participants conveyed interest in games with a bonus stage and power-up capabilities that enabled additional features leading to more points. Participants also favored power-ups as a reward for ART adherence over gaining additional points, which emphasized their preference for playing games for the challenge rather than for winning.

AYA reported that they thought the games were interesting (median 5) and satisfying (median 5). Participants in the 18- to 24-year age group were more likely to report that the games were fun (median 5.0 vs 4.0, \( P = .03 \)) and to prefer games about subjects other than the immune system (median 5.0 vs 4.0, \( P = .01 \)) (Table 3). Using a separate Likert scale with a range of 1 to 4 with 4 indicating high likelihood, participants reported that they would play the games if they were available to them (median 4.0), they would use the games to help with taking their HIV medications (median 4.0), and they would be more likely to take their medication if they could get better power-ups in the games (median 4.0). Overall, preferred game characteristics, regardless of age group, included being (1) challenging enough to keep one’s attention, (2) individually tailored based on interests, (3) competitive, (4) containing a virtual reward system to provide in-game incentives, and (5) easy to play.
Table 3. Acceptability of game prototype stratified by age group.

<table>
<thead>
<tr>
<th>Game characteristics</th>
<th>13-17 years (n=7), median response</th>
<th>18-24 years (n=5), median response</th>
<th>Total (N=12), median response</th>
</tr>
</thead>
<tbody>
<tr>
<td>The game playing was very interestingb</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>The game playing was satisfyinga</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>The game playing was funa</td>
<td>4c</td>
<td>5c</td>
<td>5c</td>
</tr>
<tr>
<td>It was easy to learn how to play the gamea</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>I liked the art and animationa</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>I liked the game environmenta</td>
<td>3.5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>I liked the game interfacea</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>I like the 7-day organizer format of the pill dispenserb</td>
<td>4.5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>The link to the pill dispenser was not a problem for mea</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I like the idea of playing games about the immune system fighting off invadersa</td>
<td>4</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>I like the idea of playing games about subjects other than the immune systema</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Games about the immune system make me more interested in the subjecta</td>
<td>3.5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>I would like to get a pop-up reminder to take my medications from the game on my phonea</td>
<td>4.5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>I would play these games if I had themb</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>I would recommend the games to a friendb</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>I would be interested in playing these games in multiplayer mode with my friendsb</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>I would use these games to help with taking my medicationsb</td>
<td>4</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td>I would be more likely to take my medication if it would get me more points in the gamesb</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>I would be more likely to take my medications if it would get me better power-ups in the gamesb</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Reviews provided on a scale of 1 to 5: strongly agree (5), somewhat agree (4), neither agree nor disagree (3), disagree (2), strongly disagree (1).
Reviews provided on a scale of 1 to 4: very likely (4), somewhat likely (3), not very likely (2), don't know (1).
Significant differences between age groups using t tests (P<.05).

Discussion

Principal Findings

This study, which explored the potential for using gaming technology to support long-term treatment of HIV among AYA, demonstrated that it would be feasible and acceptable to use a digital gaming intervention linked to Wisepill dispensers to improve ART adherence. Qualitative data collected through FGDS elicited information from both HIV care providers and AYA participants. The FGDS with HIV care providers revealed concerns regarding the level of understanding of the importance of adherence and viral resistance as well as age-related challenges such as nondisclosure, treatment fatigue, and limited capacity to foresee long-term consequences of nonadherence among their AYA patients. The challenges to daily ART adherence were echoed by the AYA FGD participants as they identified common barriers to treatment adherence and strategies to improve adherence.

Despite the providers’ concerns, participating AYA demonstrated an understanding of the concepts of resistance and adherence and acknowledged the multiple adherence barriers they face, which included changes in daily routines and fear and anxiety of disclosing their status to others, all of which are consistent with the existing literature [11,36,37]. Our discussions also indicated that AYA considered having an understanding of the chronicity of HIV and the ability to set long-term goals were facilitators of adherence. These findings are not surprising considering the complex developmental periods of adolescence and young adulthood of altered levels of identification, self-regulation, and influence from peers [11]. As adolescents mature into young adults, strategies and messages to support adherence, self-management, and well-being should reflect shifts from concrete to abstract thinking and from an invulnerable to self-preserving mindset [12]. This is reinforced by our finding of AYA expressing willingness to try more creative approaches to facilitate better medication adherence, including playing video games.
The proposed approach of using Wisepill dispensers linked to games with microtransactions and power-ups was affirmed as acceptable by both HIV care providers and AYA living with HIV. Importantly, AYA participants in our study were receptive to the concept of using Wisepill dispensers for their adherence monitoring as the majority of them had some past experience with different types of pillbox or pill dispenser use. The feasibility and acceptability of using wireless electronic device monitors for adherence among HIV-infected persons has been demonstrated in other settings including in China and Uganda. In Uganda, adult HIV-infected patients described the Wisepill device easy to use and convenient, and it was found to be noninferior to the Medication Event Monitoring System pill bottle caps [15]. Further, among Chinese children aged 10 to 15 years, Wisepill dispensers were reported to be an acceptable and rigorous method of evaluating ART adherence and were found to be potentially useful for adherence support among adolescents living with HIV [38].

Our study was also informative in gathering key information from game users and providers about the role and the design of the game within the context of health promotion. All the AYA participants reported regular technology use and unfettered access to video games and provided overall positive feedback on the proposed game prototypes. Interestingly, among HIV care providers, initial discussions regarding the potential for gaming to improve adherence indicated some hesitation and skepticism, although some of them had described previously using gaming analogies to educate patients about HIV and adherence. Even though some providers used gaming analogies to explain the role of adherence in treatment, they were clear that the games should avoid defining the youth as being HIV-infected and should instead be positive and motivational. Similarly, the AYA participants were not particularly interested in an educational or HIV-themed game. This desire for noncontent-related games among providers and AYA in our study is in contrast to some other games developed for health-related purposes, which are both educational and disease-specific [39-41]. We believe that the desire to avoid the direct topic of HIV in gaming is closely linked to the sensitive nature of HIV disease and external and internal stigma experienced by people living with HIV. Therefore, our games were not specifically HIV-related nor did they include a community-gaming aspect to intentionally avoid inadvertent disclosure. In future iterations of the game, it may be possible to incorporate a community-gaming aspect in a confidential manner as well as potentially link the Wisepill device to existing popular games to address concerns about the durability and sustained use of a particular set of games.

Through our iterative study design, AYA provided important feedback and concrete suggestions to the game development process. They preferred games that were challenging, individually tailored, included a virtual reward system, and were easy to play. In other studies, video games that were developed using input from their target population have been proven effective in the chronic disease management of health conditions in AYA such as asthma, cancer, and diabetes [42]. In fact, a video gaming approach has recently been tested among AYA living with HIV that incorporates an adherence-based app inclusive of games [39]. The developers reported that AYA enjoy the interactive and customizable app elements and found this approach to be highly acceptable [39]. Further, studies among youth with chronic diseases have also found that game interventions result in improved self-management and decreased unanticipated medical visits [43]. Based on these studies and our data, we anticipate that our versions of the games will appeal to AYA living with HIV and will be effective in supporting medication adherence and improving HIV outcomes such as viral suppression.

**Limitations**

There are several limitations to our study worth noting. Despite high numbers of initial youth recruited, the number of youth participating in the FGDs was relatively small and participants were drawn from a convenience sample recruited from a single clinical center. Although common in qualitative research, the small sample size limits the generalizability of the study findings. Further, all participants identified as black or African American, reflecting the high burden of infection among these populations in the DC region as well as in the United States [28]. While this may limit generalizability of findings to AYA of other races or ethnicities, it should be noted that none of the video game prototypes were uniquely designed for black youth and were instead based on common video game prototypes used among diverse populations. Additionally, FGD moderators can affect the outcome of an FGD by intentionally or inadvertently inserting their personal biases into the participants’ exchange of ideas. Since one of the moderators of our FGD was the developer of the video games being tested, the FGD participants could have been led into reaching certain assumptions and may have been somewhat influenced in their opinions. We attempted to reduce this potential bias by including a clinician-researcher as a moderator in each FGD to deter biased discussion and responses.

**Conclusions**

AYA living with HIV face multiple challenges with medication adherence and must develop effective strategies to help them remember to take their medications. Similarly, HIV care providers strive to support ART adherence and educate AYA about the long-term consequences of nonadherence. Given these challenges, the development of effective, youth-friendly interventions to improve adherence among AYA living with HIV is critical to achieving better health and outcomes of HIV disease. In the era of the widespread use and availability of technology including mobile phones and video games, the concept of using a technological approach to improve medication adherence is supported by both AYA and their HIV care providers. Inviting buy-in from youth to elicit their preferences as end-users of the game-based intervention and modifying the games accordingly to include non-HIV focused, interactive, adventurous, and challenging games is likely to result in stronger interest and engagement in the future intervention. The potential effectiveness of this approach will be tested in a future study among a cohort of 13- to 24-year-old HIV-infected AYA with suboptimal ART adherence at CNMC.
Acknowledgments
This work was supported by the National Institutes of Mental Health at the National Institutes of Health (R43MH105315). The authors would like to acknowledge the providers, staff, and patients at Children’s National Medical Center who participated in this study.

Conflicts of Interest
None declared.

References


Abbreviations

ART: antiretroviral treatment
AYA: adolescent and young adult
CNMC: Children’s National Medical Center
FGD: focus group discussion
SIS: Special Immunology Services

©Amanda D Castel, Saba Qasmieh, Daniel Greenberg, Nicole Ellenberger, Tyriesa Howard Howell, Caleb Griffith, Brittany C Wilbourn, Kavitha Ganesan, Nadia Hussein, Gabriel Ralte, Natella Rakhmanina. Originally published in JMIR Serious Games (http://games.jmir.org), 15.10.2018. 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.
Virtual Reality Clinical Research: Promises and Challenges

Bernie Garrett1*, RN, PhD; Tarnia Taverner1*, RN, PhD; Diane Gromala2*, BFA, MFA, PhD; Gordon Tao3*, BSc, MSc; Elliott Cordingley4*, BSc; Crystal Sun1*, BSc, MPH

1School of Nursing, University of British Columbia, Vancouver, BC, Canada
2School of Interactive Arts and Technology, Simon Fraser University, Surrey, BC, Canada
3Rehabilitation Sciences, University of British Columbia, Vancouver, BC, Canada
4Faculty of Science, University of British Columbia, Vancouver, BC, Canada
*all authors contributed equally

Corresponding Author:
Bernie Garrett, RN, PhD
School of Nursing
University of British Columbia
T201-2211 Wesbrook Mall
Vancouver, BC, V6T 2B5
Canada
Phone: 1 16047538942
Email: bernie.garrett@ubc.ca

Abstract

Background: Virtual reality (VR) therapy has been explored as a novel therapeutic approach for numerous health applications, in which three-dimensional virtual environments can be explored in real time. Studies have found positive outcomes for patients using VR for clinical conditions such as anxiety disorders, addictions, phobias, posttraumatic stress disorder, eating disorders, stroke rehabilitation, and for pain management.

Objective: This work aims to highlight key issues in the implementation of clinical research for VR technologies.

Methods: A discussion paper was developed from a narrative review of recent clinical research in the field, and the researchers’ own experiences in conducting VR clinical research with chronic pain patients.

Results: Some of the key issues in implementing clinical VR research include theoretical immaturity, a lack of technical standards, the problems of separating effects of media versus medium, practical in vivo issues, and costs.

Conclusions: Over the last decade, some significant successes have been claimed for the use of VR. Nevertheless, the implementation of clinical VR research outside of the laboratory presents substantial clinical challenges. It is argued that careful attention to addressing these issues in research design and pilot studies are needed in order to make clinical VR research more rigorous and improve the clinical significance of findings.

(JMIR Serious Games 2018;6(4):e10839) doi:10.2196/10839

KEYWORDS
virtual reality; clinical research; VR standards; VR theory; VR immersion; VR presence

Introduction

Contemporary research on computer-based virtual reality (VR) dates back to the early 1980s, although devices for presenting stereoscopic imagery (ie, using a slightly different image for each eye) such as the stereoscope started in the 1830s [1]. The exploration of VR use in clinical applications is accelerating rapidly with the advent of more powerful computer and graphics processors capable of rendering real-time three-dimensional (3D) imagery, and the availability of relatively low-cost VR headsets such as the Oculus Rift or HTC Vive (see Figure 1).

As researchers with significant experience in researching VR for clinical applications, we have identified some major issues in the development of clinical VR research. Substantial challenges remain with theoretical ambiguity and immaturity, a lack of technical standards, problems of media versus medium, practical in vivo issues, and economic feasibility.
**Background**

There has been rapid growth in the reported use of VR in the treatment of a variety of clinical conditions, such as acute and chronic pain management [2-9], anxiety disorders [10-12], phobias [13-15], posttraumatic stress disorder (PTSD) [16-18], eating disorders [19], autism [20], and rehabilitation [21-26]. Additionally, its use in professional health care education has also been expanding rapidly [22,27-32].

One early clinical application of VR was for the treatment of acrophobia [33]. Graduated exposure to virtual environments with foot bridges, balconies, and a glass elevator were used with a railing placed around the user in the real world for them to hold on to. The intervention was reported as effective. Over the last 20 years, VR clinical applications have expanded to address other phobias and anxiety disorders. The most common approaches in this field have been to model virtual environments after existing exposure therapies using graduated exposure to a VR version of the object or situation that causes distress and use of VR cognitive behavioral therapies [34-37]. For PTSD, virtual environments have been used to simulate complex traumatic scenarios under control to treat war survivors [18,38]. Similarly, VR has been used in the treatment of body image and eating disorders [39-41]. These approaches leverage education, visual feedback, and simulations of critical situations to improve body self-perception.

**Figure 1.** VR clinical application papers published by year (PubMed).
In the rehabilitation field, there has also been great interest in pairing assistive technologies (e.g., robotics, treadmills, wheelchair on rollers, wearable sensors) and VR. The primary goals have been the development of tools that support patient motivation to engage with rehabilitation and to leverage the logistical advantages of digital technology, namely performance monitoring, telehealth, and patient self-management. Additionally, translating physiotherapy exercises and activity training into the VR space allows for much greater control over and variety of scenarios.

For example, robot-assisted upper limb therapy paired with VR visual feedback allows for graded exercises contextualized in a videogame environment [56-58]. In wheelchair simulators, VR enables users to practice wheelchair navigation skills in more dangerous situations such as traffic crossings and crowds without risk [59,60].

Clinical VR research to date has generally been positive, but overall research in this field is in the early stages and faces technical and theoretical hurdles. Most studies have used non-standardized techniques and tools in small-scale pilot studies. Over the last 4 years, the authors have conducted several clinical VR research projects [6,8,49] and found a number of challenges in the field that may limit the validity and generalizability of the work.

### Challenges

#### Theoretical Ambiguity and Immaturity

As with the development of any new discipline, establishing a sound theoretical basis and standards is key to the growth of the field. However, there exists some theoretical ambiguity in the field due in part to its immaturity. Overall, VR may be considered as a growing field, defined by both its technology and its effects. The desired effect is to create an immersive experience, whereby the user is placed in a simulated environment that looks and feels as engaging as the real world. The person in this synthetic environment has a specific sense of self-location within it, can move to explore it, feels that the space surrounds them, and can interact with the objects within it. Overall, they feel a sense of presence in this environment, and their actions partially determine what happens within it [61,62].

Technically, the sense of immersion in a VR environment is largely achieved through visual and auditory stimuli that simulate 3D visual and auditory cues available in the real world. Haptic feedback can also contribute to this immersion. Visually, this is delivered to the user via a head-mounted display, which presents the computer-generated imagery (CGI) of the VR scene from the perspective of each of the user’s eyes. The literature suggests that immersion is largely influenced by both visual and audio qualities, although a universally accepted definition is yet to emerge [63-67]. Immersion has been defined as the extent to which a user feels present in the CGI environment, rather than in their actual physical environment [68,69]. In computer science, immersion has more often been defined in terms of the technology and by the extent to which the computer is able to deliver an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of the participant [67]. Therefore, immersion is often referenced by technical considerations, such as field of view and positioning of the virtual body in the CGI. The inclusion of stereoptic imagery is widely thought to be the dominant factor that enhances the immersive experience. Other technical factors, such as greater display resolution or increased field of view are also significant [52,70].

Presence, on the other hand, refers to the sense of being within an environment that is generated through technological means [68,71]. It is viewed as the sense of actually being in a constructed world [68,69,71,72]. Two experiential and technology-dependent dimensions are considered to contribute to a sense of presence. The first dimension is vividness, or the production of a sensory-rich, mediated environment. The second is interactivity, defined as a user’s ability to engage with the environment and modify its form or alter events through interaction with it. An analogy would be that you can become immersed through the text in a book, but feel a sense of presence in the story only when you feel you are actually there experiencing the events.

This differentiation of immersion from presence (which is seen as more of a subjective element) is fairly well established in computer science, but less so in clinical VR research, where the terms are often used interchangeably. For clinical use, a technical definition of immersion is limited, as it ignores the participant as a co-creator of the experience. Therefore, concepts of presence and telepresence [68] are likely more useful to clinical applications. An immersive virtual environment can be considered to be a computer-generated environment that elicits the user’s sense of presence or “being there.” It can be seen as an environment that produces an esthetic perception connected to the ideal of total immersion in virtual space involving the willing suspension of disbelief [69,71,73]. In clinical contexts, this sense of presence is likely the key element of interest that differentiates the impact of VR from other distractive and cognitive approaches. Assessment tools that separate these aspects, such as the igroup Presence Questionnaire have been developed [74]. However, clinical VR literature rarely discusses these theoretical aspects nor provides robust theoretical explanations of how VR theory applies to the specific problem under investigation. As VR is essentially a technology-mediated phenomenon, this lack of theoretical distinction, between what actually constitutes a VR experience, at the least, makes meaningful comparisons between clinical studies complex.

Adding more complexity is the issue that the actual nature of the effect of VR on the clinical problem of interest is also often unknown. For example, VR environments are hypothesized to reduce pain by mediating cognitive attentional and distractive mechanisms. The use of VR might act directly and indirectly on pain perception in different ways by altering neurological signaling pathways involving attention, emotion, concentration, memory, touch, and the auditory and visual senses. However, there are competing theoretical explanations of pain and the exact mechanisms of how VR may attenuate it remain unclear [48,75-82]. It has been theorized that VR analgesia stems from the neurobiological interactions of areas of the brain that produce analgesic effect by regulating visual, auditory, and touch sensory experiences [80]. Hoffman et al state that VR works
predominantly via distraction. Pain requires attentiveness, and humans have been found to have limited controlled attentional resources [83]. The level and impact of the distraction can depend on the level of the immersion—the more immersive the VR, the more effective in reducing pain [84]. Furthermore, using functional magnetic resonance imaging (fMRI) brain scanning, correlation in pain-related brain activity and subjective pain report was reported, thus, demonstrating the impact of VR on pain-related brain activity in all five regions of the brain [83].

VR has been shown to alter the sense of an individual’s presence to that of being in a virtual world, therefore changing features of the individuals experience associated with sensory affective and cognitive processes.

The validity of clinical VR research also needs to be considered in the context of the theory development process. Overall, there are five major processes that occur in the development and establishment of a theory: (1) creating conceptual meaning, (2) structuring and generalizing the theory, (3) generating the theoretical relationships, (4) applying the theory, and then (5) theory validation by testing in different real-world applications [85]. At this stage of VR development for clinical use, the underpinning theory has yet to reach the higher levels of established validity.

**Standardized Implementation**

The type of VR technology implemented varies greatly between clinical studies. It is arguable that the current state of the art is very much technologically led rather than theoretically led, with each new iteration of clinical research using the latest VR applications and hardware with disparate approaches for a variety of clinical conditions. As the hardware and software continue to advance rapidly, studies even a year apart may be using completely different hardware or software and, in many cases, the technology is only vaguely defined [6,12,42].

**Three-Dimensional Versus Two-Dimensional**

Many clinical studies have used the term VR to describe significantly different multimedia technologies, including two-dimensional (2D) video screen presentations, 2D-rendered images presented on screens [86] and head-mounted displays [9,46,87], 360-degree 2D presentations on head-mounted displays [88], or computer-assisted virtual environment (CAVE) room-scale projection systems [89,90]. Others used 3D-rendered VR in motion-tracked stereoscopic head-mounted displays, with a wide field of view [8,9,91-93]. There are similar differences in audio use in these studies, with some using positional stereo sound (ie, location-specific sound that moves as the user moves their head) and others using non-spatial audio. Although health outcomes may be comparable, the nature and value of 3D versus 2D applications have not been widely explored in clinical applications.

**Study Design**

In addition to the theoretical issues, the nature of VR study design itself represents another significant hurdle. Systematic reviews/meta-analyses illustrate that many of these studies are statistically underpowered, although positive statistical results are frequently claimed [6,12,18,94-96]. To establish clinical efficacy of a therapy, large-scale quality randomized controlled trials are required. Comparative clinical studies also require a suitable control environment to contrast with the VR experience. Few studies make an adequate attempt to address this and frequently neglect to differentiate the effects of the media from the medium itself (both theoretically and in practice). For example, the medium of VR could be the use of VR technology and a head-mounted display to render a 360-degree stereoscopic and stereo audio environment with which a person can interact. The media may be a puzzle-solving interactive VR computer game, a VR rollercoaster ride, or a 3D-rendered high-definition video experience of a beach environment. Failing to explore if it is the VR experience itself or the medium used that is eliciting an effect is problematic. A good design will contrast a VR experience with a non-VR equivalent of the same experience, controlling for the effects of the medium compared to the media. These issues likely reflect some degree of confirmation bias among researchers, but this illustrates the need to implement larger-scale high-quality clinical VR studies.

**Usability and Technical Proficiency**

Another more practical challenge faced by clinical researchers is the usability of VR systems and the level of technical proficiency required to run them. Although current VR iterations are designed to be more user friendly, significant technical limitations remain. The use of head-mounted displays is problematic for some patients. They are cumbersome, particularly for patients with head or neck injuries, or for those who are particularly susceptible to eye strain. Additionally, VR applications are generally not usable by people with cognitive or significant visual deficiencies, as they are unable to access existing VR interfaces. Also, prolonged exposure to a screen a few centimeters from the eyes often leads to eye strain or headaches and represents an ongoing issue with VR systems [97,98]. Users with limited head or neck mobility often reported the systems were uncomfortable to use [8]. Furthermore, most advanced head-mounted displays have a cable tether that can be a distraction from the experience or a tripping hazard for older patients.

Cybersickness, as a side effect of VR, is also well documented and limits use by many patients, particularly those taking medications that can cause nausea [8,99-103]. Newer systems that operate at room-scale (ie, where the user can walk around in a pre-determined area) have addressed this to some extent, but many patients also have limited mobility and must use the system in a seated position. This gives rise to another problem: most VR applications are currently designed to be used as either room-scale or seated, with few working well in both configurations. The issue derives from the fact that room-scale VR navigation affords the user much greater range of motion to physically approach virtual items, while the seated position requires a set visual height, longer reaching movements, and controller-based navigation of the environment. The environment design and implementation requirements generally do not transfer well from seated to room-scale and vice versa. Many VR systems have implemented teleportation navigation systems to support moving through larger distances to overcome this issue, but again those designed for room scale use do not adjust well to use from a seated position.
The design and game paradigm of many VR experiences itself can also prove challenging for older patients. For users who have grown up with computer games, the nature of VR experiences is more readily understandable: traversing 3D-rendered worlds, using menus, navigating levels, storing and retrieving items, saving progress, solving game puzzles, and relating button-presses to abstract actions are all mechanics learned through experience. This alignment of VR with recreational gaming is exemplified by the marketing and delivery of HTC Vive and Oculus Rift VR applications through the Steam online gaming platform. Most clinical users are likely to be older adults, who have no such videogame literacy and often find learning these elements frustrating and distracting to their VR experience. Little work exists exploring the VR preferences of these users and the VR market is firmly dominated by the younger consumer.

Lab Versus In Vivo Practical Issues

Much of the existing VR research has taken place in lab or clinic settings. These environments can be optimized for VR systems. However, much remains to be known about the effects of regular and prolonged VR use for real-world and home applications, where they will be used for many chronic conditions. Certainly, there are common challenges for research requiring any kind of at-home technology implementation such as logistics, remote technical support, learning curve, and compliance. However, there are a few unique challenges to consider in the implementation of VR systems outside the lab.

Current VR systems require dedicated space and are susceptible to interference. Room-scale systems require a 5 ft² space, which may be intrusive to a patient’s living space. Cables may pose tripping hazards. Infrared sensors, such as those used by systems such as the HTC Vive, may be interfered with by devices such as TV remotes, resulting in display cutting out, choppy visuals, and loss of tracking, thus disrupting the user’s experience. Other environmental factors that disrupt infrared tracking, such as climate and reflection of light off windows or mirrors, can be easily mitigated in a lab setting but can be more difficult to cope with in a home. Furthermore, calibration for motion tracking of VR equipment is sensitive and thus movement of equipment must be minimized. Effective installation of VR equipment while still maintaining the usability of the home space is challenging and may be further complicated if there are pets or children in the home.

For clinical research, where a study may take weeks or months, these technological burdens are important to negotiate with participants in advance. Despite these challenges, our experience has shown that research participants are often enthusiastic and willing to accommodate the various needs of the equipment and research study. However, these attitudes may not necessarily carry over to commercial or non-research contexts.

Costs

Finally, the cost of VR still presents a challenge to implementing large-scale trials [11]. Although costs of head-mounted displays are dropping, quality VR environments still require high-end computer systems with advanced graphics processing to run them. VR applications are also expensive to develop. The current cost of a full system to run a quality VR clinical experience is around US $2,500 per unit plus maintenance costs, making clinical research with multiple users costly. As with any information technology, attrition of value is also rapid; newer technologies rapidly make older systems obsolete. A practical assumption of minimal resale value of a VR system after 3 years is not unreasonable.

Conclusions

Although clinical VR research looks promising, significant theoretical and practical challenges remain, such as theoretical ambiguity and immaturity, lack of technical standards, differentiating effects of media versus medium, value of 2D versus 3D applications, study design, usability, conducting in vivo research, and economic feasibility. Defining the impact of presence in clinical VR studies and differentiating the concept of presence from immersion (as they are often used synonymously) is a problem, and current research designs are often ill-equipped to differentiate the role of VR from confounding factors. More robust study designs contrasting VR experience with an equivalent non-VR control are required.

Practical challenges also remain, as existing high-end VR systems remain cumbersome and require technical proficiency to use. VR systems are not always user-friendly for patients. Moreover, issues of eye and neck strain and cybersickness remain as practical barriers to wider use. For those undertaking clinical VR research, it is important to keep these issues in mind during efforts to improve the evidence base for these technologies as health interventions.

Conflicts of Interest

None declared.

References


Abbreviations

2D: two-dimensional
3D: three-dimensional
CGI: computer-generated imagery
PTSD: posttraumatic stress disorder
VR: virtual reality

©Bernie Garrett, Tarnia Taverner, Diane Gromala, Gordon Tao, Elliott Cordingley, Crystal Sun. Originally published in JMIR Serious Games (http://games.jmir.org), 17.10.2018. 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.
Implementations of Virtual Reality for Anxiety-Related Disorders: Systematic Review

Theodore Oing\textsuperscript{1}, MSc; Julie Prescott\textsuperscript{1}, PhD
School of Education and Psychology, University of Bolton, Bolton, United Kingdom

Corresponding Author:
Julie Prescott, PhD
School of Education and Psychology
University of Bolton
Deane Road
Bolton, BL3 5AB
United Kingdom
Phone: 44 0120490 ext 3676
Email: j.prescott@bolton.ac.uk

Abstract

Background: Although traditional forms of therapy for anxiety-related disorders (e.g., cognitive behavioral therapy, CBT) have been effective, there have been long-standing issues with these therapies that largely center around the costs and risks associated with the components comprising the therapeutic process. To treat certain types of specific phobias, sessions may need to be held in public, therefore risking patient confidentiality and the occurrence of uncontrollable circumstances (e.g., weather and bystander behavior) or additional expenses such as travel to reach a destination. To address these issues, past studies have implemented virtual reality (VR) technologies for virtual reality exposure therapy (VRET) to provide an immersive, interactive experience that can be conducted privately and inexpensively. The versatility of VR allows various environments and scenarios to be generated while giving therapists control over variables that would otherwise be impossible in a natural setting. Although the outcomes from these studies have been generally positive despite the limitations of legacy VR systems, it is necessary to review these studies to identify how modern VR systems can and should improve to treat disorders in which anxiety is a key symptom, including specific phobias, posttraumatic stress disorder and acute stress disorder, generalized anxiety disorder, and paranoid ideations.

Objective: The aim of this review was to establish the efficacy of VR-based treatment for anxiety-related disorders as well as to outline how modern VR systems need to address the shortcomings of legacy VR systems.

Methods: A systematic search was conducted for any VR-related, peer-reviewed articles focused on the treatment or assessment of anxiety-based disorders published before August 31, 2017, within the ProQuest Central, PsycINFO, and PsycARTICLES databases. References from these articles were also evaluated.

Results: A total of 49 studies met the inclusion criteria from an initial pool of 2419 studies. These studies were a mix of case studies focused solely on VRET, experimental studies comparing the efficacy of VRET with various forms of CBT (e.g., in vivo exposure, imaginal exposure, and exposure group therapy), and studies evaluating the usefulness of VR technology as a diagnostic tool for paranoid ideations. The majority of studies reported positive findings in favor of VRET despite the VR technology’s limitations.

Conclusions: Although past studies have demonstrated promising and emerging efficacy for the use of VR as a treatment and diagnostic tool for anxiety-related disorders, it is clear that VR technology as a whole needs to improve to provide a completely immersive and interactive experience that is capable of blurring the lines between the real and virtual world.

(JMIR Serious Games 2018;6(4):e10965) doi:10.2196/10965

KEYWORDS
virtual reality; virtual reality exposure therapy; phobic disorders; anxiety disorders
Introduction

Background

Anxiety-related disorders such as specific phobias, posttraumatic stress disorder (PTSD), and general or specific anxiety (eg, public speaking or social anxiety) disorders stand as 1 of the most common, growing mental health disorders worldwide [1]. In 2014, 19.5% of individuals above the age of 16 years had shown signs of anxiety or depression, with the United Kingdom alone experiencing a 1.5% increase from 2013 [1]. To combat the symptoms of anxiety-related disorders, 1 of the most effective treatment methods has been exposure therapy, which stems from the broader practice of cognitive behavioral therapy (CBT) [2]. In exposure therapy, patients undergo a process of systematic desensitization, where a series of systematic steps are employed to gradually expose the patient to an anxiety- or fear-inducing stimulus, with the ultimate goal of minimizing the patient’s intense and adverse behavior toward the stimulus. Furthermore, the therapists may employ methods to change the patient’s cognitions about the stimulus, such as through psychoeducation, to reinforce treatment gains from systematic desensitization [3].

Traditionally, stimuli in exposure therapy are presented through in vivo exposure (IVE) or imaginal exposure (IE), each of which carries its own set of advantages and disadvantages. IVE involves live exposure to the stimuli, often being utilized to treat specific phobias or anxieties such as arachnophobia (fear of spiders) [4], acrophobia (fear of heights) [5], and social anxiety [6]. Although IVE is considered to be the most effective method for helping the patient overcome their anxiety or fear, disorders such as aviophobia (fear of flying) and social anxiety may require sessions to be conducted in public, therefore posing a risk of breaking patient confidentiality; become too expensive to perform single or repeated exposure sessions; and introduce uncontrollable variables that may hinder the overall treatment (eg, behavior of living organisms and weather conditions) [7-9]. Even if these issues can be addressed, some individuals may feel that confronting an anxiety- or fear-inducing stimulus may be too aversive, which may lead to participants dropping out of treatment or not seeking treatment at all [5]. IE can address many of the limitations of IVE, as patients are tasked with generating the stimulus in his or her imagination rather than confronting a live version of the stimulus; however, the patient may be potentially unable or unwilling to generate a vivid imaginal representation of the stimulus [10].

Since the early to mid-1990s, therapists have attempted to seek an alternative to IVE and IE through the use of virtual reality (VR) technologies through a process known as virtual reality exposure therapy (VRET). VR technology includes a wide range of configurations, including head-mounted displays (HMDs), external projection setups such as the CAVE Automatic Virtual Environment [11], and simulators [12], all of which vary in terms of technical specifications (eg, display resolution, tracking accuracy, and field of view). Regardless of the form of VR, VRET generally follows the same treatment protocols as traditional exposure therapy but renders the anxiety or fear-inducing stimulus within a virtual environment that immerses the user with sensory stimuli. These stimuli are often limited to the user’s visual and auditory senses but may sometimes incorporate tactile stimuli through an apparatus (eg, force feedback gloves, toy spiders) to allow VR users to feel objects with their hands. By utilizing a customizable virtual environment, VRET offers an unparalleled level of control for the therapist to manipulate factors that could not be controlled in a standard IVE session and tailor the sessions based on the patient’s needs—all in the confines of the therapist’s office [9,13].

Despite a major interest in VR during its inception, VR was often expensive, uncomfortable, and required special training to operate. The computers used to run VR were barely able to do so, leading to low-quality VR experiences (eg, jagged graphics and inconsistent and low frame rates) that could lead to simulator sickness, characterized by symptoms of nausea, headaches, and dizziness [6]. In addition to simulator sickness, early HMDs were also heavy, resulting in users experiencing neck pain after prolonged use. Furthermore, without adequate software distribution systems to sell or share VR programs, special training would often be required to create VR programs to suit the research or therapists’ needs. These limitations ultimately restricted the use and research of VR-based psychotherapy to well-funded or specialized institutions [14].

Although early VR technologies have been largely inaccessible to a mass audience, recent developments in VR technologies have addressed many of the issues that plagued legacy units. Both the HTC Vive and Oculus Rift, which released in 2016, were lighter and powerful enough to render high-quality visual and auditory stimuli. Both HMDs were also integrated with major digital distribution services such as Steam, which has attracted both small, independent developers and large, professional developers alike to create high-quality VR programs. Renewed interest in VR also led to a push for mobile VR, a less powerful yet inexpensive version of computer-based VR that could run on modern mobile phones (eg, iPhone, Google Pixel, etc.).

Objective

The aim of this systematic review was to explore previously established VR studies within psychotherapy to inform future VR research. Although modern VR HMDs are still relatively new, evaluating how past studies have utilized the VR technologies of their era can serve as a comprehensive guide as to how VR-based psychotherapy programs can improve in the future as well as whether the limitations observed in past studies are still relevant with the current iteration of VR systems. Topics covered in this review will mainly cover the efficacy of VRET treatment, its uses as a diagnostic or assessment tool, and innovations in the pursuit of greater VR experiences in relation to psychological disorders in which anxiety is a key symptom, including specific phobias, PTSD and acute stress disorder (ASD), specific and general anxiety disorder, and paranoid ideations.
Methods

Databases Searched
ProQuest Central, PsycINFO, and PsycARTICLES were the databases used to conduct a comprehensive search of the past literature. Studies must have been published before August 31, 2017, peer-reviewed, published in a scholarly journal, written in English, and have full-text availability.

Search Terms
The command line used for the search was as follows: “virtual reality” AND (phobia OR anxiety) AND (treatment OR therapy). Although VR is a common referential acronym for virtual reality, the full term was exclusively used during the search to streamline the search process and avoid any other terms that may use the VR acronym (e.g., variable reward and voice recognition). An initial 2419 studies were collected from the 3 databases used to conduct this search.
Table 1. Specific phobia treatments.

<table>
<thead>
<tr>
<th>Author</th>
<th>Phobia type</th>
<th>Methodology</th>
<th>Sessions, n</th>
<th>Session length</th>
<th>Follow-up</th>
<th>Intervention (patients, n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Botella et al [17]</td>
<td>Claustrophobia</td>
<td>Case</td>
<td>8</td>
<td>35-45 min</td>
<td>1 month</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Botella et al [18]</td>
<td>Agoraphobia</td>
<td>Controlled</td>
<td>9</td>
<td>1 hour</td>
<td>12 months</td>
<td>VRET (n=1); IVE (n=12); WL (n=13)</td>
</tr>
<tr>
<td>Carlin et al [9]</td>
<td>Arachnophobia</td>
<td>Case</td>
<td>12</td>
<td>50 min</td>
<td>None</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Emmelkamp et al [19]</td>
<td>Acrophobia</td>
<td>Controlled</td>
<td>3</td>
<td>1 hour</td>
<td>6 months</td>
<td>VRET (n=17); IVE (n=16)</td>
</tr>
<tr>
<td>Garcia-Palacios et al [7]</td>
<td>Arachnophobia</td>
<td>Controlled</td>
<td>3-10 (4)d</td>
<td>1 hour</td>
<td>None</td>
<td>VRET (n=12); WL (n=11)</td>
</tr>
<tr>
<td>Maltby et al [13]</td>
<td>Aviophobia</td>
<td>Controlled</td>
<td>5</td>
<td>50 min</td>
<td>6 months</td>
<td>VRET (n=20); EGT (n=23)</td>
</tr>
<tr>
<td>Moldovan and David [20]</td>
<td>Multiple</td>
<td>Controlled</td>
<td>1</td>
<td>60 min</td>
<td>None</td>
<td>VRET (n=16); WL (n=16)</td>
</tr>
<tr>
<td>Muhlberger et al [8]</td>
<td>Aviophobia</td>
<td>Case</td>
<td>1</td>
<td>180 min</td>
<td>None</td>
<td>VRET (n=15); RT (n=13)</td>
</tr>
<tr>
<td>Rothbaum et al [21]</td>
<td>Acrophobia</td>
<td>Controlled</td>
<td>7</td>
<td>35-45 min</td>
<td>None</td>
<td>VRET (n=12); WL (n=8)</td>
</tr>
<tr>
<td>Rothbaum et al [22]</td>
<td>Acrophobia</td>
<td>Case</td>
<td>5</td>
<td>35-45 min</td>
<td>None</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Rothbaum et al [23]</td>
<td>Aviophobia</td>
<td>Case</td>
<td>6</td>
<td>35-45 min</td>
<td>1 month</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Rothbaum et al [24]</td>
<td>Aviophobia</td>
<td>Controlled</td>
<td>8</td>
<td>1 hour</td>
<td>6 months</td>
<td>VRET (n=15); IVE (n=15); WL (n=15)</td>
</tr>
<tr>
<td>Rothbaum et al [25]</td>
<td>Aviophobia</td>
<td>Controlled</td>
<td>8</td>
<td>1 hour</td>
<td>12 months</td>
<td>VRET (n=13); IVE (n=11)</td>
</tr>
<tr>
<td>Rothbaum et al [26]</td>
<td>Aviophobia</td>
<td>Controlled</td>
<td>8</td>
<td>N/Ag</td>
<td>6 and 12 months</td>
<td>VRET (n=25); IVE (n=25); WL (n=25)</td>
</tr>
<tr>
<td>Shiban et al [27]</td>
<td>Arachnophobia</td>
<td>Controlled</td>
<td>2</td>
<td>N/A</td>
<td>None</td>
<td>MCE VRET (n=15); SCE VRET (n=15)</td>
</tr>
<tr>
<td>Whitney et al [28]</td>
<td>Acrophobia</td>
<td>Case</td>
<td>8</td>
<td>N/A</td>
<td>None</td>
<td>VRET+VPT (n=1)</td>
</tr>
</tbody>
</table>

aVRET: virtual reality exposure therapy.
bIVE: in vivo exposure.
cWL: waiting list.
dMean value.
eEGT: exposure group therapy.
fRT: relaxation therapy.
gN/A: not applicable.
hMCE: multiple context exposure.
iSCE: single context exposure.
jVPT: vestibular physical therapy.

Inclusion and Exclusion Criteria

For the initial 2419 studies collected, the following inclusion and exclusion criteria were implemented. Parameters were set to limit the studies only to those whose subject was on VR (n=217) and whose document type was either an article or a case study (n=203). Studies that did not perform an experiment using an HMD for treating or examining a specific phobia or anxiety were excluded (n=177) as were other systematic reviews or meta-analyses (n=3). HMDs were chosen as the VR system of choice for this review as the systems were the most accessible compared with CAVE and simulator-type systems, alongside the notion that the most prominent modern VR systems are HMDs. In total, 23 eligible studies met the inclusion criteria.

