JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 4 (2016), Issue 2   ISSN 2291-9279   Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Review

Original Paper

A Serious Game to Increase Healthy Food Consumption in Overweight or Obese Adults: Randomized Controlled Trial

Tegan Blackburne[1*], BPsyc(Hons); Alexandra Rodriguez[1*], BPsyc(Hons); Stuart John Johnstone[1*], PhD

Brain and Behaviour Research Institute, School of Psychology, University of Wollongong, Wollongong, Australia

[*]all authors contributed equally

Corresponding Author:
Stuart John Johnstone, PhD
Brain and Behaviour Research Institute
School of Psychology
University of Wollongong
Northfields Avenue
Wollongong, 2522
Australia
Phone: 61 242214495
Fax: 61 242214163
Email: sjohnsto@uow.edu.au

Abstract

Background: Obesity is a growing global issue that is linked to cognitive and psychological deficits.

Objective: This preliminary study investigated the efficacy of training to improve inhibitory control (IC), a process linked to overeating, on consumption and cognitive control factors.

Methods: This study utilized a multisession mobile phone–based intervention to train IC in an overweight and obese population using a randomized waitlist-control design. A combination of self-assessment questionnaires and psychophysiological measures was used to assess the efficacy of the intervention in terms of improved general IC and modified food consumption after training. Attitudes toward food were also assessed to determine their mediating role in food choices. A total of 58 participants (47 female) completed 2 assessment sessions 3 weeks apart, with 2 weeks of intervention training for the training group during this time. The groups did not differ in baseline demographics including age, body mass index, and inhibitory control.

Results: Inhibitory control ability improved across the training sessions, with increases in P3 amplitude implying increased cognitive control over responses. Inhibitory control training was associated with increased healthy and reduced unhealthy food consumption in a taste test and in the week following training, as measured by the Healthy Eating Quiz and the food consumption test. Cognitive restraint was enhanced after training for the training but not the waitlist condition in the Three-Factor Eating Questionnaire, implying that attempts to avoid unhealthy foods in the future will be easier for the training group participants.

Conclusions: Inhibitory control training delivered via a purpose-designed mobile phone app is easy to complete, is convenient, and can increase cognitive restraint and reduce unhealthy food consumption.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12616000263493; http://www.ANZCTR.org.au/ACTRN12616000263493.aspx (Archived by WebCite at http://www.webcitation.org/6ioHjGING)

(JMIR Serious Games 2016;4(2):e10) doi:10.2196/games.5708

KEYWORDS

obesity; inhibition training; event-related potential; eating; mobile apps

Introduction

The prevalence of obesity has reached extreme proportions on a global scale. More than 2.1 billion people—one-third of the world's adults, and one-fourth of children and adolescents—are classified as overweight or obese [1]. Generally speaking, obesity is attributed to increased calorie intake, unhealthy changes in diet composition, and an increasingly sedentary lifestyle [2,3]. In modern society, high-calorie foods are tasty, affordable, and abundant, making them increasingly difficult to resist [4,5]. Healthy foods include fresh fruits and vegetables, as well as nuts, seeds, and lean meat, whereas unhealthy foods include processed meats, potato chips, and butter [6]. Behavioral interventions such as low-calorie meal plans are frequently implemented, but these approaches are often only effective in the short term. They also assume that individuals have the capacity to change their behavior, which is typically not the case [7,8]. An emphasis on deficits in executive functions (ie, the set of processes that manage, regulate, and control cognition and behavior) associated with excess weight, such as poor inhibitory control (IC), may add substantially to the outcomes of obesity interventions.

Inhibitory control refers to the ability to suppress dominant or automatic responses and allows for self-regulation by constraining thoughts and actions that may interfere with the completion of goal-directed actions [2,3,9]. Individuals with deficits in IC are more likely to be unsuccessful dieters and overweight or obese [4,5,10]. Individuals with poor IC tend to engage in unhealthy food consumption behaviors including overeating, excessive snacking, and binge eating [11]. According to dual-process models of obesity [2], the difference between healthy weight and overweight individuals is not the degree of positive affect they experience toward high-calorie foods but the ability to override automatic response tendencies toward consumption [5,12].

There is increasing evidence that the neural mechanisms underlying executive functions are responsive to training and that general IC may be conceptualized as a “muscle” that can be strengthened with exercise [13,14]. Go/Nogo tasks have been widely utilized to train IC, particularly in the context of reducing alcohol consumption [7,15]. These tasks involve making frequent and rapid responses to “Go” stimuli, while refraining from responding to less frequent “Nogo” stimuli [16]. Repeated pairing of inhibition responses to a stimulus creates an association between that stimulus and the goal of stopping behavior [7]. Undertaking training that attaches a stop goal to unhealthy foods may therefore help individuals with low levels of IC to inhibit unintentional impulses toward them.

Intervention studies indicate that IC training is an effective tool in reducing food intake and consumption of high-calorie foods in individuals with low trait IC [5,17]. These studies have focused on immediate taste tests in the laboratory to measure efficacy, leaving the longer-term effects of the training unexplored. Furthermore, most IC training studies in the food domain utilize only a single session of training, which may not be sufficiently intensive to produce improvements [7,10]. Few studies have used multisession, longer-term interventions. Allom and Mullan [18] utilized a stop-signal IC training, focused on weight loss, and did not include psychophysiological measures. Veling et al [19] used a Go/Nogo task delivered via the Internet in 4 weekly sessions and found that this training facilitated weight loss, but they did not assess other outcome measures. Additionally, some research has shown that delivering health interventions through mobile phones has already been studied, targeting exercise and movement, with mixed results [20,21].

Training tasks that have a food-specific, as opposed to a general, focus (using nonfood stimuli), have been found to yield greater improvements in food consumption behaviors [5,17]. Furthermore, current IC training studies have shown that priming disinhibition toward unhealthy foods can increase unhealthy food consumption after training [7,10], but no study has primed disinhibition toward *healthy* foods to determine if IC training can be used to increase healthy food consumption.

Event-related potentials (ERPs) are derived from the electroencephalogram (EEG) and provide an objective index of the sensory and cognitive processing stages involved in processing a stimulus event. In the context of this study, ERPs allow consideration of the neural correlates of inhibition elicited during Go/Nogo tasks, with the N2 and P3 ERP components showing sensitivity to inhibitory demands. The N2 component typically occurs between 200 and 300 milliseconds after stimulus onset [22] and has been positively correlated with body mass index (BMI) [23]. The P3 component typically occurs between 300 and 500 milliseconds after stimulus presentation [16], with neural sources in close proximity to motor and premotor cortices—thus, it is likely that the P3 reflects a later stage of the inhibitory process [24]. Very few studies have examined the neural correlates of inhibition in the context of obesity and food consumption. It has been reported that individuals in a healthy weight range exhibit more pronounced “Nogo” N2 and P3 components, indicative of heightened IC, toward food stimuli as opposed to general stimuli [25].

This study was a pilot study focused on researching the efficacy of an app-based intervention and its effect on short-term weight-related goals (ACTRN12616000263493). It utilized a randomized waitlist-control design and a novel IC intervention to train inhibition toward unhealthy foods and approach (or “disinhibition”) toward healthy foods in individuals identified as overweight or obese. Multisession IC training was employed, as more intensive exercise of neural inhibitory “muscles” should produce greater improvements in IC. This study also took a novel approach by (1) evaluating food consumption over a 1-week period, rather than immediately after training, (2) priming dominant automatic responses (or “disinhibition”) toward healthy foods to increase healthy food consumption, and (3) combining behavioral and psychophysiological measures. Attitudes toward dieting and food were also measured. It was hypothesized that participants in the training condition would consume less unhealthy and more healthy foods than the waitlist condition in the laboratory consumption test and over the 1-week period after training.

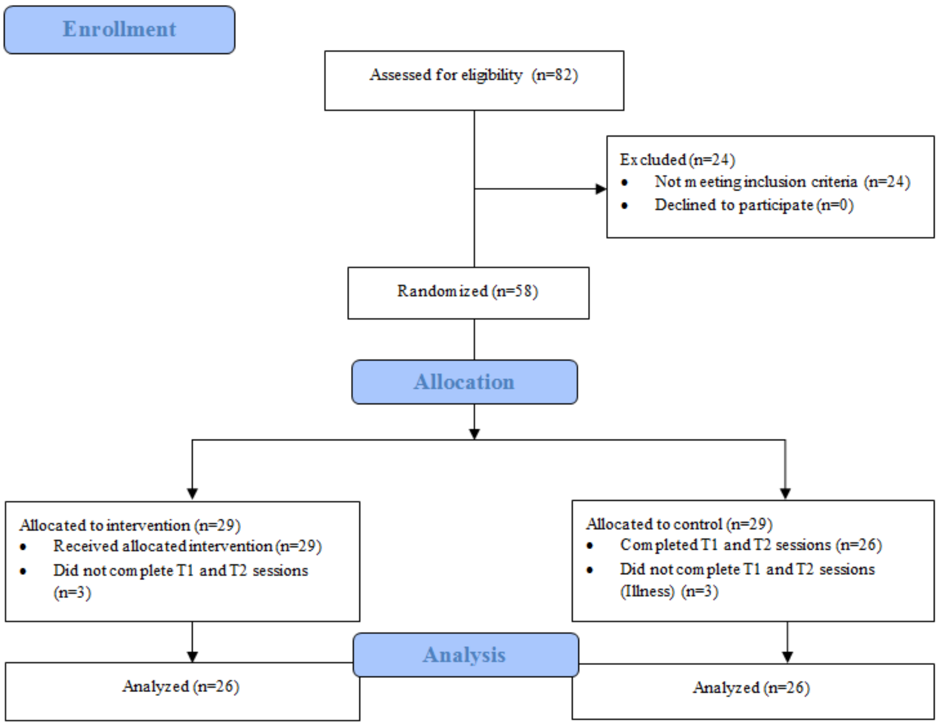
Furthermore, an exploratory hypothesis suggested that participants in the training compared with the waitlist condition would show (1) improved performance (reflected by less inhibition errors and faster reaction times) and (2) enhanced N2 and P3 amplitudes, after training in an untrained, non–food-specific Go/Nogo task. Previous research has focused on assessing the generalization of IC abilities from general to specific, or from general to general stimuli; this research is seeking to assess how this process may occur in the opposite direction. Additionally, in accordance with previous research showing changes in attitudes toward food and dieting after weight-based interventions [26,27], it was hypothesized that there would be an increase in cognitive restraint and a decrease in disinhibition and hunger after training in the training but not the waitlist condition in the Three-Factor Eating Questionnaire (TFEQ). Overall, we expected that the training would increase healthy food consumption, reduce unhealthy eating, and this would be reflected in not only eating behaviors but also psychological and neurological changes.

Methods

Participants

Participants (n=58; females=47, 81%) were recruited from the general population in the Wollongong (Australia) area through flyers distributed in the community. Flyers called for individuals who wanted to improve their eating habits, and the intervention was described broadly as "brain training," with no specific mention of IC. Participants were required to be older than 13 years (range 19-61, mean 36.48, SD 14.22), have a BMI higher than 25 (mean 29.54, SD 4.05), and possess an iOS device in order to access the training app. Participants were screened for neurological disorders and normal hearing and vision. Participants were excluded if they did not complete both experimental sessions (n=6), leaving a final sample of 52 participants (females=41). Figure 1 shows the CONSORT (Consolidated Standards of Reporting Trials) flow diagram. An a priori power analysis using the G*Power computer program [28] indicated that a total sample of 54 people would be needed to detect moderate effects (*d*=0.5) with 95% power using a *t* test between means with alpha at .05. Baseline differences between the conditions for BMI, age, and number of days between sessions were nonsignificant. Gender ratio was equal across conditions (21 females to 5 males).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) Flow diagram. Representation of the progress of participants through the trial. T1: time 1; T2: time 2.



Materials

Barratt Impulsiveness Scale

The Barratt Impulsiveness Scale (BIS) is a 30-item questionnaire rated on a 4-point Likert scale and includes 3 subscales: attentional, motor, and nonplanning [29]. It has previously been used to assess the effect of impulsivity on food consumption [30,31]. All participants completed the BIS during their first experimental session to assess potential baseline impulsivity differences between conditions, as this can affect IC [32] and unhealthy food consumption choices [33]. No significant differences were found, and thus the BIS was not included in further analysis.

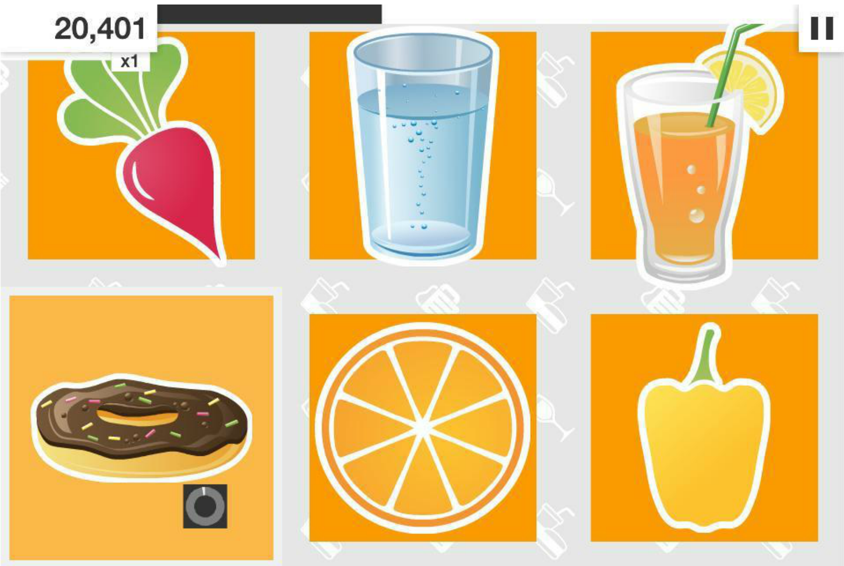
Training Task

The training was delivered via the NoGo iOS app, developed by Neurocog Solutions Pty Ltd, Australia. The NoGo app contains IC training in 3 domains (unhealthy eating, smoking, alcohol consumption)—the training mechanism and parameters were designed by SJ informed by IC training literature. The training task was based on modified Go/Nogo and stop-signal tasks, with each "game" containing (1) Go and Nogo trials in which the image remained the same when the reaction time deadline (RTD) timer appeared next to the image and started to count down and (2) stop trials in which the image could change from healthy to unhealthy after the timer appeared. Participants were instructed to tap the images of healthy food as quickly and accurately as possible before the RTD expired and to refrain from tapping images of unhealthy food. Each game consisted of 30 trials, using stimuli drawn from a pool of food-specific images. For the purpose of creating a clear divide between different food types, the presented stimuli were in 2 categories: healthy (eg, fruits, vegetables) and unhealthy (eg, chips, doughnuts) foods (Figure 2).

Participants were required to play 10 games per day for 14 consecutive days with each game taking approximately 1 minute to complete. The difficulty level of the game increased according to past performance by reducing the RTD and increasing the number of images presented at one time (maximum 12 images). Go stimuli were presented on 70%-90% of occasions, varying randomly each game to ensure individuals were responding genuinely to presented stimuli and were not prepreparing responses based on previous patterns of stimulus presentation.

A log of performance data including reaction time, game level, correct responses, and errors was stored locally and accessed by researchers at the end of the training period. For analysis purposes, and in order to assess potential improvements in NoGo app performance as game play progressed, the training data of each individual participant were split into 3 sessions. Sessions 1, 2, and 3 comprised an equal number of games for each participant (eg, if a participant completed 120 games, session 1 would consist of data collected from games 1 to 40, session 2 would consist of games 41 to 80, and session 3 would consist of games 81 to 120). Reaction time, game level, and correct/incorrect response scores were averaged for each participant across each session to create 3 data points, which were used to assess improvements in game-play performance across all participants. Although participants in the training condition were expected to play a total of 140 games, an arbitrary cutoff of 90 games was selected, reflecting a 65% adherence rate. Participants who failed to play 90 games over the 14-day period were excluded from further analysis.

Figure 2. Example of the NoGo app environment with "unhealthy" (doughnut, orange soft drink) and "healthy" (radish, capsicum, water, orange) image categories. This example shows level 6—there are 6 images shown simultaneously, with the active image (requiring a response or not) indicated by the reaction time deadline timer (below doughnut).



Healthy Eating Quiz

The Healthy Eating Quiz (HEQ) is a food frequency questionnaire containing 70 items relating to the frequency of healthy food consumption over the past 7 days, rated on a Likert scale from "never" to "5 or more times a week." It includes 8 subgroups: fruits, vegetables, meat proteins and vegetarian proteins, grains, water, dairy, and extras (ie, sauces and spreads). Participants were asked to rate approximately how many servings of items within these food groups they had consumed in the past week. Food frequency questionnaires are widely used to assess eating habits [34,35].

Food Consumption Test

The food consumption test (FCT) was used to examine immediate changes in food consumption after the IC training. It was based on the bogus taste tests used in previous research but without the associated element of deception [2,7,10]. Participants were informed that they would receive a refreshment break during the session in which snacks would be offered and that they could eat as little or as much as they liked. Participants were presented with 4 bowls containing snack-sized portions of 2 healthy (156 g of unsalted nuts and 216 g of grapes) and 2 unhealthy food options (60 g of plain potato chips and 161 g of chocolate candies, M&M's). Although nuts and grapes are still high-calorie choices, these foods were chosen as they represent "real-world" choices that participants would be faced with. For example, although grapes are still high in sugar, they are high in natural sugar and are nonprocessed. Thus, they represent a healthier choice than M&M's (which are processed and high in sugar and saturated fat) for individuals when craving a convenient, sweet snack food. Food was presented for 10 minutes while the experimenter cleaned equipment used previously in the experimental session, facing away from participants. After the participant left the laboratory, the food was weighed, consumption was recorded, and the number of calories consumed was calculated. Before completing the FCT, participants were asked covertly (questions were embedded in a larger general health questionnaire) to rate their current levels of hunger and food craving to account for differences in the FCT results between conditions. As no significant differences were found for hunger $F_{1,48}$=0.24, P=.05 or craving $F_{1,48}$=0.00, P=.99, these factors were not included as covariates in further analysis.

Three-Factor Eating Questionnaire

Participants completed the 51-item TFEQ [36] that contained a combination of true or false questions (eg, "Sometimes when

I start eating, I just can't seem to stop") and items rated on a 4-point Likert scale (eg, "How frequently do you avoid 'stocking up' on tempting foods?" Rated: not at all, slightly, moderately, or almost always). The TFEQ contains 3 subscales (cognitive restraint, disinhibition, and hunger) measuring the ability to resist foods and overcome internal and external food cues. The TFEQ has been widely used in overweight and obese populations [26,27] and is psychometrically valid in obese and nonobese populations [37].

Go/Nogo Task

Two auditory Go/Nogo tasks that varied in RTD were used to assess general IC ability. Tones were presented at 1100 Hz and 2000 Hz for 200 milliseconds, with an interstimulus interval (ISI) of either 2500 milliseconds (longer RTD) or 1250 milliseconds (shorter RTD). These specific time intervals were pilot-tested for this study in an adult population. Shorter RTDs increase task demands by inducing rapid responding to Go stimuli making the inhibition of motor responses on infrequent "Nogo" trials more difficult and increase the likelihood of errors [13]. As a result, this study used 2 separate tasks with varying task demands ("longer RTD" and "shorter RTD") to better assess potential effects of the IC training. Tasks were created in Presentation version 11, and tones were presented at 60 dB binaurally through Sennheiser HD 201 headphones. Participants were required to press a response button on a game controller upon presentation of the frequent "Go" tone (70% of presented stimuli) and to withhold response upon presentation of the infrequent "Nogo" tone. Participants completed 10 practice trials to ensure they understood task instructions and then completed 100 trials, taking approximately 2.5 minutes to complete. Assignment of each tone as Go/Nogo, and the order of tasks, was counterbalanced across sessions and participants. This task was used to assess how the IC training affected general IC ability by using a task involving similar processes but sharing few surface features and different modality (auditory rather than visual) to reduce practice effects and make comparison with the control group more equal.

Procedure

In the first experimental session (around 90 minutes) participants gave informed consent and then completed the BIS, Go/Nogo tasks, FCT, HEQ, and TFEQ. Participants also completed a passive image-viewing task, not reported here. After completing this session, participants in the training condition were given access to the training app and instructed to play 10 games a day for 14 consecutive days. Participants performed a trial game with the researchers to provide them with the opportunity to ask questions and understand the game format. Participants were then sent reminders to play the game via email on days 1, 7, and 14 to ensure compliance. During the second session, approximately 3 weeks since session 1 and 1 week since the end of the training (average time between sessions was 4.5 weeks but did not differ between conditions), individuals repeated the same procedure, game data were obtained from the training participants, and the waitlist-control condition received access to the training app. Participants were entered into a prize draw to win 1 of 4 A$100 gift vouchers after completing the study, as reimbursement for their time.

Event-Related Potential Recording and Quantification

Electroencephalographic data were recorded at 9 sites (F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4) using a 19-site cap and a Nuamps amplifier (Compumedics, Melbourne Australia). Vertical and horizontal electro-oculogram was recorded using electrodes placed above and below the left eye and near the outer canthus of each eye. Linked ears were used as a reference, and an electrode located between Fz, Fp1, and Fp2 acted as a ground electrode, with impedances kept below 5 kΩ. Signals were amplified 19 times with a band-pass down 3 dB at 0.01 and 100 Hz. Raw EEG data were visually inspected and sections of muscle artifact were removed. Data were low-pass filtered down 3 dB at 24 Hz and divided into epochs from −100 milliseconds prestimulus to 900 milliseconds poststimulus. Epochs were excluded if they contained amplitudes outside ±100 µV at any EEG site. The number of accepted epochs for Go and Nogo stimuli were compared between conditions, with no differences evident. ERPs were calculated separately for correct Go and Nogo trials. Grand mean ERPs were visually inspected to identify major peaks. Peak quantification was completed via a computer algorithm, which allows for the automatic identification of the maximum or minimum voltage occurring within a specified latency window. For N2, peak identification at all other sites was locked to the largest negativity at Fz in the 190- to 300-millisecond latency window, whereas the P3 was locked to the largest positivity at Pz in the 300- to 580-millisecond window as per Johnstone et al [38].

Statistical Analysis

One-way analysis of variance (ANOVA) was used to assess baseline differences between training and waitlist conditions and to assess NoGo training app performance data. Additionally, time (1, 2) × condition (waitlist, training) mixed design ANOVA was used to assess the FCT, HEQ, and TFEQ, with planned follow-up analysis splitting data by condition and assessing the effect of time. ERP component latency analysis was restricted to the frontal midline location (Fz), as the N2 and P3 components showed a frontal maximum. Go and Nogo ERP data were subject to a condition × time × stimulus (Go, Nogo) × RTD (longer, shorter) mixed factorial ANOVA. ERP amplitude analyses included an additional lateral (left, midline, right) factor. Planned contrasts within the lateral factor compared the right and left hemispheres. Here we focus on interactions between time and condition, and time × condition × laterality for ERP amplitude analysis. Alpha level was set to .05 for all analyses.

Results

Baseline Characteristics

Table 1 shows the analysis comparing the waitlist and training groups at session 1. As participants did not differ statistically in any of these variables, they were not included as covariates in future analysis. Both groups had equal numbers of male and female participants (19% male, 11/58).

Table 1. Baseline analysis results for training and waitlist conditions.

Demographics		Training		Waitlist		*F* (*df*)	*P* value
		Mean	SD	Mean	SD		
Age, years		35.2	14.1	38.3	14.8	0.56	.46
Body mass index		29.7	4.4	29.2	3.5	0.19	.67
Return time, days		33.5	15.5	31.2	12.5	0.34 (1,46)	.56
BIS [a] - subscale							
	Attentional	17.9	3.4	17.0	2.8	1.03	.32
	Motor	21.6	3.8	22.3	3.7	0.40	.53
	Nonplanning	23.5	3.9	22.6	5.2	0.54	.46

[a] BIS: Barratt Impulsiveness Scale.

Training Performance

For participants completing the NoGo training, 1-way repeated measures ANOVA was used to assess changes in reaction time across the 3 training session blocks. The main effect of time was significant ($F_{2,46}$=156.80, P<.001, partial η^2=.87). Planned contrasts indicated that average reaction time was reduced at session 2 (mean 330.8 milliseconds, SD 26.03) compared with session 1 (mean 401.7 milliseconds, SD 44.7), $F_{1,23}$=133.26, P<.001, partial η^2=.85, and reduced at session 3 (mean 307.1 milliseconds, SD 33.2) compared with session 2, $F_{1,23}$=35.53, P<.001, partial η^2=.61.

The main effect of time was also significant for correct Go responses ($F_{2,46}$=82.78, P<.001, partial η^2=.79). Planned contrasts indicated that session 2 (mean 53.35, SD 5.48) had more correct Go responses than session 1 (mean 48.22, SD 1.65), $F_{1,23}$=24.11, P<.001, partial η^2=.52, and session 3 (mean 62.46, SD 5.14) had more correct Go responses than session 2, $F_{1,23}$=43.37, P<.001, partial η^2=.66. The main effect was also significant for correct Nogo responses ($F_{2,46}$=93.07, P<.001, partial η^2= .79). There were more correct Nogo responses in session 2 (mean 27.83, SD 9.51) than session 1 (mean 17.57, SD 3.85), $F_{1,23}$=53.54, P<.001, partial η^2=.69, and more correct Nogo responses in session 3 (mean 42.22, SD 14.62) than session 2, $F_{1,23}$=80.19, P<.001, partial η^2=.77. Overall, these results indicate that participants in the training condition showed improved IC performance with increased training.

Healthy Eating Quiz

The total HEQ score showed a time x condition interaction ($F_{1,48}$=9.87, P=.003, partial η^2=.17). Simple effects analysis showed a significant effect of time for the training condition ($F_{1,25}$=5.21, P=.03, partial η^2=.03), with an increase in healthy food consumption from time 1 (mean 36.96, SD 8.45) to time 2 (mean 42.42, SD 9.98). The waitlist condition also showed a significant time effect ($F_{1,25}$=7.88, P=.01, partial η^2=.01), with a decrease in healthy food consumption between time 1 (mean 39.58, SD 9.68) and time 2 (mean 35.88, SD 11.24).

Food Consumption Test

Calorie consumption was calculated for grapes, mixed unsalted nuts, M&M's, and plain potato chips separately. Then, the “healthier” foods (nuts and grapes) and “less healthy” foods (chips and M&M's) were added together to create 2 food groups for analysis. The condition × food × time ANOVA was significant ($F_{1,48}$=8.88, P=.005, partial η^2= .18). Follow-up 2-way ANOVA for each condition separately showed no significant interaction for the waitlist condition but a significant food × time interaction for training condition ($F_{1,19}$=8.81, P=.008 partial η^2= .32). This showed an increase in calorie consumption of healthier foods at time 2 (mean 63.17, SD 19.08) compared with time 1 (mean 31.59, SD 9.47) and a reduction in total calorie consumption of less healthy foods from time 1 (mean 234.68, SD 74.40) to time 2 (mean 76.50, SD 25.15).