Another search was conducted based on the references detailed in each of the initial 23 eligible studies. Inclusion criteria for this search were that the reference title must have mentioned “virtual reality” alongside terms related to fear, anxiety, or a specific phobia, as well as having explicitly used an HMD within the study itself. A total of 27 additional studies were collected through these criteria, although 1 study appeared to have been published twice in 2 years with some minor differences; therefore, the most recent version of that study was kept [15], whereas the older version was excluded [16], resulting in only 26 additional studies. In total, 49 studies were examined for this review.

Information found in Figure 1 exhibits the process in which the studies in this review were obtained based on the inclusion criteria as well as the number of studies excluded based on the initial exclusion criteria.

Studies were also placed into 1 of the 5 categories for the purposes of this review: phobia treatments (see Table 1), PTSD treatments (see Table 2), anxiety treatments (see Table 3), paranoia evaluation (see Table 4), and innovations and evaluation (see Table 5).
Quality Assessment

Quality assessment of the collected studies was examined by both authors using the inclusion and exclusion criteria. The quality of each of the 49 studies was also appraised through the mixed methods appraisal tool (2011), which was designed to assess the methodological quality of quantitative (randomized, nonrandomized, and descriptive), qualitative, and mixed-methods studies used within systematic reviews [60].

Table 2. Posttraumatic stress disorder and acute stress disorder treatments.

<table>
<thead>
<tr>
<th>Author</th>
<th>Study type</th>
<th>Trauma type</th>
<th>Follow-up</th>
<th>Interventions and patients, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerardi et al [29]</td>
<td>Case</td>
<td>War</td>
<td>None</td>
<td>VRET(^a) (n=1)</td>
</tr>
<tr>
<td>Cardenas-Lopez et al [30]</td>
<td>Uncontrolled</td>
<td>Assault</td>
<td>None</td>
<td>VRET (n=6)</td>
</tr>
<tr>
<td>McLay et al [31]</td>
<td>Controlled</td>
<td>War</td>
<td>None</td>
<td>VRET (n=10); TAU(^b) (n=10)</td>
</tr>
<tr>
<td>Reger et al [32]</td>
<td>Uncontrolled</td>
<td>War</td>
<td>None</td>
<td>VRET (n=24)</td>
</tr>
<tr>
<td>Reger et al [33]</td>
<td>Controlled</td>
<td>War</td>
<td>3 and 6 months</td>
<td>VRET (n=54); IE(^c) (n=54); MA(^d) (n=54)</td>
</tr>
<tr>
<td>Cardenas Lopez and de la Rosa-Gomez [34]</td>
<td>Case</td>
<td>Assault</td>
<td>None</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Rothbaum et al [35]</td>
<td>Controlled</td>
<td>War</td>
<td>3 and 6 months</td>
<td>VRET (n=1)</td>
</tr>
<tr>
<td>Rothbaum et al [36]</td>
<td>Controlled</td>
<td>War</td>
<td>3, 6, and 12 months</td>
<td>VRET with D-cycloserine (n=53); VRET with alprazolam (n=50); VRET with placebo (n=53)</td>
</tr>
</tbody>
</table>

\(^a\)VRET: virtual reality exposure therapy.  
\(^b\)TAU: treatment as usual.  
\(^c\)IE: imaginal exposure.  
\(^d\)MA: minimal attention.

Table 3. Anxiety treatments.

<table>
<thead>
<tr>
<th>Author</th>
<th>Anxiety type</th>
<th>Study type</th>
<th>Sessions, n</th>
<th>Session length</th>
<th>Follow-up</th>
<th>Comparisons and patients, n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alsina-Jurnet et al [37]</td>
<td>Performance</td>
<td>Uncontrolled</td>
<td>1</td>
<td>90 min</td>
<td>None</td>
<td>High test anxiety (n=11); Low test anxiety (n=10)</td>
</tr>
<tr>
<td>Anderson et al [38]</td>
<td>Social</td>
<td>Case study</td>
<td>6 to 10</td>
<td>Unknown</td>
<td>Unknown</td>
<td>VRET(^a) (n=2)</td>
</tr>
<tr>
<td>Anderson et al [6]</td>
<td>Social</td>
<td>Controlled</td>
<td>8</td>
<td>Unknown</td>
<td>3 and 12 Months</td>
<td>VRET (n=25); EGT(^b) (n=25); WL(^c) (n=25)</td>
</tr>
<tr>
<td>Harris et al [39]</td>
<td>Social</td>
<td>Controlled</td>
<td>4</td>
<td>12-15 min/exposure</td>
<td>None</td>
<td>VRET (n=8); WL (n=6)</td>
</tr>
<tr>
<td>Padrino-Barrios et al [40]</td>
<td>Dental</td>
<td>Controlled</td>
<td>1</td>
<td>Unknown</td>
<td>Unknown</td>
<td>VR(^d) exposure first half (n=15); VR exposure second half (n=15)</td>
</tr>
<tr>
<td>Repetto et al [41]</td>
<td>General</td>
<td>Controlled</td>
<td>8</td>
<td>Unknown</td>
<td>Unknown</td>
<td>VRET with biofeedback (n=9); VRET without biofeedback (n=8); WL (n=8)</td>
</tr>
<tr>
<td>Tanja-Dijkstra et al [42]</td>
<td>Dental</td>
<td>Controlled</td>
<td>1</td>
<td>Unknown</td>
<td>1 week</td>
<td>Active VR (n=22); Passive VR (n=23); No VR (n=24)</td>
</tr>
<tr>
<td>Wallach et al [43]</td>
<td>Social</td>
<td>Controlled</td>
<td>12</td>
<td>1 hour</td>
<td>None</td>
<td>VRET (n=28); CBT(^e) (n=30); WL (n=30)</td>
</tr>
</tbody>
</table>

\(^a\)VRET: virtual reality exposure therapy.  
\(^b\)EGT: exposure group therapy.  
\(^c\)WL: waiting list.  
\(^d\)VR: virtual reality.  
\(^e\)CBT: cognitive behavioral therapy.
Table 4. Paranoia or paranoid ideations evaluation.

<table>
<thead>
<tr>
<th>Author and population</th>
<th>Patients, n</th>
<th>Age in years, mean (SD)</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fornells-Ambrojo et al [44]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early psychosis (clinical)</td>
<td>10</td>
<td>24.2 (2.3)</td>
<td>Qualitative</td>
</tr>
<tr>
<td>Healthy (Nonclinical)</td>
<td>10</td>
<td>23.8 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Freeman et al [45]: Assaulted 1 month before</td>
<td>106</td>
<td>34.4 (11.6)</td>
<td>Quantitative</td>
</tr>
<tr>
<td>Freeman et al [46]: Local adult</td>
<td>200</td>
<td>37.5 (13.3)</td>
<td>Quantitative</td>
</tr>
<tr>
<td><strong>Freeman et al [47]</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low nonclinical paranoia</td>
<td>30</td>
<td>44.2 (11.2)</td>
<td></td>
</tr>
<tr>
<td>High nonclinical paranoia</td>
<td>30</td>
<td>36.0 (11.7)</td>
<td></td>
</tr>
<tr>
<td>Persecutory delusions</td>
<td>30</td>
<td>44.2 (11.7)</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Virtual reality evaluations and innovations.

<table>
<thead>
<tr>
<th>Author</th>
<th>Disorder</th>
<th>Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornwell et al [48]</td>
<td>Social anxiety</td>
<td>Evaluating the relationship between trait social anxiety and startle reactivity</td>
</tr>
<tr>
<td>Geuss et al [49]</td>
<td>Acrophobia</td>
<td>Assessing perceptual estimates and actions of gaps within VR&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Hartanto et al [50]</td>
<td>Social anxiety</td>
<td>Evaluating the efficacy of various social stressors within VR</td>
</tr>
<tr>
<td>Orman [51]</td>
<td>Performance anxiety</td>
<td>Assessing effects of VR exposure on performing musicians</td>
</tr>
<tr>
<td>Owens and Beidel [52]</td>
<td>Social anxiety</td>
<td>Evaluating the efficacy of VR stimuli for social anxiety VRET&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Park et al [53]</td>
<td>Social anxiety</td>
<td>Assess the virtual interactions of patients with schizophrenia with digital avatars</td>
</tr>
<tr>
<td>Pertaub et al [15]</td>
<td>Public speaking anxiety</td>
<td>Evaluate participant responses toward positive, negative, and static virtual audiences</td>
</tr>
<tr>
<td>Powers et al [54]</td>
<td>Social anxiety</td>
<td>Evaluate a VR-based interactive dialogue system to elicit the same level of fear from an in vivo conversation</td>
</tr>
<tr>
<td>Price et al [55]</td>
<td>Social phobia</td>
<td>Evaluate the importance of presence within VR as a predictor of treatment response for social anxiety VRET</td>
</tr>
<tr>
<td>Qu et al [56]</td>
<td>Social phobia</td>
<td>Evaluate the influence of virtual bystanders on the participant’s self-efficacy, anxiety, social evaluation, vicarious experience, and cognitive consistency</td>
</tr>
<tr>
<td>Regenbrecht et al [57]</td>
<td>Acrophobia</td>
<td>Assessing the relationship between presence and fear of heights within VR</td>
</tr>
<tr>
<td>Slater et al [58]</td>
<td>Social anxiety</td>
<td>Assessing the efficacy of low-fidelity VR on social anxiety VRET</td>
</tr>
<tr>
<td>Veling et al [59]</td>
<td>Social anxiety</td>
<td>Evaluate the effects of childhood trauma on social stress reactivity and psychopathology within VR</td>
</tr>
</tbody>
</table>

<sup>a</sup>VR: virtual reality.
<sup>b</sup>VRET: virtual reality exposure theory.

Results

Quality Assessment Outcomes

The 49 studies received an average rating of 86.73% and a modal rating of 100% (n=30). A total of 10 studies were classified as qualitative, 21 studies as quantitative randomized, 15 studies as quantitative nonrandomized, and 3 studies as quantitative descriptive based on the parameters set by the mixed methods appraisal tool [60].

Summary of Papers

Specific Phobias

The implementation of VRET for the treatment of specific phobias typically mirrors traditional phobia treatment protocols; treatment rationale was explained upon or before the patient’s arrival, information-gathering procedures were used to assess the patient’s phobic level, and a stimulus hierarchy would be established based on the information gathered. Levels of the stimulus hierarchy would vary based on the phobia being treated but generally would incorporate a new level or factors as the patient progresses. For example, acrophobia patients undergoing VRET would often progress through greater heights [19,21,22], whereas aviophobia patients would experience the next stage of a flight (eg, stationary and take-off) [8,23,24,26]. In short, regardless of the research methodology used or the specific phobia examined, treatment procedures were consistent across the 16 specific phobia studies.

A few studies compared the efficacy of VRET with that of a pre-established treatment including standard IVE [18,19,24], relaxation therapy [8], or exposure group therapy [13]. In 1
study comparing treatment outcomes of VRET, IVE, and a waiting list condition for participants with agoraphobia, no significant differences were observed at the posttreatment and 12-month follow-up assessments between those that underwent VRET or IVE, but both groups did demonstrate significant improvements over those in the waiting list condition [18]. The comparison for VRET and relaxation therapy yielded similar results; however, it was found that although VRET was more effective in reducing flying avoidance in participants, it was only marginally better at reducing the participants’ fear of flying ratings compared with relaxation therapy [8]. Finally, in a comparison of VRET and exposure group therapy, more VRET participants experienced clinically significant change compared with exposure group therapy participants based on posttreatment assessments, but the significant difference disappeared between the 2 groups during the 6-month follow-up [13].

**Posttraumatic Stress Disorder and Acute Stress Disorder**

Studies investigating the efficacy of VRET on PTSD and ASD typically focused on patients who developed the disorder due to wartime combat or physical assault. Initial sessions followed the same format and components as the ones used for specific phobias, but VRET sessions were more personalized for each patient. For example, veterans were given a virtual environment that matched the war environment that they had participated in, which included a jungle for the Vietnam War [35] and a desert city for Middle Eastern wars [29, 31-33, 36]; victims of physical abuse unrelated to war were placed in an urban environment [30, 34].

A comparison of VRET, IE, and waiting list conditions found that, although VRET and IE both led to significant improvements in PTSD symptoms compared with the waiting list, IE was superior based on the Clinician-Administered PTSD Scale (CAPS), a structured interview performed by the clinician to gauge the severity of PTSD-related symptoms [61]. Follow-up assessments conducted at 3 and 6 months also indicated that those who underwent IE experienced continual improvement, whereas those who underwent VRET did not [33].

Another study sought to compare VRET with a treatment-as-usual condition, which consisted of patients performing their pre-established treatments, which included or was a combination of prolonged exposure, eye movement desensitization and reprocessing, and group therapy. A posttreatment assessment using CAPS indicated that 70% (7/10) of patients that underwent VRET showed at least a 30% improvement, whereas only 11% (1/9) of treatment-as-usual patients showed the same level of improvement. Although this difference was deemed as significant, the authors noted that a small sample size and wide variability in the treatment-as-usual condition limited the interpretations of the study’s outcomes [31].

Finally, 1 study investigated whether augmenting VRET with D-cycloserine, a glutamate receptor that had been demonstrated to improve the efficacy of exposure therapy for severe anxiety disorders, would also benefit VRET. All participants in the study underwent VRET but were given D-cycloserine, alprazolam (used primarily as a pharmacological treatment for anxiety), or a placebo pill. The study reported no significant differences in treatment outcomes for any of the groups based on CAPS scores; however, participants who were dosed with D-cycloserine experienced significant extinction learning that was not observed in the alprazolam and placebo groups, suggesting that the use of D-cycloserine helped to enhance learning effects during VRET [36].

**Anxiety**

Studies that focused on general or specific (social, public speaking, dental, or test) anxiety utilized VR as a method to deliver VRET or VR distraction interventions. Although VRET for general and specific anxiety largely mirrored the same procedural format as the VRET done for specific phobias, PTSD, and ASD, VR distraction was used to comfort patients during a dental procedure. Although VRET aims to address problematic behaviors and cognitions by exposing patients to a virtual simulation, VR distraction serves to give patients a more positive experience during an otherwise anxiety-inducing situation [40].

A comparison was conducted for the efficacy of VRET to CBT and waiting list conditions for the treatment of public speaking anxiety, and findings were largely concurrent with the specific phobia studies; both treatment groups experienced significant improvements over the waiting list, but did not significantly differ with each other based on posttreatment assessments [43]. A similar finding was reported when VRET was compared with EGT as an intervention for public-speaking anxiety across posttreatment, 3-month, and 12-month follow-up assessments; however, the study had a small sample size that limited the findings [6].

A study that evaluated VRET for the treatment of general anxiety disorder incorporated biofeedback and a mobile, rather than a computer-based, VR system. The virtual environments for the biofeedback group, which depicted various scenes associated with relaxation, could change based on the patient’s heart rate and physiological activation; a reduction in either results in a reduced intensity for certain stimuli within the virtual environment. Virtual environments for the VRET without biofeedback and waiting list groups experienced the same scenes but without the additional biofeedback features. Those who were in the biofeedback group were reported to have a significant decrease in behavioral avoidance and state anxiety, whereas the VR without biofeedback group only experienced a significant decrease in behavioral avoidance, and the waiting list group experienced no significant changes [41].

VR distraction was utilized for both dental anxiety studies in this review, which were conducted during either a simulated [40] or live [42] procedure. The stimulated dental procedure study compared active VR, passive VR, and no VR; those in the active VR condition could freely navigate around the virtual environment, whereas those in the passive VR condition could not. Those with higher levels of dental anxiety in both the active and passive VR conditions were reported to have less vivid memories of the procedure compared with those that completed the procedure without VR [40]. Similar findings were reported for the live procedure study in which an oral prophylaxis (teeth cleaning) was performed. Participants were randomly assigned to 1 group that received VR distraction during the first half of the procedure and another group that received the VR distraction...
during the second half. Participants in both groups experienced significantly greater calmness during the portion of the procedure when they received VR distraction compared with the portion when they did not [42].

**Paranoia or Paranoid Ideations**

The process of diagnosing paranoia has been difficult to do in real settings, as therapists must be able to discern whether an individual’s claims are legitimate or based on true paranoid beliefs. Through the use of VR, the diagnostic process for paranoia can be more reliable as the therapist has more control over the virtual stimuli, environment, and situational factors; avatars in the virtual environment cannot physically harm nor be harmed by the patient, and paranoid beliefs that surface during VR exposure can be verified [47]. As there were not a lot of studies dedicated to this topic, each study employed the same task within the same virtual environment: participants rode a London Underground train for a few minutes surrounded by avatars with neutral expressions and mannerisms.

In 1 study, individuals were found to be twice as likely to experience some form of persecutory thoughts during VR exposure if they reported paranoid ideations in day-to-day life [46]. This finding provided support toward the notion that neutral avatars were capable of eliciting paranoid thoughts, which was further confirmed in another study that compared the reactions of individuals belonging to clinical paranoia, high nonclinical paranoia, and low nonclinical paranoia groups [47].

**Innovations and Evaluations**

Although previous sections covered how VR has been used to treat or study certain anxiety-based disorders, it is worth noting the studies that have sought to either study VR-specific features or create innovative programs to enhance VR-based treatment. For example, an interactive dialogue system for a study on social anxiety was developed to elicit fear responses during VR exposure to match the fear response levels observed in in vivo conversations. Although the study reported that participants believed in vivo conversations were more realistic than the ones held in VR, fear ratings were found to be significantly higher for VR conversations than for in vivo ones. Although realism is an important factor, the authors considered that fear was a more important factor in the context of treating symptoms of social anxiety [54].

Another study sought to use dynamic social dialogue systems to manipulate the participant’s feelings of anxiety in real time and effectively demonstrated that different ratios of positive and negative responses could serve as effective anxiety stressors to manipulate the participant’s anxiety level in any direction (low to high) at any time [50]. Other studies aimed to evaluate changes in audience behaviors and other social stressors (eg, number of avatars present and ethnic diversity) and demonstrated similar levels of efficacy in manipulating the patient’s anxiety levels [52,59].

Several studies were also conducted to evaluate whether VR stimuli were capable of eliciting real emotions, a crucial factor for the treatment and assessment of specific phobias and other anxiety disorders. There are some mixed findings; although a study on acrophobia found evidence that participants experienced real fear when exposed to a virtual cliff [57], a study on woodwind performance anxiety found inconsistencies in subjective anxiety ratings during a performance in a virtual concert hall [51]. For the latter, the authors speculated that an increase in heart rate during VR exposure may have been due to the nature of performing on a wind instrument rather than due to the VR exposure, and the inconsistent subjective anxiety ratings could have been due to performers finding the act of performing to be psychologically calming rather than anxiety-inducing.

**Discussion**

**Principal Findings**

In relation to VRET, there appears to be an overwhelming amount of positive evidence that the VR-based treatment has an equal or greater efficacy toward the treatment of specific phobias and anxiety, but not as much for PTSD and ASD. This evidence comes from a mix of experimental designs, including case studies, controlled randomized trials, and within-group designs, with some studies also offering follow-up results as evidence of VRET’s effects beyond posttreatment. Although the use of VRET for PTSD and ASD was effective, it appeared that some patients seemed to gain continual improvement when treated with another treatment option such as IE [33]. Regardless, those who underwent VRET consistently showed significant improvement over those in the waiting list groups in the specific phobia, specific anxiety, and PTSD and ASD studies that compared the 2 together.

The use of VR to aid in the diagnosis of paranoia was also largely shown to be effective and was further reinforced due to every study related to the topic in this review using the same procedures and virtual environment to study or differentiate between individuals with varying levels of paranoia. Although VR was not used as a treatment tool for paranoia, it does provide a safe environment for the patient while simultaneously giving therapists and researchers a way to accurately identify any paranoid ideations that may arise due to VR exposure.

Finally, there have been many innovations to bolster the user’s sense of immersion, or the feeling of being present, within an environment afforded by VR technology, at least for programs focused on treating social anxiety disorders. These innovations were largely focused on making VR avatars more realistic and sociable, ranging from increasing the realism of a person-to-avatar conversation to the manipulative behaviors of multiple avatars that comprise a virtual audience. In general, these innovations achieved their purpose by eliciting a greater amount of fear within the participant [54] or providing a dynamic manipulation of participant anxiety levels [50,52,59]. As for the evaluation of VR elements, simply recreating an object that the participant fears, such as a virtual cliff for those with acrophobia [57], is enough to generate real fear, although testing VR’s efficacy on some tasks, such as performing on a woodwind instrument [51], may prove to be difficult due to the nature of the task itself and how it may conflict with common psychological or biometric measures.
Limitations
Although a large number of studies were included in this review, some topics appeared to be more researched than others, thus providing varying levels of quality and quantity. In particular, there were a small number of studies dedicated to paranoia, and although every study included in this review related to paranoia utilized the same virtual environments and procedures, the results may have been strengthened with more variety in the types of virtual environments used beyond the London Underground.

This issue also persists for the specific anxiety and PTSD and ASD studies in this review, where there was 1 clear subject that dominated, whereas there were only a few studies that ventured beyond what was commonly researched. For specific anxiety, there were more studies focused on social or public speaking anxiety, with only a couple of studies focused on dental anxiety, and the PTSD and ASD studies largely focused on war-induced trauma rather than physical assault–induced trauma.

Future Research
The landscape of modern VR has changed drastically compared with the VR systems used in most of the studies included in this review. Although legacy VR systems were expensive, required users to receive special training to operate or create VR programs, and were limited to facilities that could invest in the technology, modern VR has provided cheaper entry points, a vast library accessible through popular digital storefronts such as Steam or Google Play and Apple App Store, and user-friendly experiences. Although the most powerful VR systems available today are mostly geared toward gaming, the same systems can provide some use toward the study, diagnosis, or treatment of various anxiety-based disorders.

One area of research that would be worth pursuing is a self-directed rendition of VRET that can be done within a patient’s home with little to no therapist interaction. As there is an overwhelming amount of positive evidence toward the efficacy of VRET, at least in relation to specific phobias and anxieties, the next step toward evolving VRET may be to evaluate whether those with mild to moderate anxiety-based systems may benefit from merely exposing themselves to anxiety-inducing stimuli within a virtual environment. Self-directed interventions provide patients with care in areas with limited to no access to therapists as well as to those who may be reluctant to see a therapist [62]. By utilizing a self-directed approach to VRET, it may be possible to allow individuals with low anxiety severity to treat themselves at their own pace, within their own home, and without the need for a therapist.

Conclusions
This review evaluated a variety of topics related to the use of VR for anxiety-based disorders, including VRET for specific phobias, specific anxieties, PTSD and ASD, and paranoia, while also outlining various innovations and evaluations conducted by studies to either improve the experiences afforded by VR or investigate the various factors that contribute to its efficacy toward anxiety-based treatments. These studies provided generally positive evidence toward the diagnostic and treatment capabilities of VR for anxiety-based disorders; however, research into VR has generally been limited to institutions that had the resources to invest in it. With the advent of more affordable, user-friendly, and supported commercial VR systems, more VR research can finally be done by building on the foundation laid out by the early studies to both replicate past findings and establish new uses for VR within psychotherapy.

Conflicts of Interest
None declared.

References


Abbreviations

- ASD: acute stress disorder
- CBT: cognitive behavioral therapy
- CAPS: Clinician-Administered PTSD Scale
- EGT: exposure group therapy
- HMD: head-mounted display
- IE: imaginal exposure
- IVE: in vivo exposure
- MA: minimal attention
- MCE: multiple context exposure
- PTSD: posttraumatic stress disorder
- RT: relaxation therapy
- SCE: single context exposure
- TAU: treatment as usual
- VPT: vestibular physical therapy
- VR: virtual reality
- VRET: virtual reality exposure therapy
- WL: waiting list

©Theodore Oing, Julie Prescott. Originally published in JMIR Serious Games (http://games.jmir.org), 07.11.2018. 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.
Developing Digital Games to Address Airway Clearance Therapy in Children With Cystic Fibrosis: Participatory Design Process

Fabio Balli\textsuperscript{1,2}, GradD, MAS

\textsuperscript{1}Breathing Games, Montreal, QC, Canada
\textsuperscript{2}Concordia University, Montreal, QC, Canada

**Corresponding Author:**
Fabio Balli, GradD, MAS
Concordia University
1455 de Maisonneuve West
Montreal, QC, H3G 1M8
Canada
Phone: 1 514 848 2424
Email: info@fabioballi.net

**Abstract**

**Background:** Children affected with cystic fibrosis do respiratory exercises to release the mucus stuck in their lungs.

**Objective:** The objective of our study was to develop prototypes of digital games that use breath pressure to make this daily physiotherapy more fun.

**Methods:** We used a participatory design approach and organized short events to invite contributors from different disciplines to develop game prototypes. From the 6 prototypes, 3 were tested by 10 children during a prestudy. The source code of the games, of which 2 continue to be developed, has been released on the internet under fair use licenses.

**Results:** We discuss 7 themes of importance in designing games for health, combining our experience with a review a posteriori of literature.

**Conclusions:** This study provides examples of games and their pitfalls as well as recommendations to create games for health in a participatory approach that enables everyone to improve and adapt the work done.

*(JMIR Serious Games 2018;6(4):e18)* doi:10.2196/games.8964

**KEYWORDS**
digital games; children; cystic fibrosis; game design; treatment adherence and compliance; intersectoral collaboration; access to information; recommendations; telemedicine; mobile phone

**Introduction**

**Burden and Treatment of Cystic Fibrosis**

Cystic fibrosis (CF) is a chronic disease estimated to affect 70,000-100,000 people worldwide [1]. Although advances in therapeutic care may now provide a life expectancy of >40 years [2], respiratory diseases still represent the main cause of morbidity and mortality [3]. The expectoration of airways’ surface bacteria is normally assured by a layer of slippery mucus produced by the lungs’ epithelial cells. Because the mucus is thick and viscous in patients with CF, mucus and bacteria accumulate, which results in the obstruction, infection, and inflammation of the airways. Therefore, reducing the progression of respiratory diseases is of the utmost importance to prevent early death in patients with CF [4].

Airway clearance techniques, such as positive expiratory pressure (PEP) physiotherapy, have been highlighted as effective means to diminish an airway’s obstruction, infection, inflammation [5], and subsequently reduce pulmonary exacerbation [6]. By expiring through a tube that limits the flow of air, the device forces to make longer expirations, which help in keeping the airways open and removing the mucus stuck in the airways [7]. However, the burden of the treatment regimen with regard to CF has been highlighted numerous times [8], especially for children [9]. PEP physiotherapy presents the most time-consuming activity in the treatment, and poor adherence to this therapy is reported in almost half of the children affected with CF [10,11].
The Use of Digital Games to Increase the Treatment Adherence

Transforming the respiratory therapy into a digital game has shown promising results in increasing adherence [12,13]. Digital games are artificial systems in which players interact, which comprise conflict, rules, and quantifiable outcomes [14]. Games for health is an emerging field of “serious gaming,” that is, playing with a purpose beyond entertainment [15]. Games for health have shown promising results in increasing knowledge, delivering persuasive messages, modifying behaviors, and improving health outcomes [16,17]. Advantages of such games over other therapeutic methods appear in our practice, including the following: making respiratory therapy fun through storytelling and interaction; altering time perception through immersive experience; making the long-term impact of daily therapy visible; developing and validating knowledge through learning sequences and game levels; fostering a constructive mindset through meaningful quests and collaborative gameplay; and collecting data on a regular basis about the effectiveness of the therapy.

However, Baranowski et al [16] emphasized that games for health should be much more engaging than the ones currently available to be truly effective. They recommend involving multidisciplinary stakeholders in all development phases to ensure the usability, desirability, and feasibility of games. Involving children affected could help them take ownership of their care by designizing levels and gameplays to make their treatment fun.

This paper presents an overview of the participatory design process [18] used in the development of 6 games aimed at increasing the adherence to PEP therapy. The games were based on the traditional PEP therapy technique, in which an electronic pressure sensor was added to the PEP mouthpiece, transforming breath to a digital signal used to control the game character. In addition, this study will briefly describe each game that was developed and present the pilot evaluation we used. Because the development process of games for health and discussion of failures in their design are rarely documented [19], we will conclude by summarizing the challenges we faced and by stating what we learned, so that it might help researchers and practitioners avoid the same pitfalls.

Methods

Participatory Design Process

The project started in January 2014 between 2 game design students and the CF pediatric team at Sainte-Justine Hospital (Montreal, Canada). The clinicians were looking for a novel way to make respiratory therapy more engaging for their young patients. An initial environment scan and literature review targeting CF games showed us only a few games (AstroPEP, CFpal, Creep Frontier, Flower for all, KIBreath, LudiCross, and MyCarnival [12,20]), which were all very basic games and were not available to the public. Discussions enabled the students to understand the PEP exercise to build the game (Figure 1). Furthermore, an electronic pressure sensor that was developed previously was used so that breath could act as an input for the game (Figure 2) [21]. The project was then named “Breathing Games” [22].

Initial Developments

It took up about 800 hours to develop the 3 initial prototypes—Globule, Ange-Gardien, and PEP Hero. The key elements of the games were defined (Figure 3), including goals, gameplay, levels, obstacles, mechanics, game flow, visual universe, story, and character. The aim was to experiment with different game mechanics to test with children. Globule was designed as a simple puzzle, which could easily include a level editor so that children could create their own levels. Ange-Gardien was conceived as a tower defense, where children could manage different resources with an increasing level of difficulty. Celebrations was designed as a contemplative game, which inspired from open worlds, where children could explore different spaces without a defined goal. In Globule, the breath exercise served to collect elements that could be used in the game; in Ange-Gardien, breath activated towers that would kill enemies; in Celebrations, it gave energy to the character so that it may move. Globule and Ange-Gardien were developed part time between January and April 2014 by 2 game design and 2 sound design students. Celebrations was abandoned because of the resources required to develop a multiplayer game.

In February 2014, the team took part in a 3-day coreation event (hackathon) organized at Sainte-Justine. The team presented the project, and 5 young professionals with skills in visual arts, user experience, electronics, and communications joined. The team then created PEP Hero, a side-scrolling game. One participant worked on the user interface, another on the visuals, 3 on the design and coding, 2 on communications, and one on the electronics.

Our approach to developing the games was pragmatic; we designed the games and the CF team was available for advice. We developed the prototypes in cycles according to the value created for the end user (agile method, Figure 4). All contributions were made without funding. We used the broadly used game engine Unity 3D, which, unfortunately, is not free (libre) and open source.

Figure 1. A graph of positive expiratory pressure exercise, own representation. Red lines, inspirations; green lines, holding; blue lines, expirations. Credit: breathinggames.net.
Figure 2. A device transforming expiratory breath into digital data through a pressure sensor. Credit: breathinggames.net.

Figure 3. The document describing the key elements of the game Globule. Credit: breathinggames.net.

<table>
<thead>
<tr>
<th>Globule</th>
<th>F. Ball – Y. Genais</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Objectifs</strong></td>
<td>Renforcer une habitude de soin</td>
</tr>
<tr>
<td><strong>Plateforme</strong></td>
<td>Ordinateur, tablette et téléphone, Unity Player</td>
</tr>
<tr>
<td><strong>Genre</strong></td>
<td>1 à 12 ans, puzzle</td>
</tr>
<tr>
<td><strong>Mode</strong></td>
<td>Solo, Frizzle Fraz, Monkey go happy, Sonic</td>
</tr>
<tr>
<td><strong>Mécaniques, collectibles</strong></td>
<td>Les points de mouvement sont</td>
</tr>
<tr>
<td></td>
<td>trouver au nord, au sud, à l’est, à l’ouest, s’il s’agit d’un obstacle</td>
</tr>
<tr>
<td></td>
<td>Les obstacles sont</td>
</tr>
<tr>
<td></td>
<td>carrés, ronds, rectangulaires, etc.</td>
</tr>
<tr>
<td></td>
<td>Les collectibles sont représentés par un champ d’étoiles.</td>
</tr>
</tbody>
</table>

| Contrôles | Les points de mouvement sont gérés par la souris. |
| | L’avancement de la boule est géré par le souffle. |
| | Le zoom de la caméra peut être adapté via le clavier. |

| Audio et interface | Langue : Français, accès vers anglais |
| | Musique : Selon l’environnement |
| | Navigation : … |

| Niveaux, séquences, modes | La difficulté évolue au cours du jeu en fonction des résultats obtenus. Les facteurs : la longueur du parcours augmentent les obstacles, la vitesse qui diminue, les points de réserve deviennent plus nombreux, le temps devient limité, les points de récupération disparaissent après un temps donné. Le joueur peut commencer à partir des niveaux déjà atteints. |

| Envennez, obstacles | Le joueur peut requérir divers changements de direction pour que le joueur échappe de manière aléatoire dans le décor ou de rester bloqué contre des murs ou dans des endroits dont il ne peut pas sortir. Le joueur ne peut placer de points sur le carré où il se trouve. |

| Personnages | Le joueur est représenté par la boule. |

| USP | Environnements |
| | Architecture narrative |

**Modèle**
- **Objectif** : Contribuer au bien-être d’enfants malades.
- **Modèle** : Application open source.
- **Budget** : 0, 1 semaines, 2 personnes.
- **Revenus** : Indirect : visibilité, espérance.
- **Distribution** : Site web, promotion via réseaux sociaux.
- **Partenaires** : CHU Montbonnot, CHU Vaucluse, Afference.
- **Concurrents** : Le jeu n’est pas dans une optique de compétition.
Beyond the iterative process, people from different disciplines, including game designers, software developers, visual artists, sound composers, electronics engineers, physiotherapists, and pulmonologists, were involved. Children affected were involved in testing, but we were not able to mobilize patients with CF and family community during the cocreation process. We hypothesize that this is because of the burden already related to CF care. Involving end users is however important to develop more effective games [17]. A critical reflection on the role of serious games was also done during this period [23].

**New Developments**

It took over 1000 hours to develop 3 further prototype games: *Heritage*, *Bloïd*, and *Les aventures du Briand*. Building on the previous learning, we conceptualized *Heritage* in November 2014. The game was developed during 5 cocreation events by >20 contributors. In addition, 3 students in software development worked respectively on the breath input [24], the game design [25], and the creation of a game core [26]. *Heritage* included a settings panel so that the exercise could be adapted to practices of different countries. *Bloïd* and *Les aventures du Briand* were conceived during a game jam in March 2016 and continued at another game jam in February 2017 (Figure 5). These 3-day cocreation events were called jams to appear less technical and involved 15 contributors and collaborations with the University Hospital of Canton Vaud (CHUV), the Genevan Federation against CF (FGLM), and Lift Conference. In addition, a participatory design process similar to the one adopted for the prototypes was used. In these 3 games, the breathing exercise is fully integrated to the gameplay, and the breathing cycle is not represented as a chart.

**New Tests**

*Heritage* and *Bloïd* were tested informally during the game jams by a few children and adults, who gave a positive feedback. A mixed-methods study is being prepared to test these games. *Les aventures du Briand* requires further development before it can be tested.
## Results

### Games Developed

#### Globule

**Gameplay**

Globule (Figures 6 and 7) is a puzzle game that uses an expiratory pressure device. The breathing exercise and the gameplay are separated to potentially add the breathing component to existing games. Each game level is composed of the following 3 phases: 1) children do their breathing exercise based on the therapeutic pattern; for each breath done with the right duration and pressure, children receive an arrow pointing north, south, east, or west; 2) children have a break to expectorate mucus; and 3) they play the puzzle by placing the arrows to direct the character toward a goal. Children are also able to select among 3 environments: the desert, the jungle, a futuristic universe.

### Tests

Overall, 8 children tested the prototype [27]. After playing, most children found that the game was very fun, how to play was easy to remember, doing the therapy while playing was easy, and that the game was adapted to the treatment. All children liked to play with the arrows, path, and the level that was more difficult. One child did not understand the purpose of the game.

Regarding improvements, children proposed to add characters, have more time to set the arrows, add guitar music, correct the mistakes, correct the exhalation time, and connect it to the hardware (we had a problem with the pressure sensor in that individual test). In addition, children suggested we add levels, music, and more variety to make the game interesting over many weeks. Children mostly rated the game high; the first child testing the game gave a negative evaluation because the pace of the exercise was too fast, which was corrected.

### Strengths

Based on our observations, children were able to perform the expiratory time and constant pressure required for the therapy, and get rewarded for their performance. The puzzles are easy to customize and construct (visuals for the squares, protagonist, and background), which could enable children to create and share their own maps. The minimalist story (collecting arrows to reach a goal) seemed to positively impact children.

### Limits

The puzzle in itself has no connection to breathing. Parents mentioned that separating the breathing exercise and the game requires more time than the therapy alone, which they see as inadequate, given the time already required daily.

### Status

The development of the game was discontinued due to the aforementioned limits.
Ange-Gardien

Gameplay

Ange-Gardien (Figures 8 and 9) is a strategy tower defense game that uses an expiratory pressure device. Here, the breathing exercise and the game are combined. Children defend the little guardian angel that fell from his shoulder after being kicked by the devil on the other shoulder. The devil wants to capture the angel before the child saves it. When the game starts, the child sees a map with different paths and places to build towers. The child receives a certain number of points to build different kinds of towers, representing wind, fire, and water. Then, enemies arrive on different paths. The child clicks the tower to activate, and breathes out to make that tower shoot at the enemies around. Each devil destroyed gives points. Each demon reaching the other side of the map reduces the number of lives of the angel. After 5 seconds, the towers deactivate, prompting the child to inspire. When the breathing cycle is done, no more enemies appear because children must do their expectorations. On the next level, towers can be built or upgraded using points. The game was developed and tested on the basis of geometric shapes; visuals were prepared but not integrated into the game.

Tests

Only 2 children tested the game because of delays in the development. After playing, children found that the game was very fun, remembering how to play was easy, doing the therapy while playing was easy, and that the game was adapted to the treatment.

Children liked to set the towers and the evolution of the scene. Regarding improvements, a child proposed to correct a bug in the gameplay. In addition, children suggested adding levels, colors and other visuals, towers, enemies to make the game interesting over many weeks. Children rated the game high on a scale from 10 (best) to 1 (worst).
Strengths
The mechanics can evolve from very simple to complex strategic gameplay and breathing cycles can easily be related to different parts of the gameplay. The game can be adapted to multiple players and can also include input from the children’s relatives.

Limits
Although this was not brought up by the testers, the angel devil may not be well perceived depending on religious beliefs. In one design, the enemies are represented as mucus; this created a discussion whether integrating elements of the disease in the game would support or discourage the children in their efforts.

Status
The development was discontinued because of major changes in the game engine. The gameplay should however be reused.

**PEP Hero**

**Gameplay**

PEP Hero (Figures 10 and 11) is a side-scrolling space shooter that uses an expiratory pressure device. The game simply adds a visual universe on the top of the therapeutic exercise. The child is the captain of a spaceship that goes up when breathing out and down when breathing in. By directing the vessel, the child is able to collect items. Afterwards, the child has a break to expectorate and can then start a new cycle. The game included various vessels and items to collect, but their selection was not functional during the tests.

In the first individual test, the green items were spread over the screen; this did not fit with the exercise because the child did not know when to breathe. The items were then aligned, which helped the child to have a continuing exhalation but did not show them when to breathe in. Thus, additional items were added for the inspiration phase, that helped the child adopt a regular breathing pace.
Tests
Overall, 8 children tested the game. After playing, most children found that the game was very fun, remembering how to play was easy, doing the therapy while playing was easy, and that the game was very adapted to the treatment.

Children liked everything, including the ending, the visuals, and the first level. Regarding improvements, children proposed to increase the duration of inhalations (corrected after the individual test) to make the first expected breathing visible and solve a bug. Children suggested to add other elements such as levels and colors to make the game interesting over many weeks. Children rated the game high on a scale from 10 (best) to 1 (worst).

Strengths
The strength of the study was that simple mechanics mimicked the breathing exercise with the advantage of giving direct feedback about the correct pressure and duration of breathing. Visuals can be easily adapted.

Limits
A long-term interest could be hard to ensure because the mechanics are very simple.

Status
The development was discontinued, as Heritage was created, developing further similar mechanics.

Heritage
Gameplay
Heritage (Figures 12 and 13) is a side-scrolling game using an expiratory pressure device, in which a child defends the cultures of different countries. For Greece, for example, the character has to catch a thief who has seized Poseidon’s trident-fork. If he recovers it, the character must then find who stole Aphrodite’s potion, Zeus’ lightning, Gaia’s seed bag, Hermes’ shoes, and Athena’s rouet to collect more points. In the end, the God of the Seas will then prepare the meal for the party with his trident-fork; Aphrodite will take care of the atmosphere; etc. During the journey, a child learns facts and figures about different countries.
A table was created during a cocreation event to imagine different stages (daily exercise) with each stage having 6 levels (breathing cycles). For each level, cultural elements were defined. Peru, for example, was proposed by a participant who knew the country. For Peru, the background elements would be the sky, nonplayer supporting character would be the goddess Pachamama, the stolen object would be the earth, the transport means would be a condor, obstacles would be clouds, the asset would be an unku (piece of cloth). Graphic elements were created by different artists, who, while keeping the same line, always changed parts of already existing elements.

**Bloïd**

**Gameplay**

*Bloïd (Figures 14 and 15)* is a side-scrolling space shooter that uses an expiratory pressure device. In this game, a spaceship has to thwart the traps of an intergalactic journey to reach a mysterious planet mentioned in a prophecy. When the child breathes in, meteorites appear in front of the spaceship; when the child breathes out, projectiles are launched to destroy the meteorites. The keyboard is used to direct the ship. Over time, the number of meteorites that appear at each inspiration grows to encourage players to extend their expiration, slow down their pace, and develop their lung capacity; this enables them to gain a better score. The visuals were created to be realistic and favor immersive play.

Initially, the goal was to create a game similar to the previous ones based on the traditional CF exercise. The team of designers, developers, musicians, and patients found however that the proposed pattern was too limiting to create compelling gameplay that would really increase adhesion. Hence, they decided to create a more flexible exercise, in which breathing would be less constrained. *A posteriori*, the gameplay relates more to a heart rate variability exercise (cardiac coherence) that can help cope with stress.

**Les Aventures du Briand**

**Gameplay**

*Les aventures du Briand (Figures 16 and 17)* is an adventure game that uses an expiratory pressure device. In this game, a player is the captain of a ship, who has to breathe to solve challenges faced by the crew. The game is composed of different
mini-challenges related to breathing exercises. Similar to Bloïd, the team that conceived the game wanted to break free from a strict breathing exercise and instead, opt for a more flexible pattern. The outcomes of the game need to be further researched on.

Figure 14. A visual of the game Bloïd: introductory explanations. Credit: breathinggames.net / A D-Pierson / K Piccand.

Figure 15. A visual of the game Bloïd: breathing exercise. Credit: breathinggames.net / A D-Pierson / K Piccand.

Figure 16. A visual of the game Les aventures du Briand: start screen. Credit: breathinggames.net / J Danger.
Discussion

Principal Findings

Our approach to developing these games was pragmatic; we applied our understanding of game design to achieve a goal defined in collaboration with clinicians. Although we collectively discussed and documented key elements of different games, we did not follow any specific methodology but built on our experience and intuition. In this section, we synthesize findings from publications collected around digital games conception. Building on this new body of knowledge, we highlight key elements and learnings in the field that we had realized beforehand. The 7 themes presented are based on recurring elements mentioned in the publications selected.

People Involved in the Creation

Literature Findings

Games for health are highly influenced by the team conceiving them—the diversity of backgrounds, their perspective of health (from the reductionist biomedical management of a disease to a holistic promotion of well-being), and the level of participation in decision making.

End users should be actively involved in the creation process—their input about their pleasure of playing, interests, ways of learning, and beliefs can help make the games more relevant [28]. Multidisciplinary teams can help them translate their ideas into design elements [29]. Experts, on the other hand, can help integrate health knowledge and broaden the narrative [30]. Medical care professionals can ease the interactions with the patients while keeping ethics in mind [31] but will often not be able to evaluate the game or its safety [32].

DeSmet et al [17] shared a participatory approach that seems more adapted to create narratives, characters, and other game elements than to develop gameplay.

Learning Outcomes From the Field

The involvement of young professionals in game design, software development, visual arts, and project management was essential to realize the prototypes. In addition, electronics engineers, communications specialists, and people in the free software community contributed to building the breathing sensor, increase our Web-based outreach, and sketch the sociolegal framework, respectively.

Despite collaborations with patient associations, we were able to mobilize only one CF family through all of our events. To address this gap, patients with CF have been recruited by our partner hospital for an upcoming event in France. The contribution of medical care professionals was particularly useful in setting the goals, mechanics, and content for the games, and to answer questions during the events. Owing to their status, they may, however, overshadow the experience and needs of patients and push the focus on managing the disease versus the promotion of health and well-being of a person overall in the long term. Bringing too much medical content may reduce the fun and impact of a game.

Cocreation Process

Literature Findings

When creating a game for health, designers can either copy a successful example of gameplay, integrating the medical content into the story, or conceive a game that embeds medical information into the gameplay. The first can help in memorizing information, whereas the second allows for targeting more complex learning needs [15].

Working in short cycles gives the ability to attain better results, value rapid learning, and fine-tune the existing work in a short period. Dow et al [33] stated that contributors working in cycles on a new task are as efficient as contributors working sequentially on a task they master.

Learning Outcomes From the Field

Given that the work evolved through events, the games were developed using an agile methodology: prototypes were designed and a minimal viable product was developed and further improved by prioritizing the functions to be added. Because the aim was to make a well-defined physiotherapeutic exercise more attractive, the team worked on how to create gameplay around it. In a less binding context, creating a story...
and gameplays where health-related content can be easily embedded appears more appealing.

The immediate documentation of the work, including uploading the source code of the game, and of editable visual and sound elements is essential. The tools and the fair use licenses used are described in another paper [18].

One key learning outcome after hosting many health game jams is that such events are well adapted to raise awareness, foster collaboration, recruit long-term contributors, and create a multitude of prototypes to test different ideas. Because they are ephemeral, these events, however, do not allow the creation of high-quality games. Prototypes realized during game jams can, however, be redeveloped by a multidisciplinary team of professionals and tested during dedicated co-creation events.

Gameplay

Literature Findings

Sciart described gameplay as “a ludic experience regulated by game rules, mediated by game mechanics, and oriented to the satisfactory achievement of goals predetermined by the game and agreed upon by players.” page 103 [34]. Gameplay can be separated into the rules of the game world (mechanics), a player’s interaction with these rules (dynamics), and the emotional experience that emerges (affects) [15].

Orji et al [35] propose to either create a variety of gameplay (“competition and comparison, cooperation, customization, personalization, praise, self-monitoring and suggestion, simulation, and reward,” page 473) that can mobilize different kinds of players (“achiever, conqueror, daredevil, mastermind, seeker, socializer, and survivor,” page 473) or adopt the gameplay that better answers all players, possibly through comparison or self-monitoring. Indeed, gameplay can fill different kinds of needs—one player will like to explore, the other to compete, the third to socialize; one will like fast actions, another to solve problems, and a third one to build strategies.

To help design the player experience, the following 6 elements can be checked: voluntary participation; a commitment toward a goal; a physical involvement; the creation of meaning; the perception of feelings; and the creation of social relations [36].

Learning Outcomes From the Field

The types of games (quiz, adventure, shooter, tower defense, etc) were considered for the prototypes rather than categories of players or experiences. Given the time required for the therapy, the games that separated the respiratory exercise from the gameplay were abandoned. Even basic gameplay and simple stories were positively perceived by children. For each game project, the teams focused on developing one level of the game with a polished visual, sometimes sounds and music, and an end screen. This, however, does not allow us to test the evolution of the gameplay and challenges throughout the games.

For future events, it appears more adequate to conceive many levels with placeholders (eg, colored shapes instead of designed elements) and to develop in parallel a database of visual and sound elements that can be used in the different levels or games.

Storytelling

Literature Findings

In games for health, narratives can help players develop skills, experiment with new situations and behaviors, and change their beliefs in regard to health. For children, fantasy and nonsensical elements can increase their involvement and fun and encourage them to develop their creativity [37]. Death and other sensitive subjects can be included when learning is provided [38]. Addressing situations that affect players and creating meaningful decisions allow them to reflect on their ideals and values and possibly learn about societal issues [38]. In addition, the game creators have to decide the extent to which the story can differ from the mainstream knowledge about health.

Sciart [34] recommends that the game evolves according to the not-too-obvious decisions players have to make, and that players should not be able to replay and change these decisions. Khaled et al [39], however, states that players often expect to replay to improve their experience. The number of choices should increase during the game as players strengthen their skills, and this also increases the game effectiveness [17].

Learning Outcomes From the Field

Although a story, and sometimes a sketchboard, was written for different games, it was only appearing through the visuals and not in the initial gameplay. While this can easily be added afterwards, for example, by cropping existing characters and creating a dialogue between them, it did not allow us to really test and improve the storytelling. For Heritage, the story was built for different levels, each building on a different culture. For other games, the story generally gave some context and the quest. A reflection on what meaningful decisions players can take would certainly improve the quality of these narratives. Furthermore, games could include time references (day or night, seasons, etc) to diversify the experience. An unsolved question is the impact of including elements related to CF (lung, mucus, etc) in the game.

Character and Customization

Literature Findings

For Khaled et al [39], developing the relation between players and their character impacts the immersion and appropriation of knowledge beyond the game. Being able to customize the character can help, even if it does not add much to the gameplay. The interaction sought by the designer should define a player’s perspective; should they have a global vision of the world, should they be embodied, or both? Using nonhuman androgynous characters or fully customizable characters can help avoid stereotypes often seen in mainstream games. Guidelines to avoid marginalization, for example, providing an option to zoom, are available [40].

Learning Outcomes From the Field

The characters developed were nonstereotyped and included animals, aliens, angels, a starship, and some people. For Heritage, one idea was that players could collect and wear clothes from different cultures. No customization was implemented because of a lack of resources. Inviting visual
artists to create a database of elements to customize the characters could be valuable.

**Feedback and Rewards**

**Literature Findings**

Players need not focus on how to play but to immerse themselves in the experience. Learning how to play should be easy and should not require a tutorial or explanation. Sufficient feedback should be provided but not so much that the players get distracted [40]. Health information should be of real use for patients, and additional information should be available on demand. Visual representations and colors should be valued [41]. Rewarding players with points, material, and other awards is useful as long as it creates meaning for players. Intrinsic motivation should be sought and encourage the feelings of autonomy, competence, and belonging. Feedback may be adapted to the social context of players too [42]. For Flanagan and Nissenbaum [43], the reward system explicitly highlights the values behind the game; it can value creativity, competition, exploration, social interactions, etc. For them, players should choose whether they want to compete.

**Learning Outcomes From the Field**

The visual representation of the breathing exercise was central in different games. Players received items when doing the exercise correctly, and a gauge showing the on-time breath was also visible in some games. Although the representation helps players know what is expected, it also limits the interest in the long term. *Bloïd* hid the exercise behind its gameplay—the breath in brought meteorites, breathing out allowed the player to shoot at them, and the number of meteorites grew over time to encourage a slow breath; this allowed us to create more immersion and shift the focus away from the medical element. *Les aventures du Briand* intended to have many mini-games whose results would impact the global quest.

For all games, the rewards were closely linked to the breathing exercise and were only positive: the team did not want to penalize a wrong exercise but encourage a good one. There was a discussion about creating 2 separate reward systems, one for the health-related part and another for the quest; however, this was not implemented. Better integrating these questions while defining the story and gameplay could help ensure more variety and meaning in the rewards.

**Evaluation and Diffusion**

**Literature Findings**

Games for health should be validated with research and also be easily accessible and foster reliable information, technical efficiency, data privacy, and ease of use. The regulation regarding medical devices is changing. The French Health Authority published best practices covering these issues [44] and the Food and Drug Administration is reviewing its process [45]. Health Canada recommends checking if regulation exists, especially when the software influences decisions on health [46]. Beyond best practices, a “Mobile Application Rating Scale” was developed by the Queensland University [47]. Graafland et al [48] created a specific tool to assess the didactic and interactive experience of games for health.

**Learning Outcomes From the Field**

Access to health innovation is essential. From the beginning of the initiative, the source code of the games was released online under a license that enables every interested person to use, reproduce, and improve them. In addition, efforts were made to document the cocreation process. As mentioned earlier, the games presented were not tested extensively, but a study is planned to evaluate *Heritage*, *Bloïd*, as well as other games to be developed around CF.

**Conclusions**

The participatory design of games for health is an excellent tool to bring together people and introduce them to multidisciplinary work. Recurring short events around the same topic enable participants to prototype and test a large variety of gameplay modes, get acquainted with and raise awareness on health issues, and find contributors ready to engage in the long term. Such events are, however, not adapted to build high-quality games.

From the 6 games prototyped, 3 were abandoned because of unsatisfactory choices in their design (separation between the breathing exercise and the game in *Globule*), technical issues (major changes in the game engine for *Ange-Gardien*), and a scope that was too ambitious (the team working on *Les aventures du Briand* spread before a functional version of the game with mini-games was ready). *PEP Hero* was rewritten as *Heritage* to allow the customization of the breathing exercises. *Bloïd* was improved with additional visuals. Both *Heritage* and *Bloïd* are going to be tested and further developed.

To create further games for health, the key learning outcomes (Figure 18) from the 7 themes addressed are to develop prototypes that show the gameplay for different game levels and to separately create a database of visual and sound assets that can be used throughout the game (customizable characters, universe, rewards, etc.). Moreover, the story should encourage meaningful decision making, embed health information while not building the game around it, and foster autonomy and inclusion. As briefly mentioned, special attention should also be given to the evaluation and accessibility of the game, its source code, and its cocreation process so that a maximum number of persons can use and modify it.
Figure 18. Recommendations to design games for health. Credit: breathinggames.net.

Acknowledgments
The author thanks Dr Rilla Khaled, professor in design and computational arts, and Dr Warren Linds, professor in applied human sciences from Concordia University (Canada) as well as Marc-André Maheu-Cadotte, nurse and researcher at the University of Montreal (Canada) for their review and valuable inputs on the paper. The author thanks Scot Deeming, student in the Individualized PhD in humanities at Concordia University, for the proofreading. The author acknowledges all Breathing Games commoners who contributed to the conception, development, and test of the games and devices, namely:


The cocreation was possible thanks to the support of Sainte-Justine University Hospital (Canada), Sensorica Open Hardware Lab (Canada), the Genevan Foundation against Cystic Fibrosis (FGLM, Switzerland), the Swiss Game Center (Switzerland), Lift Conference (Switzerland), Mikorizal Software (US), and the University Hospitals of Canton Vaud (CHUV, Switzerland) and Bern (Inselspital, Switzerland).

Further information about the initiative including the source code of the games presented is available at [22].

The research received no specific funding. Publication fees were funded by Concordia University Open Access Author Fund. The proofreading was paid by the Canadian Institutes of Health Research.

Conflicts of Interest
The author cofounded and leads Breathing Games.

References


Abbreviations

CF: cystic fibrosis
PEP: positive expiratory pressure

©Fabio Balli. Originally published in JMIR Serious Games (http://games.jmir.org), 21.11.2018. 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.
Emotion-in-Motion, a Novel Approach for the Modification of Attentional Bias: An Experimental Proof-of-Concept Study

Lies Notebaert¹, PhD; Ben Grafton¹, PhD; Patrick JF Clarke¹,², PhD; Daniel Rudaizky¹, PhD; Nigel TM Chen², PhD; Colin MacLeod¹, PhD

¹Centre for the Advancement of Research on Emotion, School of Psychological Science, University of Western Australia, Crawley, Australia
²School of Psychology, Curtin University, Bentley, Australia

Corresponding Author:
Lies Notebaert, PhD
Centre for the Advancement of Research on Emotion
School of Psychological Science
University of Western Australia
35 Stirling Highway
M304
Crawley, 6009
Australia
Phone: 61 086488 ext 8080
Email: lies.notebaert@uwa.edu.au

Abstract

Background: Individuals with heightened anxiety vulnerability tend to preferentially attend to emotionally negative information, with evidence suggesting that this attentional bias makes a causal contribution to anxiety vulnerability. Recent years have seen an increase in the use of attentional bias modification (ABM) procedures to modify patterns of attentional bias; however, often this change in bias is not successfully achieved.

Objective: This study presents a novel ABM procedure, Emotion-in-Motion, requiring individuals to engage in patterns of attentional scanning and tracking within a gamified, complex, and dynamic environment. We aimed to examine the capacity of this novel procedure, as compared with the traditional probe-based ABM procedure, to produce a change in attentional bias and result in a change in anxiety vulnerability.

Methods: We administered either an attend-positive or attend-negative version of our novel ABM task or the conventional probe-based ABM task to undergraduate students (N=110). Subsequently, participants underwent an anagram stressor task, with state anxiety assessed before and following this stressor.

Results: Although the conventional ABM task failed to induce differential patterns of attentional bias or affect anxiety vulnerability, the Emotion-in-Motion training did induce a greater attentional bias to negative faces in the attend-negative training condition (P=.003, Cohen d=0.87) and led to a greater increase in stressor-induced state anxiety faces in the attend-negative training condition than in the attend-positive training condition (P=.03, Cohen d=0.60).

Conclusions: Our novel, gamified Emotion-in-Motion ABM task appears more effective in modifying patterns of attentional bias and anxiety vulnerability. Candidate mechanisms contributing to these findings are discussed, including the increased stimulus complexity, dynamic nature of the stimulus presentation, and enriched performance feedback.

(JMIR Serious Games 2018;6(4):e10993) doi:10.2196/10993

KEYWORDS
attentional bias; anxiety disorders; experimental games
Introduction

Attentional Bias Modification in Anxiety

We operate in a complex and dynamic world in which we are continuously confronted with an ever-changing stream of perceptual information. The limited capacity of our cognitive system means we can only attend to certain information, while other information is filtered out. Such filtering does not operate in the same manner across all individuals; however, it has become clear that there is a relationship between such attentional selectivity and individual differences in emotional vulnerability [1]. Specifically, research has shown that elevated anxiety vulnerability, whether indicated by elevated levels of trait anxiety or the presence of anxiety pathology, is associated with an attentional bias that favors the processing of negative information [2]. Moreover, studies that have manipulated patterns of attentional bias (using attentional bias modification [ABM] procedures) have shown that attentional bias causally contributes to anxiety vulnerability, as a change in attentional bias produces a consequent change in anxiety vulnerability [3,4-6].

The observation that ABM tasks delivered in the laboratory can exert a beneficial impact on anxiety responses to stressor has led a number of researchers to investigate whether extended exposure to such ABM training can reduce anxiety dysfunction in real-world settings. The most frequently used ABM procedure is based on the dot-probe task [7]. In this task, participants are briefly exposed to stimulus pairs, comprising 1 negative and 1 non-negative stimulus, before a small visual probe is presented, which participants are required to identify. A contingency between the position of the probes and the position of the negative stimuli is introduced, whereby probes are either always presented in the location where the non-negative stimulus was just displayed (encouraging the adoption of an attentional bias away from negative information) or else probes are always presented in the location where the negative stimulus was just displayed (encouraging the adoption of an attentional bias toward negative information). There is now a substantial body of evidence showing that such ABM tasks, configured to reduce attentional bias to negative information, can attenuate the symptoms of social anxiety disorder [8,9], generalized anxiety disorder [10,11], and subclinical obsessive-compulsive symptoms [12].

Although such encouraging findings highlight the potential clinical benefits of ABM procedures, it is important to recognize that in a number of ABM studies, the intended attentional training procedure has failed to affect emotional vulnerability [13-16]. Overall, meta-analyses show that the clinical effectiveness of the implementation of ABM procedures is small but nonetheless significant [17-21]. However, careful consideration of this literature suggests a clear pattern. In the studies where the intended ABM procedure successfully changed attentional bias, this produced a medium-sized and significant effect on emotional vulnerability. In contrast, in studies where the intended ABM procedure did not change attentional bias, no significant impact on emotional vulnerability was observed [22-26]. These results indicate that future research efforts should focus on developing more effective procedures than the dot-probe task to modify attentional bias [27-29]. Moreover, researchers have raised concerns about the suitability of the conventional probe-based training task for use with clinical cohorts because of its monotonous nature and low face validity [27].

Gamification of Attentional Bias Modification Paradigms

In recent years, some investigators sought to adapt existing ABM procedures to make them more engaging by gamifying them. Gamification refers to the use of game design elements in nongame contexts [30]. Several levels can be discerned in the gamification of ABM tasks, from simply adding game elements to existing training tasks, to adding extrinsic rewards, adding intrinsic motivators, adding a game shell, to using an existing off-the-shelf game [31]. It is thought that maintaining a close connection with validated ABM tasks and adding intrinsic motivation and a game shell may be optimal for ABM gamification [31].

Existing gamified training protocols have been variants of the original probe approach; however, all these protocols share in common the limitation that they seek to train attentional selectivity using very simple static displays that typically present only 1 or 2 stationary emotional stimuli [28,32-36]. This contrasts markedly with real-world settings, which generally require individuals to engage in patterns of attentional scanning and tracking within a complex and dynamic environment [37]. In addition, these studies have been hampered by design limitations. In some cases, these studies have lacked a control condition [33,34]. Moreover, most of these novel procedures have been delivered in studies that afford no opportunity to compare their efficacy with that of the conventional probe-based ABM approach. Without such a comparison, it remains unknown whether a novel ABM task can achieve attentional bias change under conditions where the conventional probe-based ABM task fails to do so or whether a novel task can produce change in anxiety vulnerability when such change is not elicited by a conventional probe-based ABM approach.

These studies did, however, incorporate different gamification and other (nongame) elements to enhance engagement with the tasks to improve their effectiveness in modifying attentional bias [36,38]. Some studies have included motivating feedback or goal metrics, in the form of real-time visual performance feedback or points [32,35] or block-by-block feedback on performance [28]. Others have implemented more elaborate displays or a game-shell to increase intrinsic motivation [33,35,39-43]. Another element thought to increase intrinsic motivation is goals-directed learning, which directs players to particular goals to increase targeted skills (eg, through instructing participants to attend to positive information) [44]. However, despite inclusion of these gamification elements, the majority of these studies continue to rely on either relatively sparse or mostly static stimulus displays.

This Study

The objective of this study was to develop and evaluate a novel candidate ABM procedure designed to modify attentional...
selectivity within a task setting that, like most real-world contexts, requires participants to selectively distribute attention while processing a complex and dynamic emotional environment. The task required participants to search for and track 1 particular target stimulus presented on screen among multiple moving distractors, based on its emotional valence. This screen display thus realizes the *elaborate display* gamification element, as it presents a large number of stimuli that dynamically move across the whole screen. The task also incorporated the gamification elements of motivating feedback through a game-by-game high score that participants were encouraged to try to beat as well as the element of goals-directed learning, as participants were explicitly instructed to track 1 particular emotional expression. The task was built upon the same principles as the original probe-based ABM task, in which the training condition was designed to increase attentional bias toward positive information, and the task performance would be improved to the extent that participants adopted a more positive attentional bias [6]. Similarly, in the corresponding training condition of the gamified task, performance will improve to the extent participants allocate attention to positive stimuli.

Our primary aim was to evaluate the capacity of this new candidate ABM procedure, which we have labeled the Emotion-in-Motion task, to induce a group difference in selective attentional responding to negatively and positively valenced information and to causally impact anxiety vulnerability, as evidenced by the strength of state anxiety responses to a controlled laboratory stressor. We also delivered the conventional probe-based ABM procedure to a separate cohort of similar participants under equivalent laboratory conditions. This conventional probe-based ABM task does not include any of the gamification elements introduced in the Emotion-in-Motion task. Specifically, there is no elaborate display (only 2 static images are presented), no motivating feedback after each block (only trial-by-trial feedback), and no goals-directed learning (participants are simply instructed to discriminate the identity of a probe). We chose to compare our novel Emotion-in-Motion task with the probe-based ABM task, as this is the procedure most commonly used in studies aiming to modify patterns of attentional bias [17,18,21].

This study design will enable us to determine (1) whether both the conventional probe-based ABM task and this new, complex, dynamic Emotion-in-Motion ABM task produce a group difference in attentional bias in line with the allocated attentional training condition, (2) if both tasks prove capable of so modifying attentional bias, whether the Emotion-in-Motion ABM task impacts attentional bias to an equal or greater degree than does the conventional probe-based ABM task, (3) whether both the conventional probe-based ABM task and this new, complex, dynamic Emotion-in-Motion ABM task serve to induce a group difference in anxiety vulnerability as a function of allocated training condition, and (4) if both tasks prove capable of so influencing anxiety vulnerability, whether the Emotion-in-Motion ABM task impacts anxiety vulnerability to an equal or greater degree than does the conventional probe-based ABM task.

**Methods**

**Participants**

A total of 129 undergraduate students at the University of Western Australia completed the study. In line with previous research, participants who did not show an elevation in state anxiety in response to the intended stressor were excluded before analyses [28]. This led to the exclusion of 19 participants, with 110 participants remaining. Participant characteristics are shown in Table 1.

**Conventional Probe-Based Attentional Tasks**

Overall, 55 participants completed the conventional probe-based bias training and assessment tasks. These participants were randomly assigned to either an attend-positive or attend-negative training condition. Participants assigned to these 2 conditions of the probe-based tasks did not differ significantly in age, trait anxiety scores, or gender (all *P*>.05).

**Emotion-in-Motion Attentional Tasks**

Overall, 55 participants completed our novel Emotion-in-Motion bias training and assessment tasks. Participants were randomly assigned to either an attend-positive or attend-negative training condition. Participants in these 2 conditions of the Emotion-in-Motion tasks did not differ significantly in age or trait anxiety scores (both *P*>.05). These 2 groups did differ significantly in gender ratio, *P*=.03, with a higher proportion of males in the attend-negative condition than in the attend-positive condition. Consequently, we considered gender ratio as a covariate in our analyses of the data, which provided reassurance that observed effects of this experimental manipulation remained evident when this group difference in gender ratio was accounted for.

<p>| Table 1. Age, gender, and trait anxiety scores (using the Spielberger Trait Anxiety Inventory) for participants completing the conventional probe-based and the Emotion-in-Motion attentional tasks in each of the 2 training conditions. |</p>
<table>
<thead>
<tr>
<th>Condition</th>
<th>Age, mean (SD)</th>
<th>Gender, female/male</th>
<th>Trait anxiety, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conventional probe tasks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attend-negative condition (N=27)</td>
<td>19.33 (2.86)</td>
<td>14/13</td>
<td>38.44 (8.14)</td>
</tr>
<tr>
<td>Attend-positive condition (N=28)</td>
<td>19.50 (2.60)</td>
<td>19/9</td>
<td>41.18 (11.08)</td>
</tr>
<tr>
<td><strong>Emotion-in-Motion tasks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attend-negative condition (N=28)</td>
<td>19.78 (3.62)</td>
<td>17/10</td>
<td>47.18 (8.18)</td>
</tr>
<tr>
<td>Attend-positive condition (N=27)</td>
<td>18.50 (0.95)</td>
<td>23/3</td>
<td>43.22 (9.24)</td>
</tr>
</tbody>
</table>

Materials

Attentional Tasks Stimuli

The face stimuli for the attentional tasks were selected from the Karolinska Directed Emotional Faces stimulus set [45]. These images were cropped to show only the face and the neck. The face stimuli for the training tasks were photos of 32 individuals, half of them were female and half were male. For the assessment tasks, photos of 8 different individuals were selected, half of them were male and half were female. There were 2 photographs of each individual, 1 in which they depicted a happy expression and 1 in which they depicted an angry expression. Each photograph was 258 pixels (width) by 323 pixels (height). The stimuli were the same size in the Emotion-in-Motion and probe-based tasks. For the Emotion-in-Motion training task, the 32 identities were grouped into 8 stimulus subsets, each containing the photos of 8 identities, 4 female and 4 male. Each stimulus subset was used in 1 of the 8 blocks delivered in this Emotion-in-Motion task.

Emotion-in-Motion Attentional Bias Modification Task

The aim of this task was to induce, in a complex and dynamic task environment, selective attending to angry or happy faces, depending on the assigned training condition. To provide readers with a first-hand impression of this Emotion-in-Motion task, the task can be viewed on the Web [46].

The Emotion-in-Motion ABM task consisted of 8 3.5 min blocks or games. During each block, 8 placeholder rectangles moved dynamically around the screen over a black background. Each rectangle contained an image of a face, each with a different identity. At all times, the target rectangle displayed a face with an emotional expression that differed from the emotional expressions displayed by the faces in all 7 other rectangles on screen, and participants were required to attend to and track this rectangle. In the attend-negative condition, the target rectangle displayed a face with an angry expression, whereas the other rectangles displayed faces with happy expressions. In the attend-positive condition, the target rectangle displayed a face with a happy expression, whereas the other 7 rectangles displayed faces with angry expressions. Participants were instructed to find the target rectangle and track it using the mouse cursor. All the rectangles, including the target, constantly switched faces. Participants were instructed to keep tracking the target rectangle (i.e., depicting the single face with the expression differing from that of the other 7 faces) even when the face presented within changed, as long as the emotional expression of the face presented remained the same (i.e., when the face in the target rectangle switched to a different identity, participants were required to keep tracking the rectangle as long as the emotional expression of the new face was the same as the emotional expression of the previous face). At random intervals, the emotional expression of a target face would change in addition to its identity, at which point this ceased to be the target rectangle. At that same moment, 1 of the other rectangles would assume a face depicting this emotion, and thus identifying it as the (new) target rectangle. At these points, participants had to quickly find the new target rectangle and start tracking it.

At the start of a block, each face remained constant for the first 2000 milliseconds. Thereafter, individual faces within a rectangle switched to a different identity (but same expression) randomly at any point between 1 and 2000 milliseconds throughout the block. Within each block, the target rectangle switched (thus, an expression switch occurred) 60 times, at random intervals of 5 to 10 seconds. All 8 rectangles moved with different randomly determined trajectories, at a randomly determined speed of between 30 and 50 pixels per 100 milliseconds. Thus, although the rectangles moved at different speeds, each rectangle’s speed was constant within a game. The rectangles bounced off the screen edges and other stimuli they contacted at an angle of reflection that matched their angle of incidence. The target rectangle was never indicated; however, when the mouse cursor was correctly located in the position of the current target rectangle, this cursor disappeared behind the rectangle (to not obscure the face presented within) and remained hidden as long as the participant kept it on target.

The onset of each block was preceded by a 3-second countdown presented in the center of the screen. At the end of each block, participants were presented with a tracking score (i.e., the percentage of time during that game they were tracking the target rectangle), a switching score (i.e., the average speed with which the participant was able to shift their cursor to the next target rectangle), and a total score for that block (generated by combining the tracking score and the switching score). The screen also displayed the participant’s highest prior (total) score. Participants were instructed that they would play several games of this task and were encouraged to beat their current high score in each successive game.

Emotion-in-Motion Attentional Bias Assessment Task

The training contingency was removed from the Emotion-in-Motion training task to create the assessment task used to reveal the impact of this training on attentional selectivity. Thus, participants were required to track a rectangle displaying a face with a happy expression (among 7 rectangles displaying faces with angry expressions) on half of the blocks and to track a rectangle displaying a face with an angry expression (among 7 rectangles displaying faces with happy expressions) on the other half of the blocks. This assessment task delivered 12 short blocks, each of which contained 5 target switches, resulting in a total of 60 target switches across the assessment task. In 6 of these blocks, the target rectangle displayed a face with an angry emotional expression, and in 6 blocks, the target rectangle displayed a face with a happy emotional expression. The order of these block conditions was randomly determined, with the constraint that a maximum of 2 consecutive blocks could have a target with the same valence. Each block started with a 5-second countdown.

To obtain a measure of attentional bias to negative information, an attentional bias index (ABI) was computed by subtracting the average tracking score a participant obtained in blocks where targets were happy faces from the average tracking score the participant obtained in blocks where targets were angry faces. Therefore, a higher positive score on this index reflects greater attention to negative information, as it represents more successful tracking of angry than of happy faces.
Other Experimental Tasks
The Trait Anxiety Assessment, conventional probe-based ABM, and assessment tasks as well as the anxiety reactivity assessment task are described in Multimedia Appendix 1.

Procedure
Participants were tested individually in a sound-attenuated room. Once informed consent from participants had been obtained, participants were instructed to sit at a comfortable viewing distance from the computer screen (approximately 60 cm), were given instructions, and completed the first assessment task. After completion of the training task, they completed the original assessment task again. Next, participants completed the anxiety reactivity assessment task containing an anagram stressor task preceded and followed by a measure of state anxiety. At the end of the session, participants were debriefed about the purpose of the study. The entire experimental session lasted about 1 hour. This study was approved by the University of Western Australia’s Human Research Ethics Committee, protocol RA415243.

Results
Impact of Attentional Training Procedure on Attentional Bias
The criteria to identify outliers are described in Multimedia Appendix 1. The ABI scores obtained before and after the training task are shown in Table 2.

Attentional Impact of Conventional Probe-Based Training
Application of the outlier criteria led to the exclusion of 4 participants (2 in the attend-positive training condition). To examine whether the conventional probe-based training task was capable of modifying attentional bias, a mixed-methods analysis of variance (ANOVA) was performed with the within-subjects factor attentional assessment point (pretraining assessment, postraining assessment) and the between-subjects factor training condition (attend-positive training, attend-negative training). The ABI scores obtained by participants who completed this conventional probe-based training task served as the dependent variable.

Results showed neither a significant main effect of attentional assessment point, $F_{1,48}=1.19, P=.286$, nor of training condition, $F_{1,48}=2.445, P=.119$. Most importantly, the critical interaction between attentional assessment point and training condition fell short of significance, $F_{1,48}=3.018, P=.09, \eta_{p}^2=.059$.

Attentional Impact of Emotion-in-Motion Training
Application of the outlier criteria led to the exclusion of 3 participants (1 in the attend-positive training condition). To determine whether our novel Emotion-in-Motion attentional training procedure was effective in modifying attentional responding to negative information, the ABI scores obtained by participants who completed this task were subjected to a mixed-design 2x2 ANOVA that again considered the within-group factor attentional assessment point (pretraining assessment vs postraining assessment) and the between-group factor training condition (attend-positive training vs attend-negative training). This analysis revealed a significant main effect of training condition, $F_{1,50}=4.602, P=.04, \eta_{p}^2=.084$, subsumed within a higher-order interaction of attentional assessment point x training condition, $F_{1,50}=5.629, P=.02, \eta_{p}^2=.101$. At pretraining, there was no significant difference between the ABI scores obtained by participants in the attend-positive training condition and participants in the attend-negative training condition, $F_{1,50}=1.50, P=.22, \eta_{p}^2=.057$. Cohen $d=0.87$. Although the change in attentional bias from pre- to postraining fell short of significant for participants in the attend-negative training condition, $t_{25}=-1.162, P=.26, \eta_{p}^2=0.027$, Cohen $d=0.229$, there was a significant change from pre- to postraining for participants in the attend-positive training condition, $t_{25}=2.114, P=.045, \eta_{p}^2=0.145$. Overall, this pattern of results confirms that the 2 training conditions exerted a differential impact on attentional bias to negative information, and the direction of the observed attentional training effects was as expected. When controlling for the gender, by adding this as a covariate, this interaction between attentional assessment point and training condition remained significant, $F_{1,41}=3.933, P=.04, \eta_{p}^2=.087$.

Table 2. Attentional bias index scores pre- and postraining for participants who completed the conventional probe-based attentional bias training and assessment tasks or the Emotion-in-Motion attentional bias training and assessment tasks in either the attend-positive training condition or the attend-negative training condition.

<table>
<thead>
<tr>
<th>Assessment point</th>
<th>Attend-positive condition, mean (SD)</th>
<th>Attend-negative condition, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional probe training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABI pretraining</td>
<td>$-5.669 (50.752)$</td>
<td>$-3.116 (29.591)$</td>
</tr>
<tr>
<td>ABI postraining</td>
<td>$-15.101 (45.33)$</td>
<td>$11.002 (36.9)$</td>
</tr>
<tr>
<td>Emotion-in-Motion training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABI pretraining</td>
<td>$0.449 (6.041)$</td>
<td>$1.011 (5.968)$</td>
</tr>
<tr>
<td>ABI postraining</td>
<td>$-2.739 (6.545)$</td>
<td>$2.445 (5.261)$</td>
</tr>
</tbody>
</table>

$^a$ABI: attentional bias index.
Table 3. State anxiety scores pre- and postanagram stressor for participants who previously completed the conventional probe-based attentional bias training task or the Emotion-in-Motion attentional bias training task in either the attend-positive training condition or the attend-negative training condition.

<table>
<thead>
<tr>
<th>Assessment point</th>
<th>Attend-positive condition, mean (SD)</th>
<th>Attend-negative condition, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional probe training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State anxiety pretraining</td>
<td>30.680 (10.858)</td>
<td>26.200 (12.176)</td>
</tr>
<tr>
<td>State anxiety posttraining</td>
<td>43.600 (9.734)</td>
<td>40.440 (11.623)</td>
</tr>
<tr>
<td>Emotion-in-Motion training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State anxiety pretraining</td>
<td>31.292 (10.149)</td>
<td>29.541 (11.699)</td>
</tr>
<tr>
<td>State anxiety posttraining</td>
<td>39.458 (11.026)</td>
<td>41.458 (10.879)</td>
</tr>
</tbody>
</table>

Impact of Attentional Training Procedure on Anxiety Vulnerability

The state anxiety scores obtained using the analog mood scale given before and after the final anagram stressor are shown in Table 3.

Emotional Impact of Conventional Probe-Based Training

Application of the outlier criteria on participants included in the attentional bias assessment analyses led to the additional exclusion of 1 participant (in the attend-positive condition). To examine whether the 2 training conditions had a differential impact on anxiety reactivity, state anxiety scores were subjected to a mixed-methods ANOVA with the within-subjects factor state anxiety assessment point (prestressor assessment vs poststressor assessment) and the between-subjects factor training condition (attend-positive training vs attend-negative training). Results showed a significant main effect of state anxiety assessment point, $F_{1,46}^{2} = 159.991, P<.001$, indicating that state anxiety increased from before the anagram stressor (mean 28.440, SD 11.639) to after the anagram stressor (mean 42.020, SD 10.729). However, neither the main effect of training condition, $F_{1,46}^{2} = 1.664, P=.20$, nor the critical interaction between state anxiety assessment point and training condition, $F_{1,46}^{2} = 378, P=.54, \eta_{p}^{2} = .008$, were significant.

Emotional Impact of Emotion-in-Motion Training

Application of the outlier criteria on participants included in the attentional bias assessment analyses led to the additional exclusion of 4 participants (2 in the attend-positive training condition). The same 2x2 ANOVA as reported above was conducted on state anxiety scores to examine whether in participants who completed the Emotion-in-Motion training procedure, the 2 training conditions had a differential impact on anxiety reactivity. This analysis revealed a significant main effect of state anxiety assessment point, $F_{1,46}^{2} = 125.99, P<.001, \eta_{p}^{2} = .73$, again reflecting the fact that state anxiety increased from before the stressor (mean 30.58, SD 10.87) to after the stressor (mean 40.54, SD 10.88). This main effect was now subserved within a significant two-way interaction of state anxiety assessment point and training condition, $F_{1,46}^{2} = 4.39, P=.04, \eta_{p}^{2} = .09$. When controlling gender, by adding gender as a covariate, this interaction between state anxiety assessment point and training condition remained significant, $F_{1,43}^{2} = 4.638, P=.04, \eta_{p}^{2} = .097$.

Follow-up $t$ tests revealed that immediately following the attentional training procedure but before the anagram stressor experience, participants who had received the 2 training conditions did not differ in their levels of state anxiety, $F_{1,46}^{2} = .31, P=.58, \eta_{p}^{2} = .01$. Participants in each Emotion-in-Motion attentional training condition responded to this stress manipulation by displaying an elevation in anxious mood state (attend-positive training: $F_{1,23}^{2} = 55.84, P<.001, \eta_{p}^{2} = .71$ vs attend-negative training: $F_{1,23}^{2} = 70.56, P<.001, \eta_{p}^{2} = .76$). However, the magnitude of the elevation in state anxiety evoked by this stressor was significantly attenuated in those participants who had received the Emotion-in-Motion attend-positive attentional training compared with those participants who had received the Emotion-in-Motion attend-negative attentional training condition (mean 8.17, SD 5.35 vs mean 11.92, SD 6.94; Cohen $d=0.60$). Thus, those participants who had been exposed to the Emotion-in-Motion task training contingency designed to reduce attentional bias to negative information subsequently came to display relatively attenuated elevations of anxious mood state in response to the anagram stressor experience compared with participants who had been exposed to the training condition designed to increase attentional bias to negative information. In addition, the elevation in anxiety in the positive training condition of the Emotion-in-Motion task (mean 8.17, SD 5.35) was significantly smaller than the elevation in state anxiety in the positive training condition of the conventional probe-based training task (mean 12.92, SD 8.83), $t_{47} = 2.35, P=.02$, Cohen $d=0.67$.