Three-Factor Eating Questionnaire

The disinhibition subscale did not show a time x condition interaction. A significant interaction was found for the hunger subscale ($F_{1,50}$=4.07, P=.05, partial η^2=.08). Simple effects analysis showed no time effect for the training condition, whereas the waitlist condition showed a significant time effect, $F_{1,25}$=10.74, P=.003, partial η^2=.30, with an increase between time 1 (mean 5.58, SD 3.46) and time 2 (mean 6.88, SD 3.56). The cognitive restraint subscale also showed a significant time x condition interaction ($F_{1,50}$=6.02, P=.02, partial η^2=.11). Simple effects analysis indicated no significant time effect for the waitlist condition, but a significant effect was present for the training condition, $F_{1,25}$=6.01, P=.02, partial η^2=.19, with an increase from time 1 (mean 9.27, SD 3.38) to time 2 (mean 11.5, SD 4.25).

Go/Nogo Task Performance

For the shorter RTD Go/Nogo task, the time x condition interaction was not significant for correct Go responses or reaction time, or Nogo errors. No significant interactions were found for the same measures on the longer RTD Go/Nogo task.

Event-Related Potential Components

Grand mean ERPs to Go and Nogo stimuli for each condition and time are presented separately for the longer RTD (Figure 3) and shorter RTD (Figure 4) tasks. For the latency analysis, no significant interactions were found for either component. For the amplitude analysis, there were no significant relevant interactions for the N2 component. A significant condition × time × laterality × RTD interaction was found for P3 ($F_{1,46}$=8.49, P=.005, partial η^2=.16; Figure 5). The waitlist condition showed a reduction (largest in the left frontal region) in P3 amplitude at time 2 compared with time 1 for the longer ISI task and a very minor reduction (largest in the right frontal region) at time 2 compared with time 1 for the shorter ISI task. The training condition showed a very different pattern, with a substantial increase (of similar magnitude in both frontal hemispheres) in P3 amplitude at time 2 compared with time 1 for the longer ISI task and a smaller increase (mainly in the right frontal region) at time 2 compared with time 1 for the shorter ISI task.

Figure 3. Grand mean event-related potentials at Fz for Go and Nogo stimuli in the longer reaction time deadline task. W1: waitlist time 1; W2: waitlist time 2; T1: = training time 1; T2: training time 2.

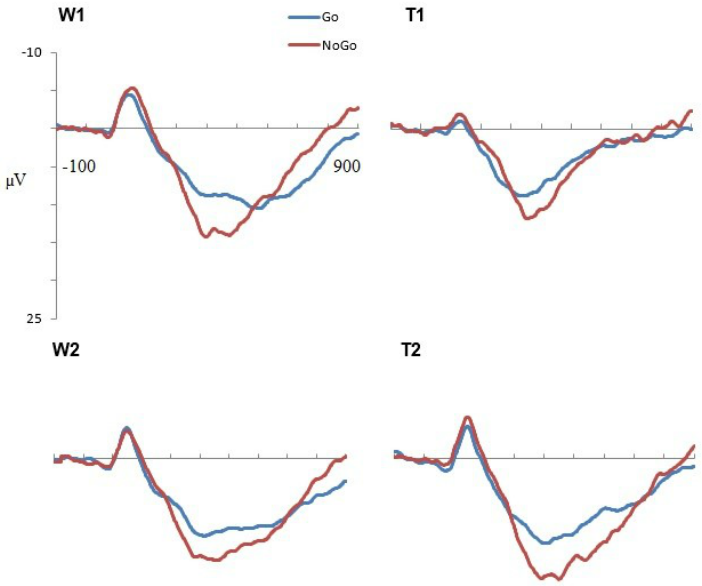


Figure 4. Grand mean event-related potentials at Fz for Go and Nogo stimuli in the shorter reaction time deadline task. W1: waitlist time 1; W2: waitlist time 2; T1: training time 1; T2: training time 2.

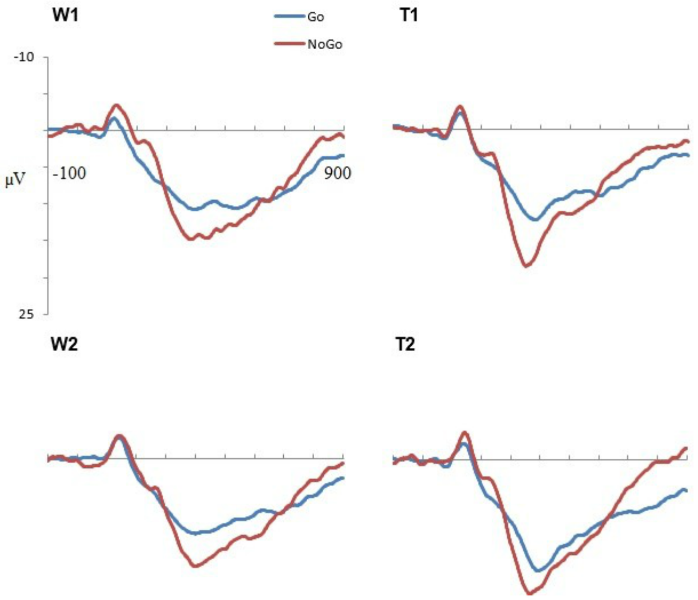
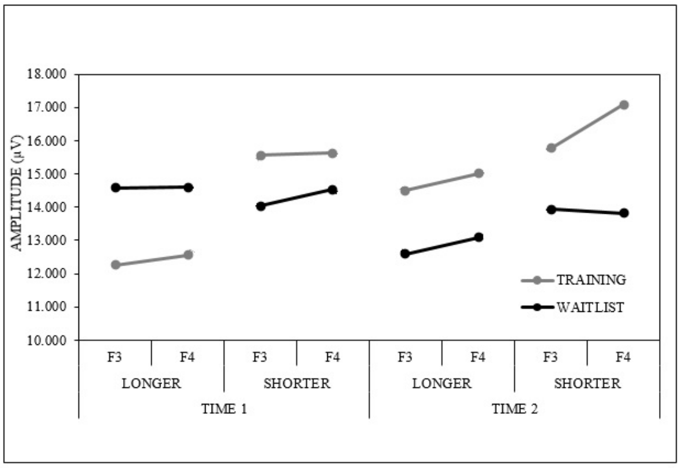


Figure 5. P3 amplitude changes in the Go/Nogo tasks. P3 amplitude at sites F3 and F4 for the Go/Nogo tasks, with each reaction time deadline and time shown separately.



Results Summary

The key significant effects for both groups are presented in Table 2, showing mean changes between time 1 and time 2.

Table 2. Mean scores for each outcome; only statistically significant results are shown.

Variable	Participant group			
	Training		Waitlist	
	Baseline	Time 2	Baseline	Time 2
Healthy Eating Quiz	36.96	42.42	39.58	35.88
FCT[a]: healthy food	63.17	31.59	-	-
FCT: unhealthy food	234.68	76.50	-	-
TFEQ[b]: hunger	-	-	5.58	6.88
TFEQ: cognitive restraint	9.27	11.50	-	-

[a] FCT: food consumption test.

[b] TFEQ: Three-Factor Eating Questionnaire.

Discussion

Principal Findings

This pilot study used a randomized waitlist-control design to examine the efficacy of food-specific IC training in overweight and obese individuals. Traditional measures were used to assess food consumption (HEQ, FCT) and attitudes toward food and dieting (TFEQ). As a measure of general IC ability, Go/Nogo performance data and ERPs were also assessed. Key findings included increased healthy food consumption outside laboratory environments, decreased unhealthy food consumption during testing, and an increase in scores on the cognitive restraint scale for the training group. Condition-specific changes in the P3 ERP component indicated that the long-term, food-specific, IC training had a generalizable effect on IC-related processing, at least in the context of this preliminary research.

This study was novel and exploratory in its linking of healthy foods with approach responses in the context of an intensive multisession training regimen. As hypothesized, a reduction in immediate unhealthy food consumption and an increase in healthy food consumption (both in the laboratory and 1 week after training) was found for participants who completed the training task, supporting previous research [5,10,17,18]. Increases in the HEQ score indicated that individuals reported an increase in healthy food consumption in the week after using the intervention, eating more fruits, vegetables, and protein or drinking more water. This provides encouraging evidence that IC training may be effective in priming approach behaviors toward healthy food, as well as strengthening resistance toward unhealthy foods. Furthermore, the results provide some evidence that these changes may have translated to the participant's day-to-day life. However, a longer-term unhealthy food consumption measure was not included in this initial research and will be necessary in the future to determine whether decreases in unhealthy food consumption are maintained.

To complement the food consumption data, attitudes toward food were also measured. Participants in the training condition showed increased cognitive restraint, a construct related to self-regulation of food intake on the TFEQ scale. Increased cognitive restraint has been linked to weight loss in dieters and improvements in avoiding unwanted food consumption [26]. Bryant and colleagues [27] showed that a weight loss intervention developed cognitive restraint in some participants, and this resulted in increased weight loss for these participants. Although our study did not monitor weight loss, this increased ability to control calorie intake may assist individuals to achieve weight loss goals, especially as it can easily be used alongside diets such as meal plans [2].

Our study incorporated psychophysiological methods to measure the neural correlates of IC training. Contrary to prediction, N2 amplitude did not increase in the training condition as a result of IC training. This indicates that the IC training did not have a broad effect on the process reflected by the N2 component at either difficulty level. This may be due to the N2 being more closely related to conflict monitoring (a process not targeted by the IC training), as opposed to IC [39,40].

Training effects were present for the P3 ERP component, with amplitude increases after training seen in the training but not the waitlist conditions. This effect could be described as "global" as it did not differ based on the stimulus type (ie, Go, Nogo) and was consistent across the longer and shorter RTD Go/Nogo tasks. However, the nature of the condition effect depended on task difficulty; the shorter ISI task places higher demands on IC as a more speeded response is required to Go stimuli and could be considered more difficult. In the easier Go/Nogo task, the training condition showed a substantial increase (of similar magnitude in left and right frontal regions) in P3 amplitude at time 2, whereas the waitlist condition showed a substantial reduction (largest in the left frontal region). In the more difficult Go/Nogo task, the training condition showed a minor increase in P3 amplitude (largest in the right frontal region), whereas the waitlist condition showed a minor reduction (largest in the right frontal region). In both tasks, rapid responding was induced and thus inhibitory "activation" was required to inhibit automatic motor responses [41]. Given the established relationship between the P3 ERP component and cancellation of a planned response [42,43], these training effects for P3 are consistent with improved IC abilities, with the degree of improvement more evident in the easier of the 2 tasks.

Limitations

The results reported here should be considered in light of several limitations, primarily because of the limited nature of this preliminary study. Although all attempts were made to recruit adolescent participants, the final sample only included participants older than 18 years, preventing analysis across different age groups. Further research investigating the ideal age to administer IC training is warranted. Adolescents may experience greater benefits from the training, as IC abilities are known to continue developing into adolescence [44]. Other variables that affect food consumption, such as beverage intake [45], mental illness comorbidities [46], and menstrual cycles for women [47], were not assessed and may have affected the outcomes of this study. Race and socioeconomic status were also not recorded, although these can affect access to food and obesity rates [48-50]. As the majority of participants were female, our results cannot be generalized across sex. Females and males have been shown to consume unhealthy food at different rates, and for different reasons, particularly when snacking [51].

Additionally, the study duration was relatively short and did not use weight loss as an outcome measure. A longer-term study that assesses weight loss and maintenance would be beneficial, as previous research has found mixed results for changes in BMI at follow-up [18,19] and additional research is needed to determine how IC training can influence health behaviors. An additional limitation is related to the stimuli used in the NoGo training app versus the foods presented in the FCT. Although grapes were featured in both the training task and the FCT, M&M's, potato chips, and nuts were not. Increases in the consumption of grapes may therefore potentially be attributed to grapes being specifically featured in the training task. Future versions of the FCT should utilize either foods that are all featured in the training task or none at all to ascertain if training on specific stimuli can transfer to general stimuli.

Most importantly, this study has focused on self-report measures, which are inherently biased and vulnerable to demand effects. Although the waitlist-control design helps to reduce the effects of random error, the nonblind randomization of participants means that individuals who know they are in a waitlist versus training group may not respond as they normally would. This may explain why the participants in the waitlist condition ate significantly less healthy food in the HEQ at time 2 and may explain the increase in the disinhibition subscale of the TFEQ. This subscale is linked to making poorer health food choices [52], which may be the result of the control group allowing themselves to "give in" to their food cravings, knowing that they will be receiving a healthy eating intervention after the experimental period. In the same way, those in the training group could easily deduce that experimenters expected them to eat more healthy foods after the training and may have been acting in accordance with this expectation, not due to the training itself. Although utilizing an active control condition (eg, having participants read information about healthy eating choices) is a potential way to reduce demand effects, nonblind assignment to active control groups may still reveal bias results, especially for gaming-based interventions [53].

Future research must be especially careful in controlling for the expectations and beliefs of both the training and waitlist groups or measure these expectations to examine potential differences that may confound results. Further studies therefore need to consider a way of incorporating a more active control group, which will help make the purpose of the intervention less apparent to participants, reducing bias effects. Alternative food consumption measures, such as food diaries or dietary interviews, would also improve the accuracy of these data, although this was beyond the scope of this study. A larger number of participants and an equal number of male and female participants would improve statistical power.

Conclusions

This study, while preliminary, replicated and extended previous food-related IC training research, by priming both inhibitory responses to unhealthy food and approach responses toward healthy foods. Inhibitory control training was delivered through a purpose-designed, mobile phone app that allowed participants to complete the training at a time and place convenient to them. Further research is required to assess if the observed changes transfer to longer-term real-life contexts, such as weight loss. As it is easy to complete and cheap to obtain, IC training is a promising alternative or addition to existing obesity interventions.

Acknowledgments

We gratefully acknowledge the individuals who participated in this study. Neurocog Solutions Pty Ltd provided the NoGo app to research participants at no cost.

Conflicts of Interest

SJ provided the training mechanism and parameters for the NoGo app to Neurocog Solutions Pty Ltd (Australia) at no cost. SJ has no financial interest in the NoGo app. SJ is a co-inventor of intellectual property licensed by the University of Wollongong (UOW) to Neurocog Solutions Pty Ltd and is entitled to a small portion of royalties received by UOW in relation to the sale of any product that uses the UOW intellectual property. This intellectual property makes up a proportion of the intellectual property used in another Neurocog Solutions product (Focus Pocus) but not NoGo. TB and AR have no financial association with Neurocog Solutions Pty Ltd.

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Larsen JK, Hermans RC, Engels RC. Food intake in response to food-cue exposure. Examining the influence of duration of the cue exposure and trait impulsivity. Appetite 2012 Jun;58(3):907-913. [doi: 10.1016/j.appet.2012.02.004] [Medline: 22326507]
3. Nederkoorn C, Smulders FT, Havermans RC, Roefs A, Jansen A. Impulsivity in obese women. Appetite 2006 Sep;47(2):253-256. [doi: 10.1016/j.appet.2006.05.008] [Medline: 16782231]
4. Jasinska AJ, Yasuda M, Burant CF, Gregor N, Khatri S, Sweet M, et al. Impulsivity and inhibitory control deficits are associated with unhealthy eating in young adults. Appetite 2012 Dec;59(3):738-747 [FREE Full text] [doi: 10.1016/j.appet.2012.08.001] [Medline: 22885454]
5. Houben K, Jansen A. Training inhibitory control. A recipe for resisting sweet temptations. Appetite 2011 Apr;56(2):345-349. [doi: 10.1016/j.appet.2010.12.017] [Medline: 21185896]
6. National Health and Medical Research Council. 2013. Australian Dietary Guidelines URL: https://www.nhmrc.gov.au/guidelines-publications/n55 [accessed 2016-07-06] [WebCite Cache ID 6ioQJX3ia]
7. Houben K. Overcoming the urge to splurge: influencing eating behavior by manipulating inhibitory control. J Behav Ther Exp Psychiatry 2011 Sep;42(3):384-388. [doi: 10.1016/j.jbtep.2011.02.008] [Medline: 21450264]
8. Del Corral P, Bryan DR, Garvey WT, Gower BA, Hunter GR. Dietary adherence during weight loss predicts weight regain. Obesity (Silver Spring) 2011 Jun;19(6):1177-1181 [FREE Full text] [doi: 10.1038/oby.2010.298] [Medline: 21164500]
9. Hwang K, Ghuman AS, Manoach DS, Jones SR, Luna B. Cortical neurodynamics of inhibitory control. J Neurosci 2014 Jul 16;34(29):9551-9561 [FREE Full text] [doi: 10.1523/JNEUROSCI.4889-13.2014] [Medline: 25031398]
10. Guerrieri R, Nederkoorn C, Jansen A. Disinhibition is easier learned than inhibition. The effects of (dis)inhibition training on food intake. Appetite 2012 Aug;59(1):96-99. [doi: 10.1016/j.appet.2012.04.006] [Medline: 22521403]
11. Allan JL, Johnston M, Campbell N. Unintentional eating. What determines goal-incongruent chocolate consumption? Appetite 2010 Apr;54(2):422-425. [doi: 10.1016/j.appet.2010.01.009] [Medline: 20100530]
12. van Koningsbruggen GM, Stroebe W, Papies EK, Aarts H. Implementation intentions as goal primes: Boosting self-control in tempting environments. Eur J Soc Psychol 2011 May 03;41(5):551-557. [doi: 10.1002/ejsp.799]
13. Benikos N, Johnstone SJ, Roodenrys SJ. Short-term training in the Go/Nogo task: behavioural and neural changes depend on task demands. Int J Psychophysiol 2013 Mar;87(3):301-312. [doi: 10.1016/j.ijpsycho.2012.12.001] [Medline: 23247193]
14. Jones A, Christiansen P, Nederkoorn C, Houben K, Field M. Fluctuating disinhibition: implications for the understanding and treatment of alcohol and other substance use disorders. Front Psychiatry 2013;4:140 [FREE Full text] [doi: 10.3389/fpsyt.2013.00140] [Medline: 24155728]
15. Bowley C, Faricy C, Hegarty B, Johnstone SJ, Smith JL, Kelly PJ, et al. The effects of inhibitory control training on alcohol consumption, implicit alcohol-related cognitions and brain electrical activity. Int J Psychophysiol 2013 Sep;89(3):342-348. [doi: 10.1016/j.psycho.2013.04.011]
16. Luijten M, Machielsen MW, Veltman DJ, Hester R, de Haan L, Franken IH. Systematic review of ERP and fMRI studies investigating inhibitory control and error processing in people with substance dependence and behavioural addictions. J Psychiatry Neurosci 2014 May;39(3):149-169 [FREE Full text] [Medline: 24359877]

17. Veling H, Aarts H, Papies EK. Using stop signals to inhibit chronic dieters' responses toward palatable foods. Behav Res Ther 2011 Nov;49(11):771-780. [doi: 10.1016/j.brat.2011.08.005] [Medline: 21906724]
18. Allom V, Mullan B. Two inhibitory control training interventions designed to improve eating behaviour and determine mechanisms of change. Appetite 2015 Jun;89:282-290. [doi: 10.1016/j.appet.2015.02.022] [Medline: 25725487]
19. Veling H, van Koningsbruggen GM, Aarts H, Stroebe W. Targeting impulsive processes of eating behavior via the internet. Effects on body weight. Appetite 2014 Jul;78:102-109. [doi: 10.1016/j.appet.2014.03.014] [Medline: 24675683]
20. Kendzor DE, Shuval K, Gabriel KP, Businelle MS, Ma P, High RR, et al. Impact of a Mobile Phone Intervention to Reduce Sedentary Behavior in a Community Sample of Adults: A Quasi-Experimental Evaluation. J Med Internet Res 2016 Jan;18(1):e19 [FREE Full text] [doi: 10.2196/jmir.5137] [Medline: 26810027]
21. Willey S, Walsh JK. Outcomes of a Mobile Health Coaching Platform: 12-Week Results of a Single-Arm Longitudinal Study. JMIR Mhealth Uhealth 2016 Jan;4(1):e3 [FREE Full text] [doi: 10.2196/mhealth.4933] [Medline: 26747611]
22. Ruchsow M, Groen G, Kiefer M, Hermle L, Spitzer M, Falkenstein M. Impulsiveness and ERP components in a Go/Nogo task. J Neural Transm (Vienna) 2008 Jun;115(6):909-915. [doi: 10.1007/s00702-008-0042-7] [Medline: 18368285]
23. Watson TD, Garvey KT. Neurocognitive correlates of processing food-related stimuli in a Go/No-go paradigm. Appetite 2013 Dec;71:40-47. [doi: 10.1016/j.appet.2013.07.007] [Medline: 23892319]
24. Drollette ES, Scudder MR, Raine LB, Moore RD, Saliba BJ, Pontifex MB, et al. Acute exercise facilitates brain function and cognition in children who need it most: an ERP study of individual differences in inhibitory control capacity. Dev Cogn Neurosci 2014 Jan;7:53-64 [FREE Full text] [doi: 10.1016/j.dcn.2013.11.001] [Medline: 24309300]
25. Sarlo M, Ubel S, Leutgeb V, Schienle A. Cognitive reappraisal fails when attempting to reduce the appetitive value of food: an ERP study. Biol Psychol 2013 Dec;94(3):507-512. [doi: 10.1016/j.biopsycho.2013.09.006] [Medline: 24056130]
26. Chaput J, Drapeau V, Hetherington M, Lemieux S, Provencher V, Tremblay A. Psychobiological impact of a progressive weight loss program in obese men. Physiol Behav 2005 Sep 15;86(1-2):224-232. [doi: 10.1016/j.physbeh.2005.07.014] [Medline: 16112692]
27. Bryant EJ, Caudwell P, Hopkins ME, King NA, Blundell JE. Psycho-markers of weight loss. The roles of TFEQ Disinhibition and Restraint in exercise-induced weight management. Appetite 2012 Feb;58(1):234-241. [doi: 10.1016/j.appet.2011.09.006] [Medline: 21983045]
28. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [Medline: 17695343]
29. Patton JH, Stanford MS, Barratt ES. Factor structure of the Barratt impulsiveness scale. J Clin Psychol 1995 Nov;51(6):768-774. [Medline: 8778124]
30. Leitch MA, Morgan MJ, Yeomans MR. Different subtypes of impulsivity differentiate uncontrolled eating and dietary restraint. Appetite 2013 Oct;69:54-63. [doi: 10.1016/j.appet.2013.05.007] [Medline: 23702263]
31. Meule A. Impulsivity and overeating: a closer look at the subscales of the Barratt Impulsiveness Scale. Front Psychol 2013;4:177 [FREE Full text] [doi: 10.3389/fpsyg.2013.00177] [Medline: 23596432]
32. Shen I, Lee D, Chen C. The role of trait impulsivity in response inhibition: event-related potentials in a stop-signal task. Int J Psychophysiol 2014 Feb;91(2):80-87. [doi: 10.1016/j.ijpsycho.2013.11.004] [Medline: 24316151]
33. Patte K. Overeating and obesity: the role of impulsivity. Canada: York University; 2006.
34. Brown D. Do food frequency questionnaires have too many limitations? J Am Diet Assoc 2006 Oct;106(10):1541-1542. [doi: 10.1016/j.jada.2006.07.020] [Medline: 17000185]
35. Hu FB, Rimm E, Smith-Warner SA, Feskanich D, Stampfer MJ, Ascherio A, et al. Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. Am J Clin Nutr 1999 Feb;69(2):243-249 [FREE Full text] [Medline: 9989687]
36. Stunkard AJ, Messick S. The three-factor eating questionnaire to measure dietary restraint, disinhibition and hunger. J Psychosom Res 1985;29(1):71-83. [Medline: 3981480]
37. Cappelleri JC, Bushmakin AG, Gerber RA, Leidy NK, Sexton CC, Lowe MR, et al. Psychometric analysis of the Three-Factor Eating Questionnaire-R21: results from a large diverse sample of obese and non-obese participants. Int J Obes (Lond) 2009 Jun;33(6):611-620. [doi: 10.1038/ijo.2009.74] [Medline: 19399021]
38. Johnstone SJ, Pleffer CB, Barry RJ, Clarke AR, Smith JL. Development of Inhibitory Processing During the Go/NoGo Task: A Behavioral and Event-Related Potential Study of Children and Adults. Journal of Psychophysiology 2005 Jan;19(1):11-23. [doi: 10.1027/0269-8803.19.1.11]
39. Dimoska A, Johnstone SJ, Barry RJ. The auditory-evoked N2 and P3 components in the stop-signal task: indices of inhibition, response-conflict or error-detection? Brain Cogn 2006 Nov;62(2):98-112. [doi: 10.1016/j.bandc.2006.03.011] [Medline: 16814442]
40. Donkers FC, van Boxtel GJ. The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. Brain Cogn 2004 Nov;56(2):165-176. [doi: 10.1016/j.bandc.2004.04.005] [Medline: 15518933]
41. Boucher L, Palmeri TJ, Logan GD, Schall JD. Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychol Rev 2007 Apr;114(2):376-397. [doi: 10.1037/0033-295X.114.2.376] [Medline: 17500631]
42. Smith JL, Johnstone SJ, Barry RJ. Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. Clin Neurophysiol 2008 Mar;119(3):704-714. [doi: 10.1016/j.clinph.2007.11.042] [Medline: 18164657]

43. Randall WM, Smith JL. Conflict and inhibition in the cued-Go/NoGo task. Clin Neurophysiol 2011 Dec;122(12):2400-2407. [doi: 10.1016/j.clinph.2011.05.012] [Medline: 21715225]
44. Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O'Reilly RC. A unified framework for inhibitory control. Trends Cogn Sci 2011 Oct;15(10):453-459 [FREE Full text] [doi: 10.1016/j.tics.2011.07.011] [Medline: 21889391]
45. Ruff RR, Akhund A, Adjoian T, Kansagra SM. Calorie intake, sugar-sweetened beverage consumption, and obesity among New York City adults: findings from a 2013 population study using dietary recalls. J Community Health 2014 Dec;39(6):1117-1123. [doi: 10.1007/s10900-014-9865-3] [Medline: 24671367]
46. Jorm AF, Korten AE, Christensen H, Jacomb PA, Rodgers B, Parslow RA. Association of obesity with anxiety, depression and emotional well-being: a community survey. Aust N Z J Public Health 2003;27(4):434-440. [Medline: 14705308]
47. Klump KL, Keel PK, Racine SE, Burt SA, Burt AS, Neale M, et al. The interactive effects of estrogen and progesterone on changes in emotional eating across the menstrual cycle. J Abnorm Psychol 2013 Feb;122(1):131-137 [FREE Full text] [doi: 10.1037/a0029524] [Medline: 22889242]
48. Wong RJ, Chou C, Ahmed A. Long term trends and racial/ethnic disparities in the prevalence of obesity. J Community Health 2014 Dec;39(6):1150-1160. [doi: 10.1007/s10900-014-9870-6] [Medline: 24715435]
49. Backholer K, Mannan HR, Magliano DJ, Walls HL, Stevenson C, Beauchamp A, et al. Projected socioeconomic disparities in the prevalence of obesity among Australian adults. Aust N Z J Public Health 2012 Dec;36(6):557-563. [doi: 10.1111/j.1753-6405.2012.00885.x] [Medline: 23216497]
50. Birch E. The Role of Socioeconomic, Demographic and Behavioural Factors in Explaining the High Rates of Obesity Among Indigenous Australians. Australian Economic Papers 2015 Dec;54(4):209-228. [doi: 10.1111/1467-8454.12055]
51. Cleobury L, Tapper K. Reasons for eating 'unhealthy' snacks in overweight and obese males and females. J Hum Nutr Diet 2014 Aug;27(4):333-341. [doi: 10.1111/jhn.12169] [Medline: 24134077]
52. Boot WR, Simons DJ, Stothart C, Stutts C. The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. Perspect Psychol Sci 2013 Jul;8(4):445-454. [doi: 10.1177/1745691613491271] [Medline: 26173122]
53. Bryant EJ, King NA, Blundell JE. Disinhibition: its effects on appetite and weight regulation. Obes Rev 2008 Sep;9(5):409-419. [doi: 10.1111/j.1467-789X.2007.00426.x] [Medline: 18179615]

Abbreviations

ANOVA: analysis of variance
BIS: Barratt Impulsiveness Scale
BMI: body mass index
EEG: electroencephalogram
ERP: event-related potential
FCT: food consumption test
HEQ: Healthy Eating Quiz
IC: inhibitory control
ISI: interstimulus interval
RTD: reaction time deadline
TFEQ: Three-Factor Eating Questionnaire

Edited by M Graafland; submitted 29.02.16; peer-reviewed by D Kendzor, P Staccini; comments to author 08.04.16; revised version received 08.05.16; accepted 15.06.16; published 13.07.16.