Discussion

Principal Findings

The objective of this study was to develop and evaluate a novel ABM procedure intended to systematically alter selective attentional responding to emotional information in a complex and dynamic task environment. Our results showed that our novel Emotion-in-Motion training procedure succeeded in modifying patterns of attentional bias, as intended. Moreover, the participants who were allocated to the attend-positive condition of the Emotion-in-Motion attentional training task showed reduced anxiety reactivity to the subsequent lab-based
stressedor as compared with participants who were allocated to the attend-negative condition of this task. These results suggest that our novel attentional training task appeared capable of modifying both patterns of attentional bias and causally influencing anxiety vulnerability.

A subsidiary aim was to permit comparison with the conventional probe-based attentional bias training procedure. Under equivalent laboratory conditions, the conventional probe-based attentional training approach failed to induce differential patterns of attentional bias, and the 2 probe-based training conditions did not lead to participant differences in anxiety reactivity to the subsequent stressor. In recent years, several studies (including 3 out of our lab) have reported similar failures of the conventional probe-based attentional training task to successfully modify patterns of attentional bias [28,47-51]; therefore, it is reasonable to conclude that the probe-based ABM procedure may be a nonoptimal way of achieving bias change.

Candidate Explanations for the Effectiveness of the Emotion-in-Motion Task

In reflecting on the reasons for the capacity of our novel Emotion-in-Motion paradigm to induce differential patterns of attentional bias, under conditions where the conventional probe-based training did not, several candidate factors can be considered. First, the Emotion-in-Motion task presents 8 stimuli simultaneously, whereas the conventional probe task displays only 2 stimuli. There is some evidence that attentional bias is more pronounced when assessed using visual displays that contain more stimuli [52,53], but as yet, it is unknown whether more robust ABM effects can be obtained using paradigms that present more stimuli. Although some training procedures that involve more complex stimulus displays already exist [41,54], so far no direct comparison between the effectiveness of training tasks using simple versus complex stimulus displays has been made. In future research, the Emotion-in-Motion paradigm can be easily adapted to present simple displays (eg, 2 rectangles) versus complex displays (eg, 8 rectangles), to enable such comparison.

A second candidate factor that could have contributed to the findings observed with the Emotion-in-Motion approach is the dynamic nature of the stimulus presentation. In the Emotion-in-Motion task, all stimuli move dynamically around the display, whereas in other attentional training paradigms, stimuli are presented in a static manner. It is possible that the dynamic nature of Emotion-in-Motion enhanced concentration and engagement with the task, thereby increasing its capacity to deliver the intended attentional bias change. In future research, the potential contribution of this dynamic component could be examined by contrasting task variants that employ the present dynamic approach with variants that instead present the same number of stimuli in static grid.

A third candidate reason for its efficacy may be the provision of enriched performance feedback in the Emotion-in-Motion task compared with the rudimentary trial-by-trial error feedback given in the conventional probe-based attentional training task. Moreover, block feedback in the Emotion-in-Motion task was encouraging, whereas trial-by-trial feedback in the probe-based task penalized participants for making errors by delaying the next trial for 3 seconds, which may have elicited increased negative mood. Block feedback of the type delivered in the Emotion-in-Motion task has been shown to enhance learning in simple repetitive tasks [55], whereas negative mood has been shown to impair learning [56]. As such, this difference in feedback may have also contributed to enhanced performance in the Emotion-in-Motion task. Future research could further examine the contribution of enriched performance feedback to the efficacy of ABM procedures by comparing conventional probe-based training with and without such block feedback or by manipulating whether or not the presently provided block feedback is delivered within the Emotion-in-Motion task.

Moreover, in the conventional probe-based training task, images depicting different emotional expression of the same identity were paired, whereas in the Emotion-in-Motion task, each image depicted a different identity. As such, participants performing the probe-based training only needed to discriminate emotional expression on the same person, whereas participants performing the Emotion-in-Motion tasks needed to discriminate emotional expressions between different identities. There is some evidence to suggest that emotion classifications are affected by variations in identity [57]. It is, therefore, possible that this increased demand on emotion classification contributed to the Emotion-in-Motion task being more challenging. The more challenging emotion classification, enhanced performance feedback, as well as the complex and dynamic nature of the task could have resulted in greater engagement with the Emotion-in-Motion task, relative to the conventional probe task. Task engagement can be conceptualized as a combination of energy, motivation, and concentration and can be measured using self-report as well as through task performance indicators [58]. In the Emotion-in-Motion task, we did not obtain self-report measures of task engagement, and the difference in the nature of the tasks leaves us unable to compare performance indicators of engagement. However, future research may usefully examine whether individuals show a difference in engagement with the Emotion-in-Motion task relative to the probe task and whether task engagement moderates the procedures’ impact on attentional bias and anxiety vulnerability [36].

An additional difference between the 2 training procedures concerns participants’ responses. The tracking response required in the Emotion-in-Motion task is continuous, whereas the probe task only requires a response every couple of seconds. It is likely that as a result, participants in the Emotion-in-Motion task spend more time attentionally engaged with the target valence (positive or negative, depending on training condition) as compared with participants in the conventional probe task. However, it is also possible that this continuous response would be harder to sustain over time as it is more motorically demanding. As such, future research may usefully examine the acceptability of this response format in multi-session training designs. It is also relevant to note that in the Emotion-in-Motion task, the mouse cursor only disappears behind the target rectangle. As such, it is possible that participants could ignore the content of the rectangle and simply see on which rectangle the mouse cursor would disappear. Although given the speed and complexity of the task...
this strategy is unlikely to have occurred, importantly, this strategy would have reduced the efficacy of the attentional training. Future research could, therefore, evaluate whether modification to the task (e.g., making the cursor disappear behind every image) would further increase the effectiveness of the task. It will be important to establish, however, whether such modifications that render participants’ awareness of the position of the cursor more uncertain cause unwanted frustration or disorientation in participants.

**Strengths and Limitations**

It is important to consider the potential limitations of this study. One such limitation is that the capacity of the Emotion-in-Motion training task and the capacity of the conventional probe-based training task to modify attentional bias were each established using a different method of assessing attentional bias. For both training tasks, the assessment approach involved delivering the same task but with the training contingency removed. This design critically allows for comparable demonstration of near transfer across the 2 training tasks. However, it does preclude direct comparison of attentional bias change observed in response to each of these 2 candidate attentional training approaches. It is possible, for example, that the assessment version of the Emotion-in-Motion task is more sensitive to individual differences in attentional bias than the probe-based attentional bias assessment task (of potential relevance to this, note that the SDs for Emotion-in-Motion ABI scores are smaller than those for the probe-based ABI scores). If this is the case, then the results of this study could be explained by the Emotion-in-Motion training task producing a greater modification of attentional bias, the Emotion-in-Motion assessment task more sensitively assessing group differences in attentional bias, or both. Future research could circumvent this limitation by employing the same attentional bias assessment approaches for all ABI tasks under evaluation.

A second potential limitation is that this study was carried out on an undergraduate nonclinical participant sample. Although this design allowed us to examine whether the Emotion-in-Motion procedure can induce differential patterns of attentional bias and consequently test the causal impact of these differential patterns of attentional bias on anxiety vulnerability, it does limit conclusion concerning either the acceptability or the efficacy of our novel Emotion-in-Motion ABM approach when used with a clinical sample. Although the complex and dynamic nature of the Emotion-in-Motion task can be expected to enhance face validity of and engagement in the task, future research using clinical cohorts will be necessary to determine whether this novel ABM task is more acceptable to patients than the conventional probe-based training task.

It is also important to consider some potential limitations of using gamification for bias modification. Some of the potential drawbacks are discussed by Boendermaker et al [31]. These authors note that some gamification elements designed to increase motivation (such as visible scores) may be distracting and impair training. Second, implementing intrinsic motivators may be costly and difficult, and the intrinsically motivating value of such elements may vary across individuals. In addition, even if a game is intrinsically motivating, it may need to be combined with a motivation to change in participants before adherence to multi-session training is improved. Most importantly, however, given the strong link between change in attentional bias and change in emotional vulnerability, it is important that in any gamified ABM procedure, the core mechanism underlying ABM (encouraging a change in attentional bias) remains intact [24,59].

**Conclusions**

In the meantime, we hope that the Emotion-in-Motion task, which this study has shown to be capable of modifying attentional bias to emotional information and altering anxiety vulnerability as indicated by anxiety reactivity to a stressor, will be of interest and potential value to researchers investigating the potential anxiolytic benefits of directly manipulating maladaptive patterns of attentional bias. To facilitate further research using this task and to encourage independent replication of the findings of this study, we made our Emotion-in-Motion task software freely available [60]. While we look forward to the future evaluation of this novel ABM approach in other cohorts and settings, we also encourage fellow researchers to develop and refine new and innovative ABM paradigms that further enhance our capacity to modify the attentional bias to negative information implicated in anxiety vulnerability and dysfunction. Such continuous improvement in our ABM approaches will optimize the prospect of developing future ABM protocols that prove capable of delivering robust and reliable therapeutic benefits within the clinic.

**Acknowledgments**

LN is supported by the Australian Research Council under Grant DP140104448. CM is supported in part by a grant from the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PNII-ID-PCCE-2011-2-0045. PJFC was supported by Australian Research Council Grant DP140103713. BG is supported in part by Australian Research Council Grant DP170104533. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Further detail on the methods and data handling.
References


36. Boendermaker J, Sanchez MS, Boffo M, Wiers RW. Attentional bias modification with serious game elements: evaluating the shots game. JMIR Serious Games 2016 Dec 06;4(2):e20 [FREE Full text] [doi: 10.2196/games.6464] [Medline: 27923780]


Notebaert L, MacLeod C, Grafton B. Emotion-in-Motion task material. The University of Western Australia 2018 [FREE Full text] [doi: 10.26182/5bbd5735031f6]

Abbreviations

ABI: attentional bias index
ABM: attentional bias modification
ANOVA: analysis of variance

Edited by TR Soron; submitted 08.05.18; peer-reviewed by W Boendermaker, M Zhang; comments to author 14.08.18; revised version received 26.08.18; accepted 27.08.18; published 28.11.18

Please cite as:
Notebaert L, Grafton B, Clarke P,JF, Rudyakzy D, Chen NTM, MacLeod C
Emotion-in-Motion, a Novel Approach for the Modification of Attentional Bias: An Experimental Proof-of-Concept Study
JMIR Serious Games 2018;6(4):e10993
doi: 10.2196/10993
PMID:30487121
© Lies Notebaert, Ben Grafton, Patrick JF Clarke, Daniel Rudaizky, Nigel TM Chen, Colin MacLeod. Originally published in JMIR Serious Games (http://games.jmir.org), 28.11.2018. 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.
Creating a Theoretically Grounded, Gamified Health App: Lessons From Developing the Cigbreak Smoking Cessation Mobile Phone Game

Elizabeth A Edwards¹,², MBBS (Hons), MRes, MRCP, MRCP (UK), DFSRH; Hope Caton³, MSc; Jim Lumsden⁴,⁵, MEng, MSc; Carol Rivas¹⁶, BSc, MSc, PhD; Liz Steed¹², BSc, MSc, PhD; Yutthana Pirunsarn³, BEng, MSc; Sandra Jumbe¹,², BSc, MSc, PhD; Chris Newby¹², BSc, MSc, PhD; Aditi Shenvi¹, MSc; Samresh Mazumdar¹,², MBBS, MRCP (UK); Jim Q Smith⁸, BSc, PhD; Darrel Greenhill³, BSc, PhD; Chris J Griffiths¹,², BA, MBBS, DPhil, FRCP (UK), FRCGP (UK); Robert T Walton¹,², BSc, MBBS, MD, FRCP (UK), FRCGP (UK)

¹Centre for Primary Care and Public Health, Blizard Institute, Bart’s and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
²Asthma UK Centre for Applied Research, Bart’s and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
³Faculty of Science, Engineering Computing, Kingston University, London, United Kingdom
⁴MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom
⁵School of Psychological Science, University of Bristol, Bristol, United Kingdom
⁶Social Science Research Unit, University College London, London, United Kingdom
⁷Centre for Complexity Science, University of Warwick, Coventry, United Kingdom
⁸Department of Statistics, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Elizabeth A Edwards, MBBS (Hons), MRes, MRCP, MRCP (UK), DFSRH
Centre for Primary Care and Public Health
Blizard Institute, Bart’s and The London School of Medicine and Dentistry
Queen Mary University of London
58 Turner Street
London, E1 2AB
United Kingdom
Phone: 44 020 7882 2498
Fax: 44 020 7882 2552
Email: dr.elizabeth.ann.edwards@gmail.com

Abstract

Background: Gaming techniques are increasingly recognized as effective methods for changing behavior and increasing user engagement with mobile phone apps. The rapid uptake of mobile phone games provides an unprecedented opportunity to reach large numbers of people and to influence a wide range of health-related behaviors. However, digital interventions are still nascent in the field of health care, and optimum gamified methods of achieving health behavior change are still being investigated. There is currently a lack of worked methodologies that app developers and health care professionals can follow to facilitate theoretically informed design of gamified health apps.

Objective: This study aimed to present a series of steps undertaken during the development of Cigbreak, a gamified smoking cessation health app.

Methods: A systematic and iterative approach was adopted by (1) forming an expert multidisciplinary design team, (2) defining the problem and establishing user preferences, (3) incorporating the evidence base, (4) integrating gamification, (5) adding behavior change techniques, (6) forming a logic model, and (7) user testing. A total of 10 focus groups were conducted with 73 smokers.

Results: Users found the app an engaging and motivating way to gain smoking cessation advice and a helpful distraction from smoking; 84% (62/73) of smokers said they would play again and recommend it to a friend.
Conclusions: A dedicated gamified app to promote smoking cessation has the potential to modify smoking behavior and to deliver effective smoking cessation advice. Iterative, collaborative development using evidence-based behavior change techniques and gamification may help to make the game engaging and potentially effective. Gamified health apps developed in this way may have the potential to provide effective and low-cost health interventions in a wide range of clinical settings.

(Keywords: smoking cessation; health behaviors; behavioral medicine; games for health; mHealth; eHealth)

Introduction

The Health App Revolution

Mobile phone use is increasing rapidly in both developed and developing countries, and by 2020, 70% of the world's population will be using mobile phones [1,2]. Three billion people globally currently use mobile health apps [3-5], with over 165,000 health apps available worldwide [2].

Health apps are in high demand. A recent study reported 800,000 downloads per month of smoking cessation apps worldwide [6]; there were 400 smoking cessation apps available on app stores in 2013 when Cigbreak development began [6]. However, most health apps are not developed from a theoretical basis that draws on evidence-based behavior change techniques (BCTs), and there is little evidence that public health practitioners or users have participated in design [6-10]. To date, there have been few rigorous evaluations of the effectiveness of health apps [2,11-14] although pilot studies and small trials have shown promising results [9,11,13,15,16].

Using Gamification to Change Health Behavior

Maintaining users’ engagement with health apps is not easy, with 77% of apps going unused only 72 hours after being installed [10,17,18]. One potential solution to this problem is gamification, which aims to harness the motivational power of gaming elements such as badges, leaderboards, competitions, rewards, and avatars to increase user engagement and hence improve effectiveness [2,19]. Gamification shares key elements with established health BCTs and behavior change theory [20,21], and there is growing evidence that gamification increases engagement with health apps [10,18,22,23]. Despite this, a recent review found that only 4% of top-rated health apps on the Apple and Android stores made use of gamification principles [2].

Health App Development

In recent years, standards have been established for health app development [24], and there are some examples of well-developed health apps that are evidence- and theory-based with expert design [9,13,15,25,26]. The British Standards Institution has formulated a code of practice for health and wellness apps, providing app developers with quality criteria to consider during the development process [24]. However, there is a lack of worked methodologies that app developers and health care professionals can follow to aid development and incorporation of appropriate features.

The Cigbreak App

In 2013, a group of clinicians, researchers, and game developers set out to build a dedicated smoking cessation app Cigbreak, developed in collaboration with potential end users. Gamification and theoretically validated BCTs were included, including those shown to be beneficial in smoking cessation [27,28], with the goal of creating an engaging, scientifically grounded health app.

The current version of Cigbreak involves players swiping their screen to break cigarettes as they dance upward from a generic cigarette pack, providing a distraction from cravings. Smokers progress along a path through a garden to a smoke-free finish line (Figure 1). Along the way, players can complete diary entries and overcome specific daily missions. Players are rewarded for both real-life smoking cessation behavior and progress through the game with health messages, coins, and trophies and are given personalized feedback. Player’s progress can be monitored and shared on Facebook, providing social support. Nicotine replacement therapy (NRT) power-ups inform players of the different kinds of NRT available, encouraging pharmacological support.

This paper documents the development process in a series of steps, providing a worked example for future development of gamified health apps.

(JMIR Serious Games 2018;6(4):e10252) doi:10.2196/10252
**Methods**

**Agile Development Process**
Principles of agile development were adopted [29] where prototypes were developed rapidly and systematically modified according to user feedback. The methods adopted in Steps 1 to 7 are described below. We aimed first to gain an understanding of the research problem (why do people still smoke?) and to identify key user preferences for app functions. Doctors, health psychologists, and researchers worked closely with app developers to incorporate established smoking cessation methods and validated BCTs. Feedback from users was obtained following each iteration to refine app functions or features. At all stages of development, the focus groups involved participants using actual prototypes of the app, having hands-on experience on mobile handsets that we provided in the 10 focus groups.

**Focus Groups**
Focus groups were held throughout the development process (step 2 and step 7). Participants were recruited through pharmacies in Tower Hamlets, London; Eurogamer (Europe’s largest gaming show); a community development charity in East London, Social Action for Health; Kick-it the smoking cessation scheme facilitated through the Tri-borough Council, London. Moreover, 10 focus groups were conducted in total with 73 participants (male [n=34] and female [n=39]). In total, 2 focus groups were conducted at Chapel Hill, University of North Carolina, to understand the differences in smoking/cessation culture between the United Kingdom and the United States, which might affect design requirements when making the app available globally. Furthermore, of the 10 focus groups, 3 (n=26) were conducted during early development as documented in step 2. There was a wide range of ages (15 to 67 years) and ethnicities (Bengali, Turkish, Russian, Polish,
Step 1: Forming an Expert Multidisciplinary Design Team

A multidisciplinary design team with appropriate experience and expertise was formed to aid in the design and construction of the app. The team included a senior app developer (HC), computer scientists (YP and DG), clinical doctors (EAE and CR), a senior health psychologist (LS), a senior medical sociologist (CR), and several smoking cessation advisors from the London Kick-it smoking cessation scheme. The membership of the design team was consistent over the development period. Extensive input from the public and potential app users was sought throughout development.

The team brainstormed extensively before any coding began, to establish key concepts and constructs to be included in the app. The design team met once per week throughout the development period. Meetings were chaired by a facilitator (RTW), and decisions were documented and logged.

Step 2: Defining the Problem and Establishing User Preferences

First, we set out to define the problem, exploring factors influencing nicotine addiction and successful quit attempts, and we then explored user preferences with our main goal to establish app components required to create user engagement, resulting in a smoking cessation app likely to be retained by users. In total, 3 focus groups with 26 smokers were conducted.

Step 3. Incorporating the Evidence Base

Cigbreak development was guided by the Medical Research Council framework for complex interventions [35] following the British Standard Institute Code of Practice for Health and Wellness app development [24]. The latest guidelines for effective smoking cessation were also considered [36-39].

Cigbreak was designed with 30 game levels, which were estimated to take an average user 28 days to complete. Evidence suggests that being abstinent for 28 days increases the chance of successfully quitting [9], and hence Cigbreak can support its users through the most critical period of their smoking cessation attempt.

Step 4: Integrating Gamification

There are several taxonomies or frameworks of gamification available [21,40-43]; however, there is a recognized lack of high-quality studies [22] and only 1 taxonomy has been validated [40,44]. Given the lack of empirical guidance in this area, we relied on the experience of our expert multidisciplinary design team aiming to create an engaging and entertaining game.

Having gamified the app, we then turned to one of the more popular taxonomies to deconstruct Cigbreak in an effort to understand how the game might address various aspects of motivation and to optimize features. Cugleman and colleagues identified 7 key gamification strategies: goal setting, capacity to overcome challenges, providing feedback on performance, reinforcement, comparing progress, social connectivity, and fun and playfulness [21].

Step 5: Incorporating Behavior Change Techniques

A BCT is “an observable, replicable and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior; that is, a technique is proposed to be an ‘active ingredient’ (eg, feedback, self-monitoring, reinforcement)” [45]. BCTs have been clearly defined and classified into an internationally recognized taxonomy [45]. It was understood that if our intervention was to be effective, it would be important to make good use of BCTs. Therefore, in addition to providing distraction from cravings, the game would become part of a causal pathway for health behavior change.

This development phase of Cigbreak involved close collaboration with a health psychologist, who advised on BCTs on the basis of strong empirical evidence for their efficacy [27,28] and feedback gained in earlier focus groups. A systematic review of BCTs in existing health apps was also conducted [2]. Self-regulatory BCTs (feedback & monitoring including self-monitoring of behavior) have been commonly used in gamified apps to promote physical activity, healthy eating, and alcohol reduction [2,12,46,47]. These specific BCTs...
were effective in achieving behavior change in previous studies [48-53], and therefore, they were an obvious choice for inclusion.

**Step 6: Developing a Logic Model and Investigating Causal Pathways**

A logic model to describe the program theory behind the intervention was formulated, and methods of representing and examining the operation of the intervention in mathematical models were researched.

**Step 7: User Testing**

Cigbreak was built in Unity for Android [54] by Healthy Games Ltd. The use of a preexisting game engine facilitated rapid development of app prototypes, which could be circulated among the team and focus groups of smokers.

**Results**

**Defining the Problem and Establishing User Preferences**

Personal experiences, concerns about ill health, financial pressures, and family/friends were key motivational themes for stopping smoking. Environmental factors and mood played an important role in whether smokers continued to smoke. Smokers identified a need for extra support and distraction from the action of smoking in certain environments. As their thoughts at these times were overwhelmingly related to smoking, a game including cigarettes that involved swiping them was felt to be helpful. It was this key finding together with brainstorming in the team that led us to develop the initial concept that the app could be used as a distraction from cravings. These focus groups drove inclusion of many other features such as giving information on health and finances and using family members to provide motivational cues. The key themes identified in motivational factors are shown in Table 1.

**Integrating Gamification**

Game elements were incorporated by game developers and computer scientists (Table 3) [21].

**Incorporating Behavior Change Techniques**

At the end of this phase, our design for Cigbreak included 36 BCTs (see Figure 2 and Table 4 for a full list). Two researchers trained in BCT coding (EAE and JL) coded Cigbreak independently in accordance to the (V1) BCT taxonomy [45].

Screenshots of the Cigbreak app are shown with examples of BCTs highlighted. Note some screenshots may show multiple BCTs, which have not been highlighted but would have been coded accordingly.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Subthemes</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial</td>
<td>Financial factors are an important motivational factor in quitting</td>
<td>● If I’m short of money one month, that definitely stops me smoking.</td>
</tr>
<tr>
<td>Mood</td>
<td>Stress/boredom/unhappiness contribute to motivation to smoke</td>
<td>● I smoke when I am stressed.</td>
</tr>
<tr>
<td>Family and friends</td>
<td>Family, especially children, can be a motivation to quit</td>
<td>● Children can be powerful.</td>
</tr>
<tr>
<td>Experience of illness and health concerns</td>
<td>Personal experiences of ill health secondary to smoking in family/friends would affect motivation to quit</td>
<td>● When my mum was unwell with her chest, it definitely made me think of stopping.</td>
</tr>
<tr>
<td>Environmental factors</td>
<td>More likely to lapse if socializing may need more support during these activities; Job stops people smoking during the day time; Being on holiday encourages people to smoke more; Alcohol encourages people to smoke more; Government bans have been very useful</td>
<td>● I would smoke if others smoked around me.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>● Interestingly when you know smoking not a possibility don’t really get cravings e.g. on 10 hour flight.</td>
</tr>
<tr>
<td>Themes</td>
<td>Subthemes</td>
<td>Quotes</td>
</tr>
<tr>
<td>------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Health information</td>
<td>To know the harms of smoking and the benefits of quitting</td>
<td>• I think what you need to hear is that stern image you have on the back of cigarettes of what it is doing to you.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maybe it could give you an explanation of how your health is improving.</td>
</tr>
<tr>
<td>Distraction from craving</td>
<td>To keep your hands busy; To help you know what to do when you get cravings; To have a craving button</td>
<td>• I am craving so idea of a craving button which could be pressed when needed and links to a page with different management strategies e.g. play game, prompt to go for walk, speak to friend.</td>
</tr>
<tr>
<td>Personalization or related to real time quitting</td>
<td>To be personalized to the individual; To receive personalized texts and messages at times of craving; To be able to enter personalized quit date and plan; To be able to record relapses</td>
<td>• Being personalized is definitely important.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Maybe you could put dates in for your stop date.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• If you had a relapses you could put it in there too, so it is tied to your real time quitting.</td>
</tr>
<tr>
<td>Bright in color</td>
<td>To be bright in color</td>
<td>• Being colorful is important.</td>
</tr>
<tr>
<td>Interactive</td>
<td>To be interactive</td>
<td>• It has to be interactive to work.</td>
</tr>
<tr>
<td>Gamification</td>
<td>To contain gaming elements; To be fun</td>
<td>• Having a game and having to do something with your hands does work because you will spend an hour on that game before you have realized it and that an hour you have not smoked a cigarette.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Games are very good, candy crush is brilliant, I was playing it at 3am when I wanted a cigarette, doing something quick with your hands takes your mind off it.</td>
</tr>
<tr>
<td>Diary</td>
<td>To contain a personal dairy</td>
<td>• Like for example I want to do something in my diary, like today I have stopped and I really fell like a cigarette, so I would write down I have saved x amount by not going out and buying cigarettes.</td>
</tr>
<tr>
<td>Financial saving</td>
<td>To contain information on the financial savings of quitting</td>
<td>• I would like to know I have saved x amount by not going out and buying cigarettes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• After a week if would be good to tot up how much money you have saved.</td>
</tr>
<tr>
<td>Links to local services, pharmacies</td>
<td>Links to local quit smoking services or phone lines or local pharmacies</td>
<td>• That would really help, or it could give locations of stop smoking places.</td>
</tr>
<tr>
<td>Information about NRT(^a)</td>
<td>To contain information about different NRT products available and how to use them</td>
<td>• I would like to know more about the different products available.</td>
</tr>
<tr>
<td>Emotional content</td>
<td>To have an emotional content, would prefer focus on positive not negative emotions; To promote feelings of relaxation and happiness</td>
<td>• It needs to have an emotional content, but I think more focus on the positive rather than the negative.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• I want it to help me relax.</td>
</tr>
</tbody>
</table>

\(^a\)NRT: nicotine replacement therapy.
Table 3. Gamification strategies in Cigbreak, including their location in the app.

<table>
<thead>
<tr>
<th>Gamification strategy</th>
<th>Description</th>
<th>Location in the app</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal setting</td>
<td>Players are asked to set or agree on goals and to set a quit date</td>
<td>Diary</td>
</tr>
<tr>
<td>Capacity to overcome challenges</td>
<td>The diary contains missions and challenges, for example calling a quit line or contacting their local pharmacy for advice on using approved medications; The game involves players increasing speed and avoiding objects when slicing cigarettes</td>
<td>Diary; Game</td>
</tr>
<tr>
<td>Feedback</td>
<td>Players are given feedback on outcomes of their behavior, for example, the number of days smoke-free and number of cigarettes smoked; Players are provided with a score for number of cigarettes sliced</td>
<td>Diary; Game</td>
</tr>
<tr>
<td>Reinforcement</td>
<td>Players are rewarded with trophies for setting quit date and for not smoking</td>
<td>Diary; Game</td>
</tr>
<tr>
<td>Comparing progress</td>
<td>Players are asked to share their scores with their friends on Facebook. Friends can compare scores and trophies</td>
<td>Game; Diary</td>
</tr>
<tr>
<td>Social connectivity</td>
<td>Players can connect with friends for social support via Facebook</td>
<td>Diary; Game</td>
</tr>
<tr>
<td>Fun and playfulness</td>
<td>Players break cigarettes that dance upward from a generic cigarette pack while trying to avoid breaking any vegetables. Players are rewarded with gold, silver, and bronze stars for breaking the required number of cigarettes</td>
<td>Game</td>
</tr>
</tbody>
</table>

Figure 2. Example of embedded behavior change techniques in Cigbreak.
<table>
<thead>
<tr>
<th>BCT</th>
<th>Description</th>
<th>Location in the game</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Goal setting (behavior)</td>
<td>Players are asked to set goals, for example, a quit date</td>
<td>Diary</td>
</tr>
<tr>
<td>1.2 Problem solving</td>
<td>Players are asked to identify specific triggers that generate the urge to smoke and develop strategies for avoiding environmental triggers</td>
<td>Diary</td>
</tr>
<tr>
<td>1.3 Goal setting (outcome)</td>
<td>Players are asked to set goals, for example, to save money as a result of the money saved from not smoking to pay for a holiday</td>
<td>Diary</td>
</tr>
<tr>
<td>1.6 Discrepancy between current behavior and goal</td>
<td>If players continue to smoke after their quit date, they are provided with feedback on the discrepancy between current smoking and their previously set outcome goals</td>
<td>Diary</td>
</tr>
<tr>
<td>2.2 Feedback on behavior</td>
<td>Players are given feedback on their behavior, for example, number of cigarettes smoked since starting the app</td>
<td>Diary</td>
</tr>
<tr>
<td>2.3 Self-monitoring of behavior</td>
<td>Players are asked to monitor their smoking daily in a diary in the app. Players are also asked daily in pop-up questions the number of cigarettes smoked and whether players have remained smoke-free</td>
<td>Diary</td>
</tr>
<tr>
<td>2.7 Feedback on outcomes of behavior</td>
<td>Players are given feedback on the number of days smoke-free and what this means for their health</td>
<td>Diary</td>
</tr>
<tr>
<td>3.1 Social support (unspecified)</td>
<td>Players are asked to share their scores and smoke free status with their friends on Facebook</td>
<td>Game</td>
</tr>
<tr>
<td>3.2 Social support practical</td>
<td>Players are provided with practical advice in the diary on how to gain support from friends or colleagues or staff to help them to quit; for example, players are provided with information on support from community pharmacies</td>
<td>Diary</td>
</tr>
<tr>
<td>4.2 Information about antecedents</td>
<td>Players are prompted in the diary to record situations or circumstances in which they are more likely to get cravings or experience the urge to smoke, for example, being at a bus stop</td>
<td>Diary</td>
</tr>
<tr>
<td>5.1 Information about health consequences</td>
<td>Players are provided written, audible, and visual information on the health consequences of smoking and the health benefits of quitting; Health reward badges inform of the health consequences of smoking</td>
<td>Game</td>
</tr>
<tr>
<td>5.2 Salience of consequences</td>
<td>The avatar’s lungs become blackened and cough as cigarettes are not destroyed, thus highlight damage to lungs from not stopping</td>
<td>Game</td>
</tr>
<tr>
<td>5.3 Information about social and environmental consequences</td>
<td>Players are provided with written and visual information regarding social and environmental consequences of smoking</td>
<td>Game</td>
</tr>
<tr>
<td>5.6 Information about emotional consequences</td>
<td>Players are provided with written information regarding emotional consequences of quitting smoking, for example, that quitting smoking increases happiness and life satisfaction</td>
<td>Game</td>
</tr>
<tr>
<td>7.1 Prompts or cues</td>
<td>Players are prompted to play the game at times of cravings via the crave button</td>
<td>Game or Diary</td>
</tr>
<tr>
<td>8.2 Behavior substitution</td>
<td>Players are prompted to substitute smoking with an alternative behavior; For example, players are prompted to play the game via the crave button when they have a craving to smoke or prompted to go for a walk</td>
<td>Diary</td>
</tr>
<tr>
<td>8.7 Graded tasks</td>
<td>Tasks or missions set to help players to quit smoking become increasingly more difficult, but achievable, until players have quit</td>
<td>Diary</td>
</tr>
<tr>
<td>9.1 Credible source</td>
<td>The app is cocreated by doctors, health psychologists, and smoking cessation advisors. Players are made aware of this on the app store, website, and in the app</td>
<td>Diary</td>
</tr>
<tr>
<td>9.2 Pros and cons</td>
<td>Players are asked to list the pros and cons for quitting smoking in the diary</td>
<td>Diary</td>
</tr>
<tr>
<td>9.3 Comparative imagining of future outcomes</td>
<td>Players are asked to imagine and compare possible outcomes following quitting smoking, imagining themselves as nonsmokers and document this in the diary</td>
<td>Diary</td>
</tr>
<tr>
<td>10.3 Nonspecific reward</td>
<td>Players are provided with trophies, stars, and virtual coins for staying smoke-free</td>
<td>Game</td>
</tr>
<tr>
<td>10.4 Social reward</td>
<td>Players are congratulated by friends on social media or Facebook</td>
<td>Game or Diary</td>
</tr>
<tr>
<td>10.6 Nonspecific incentive</td>
<td>Players are incentivized to stay smoke-free</td>
<td>Game</td>
</tr>
<tr>
<td>10.7 Self-incentive</td>
<td>Players are encouraged to plan to reward themselves in the future for staying smoke free</td>
<td>Diary</td>
</tr>
<tr>
<td>10.9 Self reward</td>
<td>Players are encouraged to reward themselves at the present time if they have stayed smoke free</td>
<td>Diary</td>
</tr>
<tr>
<td>10.10 Reward (outcome)</td>
<td>Players are provided with in-game rewards for remaining smoke-free; for example, they are provided with trophies and health reward badges</td>
<td>Game</td>
</tr>
<tr>
<td>BCT</td>
<td>Description</td>
<td>Location in the game</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>11.1 Pharmacological support</td>
<td>Players are encouraged to use NRT(^a) and to obtain this from their pharmacy. There are also NRT power-ups in the game, which allow players to progress to higher levels in the game. This encourages the use of NRT and also informs players of the different NRT available.</td>
<td>Game</td>
</tr>
<tr>
<td>11.2 Reduce negative emotions</td>
<td>Players are given advice on reducing negative emotions to quit smoking; for example, using stress management skills.</td>
<td>Diary</td>
</tr>
<tr>
<td>12.1 Restructuring the physical environment</td>
<td>Players are advised how to change their physical environment to help themselves to quit, for example, to remove cigarettes from their home and to remove ash trays.</td>
<td>Diary</td>
</tr>
<tr>
<td>12.2 Restructuring the social environment</td>
<td>Players are advised and given tips on how to change their social environment to help them to quit, for example, avoiding contact with friends who smoke.</td>
<td>Diary</td>
</tr>
<tr>
<td>12.3 Avoidance or reducing exposure to cues for the behavior</td>
<td>Players are advised and given tips on how to avoid exposure to specific social and contextual or physical cues with regard to smoking, for example, avoiding pubs and bars that have been associated with smoking.</td>
<td>Diary</td>
</tr>
<tr>
<td>12.4 Distraction</td>
<td>The gameplay involves players breaking cigarettes acting as a distraction from smoking.</td>
<td>Game</td>
</tr>
<tr>
<td>12.5 Adding objects to the environment</td>
<td>Players are provided with the app and written or visual information to aid smoking cessation.</td>
<td>Diary</td>
</tr>
<tr>
<td>13.5 Identity associated with changed behavior</td>
<td>Players are advised to construct or articulate a new self-identity as an ex-smoker.</td>
<td>Diary</td>
</tr>
<tr>
<td>15.1 Verbal persuasion about capability</td>
<td>Players are encouraged that they can quit smoking, arguing against self-doubts and asserting that they can and will succeed.</td>
<td>Diary</td>
</tr>
<tr>
<td>15.3 Focus of past success</td>
<td>Players are asked to list previous successes, for example, when they have resisted smoking.</td>
<td>Diary</td>
</tr>
</tbody>
</table>

\(^a\)NRT: nicotine replacement therapy.

**Developing a Logic Model and Investigating Causal Pathways**

The logic model (Figure 3) describes basic assumptions on which the intervention is based, together with channels through which smokers are invited to use the app, for example, social media advertisements and recommendations by clinicians. The recruitment methods are designed to reach people who find conventional health care services difficult to access.

The app is designed to record user engagement with BCTs and subsequent changes in smoking behavior to allow detailed temporal analysis of causal pathway assumptions. Frameworks and graphical models for exploring causal relations are well established [55-57], and causal methodologies to account for confounding by covariates are now sufficiently developed to enable effects of new public health interventions to be evaluated from observational studies [58-67]. The relatively unexplored approach of the Cigbreak app for improving public health holds exciting possibilities to form a defining example for the development of testing of causal assumptions and pathway analysis suited to app design. This would provide the framework for a more refined analysis of complex interventions with a dynamic longitudinal treatment regime resulting in long-term health benefit.

**User Testing**

**Changes to Cigbreak From Direct User Feedback**

Key themes were identified in a framework analysis (Table 5), and the app was redesigned accordingly. For example, players wanted the app to be more personalized. In response, a personalized diary for players was introduced, which allowed players to set personal goals and quit dates and to add pictures of the item or person motivating them to quit. At this stage of app development, the thematic analysis was oriented toward more concrete app features and design elements rather than abstract concepts.

**From Prototype to Product**

Players found the prototype app to be an engaging and motivating way to deliver smoking cessation advice, providing a useful distraction from smoking (Table 6), and 84% (62/73) said they would play again or recommend to a friend. Those with higher tobacco dependence as defined by the Fagerström test for nicotine dependence [30,31] and difficult-to-reach smokers appeared to be more engaged. Engagement was measured as both flow (focused attention and enjoyment) and usage (frequency, duration, and depth of usage) [18]. Players said they would be happy to obtain this app from their pharmacy or general practitioner.

The game was completed in 2015 with funding from the London Tri-borough Smokefree Alliance and is commissioned in 5 London boroughs. Following the final focus groups, Cigbreak was released on both Apple and Google Play stores in January 2016.
Figure 3. Cigbreak logic model describing the intended operation of the intervention.

Table 5. Key themes identified by Cigbreak users to aid app development.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Subthemes</th>
<th>Quotes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health</td>
<td>Players wanted more emphasis on health gains; Players wanted more hard health facts and graphic imagery; Players wanted the characters' lungs to be linked to reality</td>
<td>I want to know the other benefits.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I like the lungs in the corner but they need to be linked to reality. I need fear to motivate me.</td>
</tr>
<tr>
<td>Personalization</td>
<td>Players wanted the game to be more personalized and wanted a personalized avatar</td>
<td>I want the game to be personalised to me.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I like the character in the corner, but I want to personalise him.</td>
</tr>
<tr>
<td>Goals</td>
<td>Players wanted goal setting linked to real life behavior</td>
<td>I want to go on holiday so my goal is to save £x.</td>
</tr>
<tr>
<td>Levels or graphics</td>
<td>Players wanted more variation in the levels; Players wanted the game to become increasingly more difficult; Players wanted more interactive graphics; Players wanted brighter colors and graphics</td>
<td>Needs more variation to keep you interested.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>It needs to get harder and harder and needs to be more engaging.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Needs to be more interactive and enticing.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Needs to be brighter in color.</td>
</tr>
<tr>
<td>Financial</td>
<td>Players wanted the financial rewards more related to real life</td>
<td>It would be good to see what I have saved per week.</td>
</tr>
</tbody>
</table>


Table 6. Key themes identified using Cigbreak as a smoking cessation tool.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Subthemes</th>
<th>Quotes</th>
</tr>
</thead>
</table>
| Distraction     | Players felt the game would work as a distraction | - I think it could work, candy crush saga was my life when I was quitting, so if it was related to smoking that would be even better.  
- It’s a good distraction.  
- It definitely stops me thinking of cigarettes. |
| Cravings        | Players felt the app could be used at times of cravings | - I would pick this up and play when I had a craving.  
- I could use this at the bus stop when I get cravings. |
| Availability    | Players would be happy to obtain from their pharmacy or GP<sup>a</sup> | - Every GP should know about it, to give it to their patients. |

<sup>a</sup>GP: general practitioner.

Discussion

Principal Findings

A qualitative approach working directly with the target population at an early stage was key to the design process. Feedback on early prototypes helped us to gain an understanding of the problem [24], establish user preferences, and ensure desired functions were included [24,68].

Forming an expert development team was vital to ensure that our app was evidence based and potentially effective [69]. Computer scientists and app developers were included to implement game elements and health psychologists, clinicians, and researchers for incorporation of BCTs. Careful consideration of which BCTs to include was needed, with a clear understanding of which techniques are theoretically linked to the required mechanisms of action [70].