Please cite as:
Blackburne T, Rodriguez A, Johnstone SJ
A Serious Game to Increase Healthy Food Consumption in Overweight or Obese Adults: Randomized Controlled Trial
JMIR Serious Games 2016;4(2):e10
URL: http://games.jmir.org/2016/2/e10/
doi:10.2196/games.5708
PMID:27417192

Original Paper

Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy

Jim Lumsden[1,2], MSc; Elizabeth A Edwards[3], MBBS; Natalia S Lawrence[4], PhD; David Coyle[5], PhD; Marcus R Munafò[1,2], PhD

[1]UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, United Kingdom

[2]MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, United Kingdom

[3]Centre for Primary Care and Public Health, Bart's & The London School of Medicine & Dentistry, London, United Kingdom

[4]Department of Psychology, College of Life and Environmental Sciences, University of Exeter, Exeter, United Kingdom

[5]School of Computer Science, University College Dublin, Dublin, Ireland

Corresponding Author:
Jim Lumsden, MSc
UK Centre for Tobacco and Alcohol Studies
School of Experimental Psychology
University of Bristol
12a Priory rd
Bristol, BS8 1TU
United Kingdom
Phone: 44 117 92 88450
Fax: 44 1179288588
Email: jim.lumsden@bristol.ac.uk

Abstract

Background: Cognitive tasks are typically viewed as effortful, frustrating, and repetitive, which often leads to participant disengagement. This, in turn, may negatively impact data quality and/or reduce intervention effects. However, gamification may provide a possible solution. If game design features can be incorporated into cognitive tasks without undermining their scientific value, then data quality, intervention effects, and participant engagement may be improved.

Objectives: This systematic review aims to explore and evaluate the ways in which gamification has already been used for cognitive training and assessment purposes. We hope to answer 3 questions: (1) Why have researchers opted to use gamification? (2) What domains has gamification been applied in? (3) How successful has gamification been in cognitive research thus far?

Methods: We systematically searched several Web-based databases, searching the titles, abstracts, and keywords of database entries using the search strategy (gamif* OR game OR games) AND (cognit* OR engag* OR behavi* OR health* OR attention OR motiv*). Searches included papers published in English between January 2007 and October 2015.

Results: Our review identified 33 relevant studies, covering 31 gamified cognitive tasks used across a range of disorders and cognitive domains. We identified 7 reasons for researchers opting to gamify their cognitive training and testing. We found that working memory and general executive functions were common targets for both gamified assessment and training. Gamified tests were typically validated successfully, although mixed-domain measurement was a problem. Gamified training appears to be highly engaging and does boost participant motivation, but mixed effects of gamification on task performance were reported.

Conclusions: Heterogeneous study designs and typically small sample sizes highlight the need for further research in both gamified training and testing. Nevertheless, careful application of gamification can provide a way to develop engaging and yet scientifically valid cognitive assessments, and it is likely worthwhile to continue to develop gamified cognitive tasks in the future.

(JMIR Serious Games 2016;4(2):e11) doi:10.2196/games.5888

KEYWORDS

gamification; gamelike; cognition; computer games; review

Introduction

Every day, millions of people play games, on computers, consoles, and mobile devices [1]. One explanation for the massive popularity of games is that they can provide easy access to a sense of engagement and self-efficacy which reality may not deliver [2]. By their nature, games present users with difficult challenges to overcome and use narrative structure, complex graphics, strategic elements, and intuitive rules to engage the user [3]. This ability to engross users has recently begun to be leveraged for purposes beyond entertainment in the form of gamification. The aim of gamification is to use gamelike features (competition, narrative, leaderboards, graphics, and other game design elements) to transform an otherwise mundane task into something engaging and even fun.

In cognitive science, whether we are gathering data or trying to encourage behavior change, successfully engaging participants is vital. There is evidence that a lack of participant motivation has a negative impact on the quality of data collected [4], with tasks commonly being viewed as too boring and repetitive for participant attention to be sustained. This is a particular problem for Web-based studies where participants can simply close the browser window and drop out of the study the moment they decide it is not worth their time [5].

If we are attempting to alter behavior (for example, cognitive bias modification or working memory training), then it is vital that our interventions are engaging. Gamification may help in this regard. If we can successfully import game design elements into cognitive tasks without undermining their scientific value, then we may be able to improve the quality of our data, increase the effectiveness of our interventions, and improve the experience for participants. Furthermore, using games as a vehicle to deliver cognitive training may also be advantageous simply because video games appear to have positive effects on a number of outcomes, including working memory, attentional capacity, problem solving, motivation, emotional control, and prosocial behaviors [6]. In essence, delivering targeted cognitive training through a video game medium might provide a range of benefits.

Much of the literature on gamification is relatively recent, cross-disciplinary, and heterogeneous in nature. This lack of coherence in the field is partly due to poor definition of terms; for example, the gamelike tasks covered in this review could be described as "serious games," "gamelike," "gamified," "games with a purpose," "gamed-up," or simply "computer based" [7-10]. To our knowledge, there have been no systematic reviews of gamification in cognitive research. There have, however, been several reviews of the concepts core to gamification and how they can be applied successfully (see [11-13]). However, detailed discussion of these reviews is beyond the scope of this paper. Also related but beyond our scope is the body of work on the effects of video games on behavior (see [14,15] for reviews).

This paper aims to systematically review the ways in which gamification has already been used for the purposes of cognitive assessment and training. We were specifically interested in the following questions: (1) Why have researchers opted to use gamification? (2) What domains has gamification been applied in? (3) How successful has gamification been in cognitive research thus far? We deliberately used a broad search strategy so as not to miss any relevant papers. We reviewed only the peer-reviewed literature and have therefore deliberately excluded some cognitive training games available on the iTunes or Play Store (such as Luminosity and Peak) unless supported by peer-reviewed research.

Methods

Databases and Search Strategy

The following databases were searched electronically: PsycInfo, Medline, ETHOS, EMBASE, PubMed, IBSS, Francis, Web of Science and Scopus. We searched the titles, abstracts, and keywords of database entries using the search strategy (gamif* OR game OR games) AND (cognit* OR engag* OR behavi* OR health* OR attention OR motiv*), where * represents a wildcard to allow for alternative suffixes. Searches included papers published in English between January 2007 and October 2015. We searched the bibliographies of included papers to locate further relevant material not discovered in the database search.

Inclusion Criteria

Primary Research Paper

Included papers were empirical research studies, not literature reviews, opinion pieces, or design documents.

Novel Gamelike Task

Included papers focussed on newly developed gamelike tasks, created specifically for the study in question. We excluded commercially available video games (ie, "off-the-shelf" games) as well as gamelike tasks that have been in use for many years, such as Space Fortress (see [16] for a review).

Measure or Train Cognition

Included papers focussed on tasks designed to assess or train cognition. For scoping purposes, we took a narrow definition of cognition: those processes involved in memory, attention, decision making, impulse control, executive functioning, processing speed, and visual perception.

Validated or Piloted

Included papers had to involve an empirical study, either validating the task as a measure of cognition or piloting the intervention. Papers regarding usability testing alone were excluded.

Exclusion Criteria

Non–Peer-Reviewed Papers

We excluded non–peer-reviewed papers such as abstracts or conference posters.

Gamification in the Behavioral Sciences but not Involving Cognition

We excluded papers on gamification for education purposes, disease management, health promotion, exposure therapy, or rehabilitation.

Game Engines/Three-Dimensional (3D) Environments

We excluded papers that made use of virtual reality or a 3D environment without any game mechanics or gamelike framing.

Screening

We did not select papers based on whether they included "gamification." Rather, papers were selected if they were captured by our search strategy and were considered relevant based on our inclusion/exclusion criteria. We intentionally did not strictly define gamification: as has been discussed famously by Wittgenstein and is alluded to by Deterding [13], trying to precisely define what elements make a game is both difficult and limiting. Therefore, we decided that a task was gamified if its stated purpose was to increase participant motivation. Where there was insufficient detail to determine whether a paper met our inclusion criteria, we erred on the side of caution to increase our confidence in the relevance of the studies reviewed.

After screening, data were extracted from each paper using a standardized data extraction form. Data relating to the questions of interest such as application of gamification, approach taken, and efficacy were extracted from each paper. Application of gamification refers to the field of psychology in which the gamelike task was used and why a gamelike task was used. Approach taken refers to the specific game mechanics used in the task and what themes and scaffolding were applied. Finally, efficacy refers to the findings produced by the task in practice, as well as details on the participants, methods used for evaluation, and limitations of the study. Categorization of concepts (such as the cognitive domains measured) was done using the paper-authors' own words where possible. Where not possible, we mapped extracted concepts closely to existing categories.

All papers identified by our search strategy were screened by one reviewer (JL) in 3 stages, to determine whether they were relevant based on our inclusion/exclusion criteria: title, abstract, and full text. A second reviewer (EE) rescreened 20% of the papers from the title stage onward to ensure that no relevant papers were missed. Papers were only included in the review on the agreement of both JL and EE.

Results

Search Results

Our initial search yielded 33,445 papers (excluding duplicates). Of these, 23 papers from the original search and 4 papers from the manual reference search were included in the review. We repeated the search in October 2015, including papers from January 2015 until October 2015. This search produced 4448 papers (excluding duplicates) and resulted in another 4 papers being included in the review, with a further 2 also included following peer review. The total number of papers included in the review was therefore 33. See Figure 1 for a flowchart of the combined searches and Table 1 for details of all included studies. We used Cohen K to assess inter-rater reliability of paper inclusion at the 20% data check stage (7590 papers checked). There was moderate agreement between the 2 reviewers (k=.526, 95% CI, 0.416 to 0.633, P<.001). All supplementary data referenced in this paper can be found in Multimedia Appendix 1, whereas Multimedia Appendix 2 contains more detailed information on the games and game mechanics used by studies in the review.

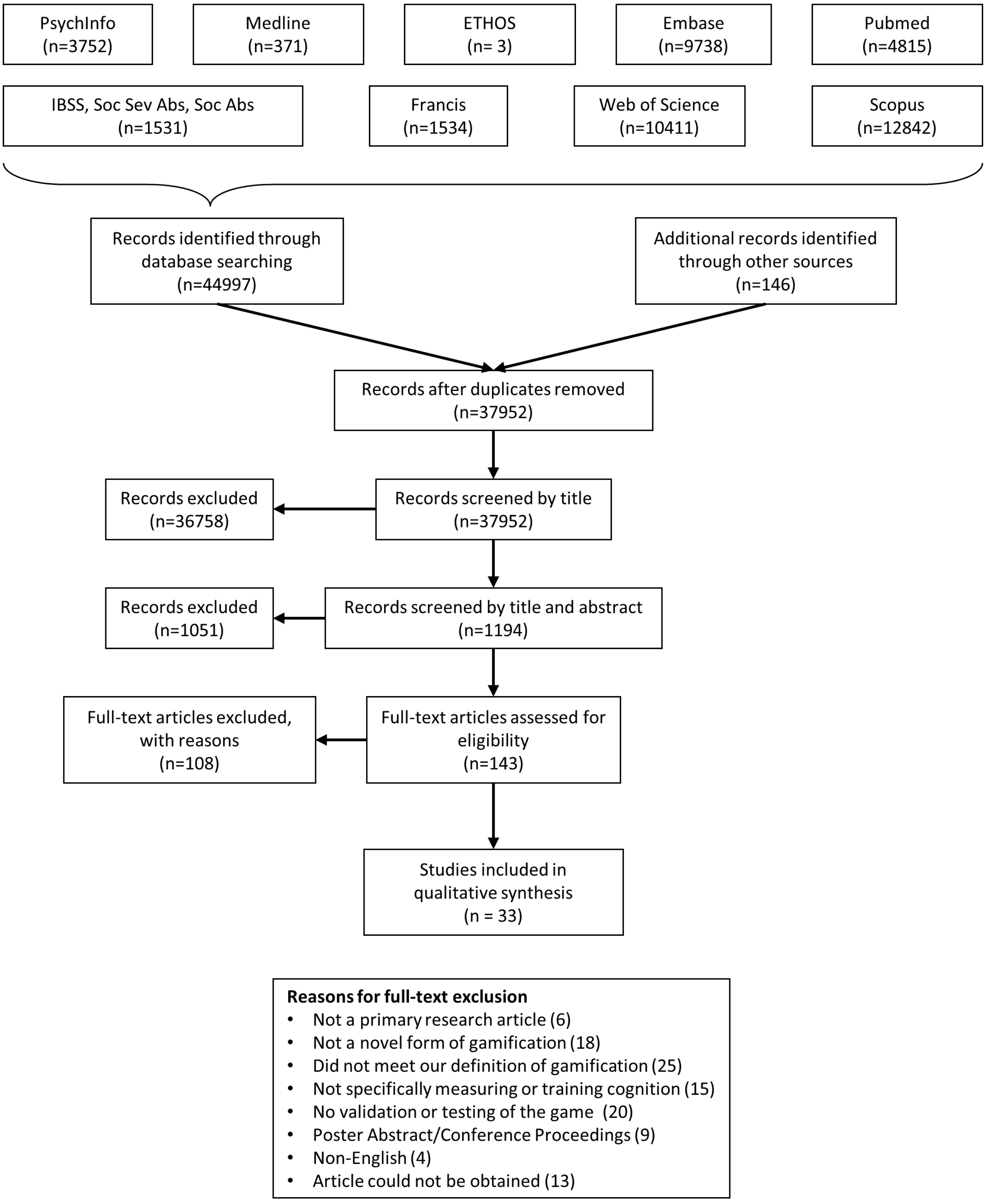
Table 1. Details of included papers.[a]

Author, year	Full title	Game	Category
McPherson and Burns, 2007 [17]	Gs invaders: Assessing a computer gamelike test of processing speed	Space Code	Testing
McPherson and Burns, 2008 [18]	Assessing the validity of computer gamelike tests of processing speed and working memory	Space Matrix/Space Code	Testing
Trapp et al, 2008 [19]	Cognitive remediation improves cognition and good cognitive performance increases time to relapse—results of a 5-year catamnestic study in schizophrenia patients	Xcog	Training
Gamberini et al, 2009 [20]	Eldergames project: An innovative mixed reality table-top solution to preserve cognitive functions in elderly people	Eldergames	Testing
Gamberini, Cardullo, Seraglia, and Bordin, 2010 [21]	Neuropsychological testing through a Nintendo Wii console	Wii Tests	Testing
Dovis, Oord, Wiers, and Prins, 2011 [22]	Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming	Megabot	Testing
Delisle and Braun, 2011 [23]	A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type)	Retirement Party	Testing
Prins, Dovis, Ponsioen, ten Brink, and van der Oord, 2011 [24]	Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD?	Supermecha	Training
Lim et al, 2012 [25]	A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder	Cogoland	Training
Heller et al, 2013 [26]	A machine learning-based analysis of game data for attention deficit hyperactivity disorder assessment	Groundskeeper	Testing
Hawkins et al, 2013 [9]	Gamelike features might not improve data	EM-Ants and Ghost Trap	Testing
Verhaegh, Fontijn, Aarts, and Resing, 2013 [27]	In-game assessment and training of nonverbal cognitive skills using TagTiles	Tap the Hedgehog	Both
Aalbers, Baars, Rikkert, and Kessels, 2013 [28]	Puzzling with online games (BAM-COG): reliability, validity, and feasibility of an online self-monitor for cognitive performance in aging adults	BAM-COG	Testing
Fagundo et al, 2013 [29]	Video game therapy for emotional regulation and impulsivity control in a series of treated cases with bulimia nervosa	Playmancer	Training
Anguera et al, 2013 [30]	Video game training enhances cognitive control in older adults	Neuroracer	Training
van der Oord, Ponsioen, Geurts, Ten Brink, and Prins, 2014 [31]	A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting	Braingame Brian	Training
Brown et al, 2014 [32]	Crowdsourcing for cognitive science—the utility of smartphones	The Great Brain Experiment	Testing
Tong and Chignell, 2014 [33]	Developing a serious game for cognitive assessment: choosing settings and measuring performance	Whack-a-mole	Testing
Katz, Jaeggi, Buschkuehl, Stegman, and Shah, 2014 [34]	Differential effect of motivational features on training improvements in school-based cognitive training	WMTrainer	Training
Dunbar et al, 2013 [8]	Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game	MACBETH	Training
Lee et al, 2013 [35]	A brain-computer interface based cognitive training system for healthy elderly: A randomised control pilot study for usability and preliminary efficacy	Card-Pairing	Training
Miranda and Palmer, 2013 [36]	Intrinsic motivation and attentional capture from gamelike features in a visual search task	Visual Search	Testing
Atkins et al, 2014 [37]	Measuring working memory is all fun and games: A four-dimensional spatial game predicts cognitive task performance	Shapebuilder	Testing
Dörrenbächer et al, 2014 [38]	Dissociable effects of game elements on motivation and cognition in a task switching training in middle childhood	Watermons	Training

Author, year	Full title	Game	Category
McNab and Dolan, 2014 [39]	Dissociating distractor-filtering at encoding and during maintenance	The Great Brain Experiment	Testing
O'Toole and Dennis, 2014 [40]	Mental health on the go: Effects of a gamified attention-bias modification mobile application in trait-anxious adults	ABMTApp	Training
Tenorio Delgado, Arango Uribe, Aparicio Alonso, and Rosas Diaz, 2014 [41]	TENI: A comprehensive battery for cognitive assessment based on games and technology	TENI	Testing
De Vries, Prins, Schmand, and Geurts, 2015 [42]	Working memory and cognitive flexibility-training for children with an autism spectrum disorder: a randomized controlled trial	Braingame Brian	Training
Dovis, Van Der Oord, Wiers, and Prins, 2015 [43]	Improving executive functioning in children with ADHD: Training multiple executive functions within the context of a computer game. A randomized double-blind placebo controlled trial	Braingame Brian	Training
Kim et al, 2015 [44]	Effects of a serious game training on cognitive functions in older adults	Smart Harmony	Training
Manera et al, 2015 [45]	"Kitchen and cooking," a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study	Kitchen and Cooking	Both
Ninaus et al, 2015 [46]	Game elements improve performance in a working memory training task	GAME	Training
Tarnanas et al, 2015 [47]	On the comparison of a novel serious game and electroencephalography biomarkers for early dementia screening	VAP-M	Testing

[a]In cases where the game was not named, we assigned a descriptive name.

Figure 1. Flowchart detailing the paper discovery and screening process.



Why Do Researchers Use Gamification?

We searched each paper for explanations as to why a gamelike task had been used and identified reasons for researchers opting to gamify their cognitive training and testing. These reasons were then coalesced into 7 categories. Some authors listed multiple reasons for gamifying their approach, whereas others gave no motivations at all. Supplementary Table 1 in Multimedia Appendix 1 provides details of which games fell into each category.

To Increase Participant Motivation

Although we assume that increasing participant motivation was a goal for every study in this review, we found 16 studies that explicitly used gamified tasks to measure or train cognition in a more motivating manner, and the majority (10 of 16) of these studies were assessment studies. Cognitive tests are typically a one off measure, and replayability is not a requirement. As such, these games tended to build simple game-archetypes, such as space invaders or whack-a-mole, around an existing cognitive task; with the goal of encouraging self-motivation, improving participant enjoyment, and even reducing test anxiety [17]. Gamification aims to increase engagement with tasks that might otherwise be perceived as demotivating [48], and the highly repetitive nature of cognitive tasks means they are ripe for improvement.

To Increase Usability/Intuitiveness for the Target Age Group

Eleven tasks were gamified specifically to enhance appeal with a given age group (see Supplementary Table 2 in Multimedia Appendix 1). The authors of these studies hypothesized that a more intuitive interface could prevent boredom and anxiety in the target age group, which might damage motivation and concentration on the tasks at hand. Six games were designed to be suitable for the elderly, who may not be familiar working with a mouse and keyboard [49]. Five other games were aimed at young children, and reframed the cognitive assessment as a game, to test the children under optimal mental conditions [41].

To Increase Long-Term Engagement

Commercial gamification is often used to create long-term interest around a user experience, product, or event [13]. Similarly, we found 11 studies (8 games) that used gamification to reduce participant dropout rates over a protracted testing or training programme. Many of these games were used in an unobserved, nonlaboratory setting; therefore, the tasks made use of motivational features that made the training and testing intrinsically more appealing and less of a burden to perform on a regular basis. A common feature was that task sessions were kept as short as possible to make them more convenient and to increase likelihood of completion [28,32]. For example, *The Great Brain Experiment* aimed to keep task sessions below 5 minutes in duration and found that the shortest task, the Stop Signal Task, was the most popular mini game.

To Investigate the Effects of Gamelike Tasks

Obviously, many of the studies in this review investigated the effects of gamelike tasks, however, only 5 studies explicitly stated that their aim was to assess the motivational and cognitive effects of gamelike features. Three of the 6 games were very simplistic, with only a few specific game mechanics and carefully designed nongame controls, to make the effects of the game mechanics on the data as apparent as possible. One game, *Watermons* [38], was designed, using Self Determination Theory [50,51], to maximize participant intrinsic motivation and make any motivational effects of game mechanics as apparent as possible.

To Stimulate the Brain

Six studies cited evidence that playing video games can be cognitively beneficial and/or stimulating as a key factor in their decision to gamify. The past decade has seen rapid growth in the investigation of the cognitive effects of video gaming, and findings have typically been positive. There is good evidence that video game players outperform nongamers on tests of working memory [52] visual attention [53,54] and processing speed [55,56]. The "active ingredient" of this cognitive enhancement effect is not yet known, but it comes as no surprise that researchers who are interested in training cognition are keen to include gamelike features in their training tasks.

To Increase Ecological Validity

Cognitive training has often suffered from a lack of transferability [57-59]. Although participants may improve at the training task, these improvements do not generalize to the real world. In a similar vein, cognitive assessment has been accused of being ecologically invalid [60]. A potential solution is to make tasks more realistic, and these tasks inevitably become gamelike as 3D graphics, sounds, and narrative are added. We identified 6 studies that used gamified tasks to test cognition in an engaging close-to-life environment or enhance transferability of learned skills. As Dunbar et al explain [8], games are uniquely suited to some forms of cognitive training as they give the player freedom to make choices and experience feedback on the effects of those choices; in other words, they provide opportunities for experiential learning [61].

To Increase Suitability for the Target Disorder

Gamified tasks may also be more appealing to patients with certain clinical conditions. Specifically, we found 6 studies (covering 4 games) designed for people with attention deficit/hyperactivity disorder (ADHD). It is commonly reported that ADHD patients are compulsive computer game players [23]. Furthermore, patients with ADHD react differently from controls to rewards and feedback: they prefer strong reinforcement and immediate feedback, as well as clear goals and objectives, all of which can easily be delivered in a gamelike environment [22,24,62].

What Are the Application Areas of Gamification?

We found comparable numbers of games used for cognitive testing (17) and training (13), with one game that can be used for both testing and training (see Supplementary Table 3 in Multimedia Appendix 1). It is also worth noting that the numbers of studies investigating cognitive testing (17) and training (15) were very similar, and the slightly smaller proportion of games used to deliver cognitive training is likely explained by the relative recency of the field.

Supplementary Table 4 in Multimedia Appendix 1 shows the cognitive domains assessed. Working memory was the most commonly tested domain. This is likely due to its ease of testing and also because working memory deficit is a common symptom in many disorders. General executive function (EF), attention, and inhibition were also commonly tested domains. Many games tested several cognitive domains and/or general EF, highlighting the difficulty of examining gamification effects on specific cognitive domains (and therefore comparing performance to standard cognitive tasks). Supplementary Table 5 in Multimedia Appendix 1 shows a breakdown of training tasks by the cognitive domains they addressed. Again, working memory was a popular target, closely followed by EF and inhibitory control training. There is a smaller overlap in domains covered by cognitive training tasks; test batteries tend to test a wide range of domains, whereas training focuses on 1 or 2 domains exclusively.

Does Gamification Work?

The studies we reviewed were generally enthusiastic about their use of gamified tasks, although given the diversity of study aims, this does not mean that all games worked as expected. Where reported, subjective and objective measures of participant engagement were positive. All studies that measured intrinsic motivation reported that the use of gamelike tasks improved motivation, compared with nongamified versions. We identified 21 of 33 studies that compared a gamelike task directly against a nongamified counterpart, and these studies can shed light on the specific effects of gamification on testing and training tasks.

Most gamified assessments were validated successfully. *Wii Tests* [21], *Shapebuilder* [37], *The Great Brain Experiment* [32,39], *BAM-COG* [28], and *Tap the Hedgehog* [27] were all found to produce output measures/scores that correlated fairly well with their non–gamelike counterparts, though mixed-domain measures were an issue (see Supplementary Table 1 in Multimedia Appendix 2 for full details of all games). Validation studies varied in their design, and some studies reported complex correlations between gamified and nongamified tasks with multiple outputs. However, sample correlations from some of the simpler validation studies suggest intertask correlations of 0.45-0.60 [28,37,39,63]. Broadly speaking, these were well-designed and well-powered studies, and together, they provide encouraging evidence that cognitive tests can be gamified and still be useful as a research tool.

Some studies found that their gamified tests were correlated with measures of multiple cognitive domains, in other words, they were mixed-domain measures. For example, use of exploratory factor analysis showed that *Whack-a-Mole's* primary output measure was correlated with 2 of the 3 EFs of interest: inhibition (r (22)=.60, P<.001) and updating (r (22) = .35, P<.05) [33]. *Space Code* [17,18] had similar problems. The initial study was successful, with *Space Code's* output measure correlating well with a conventional measure of processing speed (r (58)=.55, P<.001). However, a second paper detailing 2 experiments which aimed to replicate the previous finding found that *Space Code's* correlations with measures of working memory, visuospatial ability, and processing speed were not stable [18]. The fact that *Space Code* was thought to be a fairly pure measure in one study and then was shown to be mixed-domain in the next highlights the fact that designing gamified cognitive tasks is difficult, and multiple, well-powered validation studies may be required to ensure a task is measuring what is intended.

Gamification also has the potential to invalidate a task. For example, *Retirement Party* was compared against the Continuous Performance Task-II in healthy controls and adults with ADHD. The Continuous Performance Task-II detected more commission errors from the ADHD adults as expected (mean [M]=56, standard deviation [SD]=13 vs M=46, SD=10), but *Retirement Party* did not (M=14.4, SD=5.8 vs M=13.2, SD=4.3): this likely invalidates the game as a diagnostic tool for ADHD. However, Delisle and Braun [23] discuss the possibility that *Retirement Party* may have detected no deficit in ADHD patients as the nature of the task was such that there was no deficit: the highly structured and feedback-rich multitask environment may have normalized the ADHD patients' usual inattention. Such a performance boost resulting from gamelike elements is a disadvantage when performing cognitive assessment but is likely to be desirable in a cognitive training scenario.

Several studies in this review focused specifically on adding game mechanics to cognitive tasks to investigate the resultant changes in data, enjoyment, and motivation. Dovis et al [22] studied whether different types of incentive could normalize ADHD children's performance on working memory training. They found that regardless of incentive, ADHD children did not perform as well as healthy controls. However, ADHD children also experienced a decrease in performance over time, and the €10 condition and the gaming condition (*Megabot*) prevented this decrease. These results indicate that performance problems in ADHD training might be somewhat alleviated through the use of gamelike tasks. This is further supported by the study by Prins et al (*Supermecha* [24]), which found that ADHD children completed more training trials (M=199.48, SD=47.46 vs M=134.43, SD=34.18), with higher accuracy (69% vs 51%), when trained using a gamified working memory training task as opposed to a non–gamelike one. Children in the gamelike condition were also more engaged and enjoyed the training more, as measured by "absence time." In a similar vein, The Great Brain Experiment [39] and GAME [46] both showed gamelike tasks to be appropriate for measuring and training working memory. With Ninaus et al, presenting evidence that gamification can improve overall participant performance in a working memory training task [48] and McNab and Dolan showing that data collected from 2 very different gamified and nongamified tasks could fit similar models of working memory capacity.

In contrast, *WMTrainer*, was assessed across 7 different conditions containing different combinations of game mechanics [34]. They compared training improvement across conditions and found that the greatest training effect was caused by versions with minimal motivational features. The fully gamified condition had one of the shallowest improvement slopes and none of the conditions made any difference on subjective motivation scores. However, even the minimally gamelike version still featured simplistic graphics and displayed a player score at the end of the block. It is possible that even this minimal gamification was enough to induce increased motivation and that adding "distracting" game elements such as persistent score display may have a negative impact on performance by inducing unneeded stress or new cognitive demands [34].

One of the most theoretically driven games in our review was *Watermons* [38], which included many motivational features aimed at delivering a sense of player autonomy and competence. The task-switching training was embedded within a rich storyline and graphically enhanced theme. They found that these gamelike features increased the effect of training, reducing reaction times and switch costs, compared against a non–gamelike version of the training. Participants were also more willing to perform training when in the gamelike condition.

Miranda and Palmer [36] used a visual search task with 2 forms of reward for fast and accurate responses: sound effects and points. They found that sound effects led to increased reaction times, potentially due to attentional capture and did not improve scores of subjective engagement. Points appeared to have no effect on data and boosted subjective engagement scores. These results highlight the delicate nature of designing gamified cognitive tests because something as innocuous as a few sound effects had deleterious effects on participant performance.

Finally, Hawkins et al, [9] compared gamelike versions of 2 decision-making tasks against nongame counterparts. No difference between the data collected by the gamified versions and the nongame versions was found. Subjective ratings indicated that both versions of both tasks were equally boring and repetitive, but that the gamelike versions were more interesting and enjoyable. Given the relatively large combined sample size of these studies (N=200), they provide good evidence that game mechanics can be included in cognitive tests without invalidating the data and with the desired effect of increasing motivation.

Discussion

Principal Results

We identified 7 reasons why researchers use gamification in cognitive research. These include not only the "traditional" applications of gamification such as increasing long- and short-term engagement with a task but also more clinically related reasons such as making tasks more interactive to enhance the effect of cognitive training. Several studies aimed to reduce test anxiety and optimize performance in groups that traditionally dislike being tested, particularly electronically, such as elderly people and children. By hiding the test behind a novel interface and gameplay, the target audience might feel more comfortable.

We saw several games aimed at training and testing people with ADHD, and overall, these games appear highly engaging to users, in some cases, even increasing the time spent training. Gamified tasks may be valuable for assessing ADHD patients as computer games are particularly appealing to them: with rapid rewards, immediate feedback, and time-pressure being exactly the type of stimulus the ADHD brain craves [64,65]. The dopaminergic system is thought to be abnormal in ADHD [66,67]. However, it is thought that playing video games can facilitate the release of extrastriatal dopamine, which plays a role in focusing attention and heightening arousal [68,69]; this may improve player performance and motivation. Nevertheless, as Delisle and Braun discuss [23], we must be cautious that liberal use of game mechanics does not reverse the very deficit we are hoping to measure.

One of the primary reasons that psychologists are keen to utilize gamification is to increase performance and motivation in research populations. The results of references [9,17,18,38], show that gamified tasks can be used to improve motivation, while still maintaining a scientifically valid task. However, Katz et al [34] and Miranda and Palmer [36] highlight that this can be difficult balancing act, with several game mechanics having unforeseen deleterious effects on performance. If gamified psychological tasks are to become common in the future, further research is required to disentangle the impact of specific game mechanics on task performance, as these studies have already begun to do.

Differences Between Training and Assessment Tasks

Training games typically contained many features and were similar in appearance to commercial video games. Cognitive training normally requires several sessions to be effective, and as a result, training tasks need to be engaging enough to play for many hours. 3D graphics were quite prevalent, as was the use of avatars, points, levels, and dynamically growing game worlds (see Supplementary Figure 1 in Multimedia Appendix 2). Long-term goals which had to be completed over repeated sessions were also common and served to sustain engagement over a long period. In contrast, assessment games were simpler, predominantly using 2D graphics, sound effects, score, and theme to create the appearance of a game. Several games simply presented themselves as "puzzles" which the participant had to complete. Tasks of this nature represent gamification at its simplest, but they were well received by users, implying that minimal gamification is better than no gamification. The simplicity of gamification employed is likely due to the constant tension between creating an engaging task and the risk of undermining the task's scientific validity: including unknown game mechanics might have deleterious effects on the data collected.

Validating Gamified Tasks

We found heterogeneous standards for validating gamified tasks. Typically, cognitive assessment games were validated rigorously, using correlation with similar cognitive tasks and factor analysis to determine whether they were performing as expected. Many training games used a gamified task only, meaning the effect of gamification cannot be dissociated from the effect of the intervention. Sample sizes were small in nearly all of the studies we reviewed, and there was little consideration of statistical power when sample sizes were decided upon, with only 5 of 33 papers describing a power calculation. Gamified cognitive tests are novel scientific instruments and must be validated as such. Small pilot studies, followed by larger validation studies including assessment of test–retest reliability, and internal and external validity of the measures taken by the game are needed [70]. Regarding cognitive training, ideally gamified training should be treated as an intervention and so the current gold standard of a blinded randomized control trial is appropriate [71]. In both testing and training, we would recommend the use of at least 2 controls: a standard cognitive task designed to produce the same output measures/training effect as the newly gamified task and a nongamified version of the gamified task, built on the same software platform and identical in function/interaction, with all game mechanics removed.

Limitations

One limitation of this review is the necessity of a narrow scope. Gamification in psychology and psychiatry is a rapidly growing field, hence we decided to focus specifically on "cognitive training and testing." This has resulted in some papers being excluded on the subjective basis of not being "cognitive research." Nevertheless, to counteract this subjectivity, papers were only included in the review on consensus from both reviewers (JL and EE), and a 20% selection check was performed by EE on papers from the title-screening stage onward. An additional consideration is that many of the studies reviewed were of a preliminary nature, and as such, the findings reported here should be considered tentative.

Conclusions

As discussed by Hawkins et al [9], it has often been suggested that gamified cognitive tasks may result in higher quality data and more effective training, simply by virtue of heightened engagement. Our review found no evidence to support that gamified tests can be used to improve data quality, either by reducing between-subject noise or by improving participant performance, and there were some indications that it may actually worsen it [17,18,34,36]. We did, however, find some evidence that gamification may be effective at enhancing cognitive training, but we must take these positive training findings tentatively due to numerous methodologic problems in the studies that we reviewed.

Irrespective of whether gamification can improve data or enhance training, there are still many reasons why gamelike tasks may play an important role in the future of cognitive research. Gamified cognitive tasks are more engaging than traditional tasks, thus making the participant experience less effortful and potentially reducing drop-outs in longitudinal studies. Furthermore, despite concerns that some commonly used game mechanics might reduce participant motivation [72] (such as by having a visibly low score [73]), we saw no evidence that this was the case. Indeed, gamification was reported as both motivational and positive by study authors and participants alike. Gamelike tasks may also reduce feelings of test anxiety and allow alternative interfaces to cognitive tests that would otherwise be difficult to deliver in certain populations. The results of this review also show that it is possible to design gamified cognitive assessments that validate well against more traditional measures, providing that caution is taken to avoid developing mixed-domain measures or masking deficits of interest.

As cognitive research begins to move out of the laboratory and onto personal computers and mobile devices, engagement will be the key to collecting high-quality data. Gamification is likely to play an important role in enabling this change; but further research and more rigorous validation is needed to understand the delicate interplay between game mechanics and cognitive processes.

Acknowledgments

The authors are members of the United Kingdom Centre for Tobacco and Alcohol Studies, a UKCRC Public Health Research: Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council, and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration, is gratefully acknowledged. The UK Medical Research Council and the Wellcome Trust (092731) and the University of Bristol provide core support for ALSPAC. This work was supported by the Medical Research Council (MC_UU_12013/6), a PhD studentship to JL funded by the Economic and Social Research Council and Cambridge Cognition Limited and a NIHR funded academic clinical fellowship to EE. The funders had no role in review design, data extraction and analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

Dr Coyle is a Director of Handaxe CIC, a not-for-profit company that develops technology, including computer games, to support mental health interventions for children and adolescents.

Multimedia Appendix 1

Five tables detailing coding and categorization in the main manuscript. Table 1 lists the reasons for using gamification in cognitive training and testing. Table 2 categorizes games by the age group they were aimed at. Table 3 lists games by training or testing category. Tables 4 and 5 list games by the cognitive domains which they targeted.

[PDF File (Adobe PDF File), 102KB - games_v4i2e11_app1.pdf]

Multimedia Appendix 2

Additional information on the papers covered by this review that is highly detailed and intended for further analysis in future papers or for other researchers to assess themselves. It contains short text-based descriptions of all 31 games and a screenshot where the paper included one. There is also a detailed breakdown of the game mechanics in each task.

[PDF File (Adobe PDF File), 1MB - games_v4i2e11_app2.pdf]

References

1. State of Online Gaming Report.: Spilgames; 2013. URL: http://auth-83051f68-ec6c-44e0-afe5-bd8902acff57.cdn.spilcloud.com/v1/archives/1384952861.25_State_of_Gaming_2013_US_FINAL.pdf [accessed 2014-11-27] [WebCite Cache ID 6gt7YwEGt]
2. Cowley B, Charles D, Black M, Hickey R. Toward an understanding of flow in video games. Computers in Entertainment 2008;6(2) Article No. 20. [doi: 10.1145/1371216.1371223]
3. McGonigal J. Reality is Broken: Why games make us better and how they can change the world. New York: Vintage; 2012.
4. DeRight J, Jorgensen R. I just want my research credit: frequency of suboptimal effort in a non-clinical healthy undergraduate sample. Clin Neuropsychol 2015;29(1):101-117. [doi: 10.1080/13854046.2014.989267] [Medline: 25494327]
5. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
6. Boyle E, Hainey T, Connolly T, Gray G, Earp J, Ott M, et al. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. Computers & Education 2016 Mar;94:178-192. [doi: 10.1016/j.compedu.2015.11.003]
7. Boendermaker WJ, Boffo M, Wiers RW. Exploring Elements of Fun to Motivate Youth to Do Cognitive Bias Modification. Games Health J 2015 Dec;4(6):434-443. [doi: 10.1089/g4h.2015.0053] [Medline: 26421349]
8. Dunbar N, Miller C, Adame B, Elizondo J, Wilson S, Lane B, et al. Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game. Computers in Human Behavior 2014;37:307-318. [doi: 10.1016/j.chb.2014.04.053]
9. Hawkins GE, Rae B, Nesbitt KV, Brown SD. Gamelike features might not improve data. Behav Res Methods 2013 Jun;45(2):301-318. [doi: 10.3758/s13428-012-0264-3] [Medline: 23055169]
10. Dubbels B. Gamification, Serious Games, Ludic Simulation, and Other Contentious Categories. International Journal of Gaming and Computer-Mediated Simulations 2013 Apr;5(2):1-19. [doi: 10.4018/jgcms.2013040101]
11. Aparicio A, Vela F, Sánchez J, Montes J. Analysis and application of gamification. In: Proceedings of the 13th International Conference on Interacción Persona-Ordenador. 2012 Presented at: Interacción 2012; October 03-05, 2012; Elche, Spain. [doi: 10.1145/2379636.2379653]
12. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
13. Deterding S, Björk S, Nacke L, Dixon D, Lawley E. Designing gamification: creating gameful and playful experiences. In: Extended Abstracts on Human Factors in Computing Systems. USA: ACM; 2013 Presented at: CHI '13; April 27-May 02, 2013; Paris, France p. 3263-3266. [doi: 10.1145/2468356.2479662]
14. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010 Jul;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
15. Primack B, Carroll M, McNamara M, Klem M, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638. [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
16. Donchin E. Video games as research tools: The Space Fortress game. Behavior Research Methods, Instruments, & Computers 1995 Jun;27(2):217-223. [doi: 10.3758/BF03204735]
17. McPherson J, Burns NR. Gs invaders: assessing a computer game-like test of processing speed. Behav Res Methods 2007 Nov;39(4):876-883. [Medline: 18183904]
18. McPherson J, Burns NR. Assessing the validity of computer-game-like tests of processing speed and working memory. Behav Res Methods 2008 Nov;40(4):969-981. [doi: 10.3758/BRM.40.4.969] [Medline: 19001388]
19. Trapp W, Landgrebe M, Hoesl K, Lautenbacher S, Gallhofer B, Günther W, et al. Cognitive remediation improves cognition and good cognitive performance increases time to relapse--results of a 5 year catamnestic study in schizophrenia patients. BMC Psychiatry 2013;13:184 [FREE Full text] [doi: 10.1186/1471-244X-13-184] [Medline: 23837673]
20. Gamberini L, Martino F, Seraglia B, Spagnolli A, Fabregat M, Ibanez F, et al. Eldergames project: An innovative mixed reality table-top solution to preserve cognitive functions in elderly people. 2009 May 21 Presented at: 2nd Conference on Human System Interactions; May 21-23, 2009; Catania p. 164-169. [doi: 10.1109/HSI.2009.5090973]
21. Gamberini L, Cardullo S, Seraglia B, Bordin A. Neuropsychological testing through a Nintendo Wii console. Stud Health Technol Inform 2010;154:29-33. [Medline: 20543264]
22. Dovis S, Van der Oord Saskia, Wiers R, Prins Pier J M. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming. J Abnorm Child Psychol 2012 Jul;40(5):669-681. [doi: 10.1007/s10802-011-9601-8] [Medline: 22187093]
23. Delisle J, Braun Claude M J. A context for normalizing impulsiveness at work for adults with attention deficit/hyperactivity disorder (combined type). Arch Clin Neuropsychol 2011 Nov;26(7):602-613. [doi: 10.1093/arclin/acr043] [Medline: 21653627]
24. Prins PJ, Dovis S, Ponsioen A, ten Brink E, van der Oord S. Does computerized working memory training with game elements enhance motivation and training efficacy in children with ADHD? Cyberpsychol Behav Soc Netw 2011 Mar;14(3):115-122. [doi: 10.1089/cyber.2009.0206] [Medline: 20649448]

25. Lim, Lee, Guan, Fung Daniel Shuen Sheng, Zhao, Teng Stephanie Sze Wei, Zhang, Krishnan K Ranga Rama. A brain-computer interface based attention training program for treating attention deficit hyperactivity disorder. PLoS One 2012;7(10) [FREE Full text] [doi: 10.1371/journal.pone.0046692] [Medline: 23115630]
26. Heller M, Roots K, Srivastava S, Schumann J, Srivastava J, Hale T. A Machine Learning-Based Analysis of Game Data for Attention Deficit Hyperactivity Disorder Assessment. Games Health J 2013 Oct;2(5):291-298. [doi: 10.1089/g4h.2013.0058] [Medline: 26196929]
27. Verhaegh J, Fontijn W, Aarts E, Resing W. In-game assessment and training of nonverbal cognitive skills using TagTiles. Pers Ubiquit Comput 2012 Apr 22;17(8):1637-1646. [doi: 10.1007/s00779-012-0527-0]
28. Aalbers T, Baars Maria A E, Olde Rikkert Marcel G M, Kessels Roy P C. Puzzling with online games (BAM-COG): reliability, validity, and feasibility of an online self-monitor for cognitive performance in aging adults. J Med Internet Res 2013;15(12) [FREE Full text] [doi: 10.2196/jmir.2860] [Medline: 24300212]
29. Fagundo, Santamaría, Forcano, Giner-Bartolomé, Jiménez-Murcia, Sánchez, Granero, Ben-Moussa, Magnenat-Thalmann, Konstantas, Lam, Lucas, Nielsen, Bults, Tarrega, Menchón, de le Torre, Cardi, Treasure, Fernández-Aranda. Video game therapy for emotional regulation and impulsivity control in a series of treated cases with bulimia nervosa. European Eating Disorders Review 2013 Nov;21(6):493-499. [doi: 10.1002/erv.2259]
30. Anguera J, Boccanfuso J, Rintoul J, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 5;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
31. van der Oord S, Ponsioen A J, Geurts HM, Ten Brink EL, Prins PJ. A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: outcome on parent- and teacher-rated executive functioning and ADHD behavior. J Atten Disord 2014 Nov;18(8):699-712. [doi: 10.1177/1087054712453167] [Medline: 22879577]
32. Brown H, Zeidman P, Smittenaar P, Adams R, McNab F, Rutledge R, et al. Crowdsourcing for Cognitive Science - The Utility of Smartphones. PloS One 2014 Jul 15;9. [doi: 10.1371/journal.pone.0100662]
33. Tong T, Chignell M. Developing a Serious Game for Cognitive Assessment: Choosing Settings and Measuring Performance. In: Proceedings of the Second International Symposium of Chinese CHI. 2014 Apr 26 Presented at: Chinese CHI 2014; April 26-27, 2014; Toronto, Canada p. 70-79. [doi: 10.1145/2592235.2592246]
34. Katz B, Jaeggi S, Buschkuehl M, Stegman A, Shah P. Differential effect of motivational features on training improvements in school-based cognitive training. Front Hum Neurosci 2014;8:242 [FREE Full text] [doi: 10.3389/fnhum.2014.00242] [Medline: 24795603]
35. Lee T, Goh SJ, Quek S, Phillips R, Guan C, Cheung Y, et al. A brain-computer interface based cognitive training system for healthy elderly: a randomized control pilot study for usability and preliminary efficacy. PLoS One 2013 Nov 18;8(11):e79419 [FREE Full text] [doi: 10.1371/journal.pone.0079419] [Medline: 24260218]
36. Miranda A, Palmer E. Intrinsic motivation and attentional capture from gamelike features in a visual search task. Behav Res Methods 2014 Mar;46(1):159-172. [doi: 10.3758/s13428-013-0357-7] [Medline: 23835649]
37. Atkins S, Sprenger A, Colflesh G, Briner T, Buchanan J, Chavis S, et al. Measuring working memory is all fun and games: a four-dimensional spatial game predicts cognitive task performance. Exp Psychol 2014;61(6):417-438. [doi: 10.1027/1618-3169/a000262] [Medline: 24962121]
38. Dörrenbächer S, Müller P, Tröger J, Kray J. Dissociable effects of game elements on motivation and cognition in a task-switching training in middle childhood. Front Psychol 2014;5:1275 [FREE Full text] [doi: 10.3389/fpsyg.2014.01275] [Medline: 25431564]
39. McNab F, Dolan R. Dissociating distractor-filtering at encoding and during maintenance. J Exp Psychol Hum Percept Perform 2014 Jun;40(3):960-967 [FREE Full text] [doi: 10.1037/a0036013] [Medline: 24512609]
40. Dennis TA, O'Toole L. Mental Health on the Go: Effects of a Gamified Attention Bias Modification Mobile Application in Trait Anxious Adults. Clin Psychol Sci 2014 Sep 1;2(5):576-590 [FREE Full text] [doi: 10.1177/2167702614522228] [Medline: 26029490]
41. Delgado MT, Uribe PA, Alonso AA, Díaz RR. TENI: A comprehensive battery for cognitive assessment based on games and technology. Child Neuropsychol 2016;22(3):276-291. [doi: 10.1080/09297049.2014.977241] [Medline: 25396766]
42. de Vries M, Prins PJ, Schmand B, Geurts H. Working memory and cognitive flexibility-training for children with an autism spectrum disorder: a randomized controlled trial. J Child Psychol Psychiatry 2015 May;56(5):566-576. [doi: 10.1111/jcpp.12324] [Medline: 25256627]
43. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Improving executive functioning in children with ADHD: training multiple executive functions within the context of a computer game. a randomized double-blind placebo controlled trial. PLoS One 2015;10(4):e0121651 [FREE Full text] [doi: 10.1371/journal.pone.0121651] [Medline: 25844638]
44. Kim K, Choi Y, You H, Na D, Yoh MS, Park JK, et al. Effects of a serious game training on cognitive functions in older adults. J Am Geriatr Soc 2015 Mar;63(3):603-605. [doi: 10.1111/jgs.13304] [Medline: 25800915]
45. Manera V, Petit P, Derreumaux A, Orvieto I, Romagnoli M, Lyttle G, et al. 'Kitchen and cooking,' a serious game for mild cognitive impairment and Alzheimer's disease: a pilot study. Front Aging Neurosci 2015;7:24 [FREE Full text] [doi: 10.3389/fnagi.2015.00024] [Medline: 25852542]

46. Ninaus M, Pereira G, Stefitz R, Prada R, Paiva A, Neuper C, et al. Game elements improve performance in a working memory training task. IJSG 2015 Feb 10;2(1):3-16 [FREE Full text] [doi: 10.17083/ijsg.v2i1.60]
47. Tarnanas I, Laskaris N, Tsolaki M, Muri R, Nef T, Mosimann U. On the comparison of a novel serious game and electroencephalography biomarkers for early dementia screening. Adv Exp Med Biol 2015;821:63-77. [doi: 10.1007/978-3-319-08939-3_11] [Medline: 25416111]
48. Eriksson B, Musialik M, Wagner J. Student essay. 2012 Aug 6. Gamification - Engaging the Future URL: https://gupea.ub.gu.se/handle/2077/30037 [accessed 2016-07-08] [WebCite Cache ID 6iqtPwZyC]
49. Sirály E, Szabó Á, Szita B, Kovács V, Fodor Z, Marosi C, et al. Monitoring the early signs of cognitive decline in elderly by computer games: an MRI study. PLoS One 2015 Feb 23;10(2):e0117918 [FREE Full text] [doi: 10.1371/journal.pone.0117918] [Medline: 25706380]
50. Ryan R, Deci E. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037/0003-066X.55.1.68]
51. Ryan R, Rigby C, Przybylski A. The Motivational Pull of Video Games: A Self-Determination Theory Approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
52. Colzato LS, van den Wildenberg M, Zmigrod S, Hommel B. Action video gaming and cognitive control: playing first person shooter games is associated with improvement in working memory but not action inhibition. Psychol Res 2013 Mar;77(2):234-239. [doi: 10.1007/s00426-012-0415-2] [Medline: 22270615]
53. Durlach P, Kring J, Bowens L. Effects of action video game experience on change detection. Mil Psychol 2009;21(1):24-39.
54. Green C, Bavelier D. Enumeration versus multiple object tracking: the case of action video game players. Cognition 2006 Aug;101(1):217-245 [FREE Full text] [doi: 10.1016/j.cognition.2005.10.004] [Medline: 16359652]
55. Dobrowolski P, Hanusz K, Sobczyk B, Skorko M, Wiatrow A. Cognitive enhancement in video game players: The role of video game genre. Computers in Human Behavior 2015 Mar;44:59-63. [doi: 10.1016/j.chb.2014.11.051]
56. Dye Matthew W G, Green CS, Bavelier D. Increasing Speed of Processing With Action Video Games. Curr Dir Psychol Sci 2009;18(6):321-326 [FREE Full text] [doi: 10.1111/j.1467-8721.2009.01660.x] [Medline: 20485453]
57. Abikoff H. Efficacy of cognitive training interventions in hyperactive children: A critical review. Clinical Psychology Review 1985;5(5):479-512. [doi: 10.1016/0272-7358(85)90005-4]
58. Jaeggi S, Buschkuehl M, Jonides J, Shah P. Short- and long-term benefits of cognitive training. PNAS 2011 Jun 13;108(25):10081-10086. [doi: 10.1073/pnas.1103228108]
59. Melby-Lervåg M, Hulme C. Is working memory training effective? A meta-analytic review. Dev Psychol 2013 Feb;49(2):270-291. [doi: 10.1037/a0028228] [Medline: 22612437]
60. Burgess P, Alderman N, Forbes C, Costello A, Coates L, Dawson D, et al. The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. J Int Neuropsychol Soc 2006 Mar;12(2):194-209. [doi: 10.1017/S1355617706060310] [Medline: 16573854]
61. Kolb D. Experiential learning: experience as the source of learning and development. Englewood Cliffs, NJ: Prentice-Hall; 1984.
62. Prins PJ, Brink E, Dovis S, Ponsioen A, Geurts H, de Vries M, et al. “Braingame Brian”: Toward an Executive Function Training Program with Game Elements for Children with ADHD and Cognitive Control Problems. Games Health J 2013 Feb;2(1):44-49. [doi: 10.1089/g4h.2013.0004] [Medline: 26196554]
63. Verhaegh J, Fontijn W, Resing W. On the correlation between children's performances on electronic board tasks and nonverbal intelligence test measures. Computers & Education 2013 Nov;69:419-430. [doi: 10.1016/j.compedu.2013.07.026]
64. Kok A, Kong T, Bernard-Opitz V. A comparison of the effects of structured play and facilitated play approaches on preschoolers with autism. A case study. Autism 2002 Jun;6(2):181-196. [Medline: 12083284]
65. WebMD. 2012. Video Games and ADHD: What's the Link? URL: http://www.webmd.com/add-adhd/childhood-adhd/features/adhd-and-video-games-is-there-a-link?print=true [accessed 2016-04-19] [WebCite Cache ID 6gt9IM416]
66. Dougherty D, Bonab A, Spencer T, Rauch S, Madras B, Fischman A. Dopamine transporter density in patients with attention deficit hyperactivity disorder. Lancet 1999;354(9196):2132-2133. [doi: 10.1016/S0140-6736(99)04030-1] [Medline: 10609822]
67. LaHoste, Swanson, Wigal, Glabe, Wigal, King, Kennedy. Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. Mol Psychiatry 1996 May;1(2):121-124. [Medline: 9118321]
68. Nieoullon A. Dopamine and the regulation of cognition and attention. Prog Neurobiol 2002 May;67(1):53-83. [Medline: 12126656]
69. Nieoullon A, Coquerel A. Dopamine: a key regulator to adapt action, emotion, motivation and cognition. Curr Opin Neurol 2003 Dec;16 Suppl 2:S3-S9. [Medline: 15129844]
70. Kato P. Pam Kato's Blog. 2013 Apr 05. What do you mean when you say your serious game has been validated? Experimental vs. Test Validity URL: https://pamkato.com/2013/04/25/what-do-you-mean-when-you-say-your-serious-game-has-been-validated-experimental-vs-test-validty/ [WebCite Cache ID 6gt9POLIu]
71. Kato PM. Evaluating Efficacy and Validating Games for Health. Games Health J 2012 Feb;1(1):74-76. [doi: 10.1089/g4h.2012.1017] [Medline: 26196436]

72. Mekler E, Brühlmann F, Opwis K, Tuch A. Do Points, Levels & Leaderboards Harm Intrinsic Motivation?: An Empirical Analysis of Common Gamification Elements. In: Proceedings of the First International Conference on Gameful Design, Research, and Applications. 2013 Presented at: Gamification 2013; October 02-04, 2013; Stratford, ON, Canada p. 66-73. [doi: 10.1145/2583008.2583017]
73. Preist C, Massung E, Coyle D. Competing or Aiming to Be Average?: Normification As a Means of Engaging Digital Volunteers. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014 Presented at: ACM conference on Computer supported cooperative work & social computing; February 15-19, 2014; Baltimore, MD p. 1222-1233. [doi: 10.1145/2531602.2531615]

Abbreviations

ADHD: attention deficit/hyperactivity disorder
EF: executive function

Edited by G Eysenbach; submitted 21.04.16; peer-reviewed by C Chambers, M Ninaus; comments to author 08.06.16; revised version received 16.06.16; accepted 17.06.16; published 15.07.16.