It was found that including imagery of cigarettes and smoking-related cues did not act as a trigger to smoke. Feedback from the qualitative work suggested that a smoking-related game would be more engaging compared with a nonsmoking–related game as attentional bias is toward smoking when craving occurs. The literature also supports this finding. A randomized controlled trial compared Nicot, a video game in which players crush virtual cigarettes in a 3D game environment, with a balloon popping game. Nicot was found to improve smoking cessation rates by 13% [70]. Sanders-Jackson and colleagues found along with Due and colleagues, Hogarth and colleagues, and Meinke and colleagues that individuals who have chemical addictions are more likely to attend visually to objects associated with their addiction [71-74].

It was found that participants wanted to have positive rather than negative messaging and imagery. Gain-framed messages shift smoking-related beliefs, attitudes, and behaviors toward the direction of avoidance and cessation [75,76]. However, Mayes and colleagues found that a combination of both might be beneficial, specifically in younger smokers [77]. Moorman and Putte found that a positive frame is preferable when nicotine dependence or quitting intention is lower, and conversely, a negative frame worked better when nicotine dependence and quitting intention are both high [78].

Comparison With Prior Work

There are currently no published frameworks for production of gamified health apps with systematic inclusion of BCTs or use of behavior change theory.

Several nongamified health apps use an iterative user-centered design process [9,13,79-136]. Most apps are designed by an expert multidisciplinary team. However, only 3 explicitly used agile development [137-140], although this form of technology development is gaining in popularity, and it is anticipated that there will be greater use in the future.

There are several published frameworks for developing nongamified health apps [26,139,141-147]. The Schnall and colleagues framework uses the Information Systems Research framework [141,148]. The Goyl and colleagues framework [143] combines knowledge to action [149] and the MRC framework for complex interventions [33]. The Tombor and colleagues framework [26] is also guided by MRC framework for complex interventions and Multiphase Optimization Strategy [35,150]. Moreover, 6 frameworks also consider the incorporation of behavior change theory. Curtis and colleagues adopted the behavior change wheel (BCW) [144,151]. Tombor and colleagues also uses the BCW and the BCT taxonomy version 1 [26,45,151]. Wilhide and colleagues consider behavioral models [142], whereas Goyal and colleagues and Whitaker and colleagues incorporate social cognitive theory [143,146,152]. Mummah and colleagues combine use of both theory and a taxonomy [145]. Of the frameworks, 3 do not incorporate any form of behavior change theory [139,141,147].

As this area of research is still nascent, it is unclear if any of the above approaches are superior to the approach adopted. The framework recommend is based on our experience in having a successfully developed gamified health app rather than empirical evidence. Recent frameworks including our own are designed around guidelines for complex interventions involving a systematic development process with an iterative user-centered approach based on theory and evidence. It is likely that these elements are key for the development of health apps that may be effective at modifying health behaviors.

Strengths and Limitations

Although there are smoking cessation apps that incorporate gamification techniques, to date, there is no publicly available dedicated game to promote smoking cessation or guidance for...
the development of gamified health apps. Raiff and colleagues have developed a prototype smoking cessation mobile phone game using the concept of contingency management (delivering rewards contingent on objective evidence of smoking abstinence) with virtual in-game rewards and social connectivity. The prototype game differs from Cigbreak in concept and mechanic, with the prototype game involving swiping pollen-gems, not including imagery of cigarettes or smoking-related cues. In addition, 71% of participants in a small usability study reported that the program would help them personally to quit smoking [153]. The prototype has not yet been released on app stores and is still in development.

In this study, an example is provided of a systematic approach for developing a gamified smoking cessation app with systematic use of BCTs and evidence-based practice guided by user input. This could provide a potential methodology for development of other gamified health apps. However, the framework proposed has not been evaluated, and thus it cannot be concluded that apps developed using this framework are superior or that the methods adopted are superior to other methods. Future work could include a comparison of usability or acceptability outcomes and behavior change outcomes for apps developed using this framework compared with apps that have not used this framework. The rapidly changing technological landscape and change in sophistication of users may limit the applicability of the findings to future app development.

There are further limitations to the methods adopted. Focus groups can be subject to group bias; however, in turn, small focus groups can generate a natural open discussion, providing fruitful feedback as to usability. Participants were given different iterations of the app to use during the focus groups, which may potentially have biased participants to focus on the functionality of the app presented to them. However, this is a natural consequence of agile iterative development. Only 1 trained researcher independently analyzed the focus group transcripts, which may have led to bias; however, any uncertainties found were discussed with the study team.

Although Cigbreak has been made available to the public on a small scale, the app has not yet been formally evaluated against clinical endpoints. A rigorous evaluation to assess impact on short- and long-term quit rates is being planned.

Implications for Clinicians and Policy Makers
Mobile phone games could provide a potentially cost-effective platform for health promotion and thus have a substantial public health impact. However, developers of digital interventions need to adhere to existing regulatory frameworks and emerging standards [24,69] to develop games that health practitioners can feel confident to recommend to patients.

Questions remain as to the best way to evaluate health apps in this rapidly changing field, without stifling innovation. Michie and colleagues present recent consensus from experts at an international workshop on digital interventions in relation to health behaviors, concluding that evaluations should be made during all phases of the development cycle and need not rely solely on traditional methods [69]. New experimental methods and adaptive research designs such as A/B testing and N-of-1 studies may be used to make best use of rich data streams and assess outcomes within shorter time frames [69]. There is considerable scope for using emerging methods of analyzing observational data in this context [60].

Unanswered Questions and Future Research
Further work is needed to identify potential causal pathways and mechanisms in health apps in general and specifically for Cigbreak. Existing methods such as structured equation modeling and pathway analysis can be applied to study short-to-medium-term effects such as Cigbreak engagement. However, current methods only explore assumed causal relations between limited numbers of variables and are not well adapted to complex models or digital interventions. Thus, there are exciting possibilities for future development of more refined analysis of applications and for generating ecologically valid, real-time objective data.

Game analytics provide a rich source of data, and machine learning techniques can be used to make changes to the game design to improve the users’ experience and potentially to modify health outcomes. The key is to ensure that the correct metrics are captured during gameplay and then to use the techniques to identify patterns in the data. It should then be possible to change gameplay adaptively to optimize some chosen criteria. These methods could be used to change health behavior if this is being captured as one of the metrics. Bauckhage and colleagues [154] discuss a number of the methods in clustering game behavior data, and use of such techniques in Cigbreak is currently being explored. Future research also aims to ascertain which components of a multicomponent intervention such as Cigbreak are accounting for what effects.

Authors’ Contributions
HC was the game designer. EAE, HC, and RTW were involved in conception and design of the study. EAE, HC, CR, LS, and RTW were involved in the interpretation of the results. EAE, LS, and CR collected focus group data, and LS provided behavior change technique advice. EAE and RTW drafted the manuscript; HC, CR, LS, JL, YP, SJ, CN, AS, SM, JQS, DG, and CJG revised the manuscript critically for intellectual content. All authors approved the final version of the paper. All authors had access to all study data and take responsibility for data integrity and accuracy of the analysis. RTW is the guarantor.

RTW is an National Institute for Health Research (NIHR) Senior Investigator and Chief Investigator on NIHR Program grant RP-PG-0609-10181. EAE is an NIHR-funded in-practice Fellow. JL is conducting a PhD funded by the Economic and Social Research Council and Cambridge Cognition Limited.
Conflicts of Interest
HC is a mobile phone game developer and director of Healthy Games.

References
1. Lunden I. Techcrunch. 2015. 6.1B Smartphone Users Globally By 2020, Overtaking Basic Phone Subscriptions URL: https://techcrunch.com/2015/06/02/6-1b-smartphone-users-globally-by-2020-overtaking-basic-fixed-phone-subscriptions/ [accessed 2018-02-27] [WebCite Cache ID 6xXlgWRQw]


38. Smoking: reducing and preventing tobacco use. United Kindom: The National Institute for Health and Care Excellence; 2015. URL: https://www.nice.org.uk/guidance/fs82 [WebCite Cache ID 73YG7G0Qe3]


Abbreviations

BCT: behavior change technique
BCW: behavior change wheel
NRT: nicotine replacement therapy

©Elizabeth A Edwards, Hope Caton, Jim Lumsden, Carol Rivas, Liz Steed, Yutthana Pirunsarn, Sandra Jumbe, Chris Newby, Aditi Shenvi, Samaresh Mazumdar, Jim Q Smith, Darrel Greenhill, Chris J Griffiths, Robert T Walton. Originally published in JMIR Serious Games (http://games.jmir.org), 29.11.2018. 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 Serious Game on Delirium as an Educational Intervention for Medical Students: Randomized Controlled Trial

Kiki R Buijs-Spanjers¹, MSc; Harianne HM Hegge¹, MD; Carolien J Jansen¹, MSc; Evert Hoogendoorn², MA; Sophia E de Rooij¹, MD, PhD

¹Department of Geriatric Medicine, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
²IJsfontein, Amsterdam, Netherlands

Corresponding Author:
Kiki R Buijs-Spanjers, MSc
Department of Geriatric Medicine
University Medical Center Groningen
University of Groningen
Hanzeplein 1
PO Box 30.001 HP AA43
Groningen, 9700 RB
Netherlands
Phone: 31 503613464
Email: k.r.spanjers@umcg.nl

Abstract

Background: Adequate delirium recognition and management are important to reduce the incidence and severity of delirium. To improve delirium recognition and management, training of medical staff and students is needed.

Objective: In this study, we aimed to gain insight into whether the serious game, Delirium Experience, is suited as an educational intervention.

Methods: We conducted a three-arm randomized controlled trial. We enrolled 156 students in the third year of their Bachelor of Medical Sciences degree at the University Medical Centre Groningen. The Game group of this study played Delirium Experience. The Control D group watched a video with explanations on delirium and a patient’s experience of delirious episodes. The Control A group watched a video on healthy aging. To investigate students’ skills, we used a video of a delirious patient for which students had to give care recommendations and complete the Delirium Observations Screening Scale and Delirium Rating Scale R-98. Furthermore, students completed the Delirium Attitude Scale, the Learning Motivation and Engagement Questionnaire, and self-reported knowledge on delirium.

Results: In total, 156 students participated in this study (Game group, n=51; Control D group, n=51; Control A group, n=55). The Game group scored higher with a median (interquartile range) of 6 (4-8) for given recommendations and learning motivation and engagement compared with the Control D (1, 1-4) and A (0, 0-3) groups (P<.001). Furthermore, the Game group scored higher (7, 6-8) on self-reported knowledge compared with the Control A group (6, 5-6; P<.001). We did not find differences between the groups regarding delirium screening (P=.07) and rating (P=.45) skills or attitude toward delirious patients (P=.55).

Conclusions: The serious game, Delirium Experience, is suitable as an educational intervention to teach delirium care to medical students and has added value in addition to a lecture.

JMIR Serious Games 2018;6(4):e17 doi:10.2196/games.9886

KEYWORDS
delirium; education; medical students; serious game

Introduction

Delirium is an acute neuropsychiatric syndrome often experienced by older hospitalized patients. It is characterized by altered attention, awareness, and cognition. Delirium has serious consequences such as increased length of hospital stay, functional decline, institutionalization, and mortality [1]. Adequate delirium recognition and management are important to reduce the incidence and severity of delirium [2,3]. To improve delirium recognition and management, training of...
medical staff and students is needed [4] as timely recognition is crucial [2]. Lack of delirium awareness, knowledge, and education were the most commonly reported barriers to improving the recognition of delirium (risk) and the hospital care for delirious patients [5]. Current educational interventions focus merely on increasing knowledge and skills in recognition of delirium but do not seem to be effective enough [6,7]. It was suggested that educational interventions on delirium should have a broader scope to target (1) the attitude of the medical staff and students toward delirious patients; (2) the understanding of patients’ needs; and (3) the translation of this knowledge into the practice of offering good health care to delirious patients [7,8]. Future educational interventions on delirium should not only have a broader scope addressing these 3 objectives but also focus on teaching methods with students actively involved and supportive technologies with sufficient feedback loops [6,7].

Serious games may be an opportunity to meet this demand for new educational interventions. Serious games are games developed and intended to provide playful learning experiences, which can be transferable to or applicable in real-life settings [9]. Serious games are often more effective compared with regular health care educational interventions [10] or assessments [11]. However, there is a lack of effect studies [12] and assessment [13] of good quality on serious games.

Delirium Experience is a recently designed serious game that uses video simulation [14], which is intended to train and educate medical students on how to take better care of delirious patients. As both serious games [15] and simulation-based learning [16,17] provide learning spaces in which learners can safely practice, Delirium Experience might serve as a new educational intervention by addressing the need for a focus on caregiver attitude and the application of knowledge to the care of delirious patients.

In this study, we aimed to gain insight into whether Delirium Experience is suited as an educational intervention for medical students regarding skills in advising care for delirious patients, skills in screening and rating of delirium symptoms, and improving the attitude toward delirious patients. Additionally, we aimed to gain insight into the possible effects of Delirium Experience on learning motivation and engagement, as well as self-reported knowledge on delirium.

Methods

Design and Study Population

We conducted a three-arm randomized controlled trial. The study population consisted of undergraduate medical students at the University Medical Centre Groningen (UMCG). To be included in this study, participants had to have (1) be in their third year of preclinical education in December 2016; (2) sign up for the practical on delirium; and (3) sign the informed consent form. The UMCG offers an undergraduate program of 6 years—a 3 years of preclinical and 3 years of clinical education. Preclinical medical students at the UMCG select 1 of 4 different learning communities with different, in-depth focus during their medical education (global health, sustainable care, intramural care, and molecular medicine). At the moment, the UMCG third-year preclinical medical curriculum on delirium is based on lectures and literature. However, educators of the UMCG emphasize the need for a more practice-based education before students enter their clinical education.

Students started with the conventional lecture on delirium. Thereafter, students could voluntarily sign up for the practical on delirium, in which the study conditions took place. The practicals were given in three separate classrooms of the University of Groningen. Each study condition had a separate classroom. All students had the opportunity to join the practical on delirium, including students who did not wish to participate in the study. Students were informed about the study in the description of the practical. This practical description explained that the practical was divided over 3 different groups for research purposes but did not explain the different study conditions. Students were not aware that the serious game, Delirium Experience, was one of the study conditions, in order not to influence the motivation to sign up for the practical. All students were provided a license of Delirium Experience after the practical so they could play the serious game. Data were collected and analyzed anonymously.

We used SPSS 23.0 (IBM Inc) for stratified block randomization (block size of 6) to allocate participants into one of the three research groups [18]. Learning communities represented the 4 different strata used. All participants who signed up for the practical were randomly allocated to one of the groups. They subsequently received an email indicating the classroom in which they were expected. As our research subjects consisted of medical students who could voluntarily sign up for both the practical and the study, registration of the trial was not necessary in accordance with the ICMJE (International Committee of Medical Journal Editors) recommendations. Multimedia Appendix 1 shows the CONSORT-EHEALTH checklist.

Intervention and Control Groups

We designed three different practicals on delirium, which represented the study conditions. Only the intervention group, the Game group, played Delirium Experience [14]. Delirium Experience is a serious game focusing on delirium both from a patient’s and a caregiver’s perspective (watch the trailer in Multimedia Appendix 2). The goal of Delirium Experience is to allow players to learn how to take better care of delirious patients. The game tries to achieve this by giving players insight into what a patient experiences during delirious episodes and how your actions as a caregiver influence the experience of the patient. Delirium Experience was based on the delirium guidelines used in the United Kingdom [19] and the Netherlands [20] and on stories of patients who suffered from delirious episodes. The game was developed with personnel who were specialists in developing serious games, designing education, and treating delirium, all working closely together. Usability was tested by a group of care professionals during the development. Based on their suggestions and feedback, the final version of the game was made. Completing the game once takes approximately 20 minutes; in these 20 minutes, one experiences 4 days as a caregiver and the corresponding 4 nights as the patient. During the daytime, as a caregiver, the player has to...
take care of a delirious patient and can choose different actions. Depending on the actions one chooses, the delirious episodes of the patient differ in severity, and one gets different actions to choose from the next day. Hence, if one performs poorly as a caregiver, the severity of delirious episodes increases, and the next day, one has fewer actions to choose from compared with a caregiver who performed well. Players who perform poorly have their actions limited to only the most important actions to decrease the level of difficulty. Furthermore, players receive feedback every other day in the game on how they performed and how they could improve as a caregiver before they switch to the patient’s perspective.

We compared this Game group with two other groups, one with and one without information about delirium. The first control group, Control D group, watched a video on delirium, which explained delirium causes, symptoms, diagnosis, treatment, and pathology. Contrary to the serious game, the video did not ask for active involvement of students; thus, students were not able to try different scenarios. Furthermore, this group watched a second video of a patient’s experience explaining his suffering from delirious episodes.

The second control group, Control A group, watched a more general video on healthy aging. This video did not have any specific information on delirium and how to take care of delirious patients; each session took 20 minutes.

**Outcome Measurements**

At baseline, before the intervention started, all participants completed a form including questions on sex, age, experience with older and delirious patients, learning community, self-reported knowledge on delirium, Which mark (0-10) would you give your knowledge on delirium?; and attendance at the lecture. Primary and secondary outcome measures were assessed directly after the intervention or control condition.

The primary outcome of this research was assessment of the skills acquired by students in advising care for delirious patients, in which students describe how they would manage delirium in practice. In this outcome, students could show their understanding of patients’ needs and be able to translate this knowledge into practice [7,8]. To measure skills in advising care, all participants observed an interview of a delirious patient and were asked to give 3 written recommendations for the care of this patient. A predefined rubric-form was used to assess all given recommendations as rubric-forms can enhance the reliability of assessors’ scoring [21]. The rubric-form was based on the Dutch delirium guidelines [20]. Recommendations were assessed independently by two researchers, and a weighted kappa was calculated. To ensure blinding of the assessors, data on intervention and control groups were removed from the assessed recommendations. Each recommendation could receive 0 (incorrect or not mentioned), 1 (topic mentioned), 2 (nonspecific recommendation), or 3 (specific recommendation) points from the 10 different domains of the Dutch delirium guidelines (range, 0-9 points) [20].

Subsequently, several secondary outcomes were measured. First, use of screening and rating instruments for delirium was measured. Participants completed the Delirium Observations Screening Scale (DOSS) [22] and Delirium Rating Scale R-98 (DRS-R-98) [23,24] for the patient in the observed interview. Both scales are widely accepted and applied tests for the recognition and severity assessment of delirium. Second, attitude toward delirious patients was measured using the Delirium Attitude Scale. The Delirium Attitude Scale is based on the Dementia Attitude Scale [25]. Items regarding creativeness, enjoyment of life, and coping skills were replaced by items focusing on the experiences of delirium. This resulted in a 19-item 7-point Likert scale (range, 19-133 points). I feel confident around people with delirium and I would avoid an agitated person with delirium are examples of statements used in the Delirium Attitude Scale. Third, learning motivation and engagement were measured using the Motivation and Engagement Questionnaire to evaluate learning experiences [26], a 9-item 5-point Likert scale (range, 9-45 points). Examples of statements used in this questionnaire are as follows: It was challenging to perform well in this practical and I liked this way of learning. Finally, participants were asked to self-report their knowledge on delirium (range, 0-10 points).

**Statistical Methods**

We checked data for normality by judging histograms, skewness, and kurtosis. We analyzed discrete variables using chi-square test. Furthermore, continuous variables were analyzed using one-way analysis of variance (ANOVA) in case of normal distribution and Kruskal-Wallis in case of a nonnormal distribution. \( P < .05 \) was considered statistically significant for the results of the chi-square and one-way ANOVA or Kruskal-Wallis tests. In case of significant results regarding outcome measurements, specific post hoc or Mann-Whitney tests were performed to investigate differences between the (1) Game group and Control D group or (2) Game group and Control A group. Furthermore, a Bonferroni correction for two tests was used for the Mann-Whitney test; therefore, \( P < .025 \) was considered statistically significant for the results of the Mann-Whitney test.

**Results**

In total, 176 of 387 students subscribed for the practical on delirium in December 2016. Of these 176 students, 156 signed the informed consent form and participated in the study (Figure 1). The 20 students who declined to sign the informed consent form still participated in the practical but were not included in the study. Students did not have to give a reason why they declined to sign the informed consent form. We compared playing a serious game (Game group) to either watching a video on delirium in combination with a video of a patient’s experience (Control D group) or watching a video on healthy aging (Control A group). Data on students’ characteristics and outcome measures were not normally distributed. The median age (interquartile range [IQR] 25-75) of all participants was 20 (20-21) years, and 75% (117/156) participants were females. No differences were found between the research groups regarding baseline variables, as presented in Table 1.
Figure 1. Flowchart of approached students and participants.

![Flowchart](image_url)

Table 1. Baseline variables.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total participants (N=156)</th>
<th>Game (n=51)</th>
<th>Control D (n=50)</th>
<th>Control A (n=55)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in years**, median (IQR)**</td>
<td>20 (20-21)</td>
<td>21 (20-21)</td>
<td>20 (20-21)</td>
<td>21 (20-21)</td>
</tr>
<tr>
<td>Female**, n (%)</td>
<td>117 (75.0)</td>
<td>37 (72.5)</td>
<td>40 (80.0)</td>
<td>40 (72.7)</td>
</tr>
<tr>
<td>Experience older patients**, n (%)</td>
<td>118 (75.6)</td>
<td>36 (70.6)</td>
<td>38 (76.0)</td>
<td>44 (80.0)</td>
</tr>
<tr>
<td>Experience delirious patients**, n (%)</td>
<td>48 (30.8)</td>
<td>17 (33.3)</td>
<td>15 (30.0)</td>
<td>17 (30.9)</td>
</tr>
<tr>
<td><strong>Learning community</strong>, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global health</td>
<td>45 (28.8)</td>
<td>19 (37.3)</td>
<td>15 (30.0)</td>
<td>12 (21.8)</td>
</tr>
<tr>
<td>Molecular medicine</td>
<td>39 (25.0)</td>
<td>11 (21.6)</td>
<td>14 (28.0)</td>
<td>14 (25.5)</td>
</tr>
<tr>
<td>Sustainable care</td>
<td>31 (19.9)</td>
<td>11 (21.6)</td>
<td>7 (14.0)</td>
<td>11 (20.0)</td>
</tr>
<tr>
<td>Intramural care</td>
<td>40 (25.6)</td>
<td>10 (19.6)</td>
<td>14 (28.0)</td>
<td>18 (32.7)</td>
</tr>
<tr>
<td>Attended lecture, n (%)</td>
<td>129 (82.7)</td>
<td>44 (86.0)</td>
<td>43 (86.0)</td>
<td>43 (78.2)</td>
</tr>
<tr>
<td>Self-reported knowledge on delirium (0-10)<strong>, median (IQR)</strong></td>
<td>5 (4-6)</td>
<td>5 (4-6)</td>
<td>5 (4-6)</td>
<td>5 (4-6)</td>
</tr>
</tbody>
</table>

**Game: Delirium Experience; Control D: video on delirium with a patient experience video; Control A: video on healthy aging.

**Data compared using Kruskal-Wallis test, P>.05.

**IQR: interquartile range.

**Data compared using chi-square test, P>.05.
The primary outcome of this study, skills in advising care for delirious patients, was measured on the basis of the given care recommendations. The independently assessed recommendations, which were scored using the rubric-form, had a weighted kappa of .835. Disagreements were resolved through discussion.

Kruskal-Wallis tests showed differences between the three groups regarding given recommendations, $H(2)=54.5, P<.001$, learning motivation and engagement, $H(2)=91.5, P<.001$, and self-reported knowledge, $H(2)=26.0, P<.001$, as presented in Table 2.

Table 2. Kruskal-Wallis and Mann-Whitney U tests for primary and secondary outcomes for the Game (n=51), Control D (n=50), and Control A (n=55) groups.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Gamea</th>
<th>Control Da</th>
<th>Control Ab</th>
<th>$P$ valueb</th>
<th>$P$ value (G-D)c</th>
<th>$P$ value (G-A)d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommendations</td>
<td>6 (4-8)</td>
<td>1 (1-4)</td>
<td>0 (0-3)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DOSSc</td>
<td>10 (9-11)</td>
<td>9 (8-10)</td>
<td>9 (8-11)</td>
<td>.07</td>
<td>N/Af</td>
<td>N/A</td>
</tr>
<tr>
<td>DRS-R-98e</td>
<td>14 (12-16)</td>
<td>13 (12-15)</td>
<td>14 (11-16)</td>
<td>.45</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Attitude</td>
<td>92 (88-96)</td>
<td>94 (90-100)</td>
<td>92 (85-96)</td>
<td>.55</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Learning motivationh</td>
<td>36 (32-38)</td>
<td>27 (24-30)</td>
<td>20 (15-25)</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Delirium knowledgei</td>
<td>7 (6-8)</td>
<td>6 (6-7)</td>
<td>6 (5-6)</td>
<td>&lt;.001</td>
<td>.03</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

aData are presented as median (interquartile range 25-75); Game: Delirium Experience; Control D: video on delirium with patient experience video; Control A: video on healthy aging.

bKruskall-Wallis test to compare the three groups.

cMann-Whitney U test to compare the Game group and the Video Delirium group ($P<.025$ considered statistically significant).

dMann-Whitney U test to compare the Game group and the Video Aging group ($P<.025$ considered statistically significant).

eDOSS: Delirium Observation Screening Score.

fN/A: not applicable.

gDRS-R-98: Delirium Rating Scale R-98.

hLearning motivation and engagement.

iSelf-reported knowledge on delirium.

Discussion

Principal Findings

In this study, we investigated the effects of a serious game, Delirium Experience, as a new educational intervention. We compared playing a serious game with watching a video with delirium explanation in combination with a patient experience video or a video on healthy aging. The results showed that the serious game had a positive effect on students’ skills in advising care for delirious patients, learning motivation and engagement, and self-reported knowledge on delirium. However, the serious game did not influence skills in screening and rating the severity of delirium. In addition, it did not affect the attitude toward delirious patients.

Although students in the group playing the serious game and the group watching the video on delirium got more familiar with the behavior of delirious patients, they were not explicitly trained in the use of the DOSS and DRS-R-98 or recognizing the delirious behavior. Furthermore, the design of Delirium Experience allows players to practice caring for a delirious patient and manage delirium instead of recognizing it. The DOSS and DRS-R-98 are used and applied in a real-life setting over 24 hours by trained and experienced health care professionals [22,24]. This might explain why we did not find differences in delirium screening and rating scores, as the medical students did not get training and patient information of 24 hours, nor did they have clinical experience with delirious patients. All participants received high scores on their attitude toward delirious patients, which could have been caused by a ceiling effect or by students answering in a socially desired way. Furthermore, attitudes can be influenced by intense emotions [27]. As participants played the game as well as they could, they probably did not see the more severe delirium scenarios.
as these scenarios are only shown when the game players perform poorly.

**Implications**

Conventional simulation-based educational interventions have proven to be effective but are costly owing to a large number of teachers, role-play actors, and time and space required [17]. By using Delirium Experience as a simulation-based educational intervention, students could play without the need of teachers and role-play actors. Development of serious games is costly. Delirium Experience was developed by an unrestricted grant of NutsOhra, and development costs were covered and, therefore, not relevant for educational institutes as they only pay for a license to use the game. Delirium Experience provides students with a safe environment to practice and apply attained knowledge and can be used as an educational intervention on delirium to improve skills in advising care for delirious patients. As simulation-based assessment seems to predict the clinical performance [28], this safe simulation environment might prepare preclinical students in advance for their clinical education. In addition, this study supports earlier research on the importance of including objectively obtained measurements instead of self-reported measures [12,29]; we did not find differences in the self-reported knowledge between playing the serious game and watching the delirium video, but we did find differences between these groups regarding skills in advising care for delirious patients.

**Limitations**

This study has a number of limitations. First, skills in advising care for delirious patients were measured using a video of a delirious patient and using written answers instead of a real clinical situation, which would involve both the responsibility of caring for a delirious patient and the demonstration of the correct skills. However, simulation-based assessment seems to be a suitable tool for predicting clinical performances [30]. Second, there might have been selection bias in the recruitment of students, as more highly motivated students were more likely to sign up for the practical. However, due to the design of this randomized study, this could not have influenced the differences between the research groups. Furthermore, there is a slightly skewed number of students that declined to sign the informed consent form in the different groups. We cannot explain this, because students did not have to explain why they declined, nor do we have any information on these students because they were never included in the study. However, as there was only a small percentage of students that declined to sign the informed consent form, we do not expect this to influence the results. Finally, we did not perform a sample size calculation beforehand. We approached all third-year medical students. If we had not found statistically significant results due to a too low sample size or power, the study would have been extended in 2017. However, as we found significant results, the power was sufficient.

**Further Research**

Further research should be performed as to whether it is possible to improve attitudes toward delirious patients with Delirium Experience. If the change in attitude can be established by more emotional and intense patient scenarios [27], Delirium Experience might improve attitudes when students are allowed to play Delirium Experience several times, including “dark play.” In a dark play situation, players show behavior in the game that in a normal care situation would be problematic [31] and increases the intensity of the delirium. In Delirium Experience, this results in adverse events and scenarios with an extremely frightened patient. Showing immoral behavior, such as dark play, in video games has already been proven to lead to improved awareness of moral norms [32]. Subsequently, it would be interesting to investigate the effect of dark play on learning outcomes such as advising care for delirious patients. Because Delirium Experience increases learning motivation, it would be interesting to investigate whether students might be more motivated to use Delirium Experience as self-study material [28] and whether the increased motivation also influences learning outcomes. Furthermore, player characteristics might influence the effectiveness and use of games and should be taken into account in future studies [33]. In addition, future studies should take into account other health care professionals and trainees to generalize the results and use of interdisciplinary games, such as Delirium Experience, and investigate whether Delirium Experience can improve timely recognition of delirium. Finally, it is important to look at long-term effects of playing a serious game and ascertain our interest in whether it can influence the strain of care in experienced health care professionals working with delirious patients.

**Conclusions**

Playing Delirium Experience increases medical students’ skills in advising care for delirious patients, learning motivation and engagement, and self-reported knowledge on delirium. However, in this study, we could not show an effect on improving delirium screening and severity rating skills or on attitudes toward delirious patients after playing Delirium Experience. The serious game, Delirium Experience, is suitable as an educational intervention to teach delirium care to medical students and has added value in addition to that of a lecture.

**Conflicts of Interest**

The serious game, Delirium Experience, was developed by IJsfontein and is owned by Stichting Effectieve Ouderenzorg (a Dutch foundation for improving elderly care by research and education). EH is an employee of IJsfontein. SEdR is an unpaid member of the supervisory board of Stichting Effectieve Ouderenzorg, which waived the licensing fee required for the use of intellectual property for the purposes of this research. The game is currently commercialized, but the revenues are solely used to improve current elderly care by gamification.
Editorial notice: This randomized study was not prospectively registered. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because, according to ICMJE rules, registration of the trial was not necessary as the research subjects consisted of medical students who could voluntarily sign up for both the practical and the study. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as the lack of registration means that authors could change their outcome measures retrospectively.

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

[PDF File (Adobe PDF File), 2MB - games_v6i4e17_app1.pdf]

Multimedia Appendix 2
Trailer of Delirium Experience.

[MP4 File (MP4 Video), 217MB - games_v6i4e17_app2.mp4]

References

Abbreviations
- ANOVA: analysis of variance
- DOSS: Delirium Observations Screening Scale
- DRS-R-98: Delirium Rating Scale R-98
- ICMJE: International Committee of Medical Journal Editors
- IQR: interquartile range
- N/A: not applicable
- UMCG: University Medical Centre Groningen

©Kiki R Buijs-Spanjers, Harianne HM Hegge, Carolien J Jansen, Evert Hoogendoorn, Sophia E de Rooij. Originally published in JMIR Serious Games (http://games.jmir.org), 26.10.2018. This is an open-access article distributed under the terms of the
Potential of an Interactive Drug Prevention Mobile Phone App (Once Upon a High): Questionnaire Study Among Students

Máté Kapitány-Fövény1,2, PhD; Eszter Vagdalt1, MA; Zsófia Ruttkay4, PhD; Róbert Urbán5, DSc; Mara J Richman5, MA; Zsolt Demetrovics5, DSc

1Department of Addiction, Semmelweis University Faculty of Health Sciences, Budapest, Hungary
2Drug Outpatient Centre, Nyírő Gyula National Institute of Psychiatry and Addictions, Budapest, Hungary
3Budapest Center for Vocational Education and Training in Engineering, Budapest, Hungary
4Creative Technology Lab, Moholy-Nagy University of Art and Design, Budapest, Hungary
5Institute of Psychology, Eötvös Loránd University, Budapest, Hungary

Abstract

Background: In recent years, drug prevention networks and drug education programs have started using Web-based or mobile phone apps as novel prevention tools, testing their efficacy compared with face-to-face prevention.

Objective: The aim of this study was to assess the potential of an interactive app called Once Upon a High (VoltEgySzer).

Methods: The app approaches drug prevention from 6 different aspects, and it addresses youngsters with 6 different modules: (1) interactive comics/cartoons, telling stories of recovery; (2) quiz game; (3) roleplay game; (4) introduction of psychoactive drugs; (5) information on the somatic and psychological effects of psychoactive substances; (6) list of available treatment units, rehabs, and self-support groups in Hungary. Students of 2 vocational schools and 2 high schools filled out a questionnaire at a baseline (T0) and a 2-month follow-up (T1) data collection session. Students of 1 vocational school and 1 high school downloaded the Once Upon a High app (app group), whereas students from the other vocational school and high school did not (nonapp group). The time points of T0 and T1 questionnaires contained demographic variables, items with regard to substance use characteristics for both legal and illegal substances, including novel psychoactive substance, exercise habits, knowledge about psychoactive substances, attitudes toward substance users and validated instruments measuring the severity of tobacco (Fagerström Test for Nicotine Dependence), alcohol (Alcohol Use Disorder Identification Test), cannabis (Cannabis Abuse Screening Test), and synthetic cannabinoid consumption. Beliefs about substance use (Beliefs About Substance Abuse) and perceived self-efficacy (General Perceived Self-Efficacy) were also measured. At T1, members of the app group provided additional evaluation of the app.

Results: There were 386 students who participated in the T0 session. After dropout, 246 students took part in T1 data collection procedure. Alcohol was the most frequently consumed psychoactive substance (334/364, 91.8% lifetime use), followed by tobacco (252/386, 65.3%, lifetime use) and cannabis (43/323, 13.3% lifetime use). Decreased self-efficacy (beta=−.29, P=.04) and increased daily physical exercise frequencies (beta=.04, P<.001) predicted higher frequencies of past month energy drink consumption, whereas elevated past month alcohol consumption was mainly predicted by a decrease in negative attitudes toward substance users (beta=−.13, P=.04) in the regression models. Once Upon a High was found to be effective only in reducing energy drink consumption (beta=−1.13, P=.04) after controlling for design effect, whereas perceived utility of the app showed correlation with a decreasing alcohol use (rS(44)=.32, P=.03). The roleplay module of the app was found to be the most preferred aspect of the app by the respondents.
Conclusions: The Once Upon a High app can be a useful tool to assist preventive intervention programs by increasing knowledge and self-efficacy; however, its efficacy in reducing or preventing substance use needs to be improved and further studied. Additional potential impacts of the app need further testing.

(JMIR Serious Games 2018;6(4):e19) doi:10.2196/games.9944

KEYWORDS
secondary prevention; adolescent; mHealth; energy drinks; substance use; alcohol abuse; cannabis

Introduction

Background

The era of contemporary substance use scene is also an era in which Web-based communication influences our daily lives to an increasing extent. Spending an increasing amount of time in virtual spaces is a phenomenon that mostly affects the lifestyle of adolescents [1]. Virtual spaces—such as Web-based fora, blogs, Web markets—serve a significant role as one of the primary sources of gathering either classic or novel psychoactive substance (NPS)-related information as well. The Web-based space not only provides information and means of communication, but Web-based marketing is also responsible for increasing the availability of NPS [2].

Not only the drug market but prevention itself also transformed as one of the many consequences of the rise of information technologies. In recent years, drug prevention networks and drug education programs have started using Web-based or mobile phone apps as novel prevention tools [3]. Such virtual methods may provide an opportunity to access individuals who otherwise might not seek or receive professional help. The anonymity and accessibility of prevention apps often lead to increased self-disclosure with regard to sensitive subjects such as substance consumption [4]. The effectiveness of Web-based interventions providing personalized feedback and mobile phone apps in reducing the intake of psychoactive substances is already supported by empirical evidence [5,6]. According to the findings of previous research, such programs may result in significant decrease in alcohol consumption among heavy drinkers [7], in prolonged smoking cessation as a supplement to nicotine patch therapy [8], or in decrease of marijuana use among college students who show a high level of contemplation to be abstinent [9]. However, there are some contradictory results as well with regard to the efficacy of such programs when implemented in the adolescent population. On the basis of the systematic review of Majeed-Ariss and colleagues [10], some apps may be considered feasible health interventions that may increase self-management of chronic health conditions, but on the other hand, there are several apps available on the app market without any evidence-based background. A Web-based prevention program for ecstasy and NPS [11] was reported to be efficient in reducing adolescents’ intentions to use NPS; it increased knowledge about both ecstasy and NPS. However, changes in lifetime use of ecstasy or NPS did not differ significantly between the intervention and control conditions.

The result that tailored and interactive sites and apps are more effective than static ones [12] is also essential with respect to our project, although this principle applies not only to apps and websites but also to prevention itself. Interactivity may also increase the subjective feeling of self-efficacy and self-directedness, which might improve the commitment to participate in a prevention program. Self-directedness itself was found to be a relevant protective factor against opiate addiction [13], alcohol addiction [14], cannabis addiction [15], and smoking [16].

As opposed to personal, face-to-face prevention, preventive interventions implemented in the virtual space have distinct advantages as they are available any time, are more cost effective in the long run, are better in providing tailored feedback, and are able to reach more members of the target population [17]. Previous studies have also indicated that mobile phone apps combine the benefits of Web-based and computerized preventive interventions as they provide both interactive and static contents with and without an internet connection [18], even if some apps cannot operate properly offline. Nevertheless, updates concerning eventual content change are usually assured by automatic version refreshments in every case.

With respect to the methodological concept of our project, we followed the concept of gamification. Gamification—when utilized in prevention—is usually defined as the mixture of game design elements and traditional prevention techniques in a nongame context [19]. The aim of gamification is to increase the engagement and motivation of the target group while providing a useful method in supporting learning and problem solving. It is not essential to use any digital technology in gamification, but smartphones make the implementation easier and the outreach wider. With regard to the efficacy of gamification in preventing substance use among adolescents, Boendermaker and colleagues [20] found that game elements (authors label them as serious games) can help to motivate youngsters to do a cognitive bias modification training, which might reduce substance consumption. However, research is still lacking to draw further conclusions about the efficacy of gamification.

Objective

The aim of this study was to assess the effectiveness of a mobile phone prevention app, titled Once Upon a High (the original name of the app is VoltEgy/Szer in Hungarian). We aimed to target adolescents in an age range that may indicate the existing experiences with substance use but without clinically relevant problems. Therefore, Once Upon a High was tested as a novel tool with embedded game elements (ie, as a form of gamification) for mainly the purpose of secondary prevention. Being such an instrument, we expected this app to be effective—among other goals—in reducing or at least maintaining substance use frequencies. The app, besides being
a contemporary gadget of providing information on the risks of substance use, was expected to enhance or affect certain skills as well, which might be considered as protective factors against the increasing severity of substance consumption and are as follows: (1) self-efficacy is usually seen as a protective resource in prevention or even treatment of substance use disorders (SUD) [21] and as an important determinant of health behavior in general [22]; (2) increasing physical exercise is similarly recommended as a potential tool to be utilized in substance use prevention for the youth [23,24] or as a beneficial adjunctive treatment of SUD [25], which may lead to elevated abstinence rates or the ease of withdrawal symptoms [26].

In the study, besides the exploratory analysis of the app’s utility, the following hypotheses were tested:

Once Upon a High can be effective in:

- H1: Increasing knowledge about the risks of psychoactive substances
- H2: Enhancing perceived self-efficacy
- H3: Decreasing or maintaining the frequency of current substance use
- H4: Decreasing negative attitudes toward substance users
- H5: Increasing the frequency of physical exercise

Although H2, H4, and H5 might also be related to primary prevention, we still considered self-efficacy, physical exercise, and negative attitudes toward substance users as potential protective factors against the exacerbation of substance use among active users; and as such, preferable outcomes of secondary prevention too, potentially leading to decreased substance use frequencies. We further hypothesized that increased knowledge (H1), perceived self-efficacy (H2), physical exercise (H5), and decreased negative attitudes toward substance users (H4) might predict decreased frequencies of substance use (H3).