Please cite as:
Lumsden J, Edwards EA, Lawrence NS, Coyle D, Munafò MR
Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy
JMIR Serious Games 2016;4(2):e11
URL: http://games.jmir.org/2016/2/e11/
doi:10.2196/games.5888
PMID:27421244

Original Paper

Validation of a Computerized, Game-based Assessment Strategy to Measure Training Effects on Motor-Cognitive Functions in People With Dementia

Stefanie Wiloth[1,2], MA; Nele Lemke[1,3], MA; Christian Werner[1], MA; Klaus Hauer[1], PhD

[1]AGAPLESION Bethanien Hospital, Geriatric Centre of the University of Heidelberg, Heidelberg, Germany

[2]The Institute for the Study of Christian Social Service at the University of Heidelberg, Heidelberg, Germany

[3]Network of Aging Research (NAR), University of Heidelberg, Heidelberg, Germany

Corresponding Author:
Stefanie Wiloth, MA
AGAPLESION Bethanien Hospital, Geriatric Centre at the University of Heidelberg, Heidelberg, Germany
Research Department
Rohrbacher Straße 149
Heidelberg, 69126
Germany
Phone: 49 1739471446
Fax: 49 622131915
Email: stefanie.wiloth@gmail.com

Abstract

Background: Exergames often used for training purpose can also be applied to create assessments based on quantitative data derived from the game. A number of studies relate to these use functionalities developing specific assessment tasks by using the game software and provided good data on psychometric properties. However, (1) assessments often include tasks other than the original game task used for training and therefore relate to similar but not to identical or integrated performances trained, (2) people with diagnosed dementia have insufficiently been addressed in validation studies, and (3) studies did commonly not present validation data such as sensitivity to change, although this is a paramount objective for validation to evaluate responsiveness in intervention studies.

Objective: Specific assessment parameters have been developed using quantitative data directly derived from the data stream during the game task of a training device (Physiomat). The aim of this study was to present data on construct validity, test–retest reliability, sensitivity to change, and feasibility of this internal assessment approach, which allows the quantification of Physiomat training effects on motor-cognitive functions in 105 multimorbid patients with mild-to-moderate dementia (mean age 82.7±5.9).

Methods: Physiomat assessment includes various tasks at different complexity levels demanding balance and cognitive abilities. For construct validity, motor-cognitive Physiomat assessment tasks were compared with established motor and cognitive tests using Spearman's rank correlations (r_s). For test–retest reliability, we used intra-class correlations ($ICC_{3,1}$) and focused on all Physiomat tasks. Sensitivity to change of trained Physiomat tasks was tested using Wilcoxon statistic and standardized response means (SRMs). Completion rate and time were calculated for feasibility.

Results: Analyses have mostly shown moderate-to-high correlations between established motor as well as cognitive tests and simple (r_s=−.22 to .68, *P* ≤.001-.03), moderate (r_s=−.33 to .71, *P* ≤.001-.004), and complex motor-cognitive Physiomat tasks (r_s=−.22 to .83, *P* ≤.001-.30) indicating a good construct validity. Moderate-to-high correlations between test and retest assessments were found for simple, moderate, and complex motor-cognitive tasks (ICC=.47-.83, *P* ≤.001) indicating good test–retest reliability. Sensitivity to change was good to excellent for Physiomat assessment as it reproduced significant improvements (*P* ≤.001) with mostly moderate-to-large effect sizes (SRM=0.5-2.0) regarding all trained tasks. Completion time averaged 25.8 minutes. Completion rate was high for initial Physiomat measures. No adverse events occurred during assessment.

Conclusions: Overall, Physiomat proved to have good psychometric qualities in people with mild-to-moderate dementia representing a reliable, valid, responsive, and feasible assessment strategy for multimorbid older adults with or without cognitive impairment, which relates to identical and integrated performances trained by using the game.

(*JMIR Serious Games 2016;4(2):e12*) doi:10.2196/games.5696

KEYWORDS

serious games; computerized assessment; validation; motor-cognitive functions; elderly; older adults; cognitive impairment; dementia

Introduction

With respect to the growing elderly population, the role of innovative assessments to detect motor-functional or cognitive deficits is becoming increasingly important. Such assessment methods could help to identify appropriate interventions to delay age-related physical or mental decline.

Using Exergames for Training and Assessment Purpose

Modern computer technology has yielded creative and motivating procedures especially for training mental and physical abilities, thereby reducing fall risk in older adults and promoting independence and participation in everyday life, respectively. Recent studies could show that applications such as tablet mobile devices with integrated memory training apps can be used even by people with early stage of dementia [1,2]. Some reviews reported that exergames that combine physical activity with digital gaming have also been found wide application in healthy [3], disabled older adults [4], and in people with cognitive impairment or dementia [5-7]. Just to name a few examples, a computerized Tai Chi game using the Microsoft's motion-capture device Kinect [8] or a computer tele-rehabilitation platform that combines game-based exercises with telemonitoring [9] has the potential to improve motor and cognitive functions in the older population. To measure effects of an exergame intervention, a broad variety of outcome measures are reported in the literature (eg, [10]).

Effects on motor or cognitive functions can be assessed by using "traditional" external outcome measures after gameplay such as established paper-pencil tests or questionnaires. A lot of exergames especially commercially available games such as The Nintendo's Wii Fit provides integrated approaches to evaluate balance performance. Sensors measure bodily movements, and algorithms automatically convert sensor information into quantitative data, for example, the center of pressure (COP) (eg, [10]). The COP is used to control the game tasks, to individually adjusting the gameplay to the user, and also to generate performance scores, which provide the users with instantaneous feedback about game performance in real time (eg, [11]). Scores assessing the users performances have been validated in a few studies but show inconsistent results: One study could show that scores of a step game had good discriminant validity to differentiate between fallers and nonfallers ($P \leq .001$) as well as moderate criterion-related validity (r=−.55 to −.69) and test–retest reliability ranged from poor to good (ICC=.35-.93) in cognitively healthy older adults [12]. In contrast, Wii Fit balance activity scores from different static and dynamic balance tasks performed by "recreationally active" adults (aged 27.0±9.8 years) ranged from poor-to-moderate test–retest reliability (ICC=.29-.74), and concurrent validity was also poor (r<.50) [13]. Goble et al [14] indicate that these results suggest that game software–based assessments displayed as scores are not effective to measure balance ability.

The data flow derived from the game software can also be used to develop specific computerized tests for quantification of performances such as balance control (eg, [15-18]). Although a wide range of results in terms of psychometric properties have been reported as no uniform protocols or outcomes were used for evaluation especially of the Wii Balance Board or the Xbox Kinect [19], validation studies were successful to show that commercially available games basically provide a good basis to create reliable and valid game-based assessments: For example, the Wii Balance Board COP assessment in healthy younger adults showed good-to-excellent test–retest reliability (ICC=.66-.91), inter-rater reliability (ICC=.79-.89), intra-rater reliability (ICC=.70-.92), and concurrent validity (ICC= .73-.89) during single and double limb standing [16,18]. Another study [20] investigated test–retest reliability and construct validity of the Wii Balance Board in people after stroke (mean age 68.3±15.1 years) and showed excellent reliability (ICC=.82-.98) and poor-to-moderate correlations between the Wii Balance Board outcomes and clinical tests. The Wii Balance Board also showed excellent concurrent validity (ICC=.92-.98) with force platform–based assessments during balance tasks in people with Parkinson's disease [21]. A previous study showed excellent concurrent validity in balance tasks (r>0.75) using the Microsoft Xbox Kinect in healthy adults [17]. Schoene et al [22] have evaluated a custom-made dance mat device to assess stepping reaction times and showed excellent test–retest reliability (ICC=.90) and high correlations with other laboratory assessments (r=.86). Test–retest analyses of assessments using a force platform balance measurement and training device (Good Balance), in which participants had to move their COP along a track (a circle or a zigzag figure) shown on a computer screen, indicated also good results (ICC=.71-.83) [23].

Beyond these commercially available games used for assessment purpose, some researchers have developed and validated own game-based assessment approaches. For example, Szturm et al [24] examined a dual-task computer game–based platform that integrates head tracking and cognitive tasks with balance demands and showed moderate-to-high test–retest reliability (ICC=.55-.75) in healthy, community-dwelling individuals.

Limitations of Exergame-Assessments Found in the Literature

Although mentioned studies demonstrated that exergames can be used for reliable and valid quantification of motor performances such as balance control, assessment tasks derived from the data stream of a game that show good data on psychometric properties do commonly not represent the original game tasks used for training purpose. Some intervention studies demonstrated that dynamic aspects of COP are typical for game-based balance training requiring the participant to shift their COP to perform tasks such as catching and moving objects or popping rising balloons (eg, [11,15,25,26]). However, to measure changes in balance ability after exergaming, adequate but external instruments (neuropsychological test batteries or functional tests such as the Timed Up and Go [27] or the Berg Balance Scale [28]) have been applied in certain trials (eg, [26,29-31]). Young et al [15] have developed an interface that retrieves information from the Wii Balance Board, which can be used to create a series of balance games for both training and assessment. However, in this study, effects of a training with game tasks developed with the interface (catching apples falling from a tree and popping rising bubbles) were assessed using tasks also created with the interface but which comprised different demands as participants were instructed to maintain a static standing position for 30 seconds with eyes closed and open. Similar internal assessment approaches have also been applied, for example, by Betker et al [25].

Although we could identify 2 studies that have validated assessments based on the game's original training tasks to obtain a reliable and valid feedback of balance ability during gameplay [22,23], assessment software derived from the data stream of a game commonly includes tasks other than the original game task used for training. Therefore, some of the data might be only loosely associated (eg, use of Timed Up and Go or single or double limb standing with eyes closed to evaluate game-based training gained for shifting the COP while standing) as most validated assessment tasks relate to similar but not identical or integrated performances trained by using exergames.

Despite an increasing number of validation studies evaluating commercially available games or research grade systems not only in young participants without any injuries and history of neurological and musculoskeletal diseases [16-18,24] but also in patients with Parkinson's disease [21], patients after stroke [20], or frail nursing home residents [23], there is a lack of validation studies including people with diagnosed dementia. However, this patient sample could be a relevant target group for game-based training programs and assessment. For all identified validation studies including older adults, only a cognitive screening was performed allowing a mere classification of cognitive impairment by clinical screening tools, for example, the Mini Mental State Examination (MMSE) (eg, [32]), the Trail Making Test [22] or the Rapid Dementia Screening Test [12]. In some validation studies, mixed samples in terms of the cognitive impairment level might be examined as participants were inadequately screened for cognitive status or screening process was not described in detail (eg, [13,23,24]). We found only 1 validation study that examined feasibility and test–retest reliability of a force platform assessment in people with diagnosis of Alzheimer's disease [33]. Participants had to move their COP by shifting their weight to 8 numbered targets presented on a screen. Study results showed acceptable test–retest reliability (ICC=.48-.71). Information regarding dementia diagnoses in this study was collected using the MMSE [34] and the Frontal Assessment Battery [35], but only a small sample (n=14) was included in the study.

As outlined previously, there are a number of validation studies mainly focusing on reliability and validity analyses of assessments incorporated into commercially available exergames or of other game-based systems specifically developed for assessment purpose (16-18,20-24). However, we found no study that has conducted not only reliability and validity analyses but also targeting supplementary analyses on sensitivity to change and feasibility to guarantee high methodological quality of the assessment tools. It is striking that previous validation studies did not present especially sensitivity to change data, although this is a paramount objective for validation to evaluate responsiveness in intervention studies. Feasibility analyses based on documentations of completion rates, reasons for missing responses, and mean completion time of assessment process are also generally lacking. Commonly, questionnaires to measure motivation or subjective rating regarding difficulty of the tasks are applied.

Summary and Aim of the Present Study

In summary, there are a number of studies demonstrating that exergames can be used for reliable and valid quantification of performances such as balance control and provided good data on psychometric properties. However, (1) assessments often include tasks other than the original game task used for training and therefore relate to similar but not identical or integrated performances trained, (2) people with diagnosed dementia have insufficiently been addressed in validation studies, and 3) studies often lack additional validation analyses such as sensitivity to change to document psychometric properties. The purpose of this study was to complement the pool of validated game-based measurements that have already been reported by a number of evaluations. We have developed task-specific assessment parameters based on data directly derived from the data stream during the game task of a training device (Physiomat), which are therefore direct marker of the training tasks. Parameters test a much more complex performance including the interplay (dual task) between challenging motor and cognitive tasks. This approach much better documents the actual game performance compared with another balance performance documentation (eg, during double limb standing with eyes closed) as used in other studies. We aimed to evaluate this internal assessment approach of the training device (Physiomat) in multimorbid, frail elderly with mild-to-moderate dementia. We present data on construct validity, test–retest reliability, sensitivity to change, and feasibility.

Methods

Study Design

The validation study was part of a double-blind randomized controlled trial (ISRCTN37232817) to improve motor-cognitive functions in people with mild-to-moderate dementia. To prevent high test burden in the frail and multimorbid sample, validation measures were split. Assessments for validation were conducted before intervention (T1: construct validity and feasibility) and after the intervention period (T2: sensitivity to change) with repeated measures 2-5 days after T2 (retest: test–retest reliability). The trial was performed according to the Helsinki declaration and was approved by the Ethics Committee of the University of Heidelberg.

Recruitment

Participants were consecutively recruited including geriatric patients, nursing home residents, and community-dwelling persons. Inclusion criteria were: age>65 years, place of residence <15km from the study center, no severe neurological, cardiovascular or psychiatric disorders, or visual deficits, ability to walk 10 m without using a walking aid and written informed consent (obtained by the patient or by a legal representative). Individuals were screened for cognitive impairment using the MMSE [34]. In those with an MMSE of 17-26 indicating cognitive impairment, a comprehensive neuropsychological assessment was performed based on an established neuropsychological test battery (Consortium to Establish a Registry for Alzheimer's Disease—CERAD) [36] and the Number Connection Test (ZVT-G) [37], a modified version of the Trail Making Test (TMT) [38]. Internationally established criteria for cognitive subperformances as assessed in CERAD were used as further inclusion criteria along with amnestic reports for diagnosis of probable dementia. Patients who met predefined criteria for dementia diagnosis based on CERAD results (cognitive subperformances lay under the 10% percentile of the sample corresponding to a z-value of −1.3) were included in the study.

Measurement and Data Collection

Physiomat Assessment

Physiomat (Physio = physiological, M=medical, A=active, T=therapy, EPL medical engineering [39], Figure 1) has been developed as a training device to improve balance performance.

The operating principle of Physiomat based on a specific combination of swivel joints fixed on 2 independent levels enabling bending, tilting, and rotation movements. This device’s construction yields a special 3-dimensional (sagittal, frontal, and transverse level) movement sequences. The internal assessment approach of Physiomat is actually not comparable with the commonly used balance platforms with integrated pressure or inclination sensors. It uses 2 displacement sensors—one for the anterior–posterior motion, the other for the medio–lateral tilting and rotation motion—to record the movements of the platform. This is done by measuring the changes in resistance (measurement range 0-100 kohm, measurement accuracy 0.1 kohm, sampling rate 100 Hz). This sensor information is converted into a standard signal (normed electrical signals) by an analog-digital converter. Standard signal output acquired via the displacement sensors generates digital numerical values (digits) in a measuring range of 0-1000 digits for each motion axis. This means that the movement excursion of the platform is measured in digits/ms, and based on that sway path and sway area are determined as quantitative parameters. Movement excursion of the platform measured in digits/ms is also presented to the participants in terms of a visual feedback on the monitor in real time to control the cursor by mapping it to the target motion to solve Physiomat tasks.

The software with the training and assessment tasks as used in this study was specifically developed by the research group of the AGAPLESION Bethanien Hospital Heidelberg in cooperation with EPL to target motor-cognitive performances in patients with dementia. The assessment strategy derived from the data stream of Physiomat game tasks. The Physiomat assessment linked cognitive and motor-functional demands; concurrent dual tasks of various elements on balance ability (weight-shifting tasks and postural control while standing) with specific cognitive subperformances such as executive functions are required.

To provide a motor-cognitive task to test a complex performance including the interplay (dual task) between challenging balance and cognitive tasks, the Physiomat-Trail-Making Tasks (PTMTs) have been developed. Compared with other exergames (eg, Nintendo Wii) including virtual reality tasks where an avatar is displayed on the screen that follows the participant’s movements that do commonly not coincide with evidence-based neuropsychological tasks, we incorporated an internationally established cognitive test (the Trail-Making Test) into a balance training device. This test has been modified and successfully validated for use in older and cognitively impaired persons [37] with the introduction of a learning phase using an increasing number of digits before testing and reducing the complexity of the task by positioning of the digits. This modified version prevents frequent floor effects as compared by the original tests and is valid for the target sample of this study of cognitively impaired persons. The test is sensitive also for early deficits in the course of neurodegenerative diseases and documents cognitive subperformances such as executive functions including procedural memory, visual–spatial orientation, and attention-related performances (especially in the test setting as used in this study with the concurrent dual task of balance control with the specific cognitive subperformance of divided attention, see in the following section). These cognitive subperformances appear relatively early in the course of the disease and are therefore an adequate test for the study sample of patients with mild-to-moderate stage dementia.

PTMTs include different performance levels as defined by the number of digits to be tested. The participants were instructed to move the cursor on the screen (indicating the participant’s bending, tilting, and rotation motion) directly to each numbered target with the aim to connect an increasing number of digits (number of digits: 4, 7, 9, 14, 20) as fast as possible by weight shifting (Figure 2). Physiomat platform was not fixed and allows movements in all directions, which must be controlled by the user. The degree of movements is partly limited by rubber rings attached to the corners of the platform. We used several rings to achieve a feasible motor challenge but did not modify this for the rather homogeneous sample with impaired motor status, advanced age, and multimorbidity including cognitive impairment. Participants were instructed to use handles (see Figure 1) to control movements. For validation purpose, we only used results of the simple (4 digits), moderate (9 digits), and complex (20 digits) PTMT, as we assumed that this range of complexity levels would be sufficient for the study purpose. With the standardized motor task and the standardized but increasing challenge level of the cognitive task, we ascertained a standardized assessment procedure.

We also applied an additional standardized motor task without an increasing challenge level of cognitive task to study psychometric properties. Instructions were to move the cursor from the center of the screen directly to the targets highlighted as a moving yellow ball on the screen as fast as possible. This Follow-The-Ball Task (FTBT) was used to assess participants' ability to move their center of mass by shifting their weight to the highlighted targets (Figure 3).

We also used 3 Physiomat balance tasks (PBTs) challenging postural control while standing without using the handles (Figure 4). Tasks also differed in complexity levels (keeping postural control for 3, 10, and 30 seconds). The platform was also not fixed but contrary to the PTMTs and the FTBTs, the degree of movements was limited by a larger number of rubber rings.

During assessment, no physical assistance or cueing was allowed. For each Physiomat measurement, the best performance of 2 trials was used for statistical analyses. Temporal (test duration in seconds) and spatial (sway path in mm/s, sway area in mm^2/s) parameters have been documented as main study end points of assessment.

Furthermore, we documented the number of successfully performed Physiomat tasks for each measurement by doing dichotomous coding (1=successful; 0=not successful). Based on that we calculated a scoring for PTMT (PTMT score), for PBT (PBT score), and for the complete Physiomat assessment (total score) by summarizing the numerical codings. For PTMT score and PBT score, up to 3 points could be achieved for each as there were 3 levels (PTMTs: 4, 9, and 20 digits; PBTs: 3, 10, and 30 seconds) indicating that all complexity levels have been performed successfully. For the total score including all PTMTs, PBTs, and the FTBT, a maximum of 7 points could be achieved.

Figure 1. Physiomat including a three-dimensional moveable plate with integrated sensors for displacement measurement. It is connected with a computer and a monitor. Grab rails on each side ensure stability of the patients during training and assessment.

Figure 2. Example for complex Physiomat Trail-Making Task (PTMT). Participants were instructed to capture digits in correct order as fast as possible by shifting the weight to numbered targets.

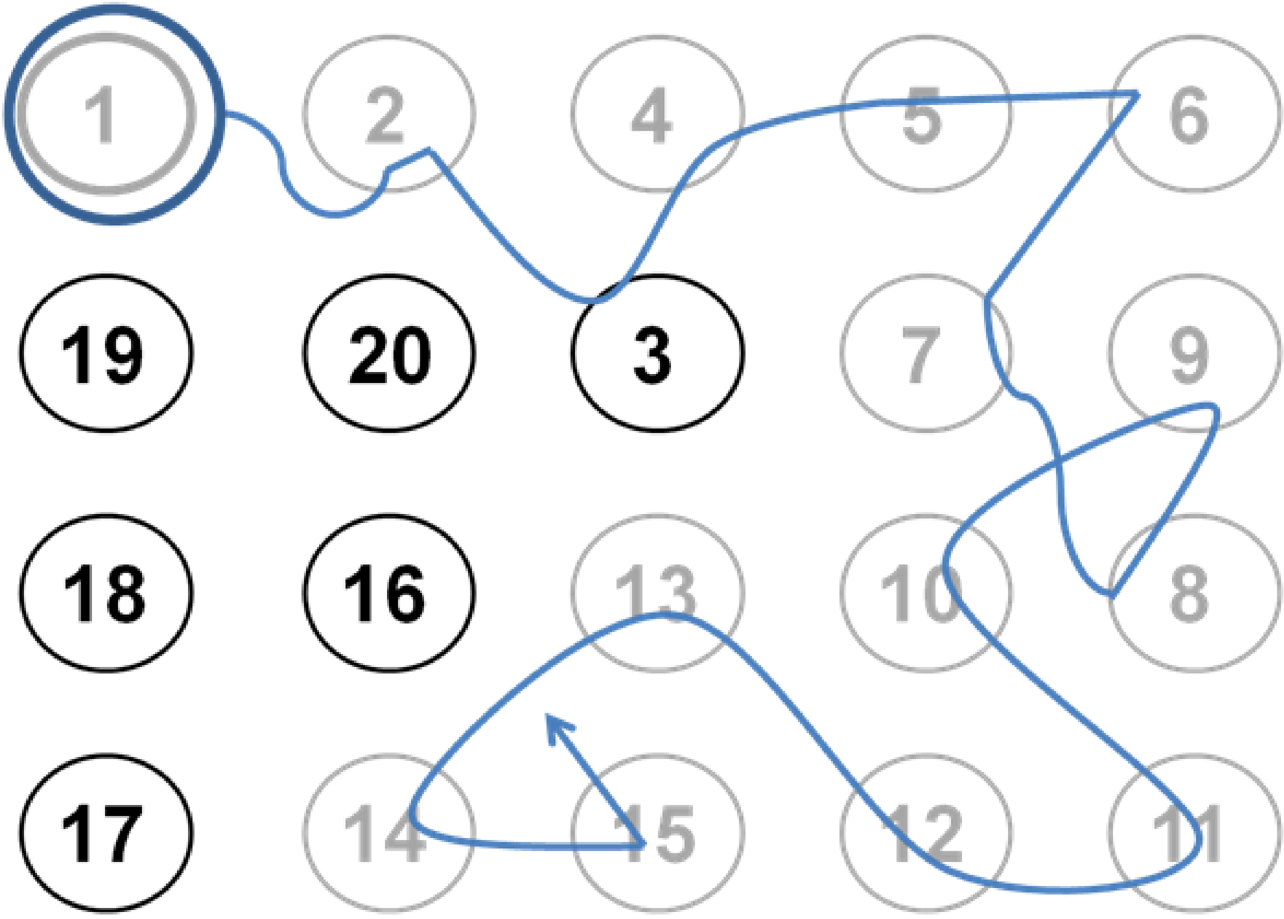


Figure 3. Physiomat Follow-the-Ball Task (FTBT). Participants were instructed to follow a yellow ball during displacement of center of mass as fast as possible using the handles.

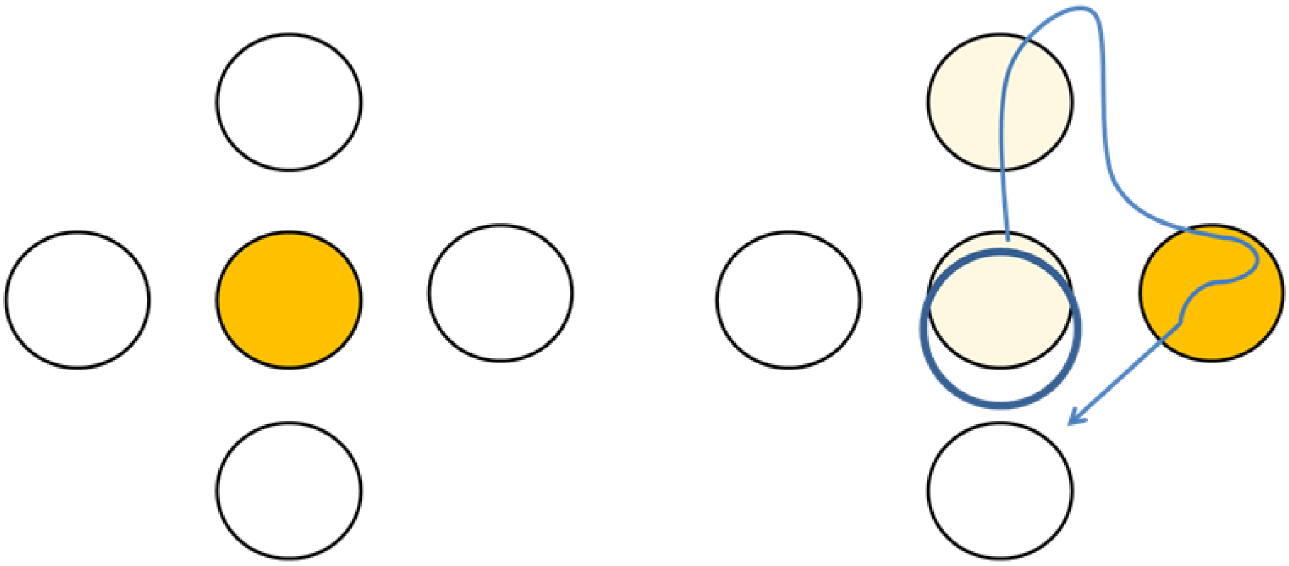
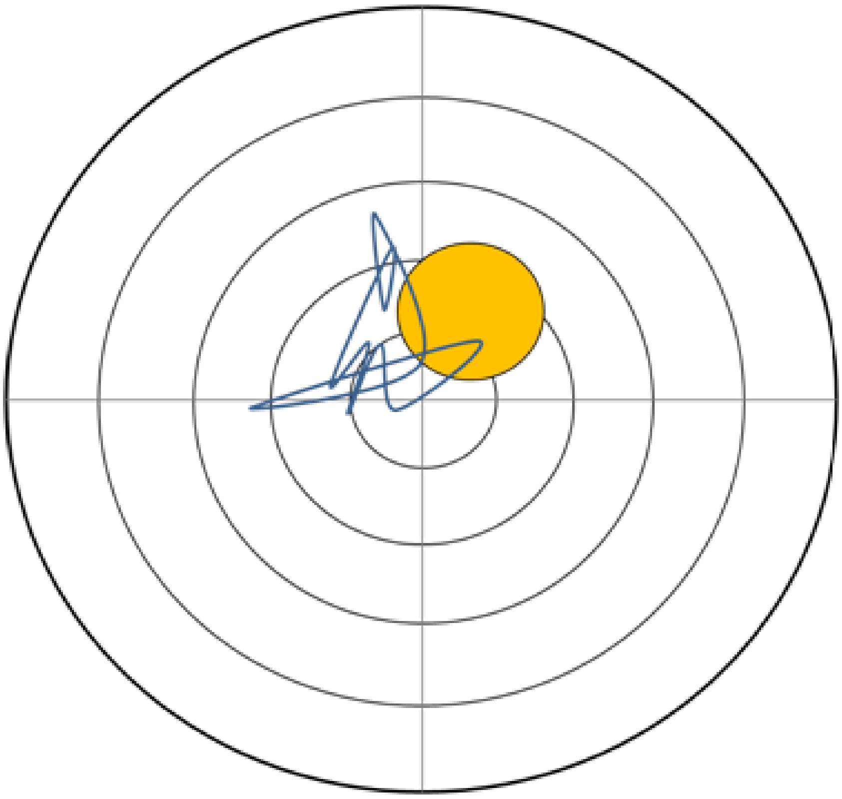


Figure 4. The Physiomat Balance Task (PBT). Participants were instructed to stand still on the plate and keep in the middle of a yellow target for 3, 10 and 30 seconds without using the grab rails.



Descriptive Measures

Demographic and clinical characteristics including age, gender, education, social status (independent or institutionalized), number of falls in the previous year, and number of medications and diagnoses were documented. Psychological status was assessed by the Geriatric Depression Scale [40] for depressive symptoms and by the Falls Efficacy Scale-International (FES-I) [41] for fall-related self-efficacy. Motor-functional status was assessed by the Timed-Up and Go (TUG) [27] and performance-oriented mobility assessment (POMA) [42]. Cognitive status was screened by the MMSE [34].

The TUG, POMA, and MMSE were also used for validity analyses. The following tests were additionally used for testing construct validity: The simple Physiomat balance task (10 seconds) and the FTBT as Physiomat balance tests (Figures 3 and 4) and word list immediate recall [43] as a subtest of the CERAD, the ZVT-G, and repeating numbers (ZN-G) [37].

Statistical Analysis

Statistical analyses were performed on SPSS 22.0 for Windows. Descriptive data are presented as means and standard deviations (SD) or number and percentages (%) as appropriate. The Kolmogorov–Smirnov test and histograms were used to analyze distribution of data. In case data were not normally distributed, nonparametric tests were used in addition to parametric tests.

Construct Validity

Spearman's rank correlations between temporal parameters (test duration in seconds) of simple, moderate, and complex PTMTs and theoretically related motor-functions, respectively, cognitive measures as described previously were calculated to test 8 predefined hypotheses [44,45]. The hypotheses are given in detail in Table 3. Our assumptions can be summarized briefly as follows:

Cognitive measures as MMSE and especially ZVT-G were expected to be associated with PTMTs because both motor-cognitive Physiomat tasks and the mentioned tests require multiple cognitive abilities. A higher correlation with ZVT-G was expected because both tests assess a similar construct where attentional abilities and executive functions are demanded. In comparison, cognitive instruments measuring domain-related cognitive functions such as memory abilities (immediate wordlist recall and number repeating) were expected to be less associated with PTMTs.

Established motor-functional assessments (TUG and POMA) measuring postural control and gait performance were expected to be associated with motor-cognitive Physiomat tasks as also balance performances are demanded in each of the tests. The FTBT was expected to be highly correlated with all PTMTs because both tasks are performed during weight shifting.

Regarding the complexity levels of motor-cognitive Physiomat tasks, we expected that complex PTMTs were strongly associated with cognitive measures and simple PTMTs with motor-functional outcomes because of an increasing cognitive challenge by accelerating complexity level.

Test–Retest Reliability

Test–retest assessments were performed within 2-5 days by the same examiner to exclude interobserver variability. For test–retest analyses, Spearman's rank correlations were calculated according to Cohen's criteria [46] for low (r_s < .2), moderate (r_s = .2-.5), or high (r_s > .5) correlations. In addition, we used ICC coefficients using a 2-way mixed model [47]. For ICCs, 95% confidence intervals were given. ICCs were considered as low (ICC <.40), moderate (.40 ≤ ICC ≤.75), or high (ICC >.75) [48].

Sensitivity to Change

To study the responsiveness of trained Physiomat tasks, we used baseline values of a RCT (ISRCTN37232817), which will be published in the near future. Progressive Physiomat training (10 minutes twice a week) in 47 participants was part of this comprehensive intervention (1.5 hours, twice a week for 10 weeks) in groups at the maximum of 7 participants including dual tasking (walking while counting) and training of compensatory sit to stand movement maneuvers to improve motor-cognitive abilities in people with dementia. Subjects of the control group (n=43) underwent a supervised, unspecific motor-functional group training for 10 weeks (1 hour, twice a week) including low-intensity strength training and flexibility exercises for the upper limbs while seated. In this paper, only the results of the intervention group that conducted the Physiomat training as one part of the overall training program was used to document sensitivity to change.

Physiomat exercise sessions were composed of the FTBT introducing the participants to Physiomat and to provide relevant strategies of balance displacement. In addition, PTMTs were trained, and complexity level was gradually increased according to the capacity of each participant. Physiomat balance tasks were not part of the intervention and responsiveness analyses. The Wilcoxon test was applied to test sensitivity to change. Effect sizes were calculated using standardized response means (SRMs) [44] according to Cohen’s criteria (small effect < .2, moderate effect .2 ≤ SRM > .5, and large effect .8 and above) [46].

Feasibility

To study feasibility of motor-cognitive (PTMTs) and motor-functional (PBT and FTBT) Physiomat measures, percentages of successfully completed tasks (completion rates), reasons of missing responses, and the mean completion time as measured by a stopwatch were documented. To assess safety of the participants, issues such as a slip or fall and any clinical events during testing were systematically documented.

It should be noted, that only 7 of 15 Physiomat tasks assessed for further analyses were used for validation purpose. Documentation of feasibility outcomes was related to the overall Physiomat assessment protocol, which additionally included a comprehensive instruction period and breaks between single performance levels for the frail, multimorbid, and cognitively impaired participants. Therefore, completion time with regard to the assessments to test feasibility will be overestimated.

Results

Participants’ Characteristics

The study sample included 105 (mean age 82.7±5.9) multimorbid and cognitively impaired subjects living at home or in nursing homes. Further demographic and clinical characteristics are summarized in Table 1.

Table 1. Descriptive characteristics of the participants.[a]

Characteristics[b]	All participants N=105
Age (years), mean (SD)	82.7 (5.9)
Gender (female), n (%)	76 (72.4)
Education (years), mean (SD)	11.8 (2.9)
Social status (institutionalized), n (%)	31 (29.5)
Cognitive status MMSE[b] (sum score), mean (SD)	21.9 (2.8)
Depression GDS[c] (sum score), mean (SD)	2.8 (2.3)
Indicated depression (GDS score >5), n (%)	19 (18.1)
Number of falls, n (%)	49 (46.7)
Fear of falling FES-I[d] (sum score), mean (SD)	9.2 (2.8)
Number of diagnosis, mean (SD)	8.2 (4.1)
TUG[e] (test duration in seconds)	18.4 (11.3)
POMA[f] (total score)	22.3 (4.0)

[a] Given are sample size (N), mean and standard deviation (SD) or percentages (%) of the sample of all characteristics.

[b] MMSE: Mini-Mental-State Examination

[c] GDS: Geriatric Depression Scale

[d] FES-I: Falls Efficacy Scale International

[e] TUG: Timed Up and Go

[f] POMA: Performance-Oriented Mobility Assessment

We separated the validation measures into 3 assessment sessions (feasibility and construct validity analyses at baseline, sensitivity to change measures after the intervention, and test–retest reliability assessment within the subsequent 2-5 days) to prevent high test burden. Assessments for feasibility and construct validity analyses were not practicable for 6 of 105 subjects (5.7%) because of serious motor-functional disability (n=3), visual impairment (n=2), and fear of assessment using Physiomat (n=1). Sensitivity to change was assessed in 47 participants in a subsample of the intervention group (n=56) as 9 participants (16.1%) dropped out owing to physical limitations (n=3), noncompliance (n=5), and death (n=1). Test–retest reliability could not be assessed in 31 of 105 participants (29.5%) because of physical limitations and pain (n=11), noncompliance (n=13), death (n=4), and increased visual impairment (n=3).

Construct Validity

PTMTs have shown a high association with established cognitive paper-and-pencil tests (ZVT-G and MMSE) and moderate associations with motor-functional instruments (TUG and POMA) indicating a good construct validity of motor-cognitive Physiomat outcomes. Correlations between simple, moderate, and complex PTMTs and cognitive as well as motor-functional paper-and-pencil tests are illustrated in Table 2.

Table 2. Construct validity of motor-cognitive Physiomat tasks.[a]

Test	Variable (unit)	Simple PTMT[b] (*P* value)	Moderate PTMT (*P* value)	Complex PTMT (*P* value)
FTBT[c]	Duration (time in seconds)	68[d] (*P* ≤.001)	.71[d] (*P* ≤.001)	.61[d] (*P* ≤.001)
PBT[e] 10 sec.	sway path (mm/second)	0.11[f] (*P*=.31)	−0.03[f] (*P*=.79)	−0.34[g] (*P*=.10)
POMA[h]	Total score	−0.22[g] (*P*=.03)	−0.40[g] (*P* ≤.001)	0.08[f] (*P*=.71)
TUG[i]	Duration (time in seconds)	0.22[g] (*P*=.03)	0.48[g] (*P* ≤.001)	0.19[f] (*P*=.35)
MMSE[j]	Total score	0.29[g] (*P*=.004)	0.35[g] (*P*=.002)	0.66[d] (*P* ≤.001)
ZVT-G[k]	Duration (time in s)	0.36[g] (*P* ≤.001)	0.44[g] (*P* ≤.001)	0.83[d] (*P* ≤.001)
ZN-G[l]	Total score	−0.25[g] (*P*=.02)	−0.19[f] (*P*=.12)	−0.22[g] (*P*=.30)
Wordlist immediate recall	Number of quoted words	−0.33[g] (*P*=.004)	−0.42[g] (*P*=.04)	−0.16[f] (*P*=.12)

[a] Given are Spearman's rank correlations (r_s) between simple (4 numbers), moderate (9 numbers), and complex (20 numbers) PTMTs and motor-functional (FTBT, moderate PBT, POMA, and TUG) and cognitive outcomes (MMSE, ZVT-G, ZN-G and wordlist immediate recall).

[b] Physiomat Trail-Making Task

[c] FTBT: Follow-The-Ball Task

[d] High correlation (r_s > .5)

[e] PBT: Physiomat-Balance-Task

[f] Low correlation (r_s < .2)

[g] Moderate correlation (.2 ≥ r_s ≤ .5)

[h] POMA: performance-oriented mobility assessment

[i] TUG: Timed Up and Go

[j] MMSE: Mini-Mental-State Examination

[k] ZVT-G: modified version of the Trail-Making-Test A

[l] ZN-G: repeating numbers

Correlations between measures assessing motor-functional performances (TUG, POMA, FTBT, moderate PBT—10 seconds) and PTMTs were low to high (r_s= −.03-.71). Highest correlations (P ≤ .001) were found for duration of FTBT (r_s=.61-.71). Correlations with PBT, POMA, and TUG were low to moderate (r_s= −.03 to .48). Highest correlations with cognitive outcomes (P ≤ .001) were found between complex PTMT and MMSE (r_s=.66) and ZVT-G (r_s=.83). Correlations with instruments measuring memory skills were low to moderate (ZN-G: r_s= −.19 to −.25; wordlist immediate recall: r_s= −.16 to −.42).

Construct validity assessed by testing 8 a priori hypotheses is presented in Table 3. Except hypothesis number 8, all assumptions could be confirmed (87.5%) for PTMTs regarding temporal parameters (time in seconds) indicating excellent construct validity [45].

Table 3. Results of 8 predefined hypotheses.

No.	Hypothesis[a]	Hypothesis confirmed?
	Expected associations with cognitive outcome measures	
1	We expected moderate-to-high associations between PTMTs[b] and MMSE[c] as both assessments measure multiple cognitive functions.	Yes
2	We expected more pronounced associations between PTMTs and ZVT-G[d] as both assessments measure a similar construct where particularly attentional abilities are demanded.	Yes
3	We expected moderate associations between PTMTs and memory tests as both cognitive tests would cover different cognitive subperformances as compared with PTMTs.	Yes
4	We expected higher associations of cognitive outcome measures with increasing complexity of PTMTs as for difficult Physiomat levels cognitive demands may predominate.	Yes (except ZN-G[e])
	Expected associations with motor-functional outcome measures	
5	We expected associations between PTMTs and TUG[f] as well as POMA[g] as also balance performances are demanded in each of the assessments, although not comparable in type of assessment.	Yes
6	We expected pronounced associations between PTMTs and FTBT[h] as FTBT is a preliminary Physiomat training task requiring similar strategies of balance performances.	Yes
7	We expected a less association between PTMTs and the moderate PBT[i] (10 seconds) as this Physiomat task requires a different strategy of balance performance.	Yes
8	We expected higher associations of motor-functional outcomes with decreasing complexity of PTMTs as for simple Physiomat levels motor-functions demands may predominate.	No

[a] Hypotheses are given for Spearman's rank correlations between PTMTs, motor-functional outcomes (hypotheses 5-8), and cognitive outcomes (hypotheses 1-4) of selected comparison measurement instruments.

[b] PTMT: Physiomat Trail-Making Task

[c] MMSE: Mini-Mental-State Examination

[d] ZVT-G: modified version of the Trail-Making-Test A

[e] ZN-G: repeating numbers

[f] TUG: Timed Up and Go

[g] POMA: performance-oriented mobility assessment

[h] FTBT: Follow-The-Ball Task

[i] PBT: Physiomat-Balance Task

Test–Retest Reliability

For almost all outcomes of Physiomat measures and for requirement level concerning all Physiomat tasks (total, PBT, and PTMT score) moderate-to-high correlations between test and retest assessment were found indicating good to excellent test–retest reliability (Tables 4 and 5).

Table 4. Test–retest results of all Physiomat tasks and requirement level (Spearman's rank correlations).

Test	Variable (unit)	N	Mean (SD) Test	Mean (SD) Retest	r_s	*P* value
PBT[a] 3 Sec						
	Sway path (mm/second)	71	134.4 (83.9)	120.6 (84.1)	.48[b]	≤.001
	Sway area (mm^2/second)		627.7(1311.3)	534.7(1289.1)	.45[b]	≤.001
PBT 10 Sec						
	Sway path (mm/second)	68	571.3 (312.2)	568.7 (292.6)	.68[c]	≤.001
	Sway area (mm^2/second)		654.5 (1164.8)	627.7 (1120.2)	.60[c]	≤.001
PBT 30 Sec						
	Sway path (mm/second)	61	1719.3 (1020.5)	1589.8 (844.1)	.78[c]	≤.001
	Sway area (mm^2/second)		750.2 (1729.2)	563.7 (1015.2)	.75[c]	≤.001
FTBT[d]						
	Sway path (mm/second)	73	3449.1 (1044.2)	3269.1 (1005.5)	.74[c]	≤.001
	Duration (time in seconds)		20.9 (5.4)	21.2 (6.8)	.69[c]	≤.001
Simple PTMT[e]	Sway path (mm/second)	73	1883.7 (558.5)	1774.7 (343.4)	.59[c]	≤.001
	Duration (time in seconds)		8.2 (2.8)	8.2 (2.9)	.60[c]	≤.001
Moderate PTMT						
	Sway path (mm/second)	69	3722.3 (910.9)	3630.9 (923.8)	.78[c]	≤.001
	Duration (time in seconds)		20.8 (5.9)	19.9 (6.2)	.74[c]	≤.001
Complex PTMT						
	Sway path (mm/second)	47	8319.4 (2220.8)	8111.1 (2170.9)	.80[c]	≤.001
	Duration (time in seconds)		51.0 (16.2)	49.1 (16.7)	.87[c]	≤.001
Total Score		74	6.3 (1.1)	6.3 (1.1)	.89[c]	≤.001
PTMT Score		74	2.6 (0.6)	2.6 (0.6)	.89[c]	≤.001
PBT Score		74	2.8 (0.7)	2.7 (0.7)	.87[c]	≤.001

[a] PBT: Physiomat-Balance Tasks

[b] moderate correlation ($.2 \geq r_s \leq .5$)

[c] high correlation ($r_s > .5$)

[d] FTBT: Follow-The-Ball Task

[e] PTMT: Physiomat Trail-Making Task

Regarding Spearman's rank correlations reliability for the total sample was moderate to high (r_s=.45-.89) for all variables. Highest correlations were found for sway path (r_s=.80) and duration (r_s=.86) of the complex PTMT as well as for requirement level (total score r_s=.89; PTMT score r_s =.89; and PBT score r_s=.87). Moderate correlations were only found for sway path (r_s =.48) and sway area (r_s=.45) of the simple PBT (3 seconds). Overall, it could be observed that correlations increased with the duration of PBT and the complexity of PTMT.

Table 5. Test–retest results of all Physiomat tasks and requirement level (ICCs[a]).

Test	Variable (unit)	N	Mean (SD) Test	Mean (SD) Retest	ICC (95%CI)	*P* value
S						
	Sway path (mm/second)	71	134.4 (83.9)	120.6 (84.1)	.50[b] (.30-.66)	≤.001
	Sway area (mm²/seconds)		627.7 (1311.3)	534.7 (1289.1)	.73[b] (.59-.82)	≤.001
PBT[c] 10 Sec						
	Sway path (mm/second)	68	571.3 (312.2)	568.7 (292.6)	.66[b] (.50-.78)	≤.001
	Sway area (mm²/s)		654.5 (1164.8)	627.7 (1120.2)	.57[b] (.38-.71)	≤.001
PBT 30 Sec						
	Sway path (mm/seconds)	61	1719.3 (1020.5)	1589.8 (844.1)	.73[b] (.59-.83)	≤.001
	Sway area (mm²/seconds)		750.2 (1729.2)	563.7 (1015.2)	.32[d] (.08-.53)	.005
FTBT[e]						
	Sway path (mm/second)	73	3449.1 (1044.2)	3269.1 (1005.5)	.84[f] (.76-.89)	≤.001
	Duration (time in seconds)		20.9 (5.4)	21.2 (6.8)	.79[f] (.68-.86)	≤.001
Simple PTMT[g]						
	Sway path (mm/second)	73	1883.7 (558.5)	1774.7 (343.4)	.47[b] (.27-.63)	≤.001
	Duration (time in seconds)		8.2 (2.8)	8.2 (2.9)	.55[b] (.37-.69)	≤.001
Moderate PTMT						
	Sway path (mm/second)	69	3722.3 (910.9)	3630.9 (923.8)	.74[b] (.61-.82)	≤.001
	Duration (time in seconds)		20.8 (5.9)	19.9 (6.2)	.79[f] (.68-.87)	≤.001
Complex PTMT						
	Sway path (mm/second)	47	8319.4 (2220.8)	8111.1 (2170.9)	.82[f] (.69-.89)	≤.001
	Duration (time in seconds)		51.0 (16.2)	49.1 (16.7)	.83[f] (.72-.91)	≤.001
Total score						
	Score	74	6.3 (1.1)	6.3 (1.1)	.92[f] (.88-.95)	≤.001
PTMT score						
	Score	74	2.6 (0.6)	2.6 (0.6)	.90[f] (.85-.94)	≤.001
PBT score						
	Score	74	2.8 (0.7)	2.7 (0.7)	.89[f] (.84-.93)	≤.001

[a] ICC: intraclass correlations
[b] moderate ICC (.40 ≤ ICC ≤.75)
[c] PBT: Physiomat-Balance Tasks
[d] Low ICC (< .40)
[e] FTBT: Follow-The-Ball Task
[f] high ICC (ICC >.75)
[g] PTMT: Physiomat Trail-Making Task

Regarding ICCs, moderate-to-high test–retest reliability (ICC=.47-.92) was found for almost all variables. Only sway path of the complex Physiomat balance task (30 seconds) was below the threshold of moderate reliability (ICC <.40). High ICCs were found for sway path (ICC=.82) and duration (ICC=.84) of not only the complex PTMT and for requirement level (total score ICC=.92; PTMT score ICC=.90; PBT score ICC=.89) but also for the duration of the moderate PTMT (ICC=.79). ICCs increase with complexity level of PTMT. For sway path (ICC=.84) and duration (ICC=.79) of FTBT, high correlations were proven, too.

Spearman's correlations and ICCs were all significantly different from zero at a.01 level (*P* ≤ .001).

To examine potential influence of deviating subsamples in different conditions (larger sample in simple condition, selection to high functioning participants in more complex conditions) on test–retest reliability, we conducted a subsequent test–retest analysis of 47 participants, which successfully conducted all complexity levels (Table 6).

Table 6. Subanalysis of test–retest reliability of motor-cognitive Physiomat tasks.

Test	Variable (unit)	N[a]	Mean (SD) Test	Mean (SD) Retest	ICC (95%CI)	*P* value
Simple PTMT[b]						
	Sway path (mm/second)	47	1799.9 (417.6)	1793.1 (313.9)	.36[c] (.09-.59)	.006
	Duration (time in seconds)		7.3 (2.2)	7.5 (2.5)	.44[d] (.18-.64)	.001
Moderate PTMT						
	Sway path (mm/second)	47	3660.4 (675.4)	3602.8 (674.5)	.75[d] (.59-.85)	≤.001
	Duration (time in seconds)		19.2 (4.2)	18.6 (4.3)	.79[e] (.66-.88)	≤.001
Complex PTMT						
	Sway path (mm/second)	47	8392.7 (2248.3)	8204.9 (2145.3)	.81[e] (.69-.89)	≤.001
	Duration (time in seconds)		51.0 (16.2)	49.1 (16.7)	.84[e] (.72-.91)	≤.001

[a] Subanalysis of test–retest reliability was conducted in a subsample of 47 participants, which successfully conducted all complexity levels of PTMTs.

[b] PTMT: Physiomat Trail-Making Task

[c] Low ICC (< .40)

[d] Moderate ICC (.40 ≤ ICC ≤.75)

[e] High ICCs (ICC >.75)

Results are comparable to the results of the nonselected group. Moderate-to-high test–retest reliability (ICC=.44-.84) was found for almost all variables. Only sway path of the simple PTMT was below the threshold of moderate reliability (ICC<.40). ICCs also increase with complexity level of PTMT.

Sensitivity to Change

All trained Physiomat tasks (FTBT and PTMTs) showed significant improvements indicating good-to-excellent sensitivity to change. Results of the Wilcoxon test and effects sizes (SRMs) are outlined in Table 7.

Table 7. Sensitivity to change for trained Physiomat tasks.