**Methods**

**Once Upon a High: Introducing the App**

Once Upon a High is an interactive drug prevention app targeting the youth and drawing attention to the importance of prevention in a novel way. The app approaches drug prevention from 6 different aspects, and it addresses youngsters with 6 different modules: (1) interactive comics/cartoons, telling stories of recovery (see Animated Comics section); (2) quiz game (see Quiz Game section); (3) roleplay game (see What If? section); (4) introduction of psychoactive drugs (see Substance Store section); (5) information on the somatic and psychological effects of psychoactive substances (see Trans-formation section); (6) list of available treatment units, rehabs, and self-support groups in Hungary (see Where to go? section). The user may choose to use any of the modules from the initial menu. The app unites the experience of usual Web education and computer games. The applied genres—cartoons, videos, and animation—are anticipated to be attractive to the youth by their nature. The app provides anonymity for its users.

The app was designed and developed in a joint project of Nyírő Gyula Hospital, National Institute of Psychiatry and Addictions and the techLab of the Moholy-Nagy University of Art and Design, Budapest. The development of the app was funded from a national tender (invited by the Hungarian Government’s Ministry of Human Capacities) supporting the establishment of novel drug prevention programs. More information on the technical details are presented in the Multimedia Appendix 1 as well as on the visual rendering of the app. To increase risk perception, yet to avoid overestimation of the dangers of specific psychoactive substances (ie, avoiding autotelic deterrence), the literature basis for each module contained current papers providing high-quality evidence (ie, mainly systematic reviews and meta-analyses or randomized controlled trials). The literature basis for each module is presented under the detailed characteristics of the app.

**Detailed Characteristics and Aims of the Modules**

**Animated Comics: Interactive Recovery Stories**

The first module comprises altogether 4 recovery stories, of which there are stories of 2 males (Adam, the beggar; Adam, the prince) and 2 females (Eve, the beggar; Eve, the princess). These 4 comic-style animated tales (Figure 1) introduce common risk factors of adolescent psychoactive substance use (eg, peer pressure, peer recognition for substance use involvement, social isolation, a dysfunctional family background indicated by symptoms such as familial substance use and either sexual, physical, or emotional victimization) based on the findings of former studies [27,28], as well as potential way outs by presenting 4 types of treatments: (1) art therapy, (2) individual psychotherapy, (3) group therapy including self-support groups and as a combination with family counseling, and (4) animal-assisted therapy. The selection of these therapeutic interventions was based on former findings with respect to their efficacy in addiction treatment. Art therapy is commonly used throughout the process of detoxification and rehabilitation, including 12-step programs and self-support groups [29]. The effectiveness of individual psychotherapy and primarily cognitive behavioral therapy in the treatment of SUD was supported by the results of numerous meta-analyses [30,31]. Group counseling and family therapies were highlighted as the most effective interventions in a study comparing the effectiveness of various treatment methods for adolescent substance use [32]. A systematic review and meta-analysis of randomized controlled trials further confirmed the efficacy of animal-assisted therapy in the treatment of SUD [33]. The interactivity of this module means that the dialogues are triggered by the user, and turns are indicated by a slight movement of the main characters. The module also aims to display different socioeconomic backgrounds of substance users by telling the stories of both socially marginalized youth and youngsters coming from high-income families.

**Quiz Game**

The second module, a quiz-game, contains a database of 50 potential questions, of which the app randomly generates 15 questions every time the user plays the game. The quiz game is structured by 3 difficulty levels (easy, moderate, and hard). Difficulty levels were determined by assigning an adjustment
weight to the questions based on the results of 98 high school students who gave either correct or wrong answers to all the 50 questions. Many of the quiz questions reflect common misbeliefs about the prevalence and risks of substance use (eg, by asking: “What percentage of high school students in your country have tried an illegal substance at least once in their lifetime?”). As a trend, adolescents often overestimate peer substance use [34], which might legitimize their own substance consumption.

Review of the current literature resulted in the following relevant areas, addressed in the quiz game: the topics of cannabis-induced psychosis [35], effects of energy drink consumption [36], dangers of adolescent alcohol use [37], peer group influence on adolescent smoking [38], or harms of NPS [39]. Less emphasis was placed on the risks associated with the use of substances with low prevalence among adolescents, such as heroin or other opioids, based on the findings of the most recent ESPAD study [40]. To every question, there are 3 potential answers (A, B, and C solutions), of which only one is correct. After every round, the user receives a feedback on his or her answers, which contains relevant citations from the scientific literature as well.

The quiz game has 2 main goals. First, to provide well-established knowledge of the epidemiology and risks of substance use, with respect to both legal and formerly banned “classic” substances (such as alcohol, tobacco, ecstasy, amphetamines, or cannabis) and NPS (in this module, mainly synthetic cannabinoids and synthetic cathinones). Second, the module aims to promote health-conscious behavior. Those users who are able to achieve a 100% quiz result are offered free admission to gyms, dance schools, laser-tag, and paintball facilities or indoor climbing centers. As such, this module might increase physical exercise frequencies as a potential benefit and, therefore, its application was considered to test our H5 hypothesis.

What If? Roleplay Game and Dialogue System

The third module titled “What if?” is an interactive dialogue system, within the confines of a roleplay situation: What if one of your close acquaintances started to use synthetic cannabinoids? What would you suggest? How would you help? During the roleplay game, the user communicates with his or her assumed acquaintance by following a structured dialogue system, in which the virtual “talk partner” makes a specific statement, whereas the user selects a response from 2 optional sentences offered (see Figure 2). The main task within the game is to persuade the virtual acquaintance to change his lifestyle by reducing his substance use, seeking professional help, or starting drug-free recreational activities. Whenever the user selects the better option, the portrayed room gets brighter and brighter and vice versa, in case of a worse option selection, the room goes darker and darker. If the user is able to convince the virtual partner to make a positive change in his life, the game ends positively. However, in case of bad choices during the dialogue (eg, by being too impatient, hypocritical, or arrogant), the game ends with a game over. In this case, users receive a feedback on how to improve their communication skills and strategy next time.

The roleplay game, therefore, aims to improve communication skills, decrease stigma, and negative attitudes against people (see H4) who use psychoactive substances and to increase willingness to help peers with substance use problems. As such it targets adolescents with or without any former substance use experiences. Communication style of the virtual partner was based mainly on clinical experiences with substance users as well as the feedback of the co-design group (see Appendix for details). The assumption that peer support might be of utmost relevance was supported by literature evidence as adolescents tend to prefer informal sources of help (eg, friends) rather than professional support [41]. Finally, the risks of synthetic cannabinoid use presented in this module were also collected from published data [42].
Figure 1. Visual appearance of the animated comics module.

Figure 2. Visual appearance of the What if? roleplay game module.
Substance Store: Information About Psychoactive Drugs

The fourth module titled “Substance-store” (Figure 3) provides information on “classic” substances (cannabis, cocaine, ecstasy and amphetamines, lysergic acid diethylamide [LSD], and heroin) and NPS or recently popularized substances (ketamine, gamma-hydroxybutyrate [GHB], Salvia divinorum, synthetic cannabinoids, and cathinones). Users are offered knowledge of the brief history of the selected psychoactive substance, information on the psychoactive compounds, both desired and unwanted or adverse effects of the substance, as well as the dangers of potential overdose. For each substance, an image
helps to recognize the common distribution form. As such, this module’s main goal is to educate the users and provide information. Cultural history of substance use and the desired and adverse effects of specific substances were collected from the literature [43-51].

**Trans-Formation: Somatic and Psychological Effects of Psychoactive Substances**

“Trans-formation” is closely related to the fourth module (Substance store). In addition to that one, this fifth module presents substance-specific somatic and psychological effects in a longitudinal aspect by offering a selection between immediate, short-term, and long-term effects of specific substances. As impulsivity—and especially sensation seeking—usually rise during adolescence [52], for this age group, it is thought to be more relevant to focus on the immediate and short-term consequences of substance use and not just the long-term risks of recurrent psychoactive substance consumption. All the presented somatic and psychological effects are based on the findings of former studies [42,53-55]. These effects are indicated by the red dots located on a picture of a schematic body with the option to touch a dot on the screen (see Figure 4) and, thus, access detailed description of the effects. Here too, we relied on visual clues and active exploration by the user as learning methodology.

**Where to Go? List of Available Treatment Units**

The aim of the final module was to create a national register of the available treatment units (both inpatient and outpatient), rehabs, and self-support groups. Thus, the purpose of this module is to help youngsters who are already at-risk substance users in finding nationwide professional help, as close to their place of residence as possible. By overviewing the existing national registries and the websites of self-support groups, we created a searchable map with the contact details of each of the treatment units and meeting points. As a result, altogether, more than 300 locations were registered.

**Sample and Procedure**

The sample of the study consisted of 386 students of 2 vocational schools and 2 high schools from Budapest, Hungary, to analyze outcome variability by school type as well. To examine the impact of age as a covariate on the outcome measures, 9th (approximately 14- to 15-years-olds), 10th (approximately 15- to 16-years-olds), and 11th (approximately 16- to 17-years-olds) graders participated in the study from each school. The age range was 14 to 18 years. Permission from the schools was obtained as well as from the parents in cases where respondents were younger than 18 years. Written informed consents were asked in every case. Our goal was to assess participants in a natural setting, that is, in their schools, with the involvement of those professionals (teachers and school psychologists) who are involved in the everyday life of these students, to model an as realistic arrangement as possible. Randomization was performed at the school level, where participants were nested in classes. Altogether 4 groups were assessed at 2 different time points.

Students of 1 vocational (VS1) and 1 high school (HS1) downloaded the Once Upon a High app, whereas students from the other vocational (VS2) and high school (HS2) did not. Those students who downloaded the app were, therefore, part of the group hereunder referred to as “app group,” whereas those who did not download the app were members of the group referred to as “nonapp group.” We wanted to avoid assigning students to both app and nonapp groups within the same schools to prevent information sharing between study groups. Schools were selected with the help and involvement of 2 school psychologists who were also involved in monitoring the data collection procedure. Therefore, a nonprobability, convenience sampling method was used. Homeroom teachers were asked to help in organizing occasions of data collection. The app itself had an option to provide further information about its use via a website. Thus, detailed psychoeducation with respect to the use of the app was not necessary. However, homeroom teachers and those colleagues who were responsible for face-to-face data collection answered any questions that might have arisen throughout the study. We additionally provided an email address to receive and answer any incoming questions. We tried to motivate students to participate (and maintain their participation) by ensuring the opportunity of winning recreational activities (ie, free admission to gyms, dance schools, laser-tag and paintball facilities, or indoor climbing centers) in case of a flawless result in the quiz-game module.

As data were collected in 2 phases (baseline and 2-month follow-up data collection), participants received a unique identifier (UID) that helped the pairing of baseline (T0) and follow-up (T1) data. UID was generated by a similar algorithm of treatment demand indicator and consisted of the following 6 characters: (1) the third letter of the participant’s surname, (2) the second letter of the participant’s first name (3) last number of the participant’s birth month, (4) last number of the participant’s birth day, (5) third letter of the maiden surname of the participant’s mother, and (6) second letter of the maiden first name of the participant’s mother. This method was found to be more effective than allowing the students to choose a UID for themselves as they were expected to forget that during the 2 months between the 2 dates of data collection.

**Measures**

The baseline and follow-up questionnaire was filled by all the participants who remained in the study. Members of the app group responded to additional questions during the follow-up measurement with respect to their experiences with the app. Participants were asked to provide an evaluation on (1) the utility and (2) subjective preference of distinct modules, as well as (3) general impression about the app using a 5-point Likert scale (from 1=not at all, to 5=absolutely). We computed 2 variables (“app preference” and “app usefulness”) as a total score of the evaluative response categories with respect to the distinct modules’ preference and the app’s perceived utility.

The baseline questionnaire contained demographic variables, questions about socioeconomic status, former experiences with other prevention programs, and the family history of either alcohol, tobacco, or illicit substance use. These items were not included in the follow-up questionnaire. Both the baseline and follow-up questionnaire comprised questions with respect to the respondents’ physical exercise and sport habits (including...
sport types, monthly and daily exercise frequencies, measured in minutes/day), psychoactive substance use experiences (including lifetime, last year, and last month frequencies of the consumption of both legal and illegal “classic” and NPS-type drugs). NPS-type substances included synthetic cathinones (mephedrone, methylenedioxyxymethamphetamine [MDMA], mephedrone), synthetic cannabinoids, and GHB, as the use of these NPS was found to be relevant among Hungarian school-aged respondents in the ESPAD study [40].

With respect to the measurement of psychoactive substance use, the questionnaire further contained validated instruments such as the 6-item Fagerström Test for Nicotine Dependence (FTND) [56,57], the 10-item Alcohol Use Disorder Identification Test (AUDIT) [58,59], and the 6-item Cannabis Abuse Screening Test (CAST) [60,61]. CAST was also used for the assessment of synthetic cannabinoid use and related problems (and referred to as sCAST). AUDIT, CAST, and sCAST were only included in the first data collection procedure as these instruments screen substance use severity at an annual rate (ie, ask questions about the past 12 months), whereas FTND measures current smoking habits. Beliefs about psychoactive substances were measured by 20-item Beliefs About Substance Abuse [62], including areas such as the ability of controlled substance use or the role of craving in relapse. Knowledge about the risks and prevalence of both legal and illegal classic and NPS-type drugs were measured by 12 items. These true/false or multiple-choice questions (eg, “What is the national prevalence of lifetime cannabis use among high school students in your country?”) were only included in the first data collection procedure as these instruments screen substance use severity at an annual rate (ie, ask questions about the past 12 months), whereas FTND measures current smoking habits. Beliefs about psychoactive substances were measured by 20-item Beliefs About Substance Abuse [62], including areas such as the ability of controlled substance use or the role of craving in relapse. Knowledge about the risks and prevalence of both legal and illegal classic and NPS-type drugs were measured by 12 items. These true/false or multiple-choice questions (eg, “What is the national prevalence of lifetime cannabis use among high school students in your country?”) A=15% to 20%, B=30% to 40%, C=50%, or more) tested the knowledge that the app aimed to provide for its users. A 5-point Likert scale was used for assessing the attitudes toward substance users. Respondents had to evaluate 10 statements on substance users and addiction itself (eg, “Addiction is not a disorder, it is a lack of will-power”), rating how strongly they agree with them (from 1=Absolutely disagree, to 5=Absolutely agree). A total score of negative attitudes was computed. Finally, perceived self-efficacy was measured by the 10-item General Perceived Self-Efficacy Scale [63,64], primarily dealing with self-observed coping skills.

Statistical Analysis

Data were analyzed by SPSS 17 (SPSS Inc) [65]. Descriptive statistics were applied to provide sample characteristics. Mean age and exercise frequencies were compared by independent sample t tests, perceived socioeconomic status as an ordinal variable was compared by Mann-Whitney U test, whereas potential differences in gender distribution and repetition of year rates were analyzed by chi-square statistics. Baseline comparison of the app versus nonapp groups, vocational versus high schools, and compliant respondents versus dropouts was implemented by using independent sample t tests, Mann-Whitney U test, chi-square test, and Fisher exact test. AUDIT, FTND, CAST, and sCAST scores were compared between the app and nonapp groups by using Mann-Whitney U test. Gender differences in substance use frequencies were analyzed by chi-square statistics and Fisher exact test.

To control for the design effect, pre- and postdata comparisons were performed by estimating treatment (intervention) effect with a series of linear regression analysis. Design effect was calculated by average cluster size and intracluster correlation. Changes in past month frequencies of less commonly consumed substances (including both “classic” and NPS-type stimulants and depressants) and the FTND scores at the postdata setting could not be examined by a regression analysis because of low response rates. The evaluation of the app was examined with descriptive statistics. Spearman rank correlation was applied to measure the connection between the repeated use of the distinct modules and the changes in attitudes, self-efficacy, knowledge, and exercise frequencies. The association between the app’s preference and perceived utility and substance use changes was also tested by Spearman rank correlation.

Results

Sample Characteristics

Altogether 386 students participated in the first data-collection (T0) session and 246 students took part in the second data collection procedure (T1). Main reasons of dropouts were lack of motivation to participate further in the study or missing identifiers to pair pre- and postdata. Tables 1 and 2 summarize detailed sample characteristics with respect to respondents with available identifiers.

With respect to both the pre-and posttest setting, the app and nonapp groups differed in age but not in gender distribution. In addition, a significant difference was found with respect to baseline exercise frequencies as members of the nonapp group showed higher means of daily exercise duration. In case of the posttest setting, this difference ceased to be significant. Vocational and high school students differed in perceived socioeconomic status as high school students reported higher living standards (mean 4.8, SD 0.9, U=14,517.5, P=.001) than students of the vocational schools (mean 4.5, SD 0.9). However, the app and nonapp groups did not show significant difference in terms of perceived socioeconomic status and rates of school year repetition because of failure (ie, poor academic performance).

Furthermore, respondents who participated in both data collection phases (ie, compliant participants) were compared with those who dropped out of the study, with regard to gender distribution, age, rates of year repetition because of academic failure, and substance use characteristics (lifetime and past month use of both legal and illegal substances, AUDIT, CAST, sCAST, and FTND scores). On the basis of this analysis, the group of students who dropped out of the study was characterized by a higher rate of male respondents (For N=384, χ²=10.4, P=.001), higher mean age (mean 16.9, SD 0.9, t=-2.72, P=.007), higher rates of lifetime cannabis (For N=323, χ²=4.4, P=.04), ecstasy (Fisher exact test, P=.006), amphetamine (Fisher exact test, P=.001), cocaine (Fisher exact test, P=.001), and LSD (Fisher exact test, P=.005) use; higher rates of last month cannabis (For N=323, χ²=4.7, P=.03), tobacco (For N=252, χ²=4.97, P=.03), LSD (Fisher exact test, P=.02), GHB (Fisher exact test, P=.02), and synthetic cannabinoid (Fisher exact test, P=.03) consumption. There were no differences in AUDIT (U=8296, P=.17), CAST (U=720.5, P=.49), and sCAST (U=914.5, P=.03).
Baseline Psychoactive Substance Use

The majority of the students have consumed alcohol at least once in their lifetimes (average rate=91.8%, 334/364), and there was no significant difference in the frequencies of lifetime and past month alcohol use between the subgroups (app vs nonapp groups; vocational vs high schools). However, vocational and high school students differed in the rates of lifetime and last month smoking. Although high school students showed a higher rate of lifetime smoking, last month frequencies were higher in vocational schools. Concerning illegal substances, respondents mainly had experiences with cannabis (average lifetime use rate=13.3%, 43/323), ecstasy (average lifetime use rate=3.1%, 10/319), and LSD or magic mushroom (average lifetime use rate=2.5%, 8/319). The nonapp group and participants from vocational schools showed higher rate of lifetime ecstasy consumption. In addition, lifetime use of cocaine was more frequent among the participants of the nonapp group. With respect to NPS-type substances, synthetic cannabinoids were found to be the most commonly used drugs (average lifetime use rate=6.6%, 21/319), whereas MDPV consumption only occurred among the students of vocational schools. Table 3 presents the results of the applied comparative analyses.

In case of baseline AUDIT, FTND, CAST, and sCAST total scores, there were no significant differences between the app and nonapp groups. However, students of the vocational schools showed higher total scores on both the CAST (mean 2.2, SD 4.3, \(U=532, P=.02\)) and the FTND (mean 3.3, SD 2.1, \(U=118, P<.001\)) compared with the CAST (mean 0.3, SD 0.7) and FTND scores (mean 1, SD 1.4) of high school students. Considering the severity of baseline substance use, 64 students (22.5%) indicated hazardous or harmful alcohol use (based on AUDIT cut-off: a score of 8 or more), 14 students (17.9%) showed moderate or severe cannabis use (based on CAST cut-off: a score of 3 or more), 3 students (5.8%) indicated moderate or severe synthetic cannabinoid use (based on CAST cut-off: a score of 3 or more), whereas 11 students (20.4%) showed risk for moderate or severe nicotine dependence (based on FTND cut-off: a score of 5 or more). With regard to potential gender differences, male students showed higher rates of lifetime ecstasy use (Fisher exact test, \(P=.04\)), whereas higher rates of lifetime smoking were found among female students (For \(N=384, \chi^2=11.95, P<.001\)). In case of AUDIT, FTND, CAST, and sCAST, no gender differences occurred.

Pre- and Postdata Comparisons

Among those participants with available and comparable pre- and postdata, potential changes in psychoactive substance use, knowledge about psychoactive substances, perceived self-efficacy, exercise frequencies, beliefs about substance use, and attitudes toward substance users were analyzed and compared between the app and nonapp group. Multimedia Appendix 2 presents pre- and posttest settings’ description with respect to the outcome measures for both the app and nonapp groups. Intervention or treatment effect was estimated with a series of linear regression analyses, while controlling for the design effect as well. Significant treatment or intervention effect was observed only in the frequency of past month energy drink consumption. Users of the app showed greater decrease in energy drink consumption after the implementation of the intervention. A trend increase in psychoactive substance–related knowledge and physical exercise frequencies could be seen in the app group but, compared with the nonapp group, these differences were not significant and could not be considered relevant.

Table 1. Sample characteristics and group differences at T0 participation.

<table>
<thead>
<tr>
<th>Sample characteristics</th>
<th>App group</th>
<th>Nonapp group</th>
<th>Significancea (app vs nonapp group)</th>
<th>Effect size, r</th>
</tr>
</thead>
<tbody>
<tr>
<td>T0 participation, N</td>
<td>VS1b</td>
<td>HS1c</td>
<td>VS2d</td>
<td>HS2e</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>16.7 (0.9)</td>
<td>16.5 (1.1)</td>
<td>16.9 (1)</td>
<td>17.2 (0.7)</td>
</tr>
<tr>
<td>Male participants, n (%)</td>
<td>131 (89.1)</td>
<td>54 (50.9)</td>
<td>70 (98.6)</td>
<td>24 (40)</td>
</tr>
<tr>
<td>Repetition of year of study because of failure, n (%)</td>
<td>13 (8.8)</td>
<td>1 (0.9)</td>
<td>5 (7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Perceived socioeconomic status, mean (SD)</td>
<td>4.5 (0.9)</td>
<td>4.9 (0.9)</td>
<td>4.5 (1.0)</td>
<td>4.6 (1.0)</td>
</tr>
<tr>
<td>Exercise frequency (minutes/day), mean (SD)</td>
<td>103.9 (91.9)</td>
<td>92.8 (51.9)</td>
<td>169.5 (250.3)</td>
<td>82.9 (47.4)</td>
</tr>
</tbody>
</table>

aIndependent sample t-test or Mann-Whitney U test or \(\chi^2\) test or Fisher exact test.
bVS1: vocational school 1.
cHS1: high school 1.
dVS2: vocational school 2.
eHS2: high school 2.
fN/A: not applicable.
Table 2. Sample characteristics and group differences at T1 participation.

<table>
<thead>
<tr>
<th>Sample characteristics</th>
<th>App group</th>
<th>Nonapp group</th>
<th>Significance(^a) (app vs nonapp group)</th>
<th>Effect size, r</th>
</tr>
</thead>
<tbody>
<tr>
<td>VS1(^b)</td>
<td>90</td>
<td>65</td>
<td>N/A(^f)</td>
<td>N/A</td>
</tr>
<tr>
<td>HS1(^c)</td>
<td>16.7 (0.9)</td>
<td>16.3 (0.9)</td>
<td>(t=−3.29; P=0.001)</td>
<td>.22</td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>38</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS2(^d)</td>
<td>16.7 (0.8)</td>
<td>17.1 (0.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS2(^e)</td>
<td>16.3 (0.9)</td>
<td>16.7 (0.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male participants, n (%)</td>
<td>79 (87.8)</td>
<td>33 (50.8)</td>
<td>(x^2_1=3.7; P=0.12)</td>
<td>.12</td>
</tr>
<tr>
<td>Repetition of year of study because of failure n (%)</td>
<td>9 (10)</td>
<td>1 (1.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perceived socioeconomic status, mean (SD)</td>
<td>4.4 (0.9)</td>
<td>5 (0.9)</td>
<td>(P=002)</td>
<td></td>
</tr>
<tr>
<td>Exercise frequency (minutes/day), mean (SD)</td>
<td>153.2 (208.9)</td>
<td>88.3 (48.5)</td>
<td>(U=6230)</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>191.3 (257.5)</td>
<td>84.6 (135.8)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Independent sample \(t\)-test or Mann-Whitney \(U\) test or \(\chi^2\) test or Fisher exact test.
\(^b\)VS1: vocational school 1.
\(^c\)HS1: high school 1.
\(^d\)VS2: vocational school 2.
\(^e\)HS2: high school 2.
\(^f\)N/A: not applicable.
Table 3. Psychoactive substance use characteristics at baseline measurement.

<table>
<thead>
<tr>
<th>Category, substance, and characteristics</th>
<th>App group</th>
<th>Nonapp group</th>
<th>Significance(^a) (app vs nonapp group)</th>
<th>Effect size, (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Legal substances, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>123 (92.5)</td>
<td>94 (88.7)</td>
<td>(\chi^2 = 0.9)</td>
<td>.05</td>
</tr>
<tr>
<td>Last month</td>
<td>65 (56)</td>
<td>53 (50.5)</td>
<td>(\chi^2 = 2.2)</td>
<td>.08</td>
</tr>
<tr>
<td>Tobacco</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>77 (52)</td>
<td>84 (77.6)</td>
<td>(\chi^2 = 2.1)</td>
<td>.07</td>
</tr>
<tr>
<td>Last month</td>
<td>35 (45.5)</td>
<td>20 (24.1)</td>
<td>(\chi^2 = 0.4)</td>
<td>.04</td>
</tr>
<tr>
<td><strong>Illegal substances, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cannabis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>14 (13.1)</td>
<td>14 (14.1)</td>
<td>(\chi^2 = 0.0)</td>
<td>.01</td>
</tr>
<tr>
<td>Last month</td>
<td>2 (1.9)</td>
<td>7 (7.1)</td>
<td>(\chi^2 = 0.0)</td>
<td>.001</td>
</tr>
<tr>
<td>Ecstasy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>3 (2.8)</td>
<td>0 (0)</td>
<td>(p = .04)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>Amphetamines</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>2 (1.9)</td>
<td>1 (1)</td>
<td>(p = .26)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>Cocaine</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .01)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>LSD(^{j})/Magic mushroom</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>3 (2.8)</td>
<td>1 (1)</td>
<td>(p = .47)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>2 (1.9)</td>
<td>0 (0)</td>
<td>(p = .63)</td>
<td>N/A</td>
</tr>
<tr>
<td>Heroin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>NPSI-type substances, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GHB(^{k})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>2 (1.9)</td>
<td>0 (0)</td>
<td>(p = .36)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .14)</td>
<td>N/A</td>
</tr>
<tr>
<td>Mephedrone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .56)</td>
<td>N/A</td>
</tr>
<tr>
<td>Pentedrone</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime use</td>
<td>2 (1.9)</td>
<td>0 (0)</td>
<td>(p = .36)</td>
<td>N/A</td>
</tr>
<tr>
<td>Last month</td>
<td>1 (0.9)</td>
<td>0 (0)</td>
<td>(p = .30)</td>
<td>N/A</td>
</tr>
<tr>
<td>MDPV(^{l})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Explaining Changes in Past Month Substance Use

As secondary outcome measures, changes in past month frequencies of psychoactive substance use were examined using the changes in beliefs about substance use, psychoactive substance–related knowledge, perceived self-efficacy, exercise frequencies, and attitudes toward substance users as explanatory variables, applying linear regression models. An increase in past month alcohol consumption was explained by a decrease in negative attitudes toward substance users (beta = −.13, \(P = .04\)). Changes in past month tranquillizer misuse or overuse, cannabis use, and FTND scores were not explained by any of the predictor variables. An increase in past month energy drink consumption was explained by an increase in exercise frequency (beta = .04, \(P < .001\)) and a decrease in perceived self-efficacy (beta = −.29, \(P = .04\)).

Evaluation of the App

Tables 4 and 5 present the results of the app’s evaluation as provided by the respondents of the app group.

### Table 4. Evaluation of the app’s distinct modules by the respondents.

<table>
<thead>
<tr>
<th>Modules</th>
<th>Utility, mean (SD)</th>
<th>Subjective preference, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animated comics</td>
<td>3.3 (1.3)</td>
<td>3.5 (1.4)</td>
</tr>
<tr>
<td>Quiz game</td>
<td>3.9 (1.2)</td>
<td>3.8 (1.3)</td>
</tr>
<tr>
<td>What if?</td>
<td>3.8 (1.3)</td>
<td>3.9 (1.3)</td>
</tr>
<tr>
<td>Substance store</td>
<td>3.9 (1.3)</td>
<td>3.8 (1.3)</td>
</tr>
<tr>
<td>Trans-formation</td>
<td>3.9 (1.3)</td>
<td>3.7 (1.3)</td>
</tr>
<tr>
<td>Where to go?</td>
<td>3.5 (1.4)</td>
<td>3.3 (1.5)</td>
</tr>
</tbody>
</table>

Respondents highlighted the overall utility of the app and the majority of the participants thought that the app might be an effective tool in providing preventive knowledge. Fewer respondents considered the app to be a useful instrument in decreasing substance use or promoting exercise. When assessing the connection between the use of the app’s distinct modules and the main outcome measures, results of the Spearman rank correlation analysis indicated that of the 6 modules of the app, repeated use of the “What if?” roleplay module showed a significant correlation with an increase in psychoactive substance–related knowledge (\(r_{S(35)} = .39, P = .02\)). A significant correlation was also observed between a decrease in negative attitudes toward substance users and the repeated use of the “Where to go?” module (\(r_{S(44)} = .31, P = .04\)). Finally, a significant correlation was found between the app’s perceived general usefulness and a decreasing past month alcohol consumption (\(r_{S(44)} = .32, P = .03\)), indicating that those respondents in the app group who found the app useful, showed lower frequencies of alcohol intake at the time of postdata collection.
This pilot study identified some potential benefits of the Once Upon a High app and also a considerable amount of flaws that need amendments to further increase the utility of the intervention. Our result that alcohol was the most commonly consumed psychoactive substance is in line with the results of epidemiological studies with respect to the substance use characteristics of the adolescent population [40]. Although we did not assess a representative sample, average lifetime use rate of NPS-type drugs was also comparable with the findings of the ESPAD study. Synthetic cannabinoids were the most popular NPS, especially in one of the vocational schools (VS2). Students of the vocational schools showed higher frequencies of tobacco, ecstasy, and MDPV use as well as higher scores on the CAST and FTND, indicating more severe problems associated with their substance use. As O’Malley and colleagues [66] emphasize in their study, school types and related socioeconomic status may have a significant impact on the students’ psychoactive substance use frequencies. Considering the assessed sample, vocational students indeed reported lower living standards than high school students.

With regard to the secondary outcome measures of the regression models, results indicate that a decrease in negative attitudes toward substance users might be a risk factor for an increasing past month alcohol consumption. This finding suggests that the subgoal of the app to destigmatize the population of substance users (those who consume either legal or illegal substances) might miss the aim: instead of facilitating social tolerance, it may indirectly promote an accepting attitude toward substance use itself. According to Hohman and colleagues [67], attitudes toward psychoactive substances—especially ambivalent attitudes—can affect substance use behavior. An increase in past month energy drink consumption was associated with higher frequencies of exercise. This result is in line with the findings of Larson and colleagues [68], namely, that pursuing sports or energy drink consumption is often related to higher rates of participation in organized sport activities among adolescents; thus, exercise is not exclusively a protective factor with respect to any substance use.

On the basis of both the respondents’ evaluation on the utility of the distinct modules and the pre-and postcomparative results of our pilot study, the Once Upon a High app might be overall ineffective in decreasing psychoactive substance use (with the exception of energy drink consumption), increasing exercise frequencies, or enhancing self-efficacy at least under the conditions of this study. When we, however, assessed repeated use of distinct modules of the app (and not the app in general), the What if? module showed significant correlation with an increase in substance-related knowledge, whereas the use of the Where to go? module was associated with decreased negative attitudes toward substance users. The What if? module was also one of the most preferred parts of the app by the responding students. As the “What if?” module contains gaming elements, the fact that it was found to be the most effective and likeable part of the app, once again highlights the benefits of gamification. Johnson and colleagues [69] also conclude that as a result of their systematic review, gamification can have a positive impact on health behaviors, although current literature lacks efficacy studies with respect to the effects of gamification on substance use. Nevertheless, Fleming and colleagues [70] also emphasize that gamification might have considerable potential for increasing the efficacy of mental health–related Web-based interventions.

In terms of hypothesis testing, our study could only partially confirm the H3 hypothesis, as the app appeared to be effective in decreasing energy drink consumption but not in reducing other substance use frequencies. Considering the emerging concerns about the harmful effects of energy drinks with high content of caffeine, sugar, or other ingredients (eg taurine or guarana) and the increasing prevalence of energy drink consumption among adolescents [71], this result might still be relevant. A correlation between the app’s perceived usefulness and a decreasing frequency of past month alcohol use was also found, indicating a more motivated subgroup of respondents.
who not just downloaded the app but also found it a utilisable tool. It further needs to be addressed that the majority of the respondents did not show severe (or even moderate) substance use problems as indicated by the cut-off scores of AUDIT, CAST, sCAST, and FTND. Furthermore, those students who dropped out from the study, showed higher frequencies of lifetime and past month substance use. Therefore, it is likely that the app could not exert its preventive effects on those students who would have needed it the most. This high-risk subgroup mainly consisted of males and older students. Assessing a sample consisting of adolescents with problematic substance use might better fit the scope of secondary prevention. On the basis of the characteristics of our sample (ie, the majority did not show at risk or hazardous substance use), it was not entirely possible to draw valid consequences in terms of the app’s secondary prevention efficacy. Although a significant increase in substance use was not found among the app users, the possibility of iatrogenic harms cannot be neglected in case of the app, as it was already noted in case of other prevention programs as well [72,73]. As Dishion and colleagues [72] pointed out, such harms might be dependent on many factors, including the adolescent target group’s homogeneity (eg, containing only at-risk individuals) or the specific phase of adolescence (eg, early vs late adolescence) at the time of the intervention.

Limitations
Our pilot study is not without limitations. We did not use a randomized controlled design. The use of the app was not monitored coherently, even if respondents had to provide answers about their experiences with the app. The sample size for the applied postanalyses—because of the relatively high dropout rate—might be considered to be small for reliable multivariate analysis. Furthermore, a nonprobability, convenience sampling method was used that doubtlessly reduced the external validity of the study. The short period of follow-up time between T0 and T1 occasions (ie, 2 months) could also encumber the valid measurement of change, particularly in terms of substance use. Behavioral and cognitive change can be perceived as the result of a complex process with many unassessed factors in the background, including, but not limited to, family dynamics, peer pressure, other sources of psychoactive substance–related information (eg, internet fora), or motivation. For similar reasons, full causality cannot be assumed in our results.

These first experiences with the Once Upon a High app highlight the relevance of the target population’s motivation or the lack of it. We tried to motivate participants by offering them free recreational activities; however, this strategy failed to fulfill its goals. The offered possibility of costless admission to gyms, paintball facilities, or indoor climbing centers, did not show association with an increased frequency of daily exercise, and the dropout rates were still high. Therefore, other strategies need to be tested in the future with respect to their efficacy in motivating adolescents to maintain the use of this or a similar app. As adolescents susceptibly react to peer opinion and judgments, effective dissemination of such an app might be crucial as it may form its perception in the target population. If prevention apps can be presented as exciting or even fashionable tools that promote and help the self-management of various health conditions, the motivation of adolescents to use and share these might increase. As far as we are concerned, tools of extrinsic rewards (eg, money or presents) do not come into question in case of such prevention programs. On the basis of the findings of El-Hilly and colleagues [74], it is intrinsic motivation that may function as a positive drive to engagement in a prevention program and to a positive health behavior as well, whereas external social influences might rather have a negative impact on the overall efficacy of gamification. Similar results were published by Habgood and Ainsworth [75], indicating that in-game intrinsic motivators are more effective than extrinsic ones in case of educational games.

Conclusions
To conclude, Once Upon a High app can be a useful and feasible tool to assist effective preventive intervention programs, as it was originally developed for such purposes as well, not necessarily as a stand-alone prevention instrument. It is likely that the beneficial effects of the app can be maximized when its use is supplemented with personal discussion. As Majeed-Ariss and colleagues [10] emphasize, similar apps that support personal self-management of chronic health conditions are usually implemented with the involvement of health professionals. School psychologists, health visitors, or social workers can also be involved in such projects to improve the effects of the app as well as to increase the target population’s commitment to use it.

Further research is needed to assess additional potentials of the app, and it is also necessary to find better ways to motivate the target population to repeatedly use this tool. Preventive effects of the app on substance use itself also need to be explored by using probability sampling and larger effect sizes and increasing the heterogeneity (eg, cross-cultural differences) to find established ways to improve the app so that it can be really called and function as a “prevention” app in the end. This pilot study might function as a basis for future research with such purposes.

Acknowledgments
The development of the app was funded by the Hungarian Ministry of Human Capacities (grant number: KAB-ME-C-22756). The study was partially supported by the Hungarian National Research, Development and Innovation Office (Grant number: K111938). The work was completed in the ELTE Institutional Excellence Program (783-3/2018/FEKUTSRAT) supported by the Hungarian Ministry of Human Capacities. The funding institution had no role in the study design or the collection, analysis, and interpretation of the data, writing the manuscript, or the decision to submit the paper for publication. The authors would like to thank the following colleagues for their immense work during the development of the app: Mónika Gyuró (graphic design),...
Dániel Karasz (programmer), Klára Berczi (project management), and Lívia Lukács (creating the app’s logo). The authors also wish to thank members of the co-design group for their contribution.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Technical details of development.

[PDF File (Adobe PDF File), 251KB - games_v6i4e19_app1.pdf]

Multimedia Appendix 2
Pre-and posttest description in point of the outcome measures.

[PDF File (Adobe PDF File), 29KB - games_v6i4e19_app2.pdf]

References


Abbreviations

AUDIT: Alcohol Use Disorder Identification Test
CAST: Cannabis Abuse Screening Test
FTND: Fagerström Test for Nicotine Dependence
GHB: gamma-hydroxybutyrate
LSD: lysergic acid diethylamide
MDPV: methylenedioxypyrovalerone
NPS: novel psychoactive substance
sCAST: Cannabis Abuse Screening Test adapted for synthetic cannabinoids
SUD: substance use disorders
UID: unique identifier

©Máté Kapitány-Fövény, Eszter Vagdalt, Zsófia Ruttkay, Róbert Urbán, Mara J Richman, Zsolt Demetrovics. Originally published in JMIR Serious Games (http://games.jmir.org), 04.12.2018. 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 Mobile App Delivering a Gamified Battery of Cognitive Tests Designed for Repeated Play (OU Brainwave): App Design and Cohort Study

Martin Thirkettle1, PhD; Jennifer Lewis2, PhD; Darren Langdridge3, PhD; Graham Pike3, PhD

1Centre for Behavioural Science & Applied Psychology, Sheffield Hallam University, Sheffield, United Kingdom
2School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom
3The Open University, Milton Keynes, United Kingdom

Corresponding Author:
Martin Thirkettle, PhD
Centre for Behavioural Science & Applied Psychology
Sheffield Hallam University
Heart of the Campus Building
42 Collegiate Crescent
Sheffield, S10 2BQ
United Kingdom
Phone: 44 01142256718
Email: m.thirkettle@shu.ac.uk

Abstract

Background: Mobile phone and tablet apps are an increasingly common platform for collecting data. A key challenge for researchers has been participant “buy-in” and attrition for designs requiring repeated testing.

Objective: The objective of this study was to develop and assess the utility of 1-2 minute versions of both classic and novel cognitive tasks using a user-focused and user-driven mobile phone and tablet app designed to encourage repeated play.

Methods: A large sample of app users (N=13,979 at first data collection) participated in multiple, self-paced sessions of classic working memory (N-back), spatial cognition (mental rotation), sustained attentional focus (persistent vigilance task), and split attention (multiple object tracking) tasks, along with the implementation of a comparatively novel action-learning task. The “OU Brainwave” app was designed to measure time-of-day variation in cognitive performance and did not offer any training program or promise any cognitive enhancement. To record participants’ chronotype, a full Morningness-Eveningness questionnaire was also included, which measures whether a person’s circadian rhythm produces peak alertness in the morning, in the evening, or in between. Data were collected during an 18-month period. While the app prompted re-engagement at set intervals, participants were free to complete each task as many times as they wished.

Results: We found a significant relationship between morningness and age (r=.298, n=12,755, P<.001), with no effect of gender (t13,539=1.036, P=.30). We report good task adherence, with ~4000 participants repeatedly playing each game >4 times each—our minimum engagement level for analysis. Repeated plays of these games allowed us to replicate commonly reported gender effects in gamified spatial cognition (F1,4216=154.861, P<.001, ηp²=.035), split attention (F1,4185=11.047, P=.001, ηp²=.003), and sustained attentional focus (F1,4238=15.993, P<.001, ηp²=.004) tasks. We also report evidence of a small gender effect in an action-learning task (F1,3988=90.59, P<.001, ηp²=.022). Finally, we found a strong negative effect of self-reported age on performance, when controlling for number of plays, in sustained attentional focus (n=1596, F6,1595=30.23, P<.001, ηp²=.102), working memory (n=1627, F6,1626=19.78, P<.001, ηp²=.068), spatial cognition (n=1640, F6,1639=23.74, P<.001, ηp²=.080), and split attention tasks (n=1616, F6,1615=2.48, P=.02, ηp²=.009).

Conclusions: Using extremely short testing periods and permitting participants to decide their level of engagement—both in terms of which gamified task they played and how many sessions they completed—we were able to collect a substantial and valid dataset. We suggest that the success of OU Brainwave should inform future research oriented apps—particularly in issues of balancing participant engagement with data fidelity.
**Introduction**

Recent advances in the performance and accessibility of Web technologies have resulted in increasing use of Web platforms to conduct cognitive psychology research. Large, diverse cohorts easily available to researchers are now accompanied by platforms capable of implementing complex tasks and accurately measuring performance [1,2]. Moving on from Web-based data collection, possibilities offered by custom built, natively coded, mobile apps include high levels of stimulus control and enormous flexibility in experimental design and data collection—both frequency of data-collecting sessions and range of data collected [3]. By collecting large sets of cognitive performance data, insights into subtle variations in cognition, both within an individual, as here, or across individuals and cultures [4], are potentially available to researchers. The aspects of tasks included here are prevalent in many everyday skills and activities—from paying attention to all potential threats when crossing a busy road (Track—multiple object tracking), to packing a suitcase efficiently (Spin—mental rotation). Understanding cognitive performance is hugely important. Even if we consider only healthy mental function, only by understanding our cognition’s fundamental properties can we design our lives [5], work [6,7], and play [8] to enable our own best performance [9,10].

A key issue for all psychology researchers is recruiting participants. While laboratory-based studies can often rely on departmental participation requirements to ensure a steady flow of—debatably—willing participants, the sample obtained is inevitably limited in demographic factors [11]. Web-based and app-based studies are one possible way of researching with a broader sample of participants, but to achieve this, researchers must ensure their task, or request, is an engaging one, especially if it requires repeated testing sessions for data collection. Embedding the experimental collecting task within an engaging, fun-to-play game is an increasingly popular way of trying to improve participant engagement and retention. A recent systematic review of gamification of cognitive tasks suggested increased engagement as one of the main reasons for gamification [12]. Moreover, this review highlighted additional benefits of gamification, such as reducing anxiety and extending the investigator’s reach, while underlining the potential that gamification has to improve data collection without necessarily impairing data’s validity.

OU Brainwave is a bespoke app, launched on multiple platforms, designed to collect research data while providing participants with understandable measures of their performances across 5 facets of cognitive ability. The app includes gamified tasks designed to measure performance on aspects of working memory, spatial cognition, sustained attentional focus, split attention, and action learning. Importantly, neither did we set out to “train” participants in any of these aspects of cognitive ability nor does the app make any promise of improvement to cognitive performance through repeated play. Instead, the app seeks to measure natural variation in performance on such tasks throughout the day [13] and in relationship to an individual’s sleep-wake cycle [14]. The app also aims to utilize a large-scale sample to answer the question of whether such variations are related to an individual’s Morningness-Eveningness score—ie, whether “Larks” perform better earlier in the day than “Owls” who perform better later [15].

Here, we present in-game data from the app, report our cohort’s broad performance across the 5 tasks in relationship to the respective task literatures and any relationships between demographic factors and performance, and discuss broader issues of gamification and task design for use in app-based testing.

**Methods**

**OU Brainwave**

The OU Brainwave app was designed and created in collaboration with an external developer (Conjure Ltd, London, UK). Each game included in the app went through numerous rounds of development, with usability and participant engagement given equal weight for essential factors of data validity and experimental design. The app was launched on Android and iOS mobile phone platforms in February 2015. The launch was publicized through blog posts and traditional media coverage. In addition, participants were encouraged to publicize the app through an built-in function of sharing a graph of their individual results to social media.

Ethical approval for this study was obtained from the Open University Human Research Ethics Committee. Immediately upon downloading and opening the app, participants were presented with an informed consent statement, with which they were required to agree via a tick box to continue using the app. Once consent was received, a unique participant ID for each participant was generated to link participant and future session data. Should individual participants wish to withdraw their consent at a later date, they could do so through a settings screen. Doing so deleted all participant data on the device and returned the participants to the app’s opening screen, where they had to agree again to the consent statement to reuse the app. At no point was any personal or potentially identifying information collected from the participants.

Participants entered simple demographic information: gender (male or female) and age in years although participants could choose not to answer either of these questions. Participants then completed the 5-item Morningness-Eveningness self-report questionnaire (MEQ) [16]. The MEQ is a well-established and validated research tool which measures whether a person’s circadian rhythm produces peak alertness in the morning, in the evening, or in between [13,17,18], and the 5-item variation of the original questionnaire was used here to move participants to the more interactive aspects of the app as quickly as possible.
Using the original scoring of this MEQ implementation, participants were coded into one of 5 types ranging from “strongly morning type” to “strongly evening type.” This result was shared onscreen with the participants, and to encourage continued and repeated participation, they were then prompted to continue to the games to “see if your performance matches your belief.”

The app also attempted to ameliorate high attrition rates from which mobile phone apps suffer by displaying the participant performance graph only once the participant had completed 3 sessions. This was made clear to participants each time they used the app until they had completed this requirement, at which point a graph of their performance, on each of the games and as an aggregate score, was shown. These graphs were designed to show participants variation in their performances on tasks across day and night, rather than to reveal absolute performance levels. As such, the performance values were normalized for each participant to highlight the best and worst scoring sessions. Accompanying the presentation of these graphs were icons encouraging the participant to share the image on social media.

At the beginning of each session (ie, on each subsequent launching of the app), participants were also asked up to 3 additional questions. A single item of mood rating [19] was included at the start of every session, “How is your mood right now?” to which they responded via a visual analog scale (VAS) slider. If a session was the first on a given day, participants were also asked what time they had woken from sleep and how many hours of sleep they had had the previous night. Participants could opt to skip answering these questions and continue to games. Each session comprised all 5 games, which were presented in a randomized order. Participants could choose to skip any game during a session, but were encouraged to complete them all through game-by-game results graphs of their individual performance within the app. These graphs were shown only after 3 full sessions to encourage a minimum level of engagement and were updated with each play after this point to promote continued play.

Games

Hotspot: Action Acquisition Task

The Hotspot game is a variation on an action discovery and acquisition task [20]. In this task, participants must discover a target area by tilting their phone or tablet to roll an onscreen ball into a target area (Figure 1). The target area is unmarked, and no feedback is given until the target area is “discovered” by the participant rolling the ball over the area, at which point the ball’s color changes. Participants must then use this color change to guide them in bringing the ball to rest within the target area. Task difficulty was adjusted in development by including a 100-millisecond delay between success (ie, entry into the target area) and a feedback signal (ie, ball’s color change). The effect of delays of this type and magnitude is to increase task difficulty [21] and was intended to prevent ceiling effects among app participants. The game consisted of 5 attempts, and each attempt presented a new, randomly chosen target area covering 5% of the game arena’s total space, with the ball covering half of that area. To succeed in the game, participants had to keep the ball within the target area for 500 milliseconds of a 1-second window. Scores were allocated so that 50% of points available were awarded for finding the target, and the remainder were apportioned according to milliseconds elapsed before the ball remained within the target area for the required time.

Figure 1. Screenshots of OU Brainwave app, showing Hotspot game instructions and play screen.
**React: Persistent Vigilance Task**

The React game was intended to be an implementation of the psychomotor vigilance task [22]. During the design process, it was decided to adjust how the task was operationalized in the game to try to increase participant engagement. This was done by including a simple choice element, which was in addition to the reaction time task and not a standard part of the classic psychomotor vigilance task. Participants were presented with 4 large, red, circular buttons (Figure 2). At a random interval from 2 to 7 seconds, one button changed color to green, and the participant had to tap the appropriate button within a 600-millisecond window. Auditory and visual feedback was given on both correct and incorrect responses. This was repeated 8 times. Participant scores were essentially simple reaction time measures, with scores reducing according to milliseconds elapsed before a correct response was recorded, after a 100-millisecond grace period. Responses made before the color change or incorrect button presses scored zero.

**Spin: Mental Rotation Task**

Spin is a gamified implementation of a spatial rotation task, using the stimulus set developed by Bethell-Fox and Shepard [23], shown in Figure 3. This stimulus set contained 18 possible patterns of filled squares within a 3×3 grid, avoiding excessive simplicity or difficulty and rotational symmetry of pattern. Each pattern contained 1, 2, or 3 groups of filled squares within the grid. While the original paper split these into levels of difficulty, all 18 patterns were presented here in a random order within a given session to provide the participant variation. In the Spin game, participants were required to match the test image with 1 of the 3 options. Feedback was given in the form of ticks and crosses in circles at the top of the screen, and a timer was shown with the remaining time for the task.

Participants were presented with a large image of the target grid and had to correctly identify the rotated version of this grid from 3 alternatives presented below. The correct version was rotated at random by 90, 180, or 270 degrees. Incorrect options consisted of the test pattern reflected either vertically or horizontally. Participants had 45 seconds to make as many correct judgments as they could, up to a maximum of 18, and correct or incorrect auditory and visual feedback was given after each response.

**Figure 2.** Screenshots of OU Brainwave app, showing React game instructions and play screen.
Super Snap: N-Back Analog

A simple implementation of the classic N-back task was used as a test of working memory (Figure 4). The N-back task has a long history of use in studies of working memory [24], see Kane [25] for a detailed discussion of the N-back’s construct validity. Here, a series of 6 brightly colored shapes (ie, circle, hexagon, rhombus, square, star, and triangle) were shown onscreen, and participants had to tap the screen to mark when the current shaped matched with previously shown two shapes. Each shape was presented onscreen for 1.5 seconds against a blank black background with an interitem delay of 1.5 seconds. Auditory and visual feedback was given after each response, along with a tick or cross at the top of the gameplay window. Participants were scored by the number of correct responses, and the game continued until 10 matches had been presented or 10 responses (including false alarm incorrect responses) had been made. Participants started each session with a score of 60 and lost 6 points for each incorrect response or miss recorded.

Track: Multiple Object Tracking

The Track game (Figure 5) is a gamified version of a multiple-object tracking task [26]. In this task, participants had to track the location of 3 members of an array of identical moving balls. A subset of balls onscreen was highlighted before the start of the trial, before reverting to white once the trial started and all balls began to move. Participants were first shown a static array of 8, 9, or 11 balls, 3 of which were highlighted in pink, rather than the color white of the other balls. After a 3-second countdown, highlighted balls reverted to white, and all balls began moving on independent, randomly assigned trajectories. Each ball’s speed and direction of movement was adjusted randomly between each frame, and collisions between balls or borders were handled such that no ball was ever overlapped or exceeded the playing area. The balls continued in motion for 5 seconds, after which time the entire array stopped, and participants were instructed to tap the 3 balls that had been highlighted at the start of the trial. Two trials of each array size were shown, with set sizes presented in an increasing order. Participants were scored on the number of balls correctly identified with nonresponses counted as incorrect. Each correctly identified ball added a score of 2.5, so a maximum score of 45 across 6 trials was possible. Auditory and visual feedback was given after each trial, along with a tick (for correct identification of all balls) or a cross for each trial, along the top of the gameplay area.
Figure 4. Screenshots of OU Brainwave app, showing Super Snap game instruction and play screen.

Figure 5. Screenshots of OU Brainwave app, showing Track game instruction and play screen.
Results

Demographics

Number of Downloads and Participants

The OU Brainwave app was launched on both Apple and Android stores on January 15, 2015. Discovery of the app peaked in its launch month, with 4394 installations during January, with the expected drop-off of installations broken only by smaller peaks in April 2015 and January 2016—both probably due to further publicity. Again, like many apps, OU Brainwave found far more users on the Apple platform than on the Android platform—with roughly two-thirds of 15,890 users across the 18.5-month data collection window using Apple devices.

Separately, and far more importantly than the raw number of downloads, is the number of active app users. As with many mobile phone apps, many downloaders either did not open the app or did not engage with the app sufficiently to be considered active users.

Of 15,890 installations between app launch and July 31, 2016, 13,979 used the app sufficiently to upload some data to the server, meaning almost 2000 downloaders did not open the app after installation. Engaging with the app for a single session only—the most popular decision among downloaders—were 3661 users, contributing at most demographic and MEQ data along with a single session's play to the dataset. Attrition among the remaining 10,319 was predictably steep, with only 5756 users playing for >3 sessions, dropping to 1435 users at ≥10 sessions. Although no contact information was deliberately collected, so precluding any survey of participants who stopped early, potential disincentives may have included technical issues, particularly with the Android app, “pestering” by app notifications, or a perception that the app’s demands were too high.

Just over 1000 users played ≥12 sessions, and just over 100 played ≥30 sessions. Additionally, each user was free to play 1, some, or all 5 games during a given session, so while 3556 users completed 5 plays of any single game, only 2780 completed 5 plays of all 5 games. For this reason, later analyses were conducted on a game-by-game level, and no overall performance measure was calculated.

Demographics of Total Participant Cohort

Of 13,979 total participants, 39.47% (5517/13,979) self-reported as male and 58.88% (8231/13,979) as female, with only 1.65% (231/13,979) declining to answer the gender question. Self-report ages of all participants are shown in Figure 6.

Distribution of participants' reported ages shown in Figure 6 reveals a shortcoming of implementation of self-report demographic data collection. While 1033 declined to answer the age question, an exceptionally large number of participants reported their age as 18 years (965, compared with 405 for 27 years, the next most common age). While this might well be an accurate figure, it could potentially be an artifact caused by the requirement for participants to confirm they are over 18 years to use the app and play the games—a stipulation necessary for ethical approval.

**Figure 6.** Age of participants by gender.
Therefore, 18 years was the lowest selectable age in the self-report question. Unless age controls on downloading and installation of apps—controlled by the developer or app store rather than the end user (or parent)—become a viable option, future apps, especially particularly gamified ones, may consider collecting and subsequently discarding (or filtering not to upload) data from particular age groups, rather than attempting to exclude by self-report of age.

**Morningness-Eveningness Questionnaire**

The 5-item MEQ [16] that each participant completed produces a score from 4 to 25, ranging between extreme evening type and extreme morning type. Respondents are then traditionally classified into 5 classes by their scores (Definitely evening type: 4-7; moderately evening type: 8-11; neither type: 12-17; moderately morning type: 18-21; and definitely morning type: 22-25). From our original sample, 13,752 participants filled in all sections of the MEQ survey. As is common in studies using the MEQ, approximately half of our overall sample scored within the “neither type,” central range of the MEQ (7172/13,752, 52.2% participants). A further third of participants scored in one of the evening type categories (4584/13,752, 33.3% participants), with the remaining 14.5% (1996/13,752) scoring in the morning type categories.

One of the stronger relationships usually found by the MEQ is that between age and morningness [18,27]—greater age is associated with greater morningness scores—and as Figure 7 shows, we replicated that finding here. We found a statistically significant correlation between age and MEQ score for participants who submitted both age and MEQ data ($r = .298$, $n = 12,755$, $P < .001$). The greater absolute number of evening types than morning types in the dataset is almost certainly a result of this relationship being expressed in our cohort, which was skewed toward younger participants.

As Figure 8 shows, there was no significant difference between MEQ scores for males and females: mean 13.20 (SD 4.00) versus mean 13.28 (SD 3.99); $F_{1,5399} = 1.036$, $P = .30$. While many studies have reported greater propensity for evening types in males than in females [17], lack of gender differences reported here is not an uncommon finding in the literature [18].

**Comparison of Results to Previous Research and Demographic Effects**

Because each participant could choose to play separate games individually, any 2 participants potentially played any given game a different number of times. Additionally, the full participant set includes participants who did not play a particular game sufficiently to become familiar with it. This makes handling data generated by the app very different from handling the usual data generated by tasks in OU Brainwave when conducted in a laboratory setting. To address the issue of participants who did not engage with a task sufficiently even to familiarize themselves with it, we implemented a cutoff minimum of 4 plays of each game for any participant to be included in analysis for that game. The intention was to remove participants who played no more than what would be considered a “practice trial” set in a laboratory-based experiment. However, this still leaves variability in the number of measures per participant (in terms of sessions played) and the possibility that those participants who played more would register a higher mean score on each game. Therefore, the effect on score of demographic variables was analyzed with the number of plays as a covariate. To remove outlier individuals, the most extreme 1% of average performance scores were identified and excluded before all analyses.

**Other Effects**

**Gender**

Gender effects were analyzed using analysis of covariance with the mean score of the participant as the dependent variable, gender as a fixed factor, and the number of plays as a covariate. To ensure gender and number of plays were not confounded, a $t$ test was conducted to confirm no significant difference between the number of plays of a particular game by each gender. For each of the 5 games, this test was nonsignificant. Table 1 shows the mean number of plays across gender for each game and the associated $t$ test statistic.

**Track: Multiple Object Tracking**

For the Track game, 4188 participants (1455 male and 2733 female) completed ≥4 sessions. Previous studies in using multiple object-tracking paradigms have shown an advantage for male participants [28], and we replicate that finding here. There was significant, but small, effect of gender on performance after controlling for number of times beyond 4 that each participant played the game ($F_{1,4188} = 11.047$, $P = .001$, $\eta^2_p = .003$). Male participants scored on average 1.2 points more than female participants: male: mean 38.9 (SD 12.0) versus female: mean 37.7 (SD 11.7). Although, it should be noted that variance in score accounted for by gender is very small (~0.3%) and only roughly half that accounted for by the significant effect of the number of plays beyond 4 by a particular participant ($F_{1,4174} = 27.524$, $P < .001$, $\eta^2_p = .007$).

**Super Snap: N-Back Analog**

For the N-back analog Super Snap, 4215 participants (1449 male and 2766 female) completed ≥4 sessions. We found no effect of gender on performance after controlling for the number of plays beyond 4 ($F_{1,4212} = 2.711$, $P = .1$, $\eta^2_p = .001$), although the number of plays beyond 4 was still found to improve performance significantly ($F_{1,4227} = 127.06$, $P < .001$, $\eta^2_p = .03$).

**React: Persistent Vigilance Task**

For React, 4241 participants (1460 male and 2781 female) completed ≥4 sessions. We found a small, but significant effect of gender on performance after controlling for the number of plays beyond 4 ($F_{1,4238} = 15.993$, $P < .001$, $\eta^2_p = .004$), accounting for 0.4% of variance in mean score. Male participants scored an average of 4 points more than female participants: male, mean 409.0 (SD 29.6) versus female, mean 405.1 (SD 29.9). The number of plays beyond 4 was not found to improve performance significantly ($F_{1,4238} = 1.264$, $P = .26$).
Figure 7. Morningness-Eveningness self-report questionnaire (MEQ) scores by age.

Figure 8. Morningness-Eveningness self-report questionnaire (MEQ) scores (color coded by MEQ category) by gender.

Table 1. Mean number of plays by game and gender.

<table>
<thead>
<tr>
<th>Game</th>
<th>Number of plays, n</th>
<th>Number of plays, mean (SD)</th>
<th>t test</th>
<th>t value</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>t value</td>
</tr>
<tr>
<td>Hotspot</td>
<td>1403</td>
<td>2640</td>
<td>8.07 (8.56)</td>
<td>8.36 (7.12)</td>
<td>t_{4041}=-1.150</td>
</tr>
<tr>
<td>React</td>
<td>1495</td>
<td>2840</td>
<td>8.09 (8.52)</td>
<td>8.32 (7.16)</td>
<td>t_{4333}=-0.933</td>
</tr>
<tr>
<td>Super Snap</td>
<td>1459</td>
<td>2773</td>
<td>8.13 (8.52)</td>
<td>8.42 (7.31)</td>
<td>t_{4230}=-1.175</td>
</tr>
<tr>
<td>Spin</td>
<td>1475</td>
<td>2784</td>
<td>8.16 (8.61)</td>
<td>8.40 (7.36)</td>
<td>t_{4257}=-0.932</td>
</tr>
<tr>
<td>Track</td>
<td>1457</td>
<td>2735</td>
<td>8.12 (8.57)</td>
<td>8.38 (7.30)</td>
<td>t_{4196}=-1.038</td>
</tr>
</tbody>
</table>
Spin: Mental Rotation Task

For Spin, 4219 participants (1455 male and 2764 female) completed ≥4 sessions. We found significant effect of gender on performance after controlling for the number of plays beyond 4 ($F_{1,4218}=154.861, P<.001, \eta^2_p=.035$), accounting for 3.5% of variance in score. Male participants scored on average 3 points more than female participants: male, mean 23.6 (SD 7.7) versus female, mean 20.6 (SD 7.5). The number of plays beyond 4 was still found to have significant, albeit smaller, benefit to performance ($F_{1,4216}=60.45, P<.001, \eta^2_p=.014$).

Hotspot: Action-Learning Task

A total of 3991 participants completed ≥4 sessions of Hotspot (1393 male and 2598 female). We found a small, but significant effect of gender on performance after controlling for the number of plays beyond 4 ($F_{1,3988}=90.59, P<.001, \eta^2_p=.022$), with male participants scoring on average 3.7 points more than female participants: male, mean 31.33 (SD 12.25) versus female, mean 27.64 (SD 12.03) and accounting for 2.2% of variance in mean scores across the cohort. However, the number of plays beyond 4 was also found to have significant effect of similar size on performance ($F_{1,3988}=105.946, P<.001, \eta^2_p=.026$).

In summary, of the 5 games, all but the N-back task showed significant effect of gender, after controlling for the number of plays beyond 4. In each game that showed an effect, male participants scored higher (React, Track, Spin, and Hotspot), with the strongest effect in the mental rotation-based game Spin.

Age Effects

For every game, there was significant positive correlation between the age of the participant and the number of plays (React: $r=4171, r=.216$; Super Snap: $n=4081, r=.221$; Spin: $n=4100, r=.216$; Hotspot: $n=3896, r=.215$; Track: $n=4035, r=.266$, all significant at $P<.001$). Older participants played far more sessions than their younger counterparts, possibly due to the cohort’s self-selecting nature. Demographically, older people are less likely to either have a mobile phone or use a mobile phone for playing games [29], so it may be the case that for an older person, downloading the app represented greater commitment to engage with it. With such a confounding relationship between the number of plays and age, it was not appropriate to adopt the same approach—analysis of covariance using mean score for each participant—as to analyze gender. Instead, a stratified approach was used to analyze the effect of age on performance, controlling for the number of plays. Rather than calculating mean scores from all sessions of a participant, the mean score for each participant from only their 4th, 5th, and 6th sessions was calculated. These early snapshots provided a measure of each participant’s performance before going on to complete differing numbers of sessions and attaining an eventual average performance level.

Using this measure, we then broke participants into 7 age groups by decade (<20, 20-29, 30-39, 40-49, 50-59, 60-69, and >70). This allowed us to analyze any potential age effects with analysis of variance and to compute effect size for each game. We found significant effect of age in all games except Hotspot. Results from React ($n=1596, F_{6,1595}=30.23, P<.001, \eta^2=.102$); Super Snap ($n=1627, F_{6,1626}=19.78, P<.001, \eta^2=.068$); and Spin ($n=1640, F_{6,1639}=23.74, P<.001, \eta^2=.080$) all align with previous findings that show strong negative associations between age and reaction time [30], age and working memory as measured by the N-back task [31], and age and mental rotation [32]. The Track game, as an implementation of the multiple object-tracking paradigm, might have been expected to similarly replicate a previously found age effect for multiple object tracking [33]. While we did find significant main effect of age on multiple object-tracking performance, it was much smaller than that for React, Super Snap, and Spin ($n=1616, F_{6,1615}=2.48, P=.022, \eta^2=.009$). There was no relationship between age and performance on the action discovery task—the Hotspot game ($n=1519, F (6,1518) = 1.78, P=.10, \eta^2=.007$). These analyses were also conducted with a correlational approach, and the pattern of results was broadly similar.

Mood

Before starting each session of games, participants were presented with a single mood rating to which they responded via a VAS slider. These additional data were collected to investigate the relationship between mood and cognition. Previous research has suggested a complex relationship between emotions, mood, and performance on cognitive tasks [34,35], with both beneficial and detrimental impacts on cognitive performance reported from a single, for example, positive mood induction [36]. Here, rather than inducing a given mood, we simply recorded the participant’s self-report of pre-existing mood, captured by a single item VAS measure (1-10, 10 being the happiest) [19].

To be as inclusive as possible, all participants completing ≥4 plays of each game were included in this analysis and collapsed across age and gender, giving group sizes from 3988 (Hotspot) to 4331 (React) for each of the 5 games. No strong relationship between participants’ average mood and average performance was found for any of the games, with all Pearson correlation coefficients for every participant were calculated individually for each game, no relationship was found, with the mean rho value being <.02 in all cases, and for Super Snap, Spin, and Track, not significantly different from 0. While future analysis may explore the possibility of nonmonotonic relationships between mood and performance or potential differences in subgroups of the cohort, this first analysis suggests that either a single item VAS measure is insensitive to impactful changes of mood or that mood and performance were unrelated on any task in the app.

Practice Effects and Learning Curve Analysis

Each participant’s freedom to play each game uncontrolled number of times and the likelihood that this would affect an individual’s performance mean that either including number of plays as a covariate or controlling this factor in analysis is the most effective way to address a gamified testing platform’s individual freedom. However, visualizing at the group level, the effect of number of plays on performance is still possible.
Plotting mean scores for all participants in each game, except for the React game, shows typical practice effects as participants familiarized themselves with tasks. Practice effects are very common in psychological and psychometric testing and almost certainly reflect some combination of task familiarization, development of ability tested by the task, development of a strategy to complete the task as set, and possibly, reduced anxiety about the task’s mechanics [37].

While 4 of the 5 games showed typical practice effects—a stabilizing of performance following an initial rise [38]—the React game, being essentially a very basic reaction time task, seems to have been too simple to produce any practice effect-driven improvement in performance across participants’ first few sessions (see Figure 9), probably because participants immediately familiarized themselves with the task on the first play, and no effective strategies can be adopted to improve performance. However, variability of the cohort as a whole—in terms of interplay interval, age, gender, and MEQ—mean that further group analysis of practice effects is unnecessary. The presence of expected stabilization of performance after a number of plays reflects the app’s intention to measure variations in performance, rather than to train or improve participants’ abilities. Most importantly, it means that future analysis of within-individual factors (eg, time-of-day of play) should have a stable performance level from which to contrast such changes.

Discussion

Success of App Approaches

The large cohort collected by the OU Brainwave app, and, moreover, the repeated measurement of this cohort in quick, engaging gamified versions of classic and novel psychological tests, is another demonstration of the promise of mobile app-based testing [39]. While we deliberately did not collect more detailed demographic data, we can safely say that with such a large sample, our testing cohort would have been extremely diverse compared with samples drawn from undergraduate participant pools that typify much laboratory-based research. This, along with the sheer size of the cohort tested, should mean that any reliable findings arising from this dataset are relatively robust and not hostage to cohort effects.

The usual caveats regarding the reliability of self-report data apply to our demographic and MEQ responses [40]. While this has resulted in potential concern regarding the high number of participants self-reporting their age as 18 years, the possibility remains that this is an accurate reflection. Moreover, as a full dataset, the cohort replicates a number of age-related findings, both in increased morningness in older participants and reductions in cognitive task performance. The very low proportion of participants who withheld demographic information—only 231 participants, <2% of our total cohort, withheld either their age or gender information—is very encouraging for future mobile phone-based research, which can expect a high level of engagement from participants who download the app.

OU Brainwave is not the most downloaded research-focused app, and since its release, a number of impressively large datasets have been collected and published using other app platforms [41] and Web apps [42]. However, while OU Brainwave suffered from the same participant attrition as all apps, it recruited an impressively engaged cohort who repeatedly played games (1400 participants played >10 times) even though the games did not vary or become more challenging with continued play. This suggests that some participant engagement features were successful.
A key intention in the development of OU Brainwave was to balance the demands of behavioral experiments, in terms of data validity and operationalization of the mechanism under study, against the enjoyment and engagement of the participants. The high levels of engagement by participants who downloaded the app suggest that an in-game narrative, characters to interact with, or even an elaborate game environment may not be necessary. Studies directly manipulating the extent of gamification have reported similar lack of effect of common gamification techniques on participant attrition [43]. The games, while offering dramatically shorter sessions than one would find in laboratory testing, did not deviate far from their experimental task heritage. Except for the React game, no significant change was made from the mechanics of underlying psychological tasks, and tasks were presented without cutesy preambles, fictional scenarios, or even in-game rewards beyond simple graphing of participant performance. Withholding individual participants’ performance graphs until they had completed 3 sessions likely had the effect of carrying more participants through the steepest part of the attrition and may have contributed to the app’s longevity for these participants. Similarly, the embedded ability to share one’s performance graph—constantly updated with continued play—encouraged both the app’s spread and the individual participant’s continued engagement. Future experimental psychology apps should potentially focus on these features during app design as potentially highly effective, simple tools to encourage participation. On the other hand, our older participants’ tendency to contribute more data in terms of sessions played may suggest this group had greater intrinsic motivation to engage with the app, or at least suggest that participants’ engagement was a function of both intrinsic motivations and app features or in-game mechanics. Future studies may find it valuable to survey users during development to isolate the most valuable engagement features.

Inclusion of the full MEQ [16] and its placement at the start of the app experience meant that morningness data collected provided possibly this early analysis’s strongest finding. Our sample of over 12,500 adults revealed strong evidence for increased age correlating with morningness—with older people being more likely to be moderate or strong morning types than younger people who, in turn, are far more likely to report themselves as moderate or strong evening types. We do not replicate the finding for a similar tendency toward morningness among female respondents compared with male respondents, but as with the strong relationship found with age, this result is in line with previous studies.

Analysis of performance in the 5 cognitive tasks that make up the OU Brainwave app’s games showed that the app produced valid data and is sensitive enough to detect small but significant effects of both age and gender on cognition. We found small effects of gender in 4 of the 5 tasks (React, Track, Spin, and Hotspot) and no effect in the other task (Super Snap). In each case that a difference was found, male participants scored slightly higher on average than female participants. The greatest difference was found in the Spin game, in which gender accounted for 3.5% of variance in average score. The Spin game is a direct implementation of a mental rotation task, which has previously been found to produce large, reliable gender effects [44]. While all gender effects reported here are small, this could well be due to participants’ freedom and resulting noise in the dataset. Furthermore, the unidirectional pattern of gender effects reported here mean that an alternative explanation for these effects of platform (ie, mobile phone app), rather than cognitive task cannot be ruled out.

The impact of age on game score was much more pronounced than that of gender. Here, we reported significant reduction in game scores for older participants compared with younger participants in all but the Hotspot game and comparatively large effects in the Super Snap, Spin, and React games, in which age accounted for 6.8%, 8%, and 10.2%, respectively, of variation in average score. That we found our largest effect of age related to decline in the task most heavily reliant on reaction time is of no surprise—increases in reaction time have long been associated with increasing age [45]. However, evidence we found for adverse effect of age on both mental rotation (Spin) and working memory (Super Snap), which involve more sophisticated constructs than simple reaction time, suggests that the app is indeed sensitive to fine-grained differences in specific aspects of cognitive performance.

Future Research
Future analysis will focus on the effect of time of day on performance. For example, those who report themselves as morning types can be compared with those who report themselves as evening types across each of the 5 tasks, although care must be taken to check and account for any bias induced by MEQ self-reporting before task performance. This approach will enable MEQ scores to be controlled or focused on in an analysis of task scores. Furthermore, recording both hours spent sleeping the previous night and time each participant awoke on each testing day will enable us to analyze the relative contribution of time spent awake, duration of preceding sleep, and time of day on any variation in cognitive performance.

Conclusion
The OU Brainwave app, with its cohort of ~14,000 active participants represents an exciting and rich dataset. User-focused features built into the app—extremely short testing durations, allowing participants to manage their participation, engaging them through in-app feedback on their performance, and encouraging them to become an active part of the recruitment process by sharing their own performance—were largely very successful. The variability this approach introduced into the resulting performance data presented a challenge to data analysis. However, the replication of expected results and the sensitivity of the app to group-level differences in performance reported here all suggest that research apps focusing on user engagement and enjoyment, even at the expense of rigid and rigorous experimental protocols, produce valid and valuable data. Future data-collecting research apps may benefit from a similar focus on participants as users, not just as data points.
Acknowledgments

The app development central to this project was supported by a grant from the Reed Foundation.

Authors' Contributions

MT designed and developed the app, analyzed data, and wrote the manuscript. JL analyzed data and contributed to the manuscript. DL and GP designed and developed the app and contributed to the manuscript.

Conflicts of Interest

None declared.

References


**Abbreviations**

- **MEQ:** Morningness-Eveningness self-report questionnaire
- **VAS:** visual analog scale
Patterns Among 754 Gamification Cases: Content Analysis for Gamification Development

Sungjin Park1*, MEng; Sangkyun Kim1*, PhD
Management of Technology Lab, Department of Industrial Engineering and Energy Research, Kangwon National University, Chuncheon, Republic Of Korea
*all authors contributed equally

Corresponding Author:
Sangkyun Kim, PhD
Management of Technology Lab
Department of Industrial Engineering and Energy Research
Kangwon National University
417, Engineering Building,
Kangwon Daehak Gil 1, Gangwon-do
Chuncheon, KS007
Republic Of Korea
Phone: 82 033 250 6287
Email: saviour@kangwon.ac.kr

Abstract

Background: Gamification is one of the techniques that applies game elements, such as game mechanics and dynamics, to a nongame context (e.g., management, education, marketing, and health care). A variety of methodologies have been published for developing gamification. However, some of these are only usable by people with a certain level of gamification knowledge. People who do not have such knowledge face difficulty in using game mechanics and experiencing enjoyment. To ease their difficulties, a gamification methodology should provide directions for using game mechanics.

Objective: This study aimed at collecting global gamification cases and determining patterns or differences among the collected cases.

Methods: In total, 754 cases were collected based on 4F process elements, such as play type, playful user experience (PLEX)–based fun factors, and game mechanics. In addition, the collected cases were classified into 6 categories. From the data analysis, basic statistics and correlation analyses (Pearson and Kendall) were conducted.

Results: According to the analysis results in PLEX-based fun factors, challenge and completion fun factors formed a large proportion among the 6 categories. In the results of the game mechanics analysis, point, leaderboard, and progress accounted for a large proportion among the 6 categories. The results of the correlation analysis showed no difference or specific patterns in game mechanics (Pearson \(r\)>.8, Kendall \(\tau\)>.5, \(P<.05\)) and PLEX-based fun factors (Pearson \(r\)>.8, Kendall \(\tau\)>.7, \(P<.05\)).

Conclusions: On the basis of the statistical findings, this study suggests an appropriate number of PLEX-based fun factors and game mechanics. In addition, the results of this study should be used for people who do not have gamification knowledge and face difficulty using game mechanics and PLEX-based fun factors.

(JMIR Serious Games 2018;6(4):e11336) doi:10.2196/11336

Keywords
gamification; PLEX; game mechanics; gamification correlation analysis; 4F process

Introduction

Gamification is one of the techniques that applies game elements, such as game mechanics and dynamics, to a nongame context (e.g., management, education, marketing, and health care) for motivation, engagement improvement, and cooperation promotion\[1,2]. Huotari and Hamari\[3\] defined gamification as the process of providing gameful experiences to customers and promoting customer affordance. Gamification can be applied to company operations, human resources training, marketing, and online portal to expand the loyalty, engagement, and participation of employees and customers. In 2012, Gartner Group referred to gamification as a mediator that connects...
Among the collected global gamification cases, some memorable cases are summarized as follows. The IBM (International Business Machines Co) Research Center applied gamification elements, such as point, leaderboard, and feedback, to their in-house social network service (SNS) and conducted before and after comparison studies. The study results showed that the activity in the in-house SNS increased by approximately 2.5 times after the application of gamification [7]. Team Maker (Figure 1) is a case of gamification in the form of a board game developed for leadership training based on Situational Leadership II developed by Blanchard and Zigarmi [8]. Situational Leadership II is a theory indicating that team members (followers) should be classified into 4 categories according to their sociability and job-processing ability, and appropriate leadership should be applied to each type. Team Maker consists of components such as team member cards, team leader cards, and virtual currency. Team member cards are composed of cards reflecting the characteristics of members of South Korean society. Compared with general leader cards, the legend leader cards are cards comprising existing world-famous great persons; these correspond to Situational Leadership II. Before starting the game, the players select team member cards that fit the characteristics of the members of the team to which they belong. After selecting the team member cards, the team leader cards are purchased through auction during the course of the gameplay. The player that selects the team leader card that fits the team member card wins the game. The gameplay takes approximately 60 min. Figure 2 shows the case of The Lost City, which promotes and helps in the understanding of the financial products. The Lost City is a case of marketing gamification developed with the support of 5 major Korean banks. Each player participating in The Lost City receives a tablet personal computer to play the game, in which virtual resources are created according to game scenarios, and the team that has accumulated the largest amount of money through trading wins the game. The Lost City was designed to enable the indirect experience of real financial products, such as mortgages and loans, in the process of using financial products provided by nonplay characters in the game.

Each player should build the facilities necessary for resource supply or demand by using borrowed virtual currency and collect the virtual currency through trade among each player. The Lost City is a game played at the financial center of Yeouido, a financial city in the Republic of Korea. Kang [9] verified that The Lost City had positive effects on the understanding of financial products, understanding of how to use financial products, and financial knowledge. Moreover, Kang proposed that to maximize the effect of experience-based learning, learner’s participation promotion, and use in other educational context, gamification should be applied to learning content. With the increase in the interest in gamification, many researchers published gamification development methodology. Mora et al [10] conducted an empirical study on the gamification development methodology published from 2011 to 2015. The following are common elements of the published methodology:

- Economic (about funding and operating)
- Logic (about rule and game mechanics setting)
- Measure (about performance indication)
- Psychology (about motivating and social acting)
- Interaction (about user experience, user interface, and technology).

Bockle et al [11] published the following 4 criteria for the gamification development methodology, which allows users to develop a gamification in the order they want:

1. Purpose of adaptivity (about user’s or participant’s analysis)
2. Adaptivity criteria (about player type, context, goal setting, and level of contents)
3. Adaptive interventions (about user experience or user interface and guideline for playing)
4. Adaptive game mechanics and dynamics (about game mechanics and dynamics).

There are many ways to develop gamification; however, the published methodologies are difficult to use as users with no experience face difficulty in using and applying game mechanics and fun experience. In addition, the number of these factors to be applied is a know-how factor. To solve some of these problems, this research team studied case studies to determine which game mechanics account for a large portion and what fun elements are used. For the case analysis, the 4F process, which is a gamification development methodology developed by Kim et al [12], was applied to analyze the game elements in collected cases. For the systematic study, the following 3 research questions were set, and 754 global gamification cases were collected and classified into 6 categories:

RQ 1: What is the distribution of global cases in terms of application category, play type, and published year?
RQ 2: What is the distribution of playful user experience (PLEX)--based fun factors in global cases, and is there a specific pattern noticeable in the applied factor among categories?
RQ 3: What is the distribution of game mechanics in global cases, and is there a specific pattern noticeable in the applied factor between among categories?
Figure 1. Team Maker.

Figure 2. The Lost City.
Methods

To collect global gamification cases, Google search and the gamification website enterprise gamification were used and related books were referred. Only cases that can be played or whose videos or screenshots could be referenced were selected. To analyze the collected data, Microsoft Excel 2010 and R studio (Psych Package) were used. The cases were collected from October 1, 2017, to January 15, 2018. The search keywords were Enterprise Gamification (Gamification examples in Enterprise) and Gamification in (of) Marketing, Recruiting, Management, Loyalty, Engagement, and Participant.