Test	Variable (unit)	N	Mean (SD) T1 – before intervention period	Mean (SD) T2 – after intervention period	*P* value[a]	SRM[b]
FTBT[c]						
	Sway path (mm/second)	47	4356.5 (3064.8)	3135.4 (569.6)	≤.001	0.4[d]
	Duration (time in seconds)		19.3 (4.6)	18.6 (4.3)	≤.001	0.7[e]
Simple PTMT[f] (4 numbers)						
	Sway path (mm/second)	45	2944.3 (4597.5)	1732.5 (307.3)	≤.001	0.3[d]
	Duration (time in seconds)		17.6 (21.9)	7.2 (1.9)	≤.001	0.5[e]
Moderate PTMT (9 numbers)						
	Sway path (mm/second)	37	4296.5 (1482.6)	3472.5 (643.1)	≤.001	0.7[e]
	Duration (time in seconds)		28.6(10.6)	18.5(4.1)	≤.001	1.1[g]
Complex PTMT (20 numbers)						
	Sway path (mm/second)	14	8361.7 (2269.5)	6850.6 (1341.2)	.01	0.8[e]
	Duration (time in seconds)		57.6 (11.7)	37.5 (7.8)	.001	2.0[g]
PTMT Score		47	2.0 (0.8)	2.8 (0.6)	≤.001	1.1[g]

[a]*P* values for Wilcoxon test applied to test differences between T1 and T2.

[b] SRM: standardized response mean (difference between the mean scores at assessments, divided by the mean scores of the standard deviation).

[c] FTBT: Follow-The-Ball Task

[d] Small effect size (SRM=0.2-0.5)

[e] Moderate effect size (SRM=0.5-0.8)

[f] PTMT: Physiomat Trail-Making Task

[g] Large effect size (SRM >0.8)

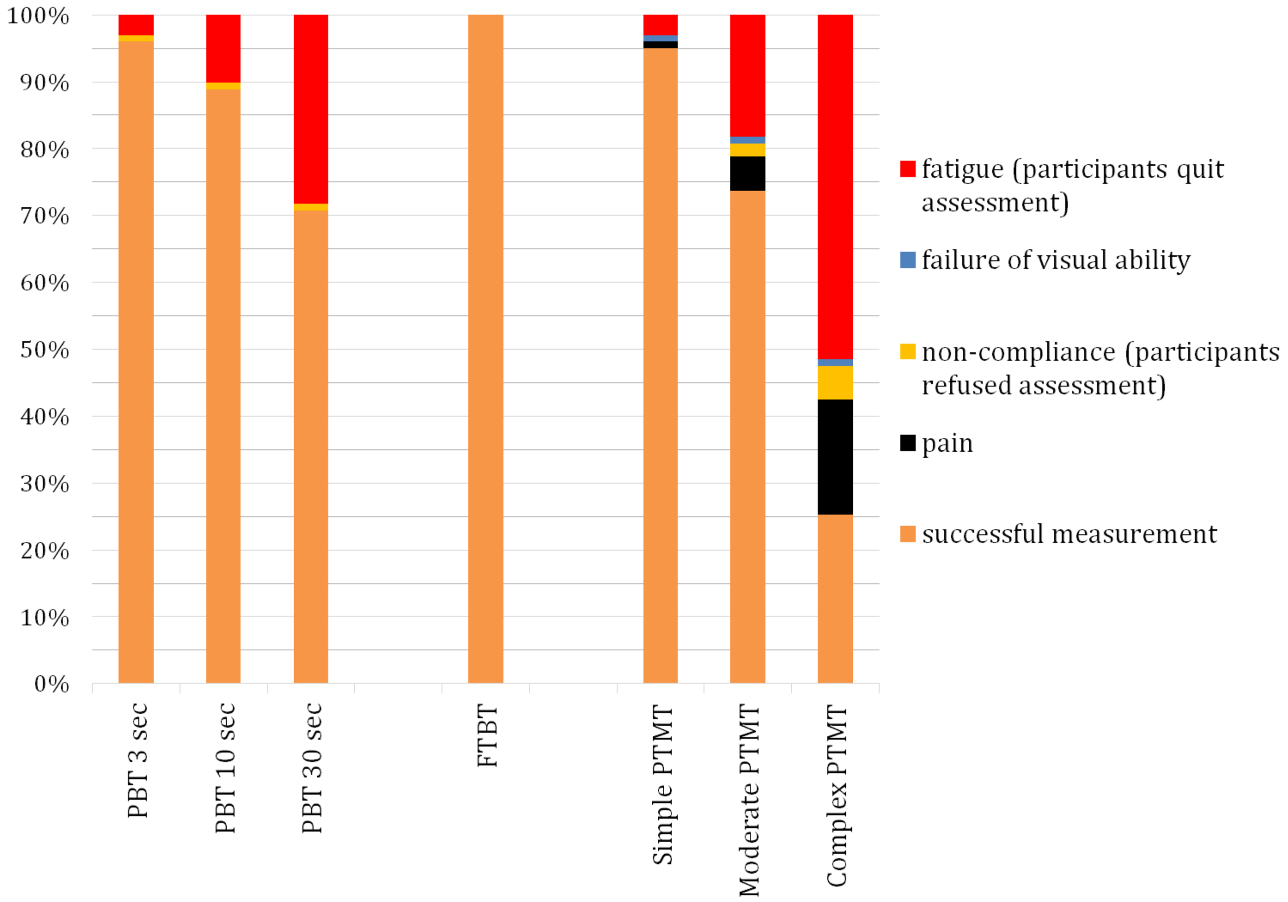
Results showed significant changes ($P \leq .001$) of both spatial and temporal parameters of PTMTs and the FTBT after a 10-week Physiomat intervention (twice a week/10 minutes). Large effect sizes were evident especially for duration (SRM=2.0) of complex PTMT and moderate PTMT (SRM=1.1). There were also significant changes ($P \leq .001$) for requirement level concerning motor-cognitive Physiomat tasks (PTMT score) with a large effect size (SRM=1.1).

Feasibility

Physiomat assessment was found to be practicable by frail, old, and multimorbid persons with mild to moderate–stage dementia. There were no clinical events, slips, or falls during training or assessment. A total of 99 of 105 participants (94.3%) could be tested at baseline. Five subjects could not be assessed due to severe motor-functional (hemiparesis) and visual (blindness) limitations. Willingness of the participants to attend to the computerized Physiomat assessment was excellent, and only 1 very frail and fearful person refused assessment.

Duration of assessment averaged 25.8 minutes (range: 6-55 minutes). No technical problems in assessing and analyzing the data occurred. Results of response rates are illustrated in Figure 5. For most of the participants, PBT (3, 10, and 30 seconds) was feasible. 30 of 99 subjects (30.3%) could not execute the complex PBT (30 seconds) because of self-reported fatigue. All participants could perform the FTBT. Regarding the motor-cognitive Physiomat tasks, 94 of 99 participants (94.9%) could conduct the simple, 73 subjects (73.7%) the moderate and 25 subjects (25.3%) the complex PTMT. Reasons for discontinuation were also predominantly fatigue reported by almost half of the participants (51.5%) followed by pain (17.2%) and by noncompliance (5.1%).

Figure 5. Feasibility analysis including response rates during a consecutive Physiomat assessment.



Discussion

Principal Findings

In this study, we validated an internal assessment approach of a game-based training device (Physiomat) to obtain a reliable and valid feedback of motor-cognitive abilities during gameplay. In contrast to recent studies focusing on computerized game-based assessments, which often conducted only reliability and validity analyses, we investigated multiple psychometric properties to allow a more comprehensive evaluation of the methodological quality of the assessment tested. In this study, validation was performed in frail, older persons with mild-to-moderate dementia who had not been addressed in most previous validation studies. Despite the crucial problems to assess persons with dementia [49], results indicated good construct validity, test–retest reliability, sensitivity to change, and feasibility of the tested device.

Construct Validity

In this study, construct validity was analyzed by testing 8 a priori hypotheses. For this purpose, Spearman‘s rank correlations between temporal parameters (test duration in seconds) of simple, moderate, and complex PTMTs and theoretically related motor-functional as well as cognitive measures were calculated. Almost all predefined hypotheses could be confirmed indicating a good construct validity of Physiomat assessment.

Expected higher correlations of cognitive instruments with increasing complexity of PTMTs were found for all cognitive outcome measures except the ZN-G. Correlations with ZN-G and wordlist immediate recall were partly not significant. As highest correlations were found between the complex PTMT and ZVT-G, more difficult motor-cognitive tasks seem to be associated with increased cognitive demands especially including attentional abilities and information processing. These results showed that particularly domain-related cognitive functions could rather be assessed when participants perform more complex PTMTs.

To our knowledge, there are no studies including cognitively impaired participants that investigated the relationship between interactive computerized assessment strategies comparable to Physiomat and cognitive test batteries, which do not use technological devices, although game-based assessments require a combination of motor and cognitive abilities. Expected higher associations of motor-functional instruments with decreasing complexity of PTMTs could not have conclusively identified for all tests (Table 2). Except the Physiomat balance task, highest correlations of motor-functional tests (FTBT, TUG, and POMA) were actually found with moderate and simple PTMTs, and expected lowest associations between complex PTMTs and TUG as well as POMA were not significant. However, when balance performances are challenged, rather an additional complex cognitive task than a simple or moderate task may require a higher level of attention in patients with dementia, which leads to a decrease in postural control. A previous study

[50] found a further decline of up to 15% in postural control during a more complex task in a cognitively impaired sample. Although findings were not significant, which might be due to a small study sample, balance performances seem to be determined by the complexity level of additional cognitive tasks, which could explain stronger correlations between motor-functional outcome measures and easier levels of PTMTs in our study.

We could not find any studies that have investigated the relationship between interactive assessment methods and established motor tests in people with dementia. Therefore, the comparability of our results with recent validity studies is limited. One study [20] examined associations of the Wii Balance Board and clinical tests in patients after stroke. The study could show moderate Spearman's correlations (r_s=−0.57) between a Wii Balance Board dynamic balance task measuring shifting of body weight to follow a visual feedback target and the TUG, which is comparable to our findings. A further study examining the association of TUG scores and different levels of a computer-based balance board test using the Biodex Balance System in healthy adults (mean age 48.9±15.4) showed stronger associations between relatively easy levels of machine-based assessments and manual balance tests [51]. In this study, an assessed stability index on the Biodex Balance System indicating the degree of body movement during the balance test was highly correlated with TUG scores especially at relatively easy Biodex Balance System levels (higher stability of the foot platform). This is comparable to our results including higher correlations between the simple and moderate PTMTs (less digits to connect/less path of movement) and TUG. Significant stronger associations with relatively easy tasks (simple and moderate PTMT) indicate that the assessment of balance performance using motor-cognitive Physiomat tasks should be conducted at a simple or moderate level.

Test–Retest Reliability

Almost all Physiomat outcome measures showed moderate-to-high correlations between test and retest assessment indicating good test–retest reliability. Both temporal (speed of task performance) and spatial (accuracy of task performance) parameters of Physiomat tasks showed similar test–retest reliability. These findings are comparable to results of test–retest analyses of temporal (time used to complete the test) and spatial (the extent of the path moved by the COP during the test) variables of different dynamic balance tasks using a force platform with visual biofeedback in nondemented female nursing home residents [23].

Most previous studies excluded patients with dementia. Exclusion might be based on the assumptions that cognitively impaired persons show an increasing variability of test performance due to illness-related symptoms such as attentional deficits, inability to follow instructions, and impaired executive function. Such dementia-related characteristics challenge an accurate assessment and substantially restrict the reproducibility of specific performances (eg, [52,53]). The only study we found including persons with Alzheimer's disease showed similar results [33] analyzing test–retest reliability for temporal (reaction time) and spatial (maximum excursion during test performance) variables.

Regarding test–retest reliability of motor-cognitive Physiomat tasks (PTMTs), we found an association between the complexity level and reproducibility. Spearman's rank correlations and ICCs were lower for simple PTMT compared with moderate and complex motor-cognitive tasks. We could exclude effects of deviating subsamples in different conditions (larger sample in simple condition, selection to high functioning participants in more complex conditions) by subsequent statistical test–retest analyses of 47 participants that successfully conducted all complexity levels. Results of subanalyses confirmed results of the nonselected group as ICCs were lower for simple PTMT compared with moderate and complex motor-cognitive tasks. Referring to this, we suggest a task-specific learning effect from simple to complex PTMTs, which may have led to smaller test performance variability and increased reproducibility regarding complex tasks. Such task-specific learning effects from simple to complex tasks have been reported by Lezak et al [54] attributed to the results of Oliveira et al [55]. The scientists argued that during an initial test, strategies to manage the task might have been developed, which facilitate performing subsequent tasks [54]. Such training mechanisms might have contributed to more reliable test results of complex PTMT in this study, as participants may have felt easier, and more competent in executing the tasks while gaining confidence and stability in performance by prolonged testing.

Sensitivity to Change

Sensitivity to change was good to excellent for Physiomat assessment as it reproduced significant improvements regarding all trained Physiomat tasks (FTBT as well as simple, moderate, and complex PTMTs) after a 10-week (twice a week/10 minutes) intervention period. In this study, temporal parameters of PTMTs and FTBT appeared to be more sensitive to change, as effect sizes of test duration (time) were larger than those of accuracy (sway path). Differences might be the result of the test instruction to "perform the FTBT and the PTMTs as fast—not as accurate—as possible," which focused on speed rather than accuracy of action. Results refer to a "speed-accuracy tradeoff" also reported in an intervention study [56] showing that participants were able to navigate quicker through a test path to measure foot placement but suffered the loss of accuracy after dance video game training. Results may also be influenced by variance of measurement as spatial outcomes showed higher SD compared with temporal outcomes leading to smaller effect sizes.

The complexity level of PTMTs seems to be relevant for responsiveness of Physiomat assessment. Whereas participants showed significant changes with low to high effect sizes in simple and moderate PTMTs, highest effect sizes were found under more challenging conditions (complex PTMT). Results confirmed previous findings from our research group in patients with dementia that more challenging tasks showed higher training gains, in case challenging tasks were still feasible for participants [57]. It is the very large effect sizes documented in highly trained outcomes representing the maximal potential change to be achieved, which are of paramount methodological interest in this study. These large effect sizes indicated the excellent sensitivity to change for the Physiomat assessment. Results supported the task-specific assessment approach as developed for the computerized game-based training and assessment program to document task-specific training gains.

Feasibility

Physiomat assessment was feasible even in an old and frail sample with mild-to-moderate dementia. Willingness of the participants to attend in the computerized game-based Physiomat assessment was excellent as only 1 very frail subject, who expressed fear, refused assessment. Results were in line with other reports in force platform–based assessment strategies that are comparable with Physiomat measures, which indicated adequate participation in machine-based computerized tests [58] and a high response rate in patients with dementia [33].

In this present study, all subjects could cope with the FTBT, and most participants could perform lower complexity levels of simple and moderate PBTs and simple and moderate PTMTs. As expected, response rate of the complex tasks was lower based on the higher request on motor-cognitive performance. According to the participants' reports, fatigue based on motor-functional or cognitive limitations was the primary reason for discontinuation. Unfortunately, based on the participants' reports, we could not further specify results. Report of fatigue may have been caused by advanced motor impairment and frailty in the study sample or by psychological mechanisms. Previous results of the working group documented that repressive coping strategies or denial of events were significantly associated with inadequate reports on anxiety-related events such as falls in old age [59]. As denial is distinctive of types of dementia [60], we supposed that repressive coping strategy may have led to underreporting of cognitive limitations causing fatigue.

Completion time of Physiomat assessment averaged 25.8 minutes for the comprehensive Physiomat test protocol including a detailed and clear instruction, several trials and breaks between single tasks, and performance levels. As in the original test protocol, more than 3 tasks (simple, moderate, and complex PTMTs) as documented in this validation study were assessed and the completion time to perform those will be substantially reduced. Time to complete ranged from 6 to 55 minutes because of a large heterogeneity of the participants with respect to motor-functional and cognitive status. Completion time of a variety of noninteractive computer-based cognitive tests or batteries to assess or detect age-related changes in cognition ranged from 15 to 60 minutes [61]. Time to complete assessments that is directly comparable to the presented Physiomat measures such as force platform–based assessments [23,33] was not mentioned in the papers.

Participants' safety was a clear focus in this study as training and assessment was tightly supervised, and clear and brief instructions were provided. As a result, all Physiomat tasks included in the study could be performed safely in this challenging sample of cognitively impaired older adults as no clinical events, falls, and slips could be documented. Safety outcomes are in line with a comparable study examining test–retest reliability of a force platform assessment in people with dementia [33].

Limitations

Increasing complexity levels in different task conditions led to decreasing sample sizes for each condition. Although we confirmed results of the whole group in the subsample of persons who participated in all test, comparability of psychometric quality may be influenced by slight change of samples between conditions.

Conclusions

Study results confirm good-to-excellent psychometric quality of an internal assessment approach using quantitative data derived from a computerized game-based training program (Physiomat) in frail persons with mild-to-moderate stage of dementia. This approach provides quantitative parameters that relate to identical and integrated performances trained by using Physiomat and are therefore direct marker of motor-cognitive Physiomat training tasks. Physiomat assessment may represent an evaluation strategy to document game performances and training-associated effects in a rapidly increasing research field including serious games, virtual reality, and machine-based, computerized training.

Acknowledgments

The study was supported by the Robert Bosch Foundation, the Network of Aging Research (NAR) at the University of Heidelberg, and the Dietmar Hopp Foundation. We kindly thank M. Günther for her assistance in training and supervising participants and S. Schnaidt, and O. Schilling for supporting statistical analysis. J. Lautenschläger (EPL engineering, Physiomat) is acknowledged for technical support during the study and cooperation in the development of the training program.

Conflicts of Interest

None declared.

References

1. Zmily A, Mowafi Y, Mashal E. Study of the usability of spaced retrieval exercise using mobile devices for Alzheimer's disease rehabilitation. JMIR Mhealth Uhealth 2014;2(3):e31 [FREE Full text] [doi: 10.2196/mhealth.3136] [Medline: 25124077]
2. Lim FS, Wallace T, Luszcz MA, Reynolds KJ. Usability of tablet computers by people with early-stage dementia. Gerontology 2013;59(2):174-182. [doi: 10.1159/000343986] [Medline: 23257664]

3. Schoene D, Valenzuela T, Lord SR, de Bruin ED. The effect of interactive cognitive-motor training in reducing fall risk in older people: a systematic review. BMC Geriatr 2014;14:107 [FREE Full text] [doi: 10.1186/1471-2318-14-107] [Medline: 25240384]
4. Webster D, Celik O. Systematic review of Kinect applications in elderly care and stroke rehabilitation. J Neuroeng Rehabil 2014;11:108 [FREE Full text] [doi: 10.1186/1743-0003-11-108] [Medline: 24996956]
5. Kazmi S, Ugail H, Lesk V, Palmer I. Interactive Digital Serious Games for the Assessment, Rehabilitation, and Prediction of Dementia. International Journal of Computer Games Technology 2014;2014:1-11. [doi: 10.1155/2014/701565]
6. McCallum S, Boletsis C. Dementia Games: A Literature Review of Dementia-Related Serious Games. In: Ma M, Oliveira MF, Petersen S, Hauge JB, editors. Serious games development and applications. Berlin Heidelberg: Springer Berlin Heidelberg; 2013:15-27.
7. Robert PH, König A, Amieva H, Andrieu S, Bremond F, Bullock R, et al. Recommendations for the use of Serious Games in people with Alzheimer's Disease, related disorders and frailty. Front Aging Neurosci 2014;6:54 [FREE Full text] [doi: 10.3389/fnagi.2014.00054] [Medline: 24715864]
8. Kayama H, Okamoto K, Nishiguchi S, Yamada M, Kuroda T, Aoyama T. Effect of a Kinect-based exercise game on improving executive cognitive performance in community-dwelling elderly: case control study. J Med Internet Res 2014 Jan;16(2):e61 [FREE Full text] [doi: 10.2196/jmir.3108] [Medline: 24565806]
9. Szturm T, Hochman J, Wu C, Lisa L, Reimer K, Wonneck B, et al. Games and Telerehabilitation for Balance Impairments and Gaze Dysfunction: Protocol of a Randomized Controlled Trial. JMIR Res Protoc 2015;4(4):e118 [FREE Full text] [doi: 10.2196/resprot.4743] [Medline: 26490109]
10. van Diest DM, Lamoth CJ, Stegenga J, Verkerke GJ, Postema K. Exergaming for balance training of elderly: state of the art and future developments. J Neuroeng Rehabil 2013;10:101 [FREE Full text] [doi: 10.1186/1743-0003-10-101] [Medline: 24063521]
11. Betker AL, Desai A, Nett C, Kapadia N, Szturm T. Game-based exercises for dynamic short-sitting balance rehabilitation of people with chronic spinal cord and traumatic brain injuries. Phys Ther 2007 Oct;87(10):1389-1398 [FREE Full text] [doi: 10.2522/ptj.20060229] [Medline: 17712036]
12. Yamada M, Aoyama T, Nakamura M, Tanaka B, Nagai K, Tatematsu N, et al. The reliability and preliminary validity of game-based fall risk assessment in community-dwelling older adults. Geriatr Nurs 2011;32(3):188-194. [doi: 10.1016/j.gerinurse.2011.02.002] [Medline: 21501899]
13. Wikstrom EA. Validity and reliability of Nintendo Wii Fit balance scores. J Athl Train 2012 Jun;47(3):306-313 [FREE Full text] [doi: 10.4085/1062-6050-47.3.16] [Medline: 22892412]
14. Goble DJ, Cone BL, Fling BW. Using the Wii Fit as a tool for balance assessment and neurorehabilitation: the first half decade of “Wii-search”. J Neuroeng Rehabil 2014 Feb;11:12 [FREE Full text] [doi: 10.1186/1743-0003-11-12] [Medline: 24507245]
15. Young W, Ferguson S, Brault S, Craig C. Assessing and training standing balance in older adults: a novel approach using the 'Nintendo Wii' Balance Board. Gait Posture 2011 Feb;33(2):303-305. [doi: 10.1016/j.gaitpost.2010.10.089] [Medline: 21087865]
16. Clark RA, Bryant AL, Pua Y, McCrory P, Bennell K, Hunt M. Validity and reliability of the Nintendo Wii Balance Board for assessment of standing balance. Gait Posture 2010 Mar;31(3):307-310. [doi: 10.1016/j.gaitpost.2009.11.012] [Medline: 20005112]
17. Clark RA, Pua Y, Oliveira CC, Bower KJ, Thilarajah S, McGaw R, et al. Reliability and concurrent validity of the Microsoft Xbox One Kinect for assessment of standing balance and postural control. Gait Posture 2015 Jul;42(2):210-213. [doi: 10.1016/j.gaitpost.2015.03.005] [Medline: 26009500]
18. Park D, Lee G. Validity and reliability of balance assessment software using the Nintendo Wii balance board: usability and validation. J Neuroeng Rehabil 2014;11:99 [FREE Full text] [doi: 10.1186/1743-0003-11-99] [Medline: 24912769]
19. Ruff J, Wang TL, Quatman-Yates CC, Phieffer LS, Quatman CE. Commercially available gaming systems as clinical assessment tools to improve value in the orthopaedic setting: a systematic review. Injury 2015 Feb;46(2):178-183. [doi: 10.1016/j.injury.2014.08.047] [Medline: 25441576]
20. Bower KJ, McGinley JL, Miller KJ, Clark RA. Instrumented static and dynamic balance assessment after stroke using Wii Balance Boards: reliability and association with clinical tests. PLoS One 2014;9(12):e115282 [FREE Full text] [doi: 10.1371/journal.pone.0115282] [Medline: 25541939]
21. Holmes JD, Jenkins ME, Johnson AM, Hunt MA, Clark RA. Validity of the Nintendo Wii balance board for the assessment of standing balance in Parkinson's disease. Clin Rehabil 2013 Apr;27(4):361-366. [doi: 10.1177/0269215512458684] [Medline: 22960241]
22. Schoene D, Lord SR, Verhoef P, Smith ST. A novel video game--based device for measuring stepping performance and fall risk in older people. Arch Phys Med Rehabil 2011 Jun;92(6):947-953. [doi: 10.1016/j.apmr.2011.01.012] [Medline: 21549352]
23. Sihvonen SE, Sipilä S, Era PA. Changes in postural balance in frail elderly women during a 4-week visual feedback training: a randomized controlled trial. Gerontology 2004;50(2):87-95. [doi: 10.1159/000075559] [Medline: 14963375]

24. Szturm T, Sakhalkar V, Boreskie S, Marotta JJ, Wu C, Kanitkar A. Integrated testing of standing balance and cognition: test-retest reliability and construct validity. Gait Posture 2015 Jan;41(1):146-152. [doi: 10.1016/j.gaitpost.2014.09.023] [Medline: 25455701]
25. Betker AL, Szturm T, Moussavi ZK, Nett C. Video game-based exercises for balance rehabilitation: a single-subject design. Arch Phys Med Rehabil 2006 Aug;87(8):1141-1149. [doi: 10.1016/j.apmr.2006.04.010] [Medline: 16876562]
26. Szturm T, Betker A, Moussavi Z, Desai A, Goodman V. Effects of an interactive computer game exercise regimen on balance impairment in frail community-dwelling older adults: a randomized controlled trial. Phys Ther 2011;91(10):1449-1462. [doi: 10.2522/ptj.20090205] [Medline: 21799138]
27. Podsiadlo D, Richardson S. The timed “Up & Go”: a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [Medline: 1991946]
28. Berg K, Wood-Dauphinee SL, Williams J, Maki B. Measuring balance in the elderly: validation of an instrument. Can J Public Health 1992;83 Suppl 2:S7-11. [Medline: 1468055]
29. Padala KP, Padala PR, Malloy TR, Geske JA, Dubbert PM, Dennis RA, et al. Wii-fit for improving gait and balance in an assisted living facility: a pilot study. J Aging Res 2012;2012:597573 [FREE Full text] [doi: 10.1155/2012/597573] [Medline: 22745909]
30. Maillot P, Perrot A, Hartley A. Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. Psychol Aging 2012 Sep;27(3):589-600. [doi: 10.1037/a0026268] [Medline: 22122605]
31. Lai C, Peng C, Chen Y, Huang C, Hsiao Y, Chen S. Effects of interactive video-game based system exercise on the balance of the elderly. Gait Posture 2013 Apr;37(4):511-515. [doi: 10.1016/j.gaitpost.2012.09.003] [Medline: 23177921]
32. Koslucher F, Wade MG, Nelson B, Lim K, Chen F, Stoffregen TA. Nintendo Wii Balance Board is sensitive to effects of visual tasks on standing sway in healthy elderly adults. Gait Posture 2012 Jul;36(3):605-608. [doi: 10.1016/j.gaitpost.2012.05.027] [Medline: 22748469]
33. Suttanon P, Hill KD, Dodd KJ, Said CM. Retest reliability of balance and mobility measurements in people with mild to moderate Alzheimer's disease. Int Psychogeriatr 2011 Sep;23(7):1152-1159. [doi: 10.1017/S1041610211000639] [Medline: 21489342]
34. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975 Nov;12(3):189-198. [Medline: 1202204]
35. Dubois B, Slachevsky A, Litvan I, Pillon B. The FAB: a Frontal Assessment Battery at bedside. Neurology 2000 Dec 12;55(11):1621-1626. [Medline: 11113214]
36. Morris JC, Mohs RC, Rogers H, Fillenbaum G, Heyman A. Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. Psychopharmacol Bull 1988;24(4):641-652. [Medline: 3249766]
37. Oswald W, Fleischmann U. Das Nürnberger-Alters-Inventar (NAI) - Testinventar & NAI-Testmanual und Textband. 4th edition. Göttingen: Hogrefe; 1997.
38. Reitan R. Validity of the trail making test as an indicator of organic brain damage. Perceptual and Motor Skills 1958;8:271-276.
39. EPL medical engineering. URL: http://www.epl.de/ [accessed 2016-02-21] [WebCite Cache ID 6fSbU22cu]
40. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [Medline: 7183759]
41. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). Age Ageing 2005 Nov;34(6):614-619 [FREE Full text] [doi: 10.1093/ageing/afi196] [Medline: 16267188]
42. Tinetti ME. Performance-oriented assessment of mobility problems in elderly patients. J Am Geriatr Soc 1986 Feb;34(2):119-126. [Medline: 3944402]
43. Atkinson RC, Shiffrin RM. The control of short-term memory. Sci Am 1971 Aug;225(2):82-90. [Medline: 5089457]
44. Carmine E, Zeller R. Reliability and validity assessment. Newbury Park, London, New Delhi: SAGE publications; 1979.
45. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
46. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. New York: Academic Press; 1988.
47. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
48. Fleiss J. Design and analysis of clinical experiments. New York: John Wiley & Sons Inc; 1986.
49. Hauer K, Oster P. Measuring functional performance in persons with dementia. J Am Geriatr Soc 2008 May;56(5):949-950. [doi: 10.1111/j.1532-5415.2008.01649.x] [Medline: 18454757]
50. Hauer K, Pfisterer M, Weber C, Wezler N, Kliegel M, Oster P. Cognitive impairment decreases postural control during dual tasks in geriatric patients with a history of severe falls. J Am Geriatr Soc 2003 Nov;51(11):1638-1644. [Medline: 14687396]