The collected cases were classified into the 6 categories to conduct a systematic study. There is an academical systematic study method for the paper such as the Preferred Reporting Items for Systematic Reviews and Meta-Analysis. However, no systematic study method has been proposed for general cases. An academic basis or criterion for general cases was not found. Most of the previous studies were classified according to a self-developed criterion. This study also approached in detail to categorize the collected cases according to self-developed criterion. This study was categorized from the perspectives of gamification value and player. The criteria for classification of the categories are shown in Figure 3 in which the terms individual and group were set on the x-axis from the gamification-player perspective and human-oriented and profit-related were set on the y-axis from the perspective of the gamification value. On the basis of these criteria, a total of 6 categories were derived:

- Commercialization (C): related to marketing, loyalty improvement, business management, and supply chain management
- Education (E): to obtain academic knowledge related to mathematics, science, and programming
- HRM (H): related to incentive, engagement improvement, and organization management
- Lifestyle (L): personal issue related to time management, money saving, and health care
- Social Issue (S): related to charity, car overspeeding problem, and hunger problem
- Training (T): to train technique required for specific jobs and tasks.

The case collection was based on the 4F process, as shown in Figure 4, and developed by Kim et al [12] for the development of gamification. The elements that constitute the 4F process were developed in consideration of 13 requirements for gamification development proposed by Morchhauser et al [13]. The 4F process is a development methodology to systematically manage the design process and improve the quality of gamification. The 4F process progresses in the order of Figure Out, Focus, Fun Design, and Finalize. In the Figure Out step, a targeting program or content is analyzed, and the personal identification (eg, age, gender, and favorite game genre), emotional state, and player type of participants are checked. To check the emotional state of participants in this step, the flow theory [14] for the player’s in-game emotion was applied. Player types were analyzed based on Bartle’s 8 player types [15]: opportunists, planners, scientists, hackers, networkers, friends, griefers, and politicians. Player-type analysis for fundamental problem solving is essential because the preferred game or play types depend on the player types [16]. The next step, Focus, sets the goal of the gamified program or content to be developed, specifically about the part of content to be gamified, cooperation possibility of other field experts, play time, and play type.

The play type is set into software (mobile app and program), hardware (board game), and Big game or alternative reality game. The fun factors set up the experience of fun to be delivered. The fun types are applied based on the PLEX model (Table 1), which is a theory that organizes fun experienced by humans into 22 types [17]. The Fun Design step sets the option of gamification components, such as storytelling, main game mechanics, dynamics, and winning or losing elements ratios. The storytelling compresses the 12 steps of the characteristics of the story structure of Hero’s developed by Vogler and Montez [18] into 4 steps.

The game mechanics and dynamics in the 4F process were developed by referring to the framework of Mechanics, Dynamics, and Aesthetics provided by Hunnicke et al [19]. Game mechanics was applied in the same or similar sense in previous studies. Kim et al [12] summarized 25 game mechanics from previous studies through empirical research. Overall, 6 large frame types were considered: reward, reward planning, avoidance, leaderboard, identification, and quest. In this study, only 18 game mechanics that can be experienced are used. The game mechanics used are listed in Table 2. The game dynamics was not defined in the form of academic rules and should be applied appropriately to match the characteristics of the player and content. The winning or losing element ratios should be set after appropriately adjusting 3 factors: player’s knowledge and skill, experience, and luck (lottery).

The Finalization step involves the completion of the prototyping of previously designed gamification and reflection through the pretester’s feedback. For the prototyping, a step-by-step manager should be determined to complete and frequently check the overall process of gamification development. After the prototyping is complete, the game should be pretested by the development team, and their feedback is supplemented. These steps are repeated to improvise the prototype’s quality, defined as good. This study applied the focus and fun design steps of the 4F process because of the following reasons. The 4F process is an academically proven model and has been applied to the proven gamification development methodology to ensure the reliability of the collected cases. In this study, the authors did their best to increase the reliability of the analysis results by using recently published gamification development methodologies.

The 4F process is a development methodology published in 2017 and has been compared with previous studies; it was thus used in this study.
Figure 3. The Criteria of Six Categories. HRM: human resource management.

Figure 4. 4F Process.
<table>
<thead>
<tr>
<th>Playful user experience fun factors</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Captivation</td>
<td>Forgetting about one’s surroundings and flow in it</td>
</tr>
<tr>
<td>Challenge</td>
<td>Testing one’s ability in specific tasks</td>
</tr>
<tr>
<td>Competition</td>
<td>Competing himself or herself or other player (or nonplay character)</td>
</tr>
<tr>
<td>Completion</td>
<td>Finishing what want to do</td>
</tr>
<tr>
<td>Control</td>
<td>Dominating the surroundings with one’s ability</td>
</tr>
<tr>
<td>Cruelty</td>
<td>Causing others mental or physical pain</td>
</tr>
<tr>
<td>Discovery</td>
<td>Finding something new information of unknown</td>
</tr>
<tr>
<td>Eroticism</td>
<td>Having personal feelings for others</td>
</tr>
<tr>
<td>Exploration</td>
<td>Investigating a new event or situation</td>
</tr>
<tr>
<td>Expression</td>
<td>Manifesting oneself using item or object</td>
</tr>
<tr>
<td>Fantasy</td>
<td>An imagined experience in the game</td>
</tr>
<tr>
<td>Fellowship</td>
<td>Communicating with others and to make friend in the game</td>
</tr>
<tr>
<td>Humor</td>
<td>Fun, joy, joke, and gags</td>
</tr>
<tr>
<td>Nurture</td>
<td>Taking care of oneself or to help others to be growing</td>
</tr>
<tr>
<td>Relaxation</td>
<td>Relaxing and healing their mental or body in the game</td>
</tr>
<tr>
<td>Sensation</td>
<td>Exciting by play using 5 senses</td>
</tr>
<tr>
<td>Simulation</td>
<td>Testing or making something that is impossible in real world</td>
</tr>
<tr>
<td>Submission</td>
<td>Being part of a large group of people</td>
</tr>
<tr>
<td>Subversion</td>
<td>Breaking social rules or laws</td>
</tr>
<tr>
<td>Suffering</td>
<td>Anger, loss, and frustration</td>
</tr>
<tr>
<td>Sympathy</td>
<td>Sharing their emotional feelings</td>
</tr>
<tr>
<td>Thrill</td>
<td>Exciting derived from risk and danger</td>
</tr>
</tbody>
</table>
Table 2. Description of game mechanics used in this study.

<table>
<thead>
<tr>
<th>Game mechanics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority</td>
<td>The Power to control other players, town, and item ships in the game</td>
</tr>
<tr>
<td>Avatar</td>
<td>Displaying the player’s character visually in the game</td>
</tr>
<tr>
<td>Badge</td>
<td>Displaying something such as trophy and flag that player’s achievement visually</td>
</tr>
<tr>
<td>Codiscovery</td>
<td>Completing the quest (mission) using collective ability or collaboration with others</td>
</tr>
<tr>
<td>Competition</td>
<td>Competing with other players’ records or competing with their own shadow, nonplay characters</td>
</tr>
<tr>
<td>Countdown</td>
<td>A time when player must complete a specific quest (mission)</td>
</tr>
<tr>
<td>Leaderboard</td>
<td>Showing the player’s level, point, and ranks and providing the feedback</td>
</tr>
<tr>
<td>Level</td>
<td>Displaying player’s achievement, ability, and power to number in the game</td>
</tr>
<tr>
<td>Lottery (luck)</td>
<td>The gameful tool in which the winner, the course, score, and item acquisition by probability</td>
</tr>
<tr>
<td>Point</td>
<td>Rewarding about player’s action such as quest (mission) success</td>
</tr>
<tr>
<td>Progress (bar)</td>
<td>Providing real-time information about the player’s current play situation</td>
</tr>
<tr>
<td>Quest (mission)</td>
<td>A specific goal for the player’s growth and providing reward when solved</td>
</tr>
<tr>
<td>Real goods</td>
<td>Rewarding in the real world about the achievement in the game</td>
</tr>
<tr>
<td>Scaffolding</td>
<td>A device that reduces difficulty when faced with difficulties</td>
</tr>
<tr>
<td>Social network</td>
<td>Linking player to player and displaying other’s progress</td>
</tr>
<tr>
<td>Unlocking</td>
<td>Providing new contents and function when player clear quest (mission) or level up</td>
</tr>
<tr>
<td>Virtual goods</td>
<td>The item that can be purchased, acquired, and traded</td>
</tr>
<tr>
<td>Virtual money</td>
<td>The currency used in the game</td>
</tr>
</tbody>
</table>

Results

Case Features

Table 3 shows the results of basic statistics and classification applied to the elements of the Focus step of the 4F process. The table shows the analysis of the categories defined by the research team of this study; play types, and distribution of published year. According to the results of the basic statistics analysis and classification, HRM (148/754, 19.7%) had a relatively large proportion among the categories, followed by training (143/754, 19.0%), social issue (133/754, 17.7%), education (127/754, 16.9%), commercialization (108/754, 14.3%), and lifestyle (95/754, 12.6%).

In the results of play types, software relatively accounted for a large proportion of approximately 80.0% (603/754) among play types, followed by hardware (89/754, 11.8%) and Big game or Alternative Reality game (62/754, 8.2%). According to the results of distribution of published year, the most cases were collected in 2014 (291/754, 38.6%) and 2015 (208/754, 27.6%).

Result of Game Mechanics and Playful User Experience–Based Fun Factors in the Cases

Tables 4 and 5 show the distribution of PLEX-based fun factors and game mechanics for each category defined in this study. Cases were recorded in the tables in order of commercialization (C), education (E), HRM (H), lifestyle (L), social issue (S), and training (T). For example, captivation of C was used 16 times out of the 108 cases (16/108, 15.0%).

The analysis results of PLEX-based fun factors are mentioned below. The factors of challenge and completion accounted for 90% in all categories. In contrast, cruelty, submission, subversion, suffering, and thrill were not found at all or only found in a few cases. In the commercialization category, competition was relatively higher than in other fun factors, and most of the PLEX-based fun factors were similarly applied compared with other categories. With respect to the education category, discovery relatively accounted for a large proportion, followed by simulation. Apart from the first 2 PLEX-based fun factors in education, the remaining fun factors were similarly applied in other categories. In the HRM category, nurture accounted for a large proportion and a specific feature pattern was not found. In the lifestyle category, expression, fellowship, relaxation, and sensation accounted for a relatively large proportion and other fun factors were similarly applied; no specific pattern was observed. In the social issue category, a specific pattern was not determined because game mechanics in social issue were applied similar to those in other categories. In the training category, discovery and simulation accounted for a large proportion, whereas no specific pattern was determined for the other fun factors.

The analysis results of game mechanics are as follows. Point, progress, and leaderboard accounted for a high proportion in all categories. In the commercialization category, the proportion of real goods was relatively higher than that in other categories. A specific pattern of game mechanics was not found in commercialization compared with other categories. In the education category, avatar, level, unlocking, and quest (mission) accounted for a large proportion. Other game mechanics were similarly applied as in other categories. In the HRM category, badge and virtual goods accounted for a large proportion. In the lifestyle category, most of the game mechanics were similarly
applied. Thus, a specific pattern was not found. In the social issue category, the proportion of codiscovery was highest, and in the training category, authority, avatar, quest (mission), and unlocking accounted for a large proportion; this is similar to game mechanics distribution of the education.

Table 3. Case features (N=754).

<table>
<thead>
<tr>
<th>Variables</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Categories</strong></td>
<td></td>
</tr>
<tr>
<td>Commercialization</td>
<td>108 (14.3)</td>
</tr>
<tr>
<td>Education</td>
<td>127 (16.9)</td>
</tr>
<tr>
<td>Human resource management</td>
<td>148 (19.7)</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>95 (12.6)</td>
</tr>
<tr>
<td>Social issue</td>
<td>133 (17.7)</td>
</tr>
<tr>
<td>Training</td>
<td>143 (19.0)</td>
</tr>
<tr>
<td><strong>Play types</strong></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td>603 (80.0)</td>
</tr>
<tr>
<td>Hardware</td>
<td>89 (11.8)</td>
</tr>
<tr>
<td>Big game or Alternative Reality Game</td>
<td>62 (8.2)</td>
</tr>
<tr>
<td><strong>Distribution of published year</strong></td>
<td></td>
</tr>
<tr>
<td>Before 2010</td>
<td>28 (3.8)</td>
</tr>
<tr>
<td>2011</td>
<td>22 (3.0)</td>
</tr>
<tr>
<td>2012</td>
<td>34 (4.6)</td>
</tr>
<tr>
<td>2013</td>
<td>54 (7.1)</td>
</tr>
<tr>
<td>2014</td>
<td>291 (38.6)</td>
</tr>
<tr>
<td>2015</td>
<td>208 (27.6)</td>
</tr>
<tr>
<td>2016</td>
<td>86 (11.4)</td>
</tr>
<tr>
<td>2017</td>
<td>31 (4.1)</td>
</tr>
</tbody>
</table>
Table 4. Playful user experience (PLEX) fun factors distribution of each category.

<table>
<thead>
<tr>
<th>Playful user experience</th>
<th>C(^a) (108 cases), n (%)</th>
<th>E(^b) (127 cases), n (%)</th>
<th>H(^c) (148 cases), n (%)</th>
<th>L(^d) (95 cases), n (%)</th>
<th>S(^e) (133 cases), n (%)</th>
<th>T(^f) (143 cases), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Captivation</td>
<td>16 (15.0)</td>
<td>15 (12.0)</td>
<td>11 (7.4)</td>
<td>17 (18)</td>
<td>16 (12.0)</td>
<td>19 (13.3)</td>
</tr>
<tr>
<td>Challenge</td>
<td>108 (100.0)</td>
<td>127 (100.0)</td>
<td>147 (99.3)</td>
<td>94 (99)</td>
<td>132 (99.2)</td>
<td>143 (100.0)</td>
</tr>
<tr>
<td>Competition</td>
<td>43 (40.0)</td>
<td>36 (28.3)</td>
<td>50 (34.0)</td>
<td>27 (28)</td>
<td>34 (26.0)</td>
<td>53 (37.1)</td>
</tr>
<tr>
<td>Completion</td>
<td>101 (94.0)</td>
<td>127 (100.0)</td>
<td>148 (100.0)</td>
<td>94 (99)</td>
<td>132 (99.2)</td>
<td>143 (100.0)</td>
</tr>
<tr>
<td>Control</td>
<td>6 (6.0)</td>
<td>17 (13.4)</td>
<td>19 (13.0)</td>
<td>12 (13)</td>
<td>16 (12.0)</td>
<td>27 (19.0)</td>
</tr>
<tr>
<td>Cruelty</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0)</td>
<td>0 (0.0)</td>
<td>1 (1.0)</td>
</tr>
<tr>
<td>Discovery</td>
<td>63 (58.3)</td>
<td>98 (77.2)</td>
<td>67 (45.3)</td>
<td>48 (51)</td>
<td>74 (56.0)</td>
<td>90 (63.0)</td>
</tr>
<tr>
<td>Eroticism</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>1 (1)</td>
<td>2 (2.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Exploration</td>
<td>38 (35.2)</td>
<td>35 (29.0)</td>
<td>13 (9.0)</td>
<td>18 (19)</td>
<td>28 (21.0)</td>
<td>31 (22.0)</td>
</tr>
<tr>
<td>Expression</td>
<td>58 (54.0)</td>
<td>47 (37.0)</td>
<td>65 (44.0)</td>
<td>62 (65)</td>
<td>69 (52.0)</td>
<td>74 (52.0)</td>
</tr>
<tr>
<td>Fantasy</td>
<td>24 (22.2)</td>
<td>44 (35.0)</td>
<td>13 (9.0)</td>
<td>21 (22)</td>
<td>20 (15.0)</td>
<td>33 (23.1)</td>
</tr>
<tr>
<td>Fellowship</td>
<td>48 (44.4)</td>
<td>36 (28.3)</td>
<td>66 (45.0)</td>
<td>52 (55)</td>
<td>64 (48.1)</td>
<td>52 (36.4)</td>
</tr>
<tr>
<td>Humor</td>
<td>11 (10.2)</td>
<td>3 (2.4)</td>
<td>7 (5.0)</td>
<td>11 (12)</td>
<td>10 (8.0)</td>
<td>4 (3.0)</td>
</tr>
<tr>
<td>Nurture</td>
<td>32 (30.0)</td>
<td>53 (42.0)</td>
<td>75 (51.0)</td>
<td>46 (48)</td>
<td>60 (45.1)</td>
<td>55 (38.5)</td>
</tr>
<tr>
<td>Relaxation</td>
<td>50 (46.3)</td>
<td>33 (26.0)</td>
<td>53 (36.0)</td>
<td>46 (48)</td>
<td>62 (47.0)</td>
<td>33 (23.1)</td>
</tr>
<tr>
<td>Sensation</td>
<td>25 (23.1)</td>
<td>46 (36.2)</td>
<td>27 (18.2)</td>
<td>45 (47)</td>
<td>31 (23.3)</td>
<td>53 (37.1)</td>
</tr>
<tr>
<td>Simulation</td>
<td>29 (27.0)</td>
<td>86 (68.0)</td>
<td>45 (30.4)</td>
<td>30 (32)</td>
<td>63 (47.4)</td>
<td>107 (75.0)</td>
</tr>
<tr>
<td>Submission</td>
<td>1 (1.0)</td>
<td>1 (1.0)</td>
<td>0 (0.0)</td>
<td>1 (1)</td>
<td>2 (2.0)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Subversion</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Suffering</td>
<td>1 (1.0)</td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>0 (0)</td>
<td>0 (0.0)</td>
<td>4 (3.0)</td>
</tr>
<tr>
<td>Sympathy</td>
<td>28 (26.0)</td>
<td>14 (11.0)</td>
<td>37 (25.0)</td>
<td>28 (29)</td>
<td>65 (49.0)</td>
<td>37 (26.0)</td>
</tr>
<tr>
<td>Thrill</td>
<td>3 (3.0)</td>
<td>1 (1.0)</td>
<td>0 (0.0)</td>
<td>4 (4)</td>
<td>4 (3.0)</td>
<td>6 (4.2)</td>
</tr>
</tbody>
</table>

\(^a\)C: commercialization.
\(^b\)E: education.
\(^c\)H: human resource management.
\(^d\)L: lifestyle.
\(^e\)S: social issue.
\(^f\)T: training.
Correlation of Game Mechanics and Playful User Experience Fun Factors Distribution

The results of the correlation analysis are summarized in Tables 6 and 7. In this study, Pearson and Kendall correlation coefficients were used in the correlation analysis. Both parametric and nonparametric analyses should be conducted to check the statistical significance [20,21]. Pearson correlation coefficient is a type of parametric correlation analysis, and Kendall correlation analysis is a type of nonparametric correlation analysis. This study has tried to ensure the reliability of the results through 2 correlation methods.

According to Gustavo et al [22], the standard Pearson and Kendall correlation coefficients are as follows:

- small effect: $\tau=.20$ is approximately equal to $r=.30$
- medium effect: $\tau=.34$ is approximately equal to $r=.50$
- large effect: $\tau=.50$ is approximately equal to $r=.70$.

The results of the game mechanics correlation for each category are as follows. There is a statistically large correlation among each category (Tables 6 and 7). The category of commercialization and other categories are significantly correlated with education ($r=.822$, $P<.001$, $\tau=.518$, $P=.003$), HRM ($r=.949$, $P<.001$, $\tau=.598$, $P<.001$), lifestyle ($r=.951$, $P<.001$, $\tau=.678$, $P<.001$), social issue ($r=.956$, $P<.001$, $\tau=.713$, $P<.001$), and training ($r=.877$, $P<.001$, $\tau=.607$, $P<.001$). The categories of HRM ($r=.882$, $P<.001$, $\tau=.682$, $P<.001$), lifestyle ($r=.927$, $P<.001$, $\tau=.801$, $P<.001$), social issue ($r=.846$, $P<.001$, $\tau=.507$, $P=.003$), and training ($r=.972$, $P<.001$, $\tau=.862$, $P<.001$) are significantly correlated with education. The HRM category is significantly correlated with lifestyle ($r=.969$, $P<.001$, $\tau=.742$, $P<.001$), social issue ($r=.920$, $P<.001$, $\tau=.559$, $P=.001$), and training ($r=.913$, $P<.001$, $\tau=.743$, $P<.001$). The categories of lifestyle and other categories are significantly correlated with social issue ($r=.958$, $P<.001$, $\tau=.665$, $P<.001$) and training ($r=.952$, $P<.001$, $\tau=.849$, $P<.001$). The categories of social issue and training are significantly correlated by $r=.897$ ($P<.001$, $\tau=.608$ $P<.001$).

The results of the correlation analysis of PLEX fun factors are as follows. There is a statistically large correlation among each category (Tables 8 and 9). The category of commercialization and other categories are significantly correlated with education ($r=.898$, $P<.001$, $\tau=.735$, $P<.001$), HRM ($r=.951$, $P<.001$, $\tau=.828$, $P<.001$), lifestyle ($r=.956$, $P<.001$, $\tau=.841$, $P<.001$), social issue ($r=.953$, $P<.001$, $\tau=.841$, $P<.001$), and training ($r=.905$, $P<.001$, $\tau=.773$, $P<.001$). The categories of education

---

### Table 5. Game mechanics distribution of each category.

<table>
<thead>
<tr>
<th>Mechanics</th>
<th>C (108 cases), n (%)</th>
<th>E (127 cases), n (%)</th>
<th>H (148 cases), n (%)</th>
<th>L (95 cases), n (%)</th>
<th>S (133 cases), n (%)</th>
<th>T (143 cases), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authority</td>
<td>20 (19.0)</td>
<td>33 (26.0)</td>
<td>30 (20.3)</td>
<td>21 (22)</td>
<td>36 (27.0)</td>
<td>44 (31.0)</td>
</tr>
<tr>
<td>Avatar</td>
<td>41 (38.0)</td>
<td>76 (60.0)</td>
<td>65 (44.0)</td>
<td>38 (40)</td>
<td>33 (25.0)</td>
<td>84 (59.0)</td>
</tr>
<tr>
<td>Badge</td>
<td>78 (72.2)</td>
<td>73 (57.5)</td>
<td>125 (84.5)</td>
<td>73 (77)</td>
<td>84 (63.1)</td>
<td>79 (55.2)</td>
</tr>
<tr>
<td>Codiscovery</td>
<td>54 (50.0)</td>
<td>56 (44.1)</td>
<td>54 (36.5)</td>
<td>44 (46)</td>
<td>76 (57.1)</td>
<td>64 (45.0)</td>
</tr>
<tr>
<td>Competition</td>
<td>42 (39.0)</td>
<td>38 (30.0)</td>
<td>50 (34.0)</td>
<td>26 (27)</td>
<td>34 (26.0)</td>
<td>55 (38.5)</td>
</tr>
<tr>
<td>Countdown</td>
<td>29 (27.0)</td>
<td>49 (39.0)</td>
<td>27 (18.2)</td>
<td>32 (34)</td>
<td>45 (34.0)</td>
<td>58 (41.0)</td>
</tr>
<tr>
<td>Leaderboard</td>
<td>102 (94.2)</td>
<td>124 (98.0)</td>
<td>146 (99.0)</td>
<td>94 (99)</td>
<td>128 (96.2)</td>
<td>140 (98.0)</td>
</tr>
<tr>
<td>Level</td>
<td>20 (19.0)</td>
<td>63 (50.0)</td>
<td>32 (22.0)</td>
<td>28 (29)</td>
<td>19 (14.3)</td>
<td>46 (32.2)</td>
</tr>
<tr>
<td>Lottery (luck)</td>
<td>30 (28.0)</td>
<td>31 (24.4)</td>
<td>19 (13.0)</td>
<td>24 (25)</td>
<td>31 (23.3)</td>
<td>35 (24.5)</td>
</tr>
<tr>
<td>Point</td>
<td>87 (81.0)</td>
<td>119 (94.0)</td>
<td>144 (97.3)</td>
<td>84 (88)</td>
<td>112 (84.2)</td>
<td>132 (92.3)</td>
</tr>
<tr>
<td>Progress</td>
<td>105 (97.2)</td>
<td>126 (99.2)</td>
<td>146 (99.0)</td>
<td>94 (99)</td>
<td>132 (99.2)</td>
<td>143 (100.0)</td>
</tr>
<tr>
<td>Quest (mission)</td>
<td>48 (44.4)</td>
<td>90 (71.0)</td>
<td>53 (36.0)</td>
<td>48 (51)</td>
<td>61 (46.0)</td>
<td>82 (57.3)</td>
</tr>
<tr>
<td>Real goods</td>
<td>38 (35.2)</td>
<td>6 (5.0)</td>
<td>16 (11.0)</td>
<td>12 (13)</td>
<td>40 (30.1)</td>
<td>10 (7.0)</td>
</tr>
<tr>
<td>Social network</td>
<td>35 (32.4)</td>
<td>14 (11.0)</td>
<td>27 (18.2)</td>
<td>16 (17)</td>
<td>22 (17.6)</td>
<td>15 (10.5)</td>
</tr>
<tr>
<td>Scaffolding</td>
<td>20 (19.0)</td>
<td>46 (36.4)</td>
<td>20 (14.0)</td>
<td>16 (17)</td>
<td>28 (21.6)</td>
<td>36 (25.2)</td>
</tr>
<tr>
<td>Unlocking</td>
<td>21 (19.4)</td>
<td>45 (35.4)</td>
<td>24 (16.2)</td>
<td>22 (22)</td>
<td>29 (22.0)</td>
<td>50 (35.0)</td>
</tr>
<tr>
<td>Virtual goods</td>
<td>59 (55.0)</td>
<td>76 (60.0)</td>
<td>90 (61.0)</td>
<td>45 (45)</td>
<td>59 (44.4)</td>
<td>72 (50.3)</td>
</tr>
<tr>
<td>Virtual money</td>
<td>14 (13.0)</td>
<td>31 (24.4)</td>
<td>34 (23.0)</td>
<td>19 (19)</td>
<td>30 (23.0)</td>
<td>25 (17.5)</td>
</tr>
</tbody>
</table>

a: C: commercialization.
b: E: education.
c: H: human resource management.
d: L: lifestyle.
e: S: social issue.
f: T: training.
and other categories are significantly correlated with HRM ($r=.893, P<.001$, $τ=.774, P<.001$), lifestyle ($r=.88, P<.001$, $τ=.77, P<.001$), and training ($r=.975, P<.001$, $τ=.871, P<.001$). The categories of lifestyle ($r=.961, P<.001$, $τ=.841, P<.001$), social issue ($r=.9, P<.001$, $τ=.787, P<.001$), and training ($r=.975, P<.001$, $τ=.882, P<.001$) are significantly correlated with HRM. Lifestyle is significantly correlated with social issue ($r=.959, P<.001$, $τ=.889, P<.001$) and training ($r=.906, P<.001$, $τ=.803, P<.001$). The categories of social issue and training are significantly correlated by $r=.935 (P<.001)$, $τ=.812 (P<.001)$.

### Table 6. Pearson correlation results of game mechanics.

<table>
<thead>
<tr>
<th>Game mechanics</th>
<th>Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>E</td>
<td>.822&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>H</td>
<td>.949&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>L</td>
<td>.951&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>S</td>
<td>.956&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>T&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.877&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>C: commercialization.  
<sup>b</sup>E: education.  
<sup>c</sup>H: human resource management.  
<sup>d</sup>L: lifestyle.  
<sup>e</sup>S: social issue.  
<sup>f</sup>P<.001.

### Table 7. Kendall correlation results of game mechanics.

<table>
<thead>
<tr>
<th>Game mechanics</th>
<th>Kendall correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>E</td>
<td>.518&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>H</td>
<td>.598&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>L</td>
<td>.678&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>S</td>
<td>.713&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
<tr>
<td>T&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.607&lt;sup&gt;g&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>C: commercialization.  
<sup>b</sup>E: education.  
<sup>c</sup>H: human resource management.  
<sup>d</sup>L: lifestyle.  
<sup>e</sup>S: social issue.  
<sup>f</sup>P<.003.  
<sup>g</sup>P<.001.  
<sup>h</sup>P<.001.
### Table 8. Pearson correlation results of playful user experience (PLEX) fun factors.

<table>
<thead>
<tr>
<th>Game mechanics</th>
<th>Pearson correlation</th>
<th>C&lt;sup&gt;a&lt;/sup&gt;</th>
<th>E&lt;sup&gt;b&lt;/sup&gt;</th>
<th>H&lt;sup&gt;c&lt;/sup&gt;</th>
<th>L&lt;sup&gt;d&lt;/sup&gt;</th>
<th>S&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td>.898&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>.951&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.893&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>L</td>
<td></td>
<td>.956&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.880&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.961&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td>.953&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.900&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.970&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.959&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
</tr>
<tr>
<td>T&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td>.905&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.975&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.921&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.906&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.935&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>C: commercialization.  
<sup>b</sup>E: education.  
<sup>c</sup>H: human resource management.  
<sup>d</sup>L: lifestyle.  
<sup>e</sup>S: social issue.  
<sup>f</sup><i>P</i> < .001.  
<sup>g</sup>T: training.

### Table 9. Kendall correlation results of playful user experience (PLEX) fun factors.

<table>
<thead>
<tr>
<th>Game mechanics</th>
<th>Kendall correlation</th>
<th>C&lt;sup&gt;a&lt;/sup&gt;</th>
<th>E&lt;sup&gt;b&lt;/sup&gt;</th>
<th>H&lt;sup&gt;c&lt;/sup&gt;</th>
<th>L&lt;sup&gt;d&lt;/sup&gt;</th>
<th>S&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td>.735&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>.828&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.774&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>L</td>
<td></td>
<td>.841&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.787&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.84&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>S</td>
<td></td>
<td>.841&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.770&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.832&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.889&lt;sup&gt;f&lt;/sup&gt;</td>
<td>—</td>
</tr>
<tr>
<td>T&lt;sup&gt;g&lt;/sup&gt;</td>
<td></td>
<td>.773&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.871&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.794&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.803&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.812&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>C: commercialization.  
<sup>b</sup>E: education.  
<sup>c</sup>H: human resource management.  
<sup>d</sup>L: lifestyle.  
<sup>e</sup>S: social issue.  
<sup>f</sup><i>P</i> < .001.  
<sup>g</sup>T: training.

### Table 10. Average, minimum, and maximum number of applied game mechanics and playful user experience (PLEX) fun factors.

<table>
<thead>
<tr>
<th>Factors and mechanics</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Playful user experience fun factors</strong></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>6</td>
</tr>
<tr>
<td>Minimum</td>
<td>2</td>
</tr>
<tr>
<td>Maximum</td>
<td>18</td>
</tr>
<tr>
<td><strong>Game mechanics</strong></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>8</td>
</tr>
<tr>
<td>Minimum</td>
<td>3</td>
</tr>
<tr>
<td>Maximum</td>
<td>18</td>
</tr>
</tbody>
</table>
Table 11. Number of applied Game mechanics and playful user experience (PLEX) fun factors.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of playful user experience fun factors</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19 (2.6)</td>
</tr>
<tr>
<td>3</td>
<td>59 (8.0)</td>
</tr>
<tr>
<td>4</td>
<td>91 (12.1)</td>
</tr>
<tr>
<td>5</td>
<td>145 (19.2)</td>
</tr>
<tr>
<td>6</td>
<td>129 (17.1)</td>
</tr>
<tr>
<td>7</td>
<td>88 (12.0)</td>
</tr>
<tr>
<td>8</td>
<td>81 (10.8)</td>
</tr>
<tr>
<td>9</td>
<td>46 (6.1)</td>
</tr>
<tr>
<td>10</td>
<td>31 (4.1)</td>
</tr>
<tr>
<td>11</td>
<td>31 (4.1)</td>
</tr>
<tr>
<td>12</td>
<td>15 (2.0)</td>
</tr>
<tr>
<td>13</td>
<td>9 (1.2)</td>
</tr>
<tr>
<td>14</td>
<td>5 (1.0)</td>
</tr>
<tr>
<td>15</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>16</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td>18</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Total</td>
<td>754 (100.0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of game mechanics</strong></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14 (2.0)</td>
</tr>
<tr>
<td>3</td>
<td>50 (7.0)</td>
</tr>
<tr>
<td>4</td>
<td>81 (11.0)</td>
</tr>
<tr>
<td>5</td>
<td>115 (15.2)</td>
</tr>
<tr>
<td>6</td>
<td>116 (15.4)</td>
</tr>
<tr>
<td>7</td>
<td>103 (14.0)</td>
</tr>
<tr>
<td>8</td>
<td>84 (11.1)</td>
</tr>
<tr>
<td>9</td>
<td>61 (8.1)</td>
</tr>
<tr>
<td>10</td>
<td>55 (7.3)</td>
</tr>
<tr>
<td>11</td>
<td>29 (4.0)</td>
</tr>
<tr>
<td>12</td>
<td>18 (2.4)</td>
</tr>
<tr>
<td>13</td>
<td>10 (1.3)</td>
</tr>
<tr>
<td>14</td>
<td>13 (2.0)</td>
</tr>
<tr>
<td>15</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>16</td>
<td>3 (0.3)</td>
</tr>
<tr>
<td>18</td>
<td>1 (0.3)</td>
</tr>
<tr>
<td>Total</td>
<td>754 (100.0)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

The purpose of this study was to collect 754 gamification cases and derive their implication through an empirical study and a statistical data analysis. The 4F process was applied for analyzing gamification elements. Table 12 summarizes the answers to the 3 research questions. The results of this study would help people who have difficulty in developing gamification. When developing gamification, we expect that the results of this study will have a positive impact on people who have difficulty using game mechanics and experiencing fun.
Table 12. Summary of results.

<table>
<thead>
<tr>
<th>Research question and question keywords</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Categories</td>
<td>Category classification: 6 categories</td>
</tr>
<tr>
<td>Play types</td>
<td>Play types: software (80%), hardware (12%), and Big game or ARG² (8%)</td>
</tr>
<tr>
<td>Published year</td>
<td>High proportion of published year: 2014 and 2015</td>
</tr>
<tr>
<td>2 PLEX b fun factors distribution</td>
<td>High proportion of PLEX fun factors: challenge and completion; low proportion of PLEX fun factors: cruelty, submission, subversion, suffering, and thrill</td>
</tr>
<tr>
<td>Correlation each category</td>
<td>Correlation analysis: distribution of PLEX fun factors was similar in each category (a positive correlation for each category)</td>
</tr>
<tr>
<td>3 Game mechanics distribution</td>
<td>High proportion of game mechanics: point, progress, and leaderboard; proportion of game mechanics was similarly distributed and used</td>
</tr>
<tr>
<td>Correlation each category</td>
<td>Correlation analysis: distribution of game mechanics was similar in each category (a positive correlation for each category)</td>
</tr>
</tbody>
</table>

aARG: alternative reality game.
bPLEX: playful user experience.

Discussion of Research Question 1

Among the 3 types of gameplay, software accounted for 80% of overall cases. It is related to recent advances in information technology such as smartphones, smartpads, and improvement of internet environment. As technology advances, because access to software type of gamification has been strengthened, the player is possible to play the variety of type of gamification contents without time and space constraints. Furthermore, software type of gamification is used to secure marketing and potential customer and their loyalty, and it is possible to solve the inconvenience of accessing services such as online banking system and stock trading system.

Baptista and Oliveira [23] applied gamification to a software type of mobile banking system to improve its usability and accessibility to customers. For validating the results of their study, a structural equation was used. According to the results of their study, a software type of gamification was possible to improve customers’ engagement, accessibility, and usability. Baptista and Oliveira [23] suggested using gamification for changing customers’ behavior through gaming experience.

According to the results of this study, the largest number of gamification cases are published in 2014 and 2015. It is assumed that the collected cases are maintained through continuous updating. However, this does not imply that new cases have not been published. Global gamification cases have increased steadily since gamification was activated, with a significant increase observed in 2014 and 2015, and these cases are being kept. Dicheva et al [24] conducted an empirical research of an advanced study on gamification in education. The number of education gamification cases began increasing from 2013, and the distribution of overall cases was similar to that obtained in this study. Additional studies should be conducted to infer clear conclusions about the year of distribution for published cases.

Discussion of Research Question 2

In the analysis results of PLEX fun factors from among the 6 categories defined in this study, challenge, completion, and competition constituted a large proportion of overall categories. The distribution of PLEX fun factors in this study resembles that of Kim’s study [25]. The reason why the fun factors of challenge and completion accounted for a large proportion is assumed to be related to the winning condition. The structure of gamification is also designed to be similar to that of the game. Players must challenge and solve problems to win the game. Gamification also provides a resolvable problem such as a game for players. In the process of problem solving, the fun factors of challenge and completion are added. For this reason, both these fun factors accounted for a large proportion among the fun factors. The fun factor of competition was used assuming that it will allow more players to participate. On the other hand, the fun factors of cruelty, submission, subversion, suffering, and thrill were either found in a few cases or not found in the collected cases. These 5 fun factors were found to be very difficult to implement and apply within game-based content.

The correlation analysis results of PLEX fun factors showed similar patterns among the 6 categories. This means that there is no particularly preferred pattern according to the application field of the PLEX fun factors. Having no particularly preferred pattern among the PLEX fun factors implies that there was no difference among the fun factors. The correlation analysis results can be shown as the grounds for the use of suitable PLEX fun factors regardless of the applied target. However, reckless use of fun factors is harmful to the quality of gamification. Depending on the results of this study, it is recommended to use 4 to 8 PLEX fun factors.

Discussion of Research Question 3

In the analysis results of game mechanics among the 6 categories defined in this study, points, progress, and leaderboard constituted a large proportion of the overall categories. The
distribution of game mechanics in this study resembles that of Dicheva et al’s study [24]. In most cases, points have been shown to be used as concept of reward. Progress and leaderboard have been using similar concepts in most cases. However, both these game mechanics are of a different nature and should be analyzed separately. On the other hand, badge, a Point, Badge, and Level system [26], found its utilization to be lower than points and leaderboard. It is assumed to have been caused by regional or cultural differences. In North America, many people are found collecting badges from local universities. In Asia, badge collection is quite strange. Due to these regional or cultural differences, it is assumed that the use of badge is relatively low. Specially, among the 6 categories of social issues, the game mechanics of codiscovery was found to be relatively higher than that in other categories. It is important to address social issues, but it can be thought of as using collective intelligence.

In the correlation analysis results of game mechanics, the correlation among the 6 categories was significant. This indicates that most categories are based on similar game mechanics. However, reckless use, such as of PLEX fun factors, can harm players. This study recommends using 5 to 9 game mechanics when exploring game mechanics. A designer or developer would like to use these many game mechanics during the gamification development process.

Limitations and Future Direction
The limitations of this study are as follows:

- Criteria for case classification without Mutually Exclusive and Completely Exhaustive (MECE)
- Absence of a differentiated method of data analysis
- Validity and reliability problems caused by applying a single gamification development methodology
- Limits of the restricted search range centered on the Google search engine and enterprise gamification websites.

Before conducting this study, criteria were set to classify the cases. However, the category could not fully classify the cases because MECE was not applied. Using MECE and a proven taxonomy to analyze the cases based on academic grounds, more relevant study results must be obtained.

Recent developments in big data have begun to refine data analytics. This study only performed basic statistical analyses on the correlation between PLEX fun factors and game mechanics. However, further studies should be conducted considering not only the number used in the data analysis techniques but also the relevant factors such as player type. For this, data should be collected more systematically, and analysis techniques should be applied accordingly to conduct further studies.

This study only used the 4F process to analyze the collected cases. Therefore, analytical factors from different perspectives should be obtained. Researchers in related fields are constantly developing methodologies to develop systematic gamification such as 6D process [27] and the 13 requirements for gamification development [13]. In addition, player-type classification procedures should be added through data analysis. Examples of representative player type classifications include Bartle’s 8 player types [15], BrainHex [28], and Octalysis [29].

The cases used in this study were collected only through Google search and enterprise gamification Web page. However, future study should be conducted on a wider range of cases, including cases analyzed in previous papers. Future study should be conducted through diverse case collection for verifying the reliability and feasibility of study results.

Acknowledgments
This study was supported by the National Research Foundation of Korea grant funded by the Korean Government (MSIT; No. 2017R1A2B2002798). This study was supported by 2016 Research Grant from Kangwon National University (No. 520160385).

Conflicts of Interest
None declared.

References


Abbreviations

HRM: human resource management
MECE: Mutually Exclusive and Completely Exhaustive
PLEX: playful user experience
SNS: social network service