51. Oh KY, Kim SA, Lee SY, Lee YS. Comparison of manual balance and balance board tests in healthy adults. Ann Rehabil Med 2011 Dec;35(6):873-879 [FREE Full text] [doi: 10.5535/arm.2011.35.6.873] [Medline: 22506217]
52. Ries JD, Echternach JL, Nof L, Gagnon BM. Test-retest reliability and minimal detectable change scores for the timed "up & go" test, the six-minute walk test, and gait speed in people with Alzheimer disease. Phys Ther 2009 Jun;89(6):569-579 [FREE Full text] [doi: 10.2522/ptj.20080258] [Medline: 19389792]
53. Tappen RM, Roach KE, Buchner D, Barry C, Edelstein J. Reliability of physical performance measures in nursing home residents with Alzheimer's disease. J Gerontol A Biol Sci Med Sci 1997 Jan;52(1):M52-M55 [FREE Full text] [Medline: 9008669]
54. Lezak M, Howieson D, Bigler E, Tranel D. Neuropsychological assessment. New York: Oxford University Press; 2012.
55. Oliveira RS, Trezza BM, Busse AL, Jacob-Filho W. Learning effect of computerized cognitive tests in older adults. Einstein (Sao Paulo) 2014 Apr;12(2):149-153 [FREE Full text] [Medline: 25003917]
56. Pichierri G, Murer K, de Bruin ED. A cognitive-motor intervention using a dance video game to enhance foot placement accuracy and gait under dual task conditions in older adults: a randomized controlled trial. BMC Geriatr 2012;12:74 [FREE Full text] [doi: 10.1186/1471-2318-12-74] [Medline: 23241332]
57. Schwenk M, Zieschang T, Oster P, Hauer K. Dual-task performances can be improved in patients with dementia: a randomized controlled trial. Neurology 2010 Jun 15;74(24):1961-1968. [doi: 10.1212/WNL.0b013e3181e39696] [Medline: 20445152]
58. Piirtola M, Era P. Force platform measurements as predictors of falls among older people - a review. Gerontology 2006;52(1):1-16. [doi: 10.1159/000089820] [Medline: 16439819]
59. Hauer K, Tremmel A, Ramroth H, Pfisterer M, Todd C, Oster P, et al. Repressive coping in geriatric patients' reports - impact on fear of falling. Z Gerontol Geriatr 2009 Apr;42(2):137-144. [doi: 10.1007/s00391-008-0552-9] [Medline: 18560787]
60. Bahro M, Silber E, Sunderland T. How do patients with Alzheimer's disease cope with their illness? A clinical experience report. J Am Geriatr Soc 1995 Jan;43(1):41-46. [Medline: 7806738]
61. Wild K, Howieson D, Webbe F, Seelye A, Kaye J. Status of computerized cognitive testing in aging: a systematic review. Alzheimers Dement 2008 Nov;4(6):428-437 [FREE Full text] [doi: 10.1016/j.jalz.2008.07.003] [Medline: 19012868]

Abbreviations

CERAD: Consortium to Establish a Registry for Alzheimer's disease
CI: confidence interval
COP: center of pressure
FES-I: Falls-Efficacy-Scale-International
FTBT: Physiomat Follow-The-Ball Task
GDS: Geriatric Depression Scale
ICC: intraclass correlations
MMSE: Mini-Mental-State Examination
POMA: performance-oriented mobility assessment
PTMT: Physiomat Trail-Making Task
PBT: Physiomat-Balance Task
SD: standard deviation
SRM: standardized response mean
TMT: Trail-Making-Test
TUG: Timed-Up-and-Go
ZN-G: Zahlen-Nachsprechen-G (Repeating-Number-Test)
ZVT-G: Zahlen-Verbindungs-Test-G (Connecting Number-Test)

Edited by G Eysenbach; submitted 25.02.16; peer-reviewed by T Szturm, M D. Patterson; comments to author 14.04.16; revised version received 26.05.16; accepted 04.07.16; published 18.07.16.

Please cite as:
Wiloth S, Lemke N, Werner C, Hauer K
Validation of a Computerized, Game-based Assessment Strategy to Measure Training Effects on Motor-Cognitive Functions in People With Dementia
JMIR Serious Games 2016;4(2):e12
URL: http://games.jmir.org/2016/2/e12/
doi:10.2196/games.5696
PMID:27432746

Original Paper

A Video Game Promoting Cancer Risk Perception and Information Seeking Behavior Among Young-Adult College Students: A Randomized Controlled Trial

Georges Elias Khalil[1], MPH, PhD; Ivan L Beale[2], MSc, PhD; Minxing Chen[3], MS; Alexander V Prokhorov[1], PhD, MD

[1]The M.D. Anderson Cancer Center, Department of Behavioral Science, University of Texas, Houston, TX, United States

[2]School of Psychology, University of New South Wales, New South Wales, Australia

[3]The M.D. Anderson Cancer Center, Division of Quantitative Sciences, University of Texas, Houston, TX, United States

Corresponding Author:
Georges Elias Khalil, MPH, PhD
The M.D. Anderson Cancer Center
Department of Behavioral Science
University of Texas
1155 Pressler St. Unit 1330
Houston, TX,
United States
Phone: 1 713 563 2605
Fax: 1 713 745 4286
Email: gekhalil@mdanderson.org

Abstract

Background: Risky behaviors tend to increase drastically during the transition into young adulthood. This increase may ultimately facilitate the initiation of carcinogenic processes at a young age, highlighting a serious public health problem. By promoting information seeking behavior (ISB), young adults may become aware of cancer risks and potentially take preventive measures.

Objective: Based on the protection motivation theory, the current study seeks to evaluate the impact of challenge in a fully automated video game called *Re-Mission* on young adult college students' tendency to perceive the severity of cancer, feel susceptible to cancer, and engage in ISB.

Methods: A total of 216 young adults were recruited from a university campus, consented, screened, and randomized in a single-blinded format to 1 of 3 conditions: an intervention group playing *Re-Mission* at high challenge (HC; n=85), an intervention group playing *Re-Mission* at low challenge (LC; n=81), and a control group with no challenge (NC; presented with illustrated pictures of *Re-Mission*; n=50). Measurement was conducted at baseline, immediate posttest, 10-day follow-up, and 20-day follow-up. Repeated-measures mixed-effect models were conducted for data analysis of the main outcomes.

Results: A total of 101 young adults continued until 20-day follow-up. Mixed-effect models showed that participants in the HC and LC groups were more likely to increase in perceived susceptibility to cancer (*P*=.03), perceived severity of cancer (*P*=.02), and ISB (*P*=.01) than participants in the NC group. The LC group took until 10-day follow-up to show increase in perceived susceptibility (B=0.47, standard error (SE) 0.16, *P*=.005). The HC group showed an immediate increase in perceived susceptibility at posttest (B=0.43, SE 0.14, *P*=.002). The LC group exhibited no changes in perceived severity (B=0.40, SE 0.33, *P*=.24). On the other hand, the HC group showed a significant increase from baseline to posttest (B=0.39, SE 0.14, *P*=.005), maintaining this increase until 20-day follow-up (B=−0.007, SE 0.26, *P*=.98). Further analyses indicated that perceived threat from virtual cancer cells in the game is related to the increase in perceived severity (B=0.1, SE 0.03, *P*=.001), and perceived susceptibility is related to changes in ISB at 10-day follow-up (B=0.21, SE 0.08, *P*=.008).

Conclusions: The feature of challenge with cancer cells in a virtual environment has the potential to increase cancer risk perception and ISB. The results are promising considering that the *Re-Mission* intervention was neither designed for cancer risk communication, nor applied among healthy individuals. Further research is needed to understand the theoretical framework underlying the effects of *Re-Mission* on ISB. The findings call for the development of a Web-based, game-based intervention for cancer risk communication and information seeking among young adults.

ClinicalTrial: International Standard Randomized Controlled Trial Number (ISRCTN): 15789289; http://www.controlled-trials.com/ISRCTN15789289 (Archived by WebCite at http://www.webcitation.org/6jGYZC3lZ)

(JMIR Serious Games 2016;4(2):e13) doi:10.2196/games.5793

KEYWORDS

cancer; risk; information seeking; perceived risk; perceived susceptibility; perceived severity; cancer prevention; games for health; serious games

Introduction

Cancer-related information seeking behavior (ISB) is a goal-directed behavior adopted as a response to threatening situations, and it assists in uncertainty reduction concerning cancer [1]. Despite the importance of information seeking, the Health Information National Trends Survey has reported that less than half of Americans look for cancer information [2]. By promoting ISB [3], young adults may become aware of cancer risks and potential preventive measures. In particular, their active search for information about cancer may increase their cancer knowledge and equip them with ways to get protected from cancer [4,5].

One way to encourage ISB is by helping young adults perceive cancer risk [6]. Cancer risk perception is mainly characterized by two dimensions. First, perceived susceptibility to cancer explains one's beliefs about the likelihood of being diagnosed with cancer. The second dimension is perceived severity of cancer, which explains one's perception of the seriousness of cancer diagnosis. Previous research has found that young adults' perceived susceptibility to and severity of cancer may moderate engagement in healthy behaviors such as breast self-examination [7], mammography [8,9], skin protection [6], and smoking cessation [10].

Ultimately, the lack of ISB and cancer risk perception may delay cancer prevention and control at long term. As a result, the World Health Organization has emphasized the need to design interventions that successfully raise awareness about cancer risks and ISB [11]. While previous research has well examined processes by which individuals engage in health-related information seeking [12-14], little is known about the role of game play features in driving ISB. Responding to this need, video games have been designed as innovative tools for health promotion and disease prevention. The current study seeks to evaluate the impact of a video game called *"Re-Mission"* [15] on young adult college students' tendency to perceive the severity of cancer, feel susceptible to cancer, and seek cancer-related information.

Re-Mission is a fully automated game in which players control a virtual nanorobot that goes inside virtual cancer patients' bodies to fight cancer cells [16]. The *Re-Mission* intervention was designed primarily to encourage pediatric cancer patients to adhere to their medication [17]. However, recent exploratory research has shown that *Re-Mission* may have an impact on healthy young adults' risk perception [18,19]. As a result, the evaluation of *Re-Mission* in the context of risk communication deserves attention. We conducted the current experimental evaluation [ISRCTN15789289] to verify whether and how *Re-Mission* might modify risk perception in healthy young adults. This would indicate whether there is a potential to design a digital game for the promotion of cancer preventing behaviors among young adults.

One theory supporting an association between game-play and cancer preventing behaviors is the protection motivation theory (PMT) [20,21]. According to PMT, threatening health messages can stimulate risk perception and encourage protective behavior. By experiencing threat, individuals may become motivated to take actions that can protect them, such as seeking information about the health topic [22-26]. Therefore, we hypothesized that young adults who play *Re-Mission* at 'high challenge' (HC) are more likely to increase perceived severity of cancer, perceived susceptibility to cancer, and cancer-related ISB, compared with young adults who play *Re-Mission* at 'low challenge' (LC) or do not play *Re-Mission*. We also hypothesized that (1) perceived threat in the gaming intervention is related to perceived susceptibility and perceived severity, and (2) such secondary outcomes are related to ISB.

Methods

Game Intervention Format and Key Features

The story in *Re-Mission* revolves around Roxxi, a nanorobot designed by a doctor and injected into the body of virtual cancer patients to help them fight cancer cells. Players are first presented with the narrative, the characters, and main game objective. Then, they are asked to choose a virtual cancer patient that needs assistance. The game gives the players control over the movement of Roxxi, who undertakes missions to fight cancer cells in a three-dimensional environment, within the bodies of cancer patients (See Multimedia Appendix 1).

Beyond mere exposure to information, *Re-Mission* involves a first-hand experience of cancer threat that ultimately allows players to perceive cancer risk. In *Re-Mission*, players are able to witness cancer cell behavior, from cell division to invasion, and ultimately find themselves in conflict with cancer cells. Conflict with cancer is a key feature of challenge in *Re-Mission*. A highly challenging environment with cancer cells (ie, presence of obstacles at high difficulty when fighting cancer cells) may facilitate perceived severity of cancer and intentions to seek cancer-related information [19].

Study Design

The components of this controlled trial adhere to the CONSORT and CONSORT-EHEALTH guidelines [27,28] on information to include when reporting trials in general and eHealth in particular. This efficacy trial used a three-arm, single-blinded randomized controlled (Time × Condition) design with assessments at baseline, immediate posttest, 10-day, and 20-day follow-ups. The trial was registered at the Current Controlled Trials [ISRCTN15789289].

Sample

We assessed the eligibility of all interested undergraduate students through a screening conducted before participation. Inclusionary criteria were being aged 18 to 35 years, attending college, consenting to play video games, and speaking English, Spanish, or French (the game was available in all three languages). Exclusionary criteria involved having a medical or mental condition that hindered the ability to play games or complete questionnaires. All participants were informed about the aim of the study and their consent for participation was recorded. The institutional review board at the University of Texas MD Anderson Cancer Center and the University at Buffalo, the State University of New York approved this study.

Intervention Groups and Control Group

Young adults were randomly assigned to 1 of 3 conditions: an intervention group playing *Re-Mission* at HC (*n*=85), an intervention group playing *Re-Mission* at LC (n=81), and a control group with no challenge (NC; presented with illustrated pictures of *Re-Mission*; n=50).

Manipulation of challenge was conducted as suggested by previous research on conflict manipulation in *Re-Mission* [19]. HC was conceptualized as a condition that arises from a set of obstacles (eg, invasion by and multiplication of cancer cells) that prevent the players from attaining their goals in the game (eg, killing cancer cells and helping the patient recover from cancer) [29,30]. Three characteristics of challenge were manipulated in the game-mechanics to form an HC environment: (1) difficulty level (ie, amount of ammunition needed to destroy a cancer cell), (2) vulnerability to cancer cells (ie, cancer cells can put the nanorobot to sleep), and (3) limited ammunition (limited virtual medication to kill cancer cells). See Multimedia Appendix 2 for a pictorial depiction of manipulation.

The NC condition involved the presentation of several illustrations from the video game with a description of the conflict occurring between the nanorobot and the cancer cells. In particular, they were presented with illustrations of *Re-Mission* that represent steps of a conflict event with cancer cells (ie, exposure to cancer cells, cancer cells multiplying, Roxxi approaching cancer cells, cancer cells attacking Roxxi, and Roxxi fighting cancer cells; see examples in Multimedia Appendix 3). Each illustration included a textual description of the event and the context of the game. As a result, this condition lacked a first-hand experience of challenge, while keeping a presentation of the game available to participants. Instead of using a noneducational video game, the control condition of this study preserved the context of cancer prevention in order to highlight the role of conflict experience as a driver of health outcomes.

Implementation

A verbal announcement was made in 3 undergraduate classes at a northeastern university. Each class included approximately 500 young adults. A Web-based announcement was also posted through the course material announcement page, and interested students were able to contact the research team for participation. Recruitment continued for a period of 2 months, or until reaching the target sample size. All interested students provided verbal and written informed consent and were invited for baseline assessment.

One week after completing a baseline survey, participants arrived at the intervention site and were randomly assigned to 1 of 3 conditions: LC, HC, or NC. The principal investigator generated the random allocation sequence. The research assistant enrolled participants and assigned them to groups. Concealed envelopes were used to implement the random allocation while concealing the sequence until intervention assignment. Participants were not told which intervention was the intervention of interest. Intervention implementation occurred in a noise-protected room at the university campus. Before playing *Re-Mission*, participants in the LC and HC conditions were seated in front of computers of the same brand and size, and they completed a tutorial that allowed them to practice using the controls in the game when attempting to move the avatar Roxxi (approximately 7 minutes). After the tutorial, all participants were seated in front of computers of the same brand and size, and were provided with headphones for privacy and maximum immersion. Then, participants were invited to start the first mission of the game and played for 35 minutes. Every time the players completed the mission and every time they lost in the game, they were asked to play it again, until the session was over. This method of intervention implementation with game interventions has been previously applied and validated with *Re-Mission* [19]. Participants in the NC group were also seated in front of computers, but they were presented with the NC condition of *Re-Mission*. For all conditions, the research assistant monitored progress from a different room.

Participants in all 3 conditions were invited to complete a survey immediately after implementation, 10 days later, and 20 days later. As an ethical consideration, after the 20-day follow-up, the NC group received information about *Re-Mission* and ways to access the game, if interested.

Compensation

Participants were offered credits for their respective classes from their professors. Credits were provided for each assessment, and as a result of intervention participation (ie, 0.5 credit points for baseline survey, 1 credit point for intervention participation, 0.5 credit points for immediate post-test survey, and 1 credit point for each of the follow-up surveys).

Measures

Outcome measures were assessed through Web-based closed surveys. The surveys were pretested for validity, reliability, usability, and technical functionality during the pilot study [19]. Adherence to the checklist for reporting results of Internet e-surveys [31] is provided as Multimedia Appendix 4. The post-test survey is the only survey that occurred in the presence of a research assistant, who was only available for technical assistance. The primary outcome ISB was assessed at baseline, 10-day, and 20-day follow-ups. The secondary outcomes, perceived severity of cancer and perceived susceptibility to cancer, were assessed at baseline, immediate posttest, 10-day, and 20-day follow-ups. In addition to the endpoints, other variables that might affect play behavior in a challenging virtual environment were measured (eg, frequency of weekly game play, perceived skills with video games, and perceived control over stress during game play). Perceived control over *Re-Mission* was measured to tap on players' control over gaming events, including cancer cells during the challenge. Perceived threat from virtual cancer cells was measured at immediate posttest. Means, standard deviations, measure descriptions, and Cronbach's α values are reported in Table 1. Previous work and the pilot study have tested the measures for validity and reliability [6,19,32-37].

Table 1. Main study measures.

Measures	T1[a] Mean (SD)	T2[a] Mean (SD)	Description	α[b]
ISB	2.12 (1.50)	2.16 (1.50)	Two items: "Have you paid attention to any cancer information in the past week or so?" and "Have you attempted to look for information about cancer in the past week or so?" (from 1=not at all to 9=a whole lot).	.71[d]
Perceived susceptibility	2.92 (1.22)	3.17 (1.20)	Participants were asked how possible they were to contract cancer in the next year, in 5 years, in 10 years, and in their life-time (from 1=not at all possible to 7=extremely possible).	.91
Perceived severity	5.71 (1.28)	5.67 (1.13)	Four items such as "Cancer is a serious disease that can kill" (from 1=very strongly disagree to 7=very strongly agree).	.80
General control over stress	5.00 (0.98)	-	Ten items such as "I am able to control my level of anxiety while playing a video game" (from 1=very strongly disagree to 7=very strongly agree).	.79
General reaction to threat	4.64 (0.99)	-	Five items such as "There is little I can do to change threatening events" (Reverse coded; from 1=very strongly disagree to 7=very strongly agree).	.77
General perceived skills in game play	3.72 (1.34)	-	Six items such as "I am very skilled at playing shooting games" (from 1=very strongly disagree to 7=very strongly agree).	.85
Frequency of game play	2.71 (4.98)	-	One open-ended question: "How many hours per week do you spend playing computer games?"	-
Perceived control over *Re-Mission*	-	4.11 (1.84)	An adapted scale with 9 items such as "For me to feel in control over all cancer cells was difficult" (from 1=very strongly disagree to 7=very strongly agree).	.95
Perceived challenge	-	3.69 (1.39)	Four items such as "Playing *Re-Mission* has challenged me to perform to the best of my abilities" (from 1=very strongly disagree to 7=very strongly agree).	.90
Perceived threat from virtual cancer cells	-	4.70 (2.27)	Four 9-point semantic differential items such as "While playing *Re-Mission*, how threatening did you feel cancer cells to be?" (from 0=not at all threatening to 8=extremely threatening).	.92
Attitude toward *Re-Mission*	-	4.22 (1.70)	Eight 9-point semantic differential items (e.g., dislike/like and not worth owning/worth owning).	.91

[a]T1 and T2 indicate measures at pretest and post-game play respectively for all participants. T2 for ISB indicates 20-day follow-up.

[b]Coefficients for Cronbach's α were calculated from post-test data, with the exception for measures with data collected at T1 only.

[c]Standard deviations appear in parentheses below the mean.

[d]Indicates Pearson's correlation between 2 items, instead of Cronbach's α.

Sample Size and Statistical Analysis

Sample size was estimated on the basis of a previous pilot study of 44 young adults [19]. Analyses targeted detection of an effect size of 0.15 (Cohen's d) with 85% power and α=0.05 (two-sided), with adjustment for an anticipated 70.0% (n=102/145) retention rate.

Statistical analyses were conducted using STATA 12. First, a series of chi-squared analyses were conducted to check for any sociodemographic differences between the groups. Then, manipulation checks were conducted in order to check if the manipulation appropriately reflects levels of challenge. This involved a series of one-way analyses of variance (ANOVAs), checking for group differences with respect to prior gaming experience (ie, skills in digital game play, general reaction to threat during game play, control over stress, or prior history of game play in hours per week). The ANOVAs also checked for any group difference in attitude toward *Re-Mission*, perceived control over game play, and experience of positive challenge during game play. When warranted, Bonferroni adjustment was performed, correcting for alpha over repeated comparisons and guarding against Type 1 error [38].

Repeated-measures, mixed-effect linear models were used, testing differences between the three treatment groups at three time-points in a 3 (treatment) × 3 (time; baseline, 10-day and 20-day follow-up) factorial design for ISB, and at four time-points in a 3 (treatment) × 4 (time; baseline, post-test, 10-day, and 20-day follow-up) factorial design for perceived cancer severity and susceptibility. ISB was not measured at immediate posttest because participants did not yet have the chance to seek cancer information. Intervention effects on the change in outcomes over time were determined by the treatment × time interaction term and *P* values are reported. Change over time is analyzed using post-hoc tests of significant difference in scores between time points, and *P* values are reported.

One-way ANOVA was conducted to compare the groups with respect to ISB change. ISB change is a variable measured by subtracting the ISB score at baseline from the ISB score at 10-day follow-up.

Logistic regression analysis was also conducted to determine group differences in ISB at three time-points (baseline, 10-day, and 20-day follow-up). In this case, the ISB measure was treated as a dichotomous variable with "not at all" indicating no information seeking (coded 0), and all other answer choices indicating information seeking (coded 1). Results were determined with the odds ratio (OR) and 95% confidence interval (CI).

For mixed-effect and logistic regression models, adjustment for effects of gender, age, ethnic group, prior cancer history, gaming skills, and usual frequency of gameplay did not alter primary conclusions (analyses not shown).

To check for potential demographic confounders, we analyzed mixed-effect models with the interactions (1) gender × condition × time, (2) ethnicity × condition × time, (3) prior cancer history × condition × time, and (4) general perceived skills in game play × condition × time predicting ISB.

To test whether perceived threat in the gaming experience is related to the secondary outcomes (ie, perceived susceptibility and perceived severity), two repeated-measures, mixed-effect models were conducted controlling for the intervention effect, age, gender, and ethnicity. Also, to test whether perceived susceptibility and perceived severity are related to the primary outcome (ie, ISB), a repeated-measures, mixed-effect model was conducted controlling for the intervention effect, age, gender, and ethnicity.

Results

Attrition and Intervention Adherence

A total of 220 college students responded to the advertisement. After screening, we excluded 2 of the respondents who did not meet the young-adult age criterion (ages 18 through 35). A total of 216 young adults took the baseline survey, were randomized, participated in the intervention, and completed a post-test survey. At the intervention site, all participants played *Re-Mission* as prescribed. Then, 81.02% (175/216) of participants continued to 10-day follow-up (retention rate from baseline), and 46.76% (101/216) continued to 20-day follow-up (Figure 1).

There was no significant difference between participants who did and those who did not continue to 10-day follow up assessment with respect to age ($F_{1,201}$=3.40, *P*=.07), gender (χ^2_1=0.17, *P*=.68), ethnicity (χ^2_4=7.08, *P*=.13), usual frequency of gameplay at preintervention ($F_{1,203}$=0.09, *P*=.76), or control over stress ($F_{1,180}$=0.02, *P*=.90). Similarly, there were no differences between participants who did and those who did not continue to 20-day follow-up with respect to age ($F_{1,201}$=0.08, *P*=.78), gender (χ^2_1=0.03, *P*=.85), ethnicity (χ^2_4=4.65, *P*=.32), usual frequency of gameplay at preintervention ($F_{1,203}$=0.07, *P*=.79), or control over stress ($F_{1,180}$=0.17, *P*=.68). Sociodemographic characteristics of the 3 groups did not differ significantly at baseline (Table 2).

Table 2. Baseline participants' characteristics.

Characteristics	Participants, n (%)[a]				P [b]
	No challenge (control) (n=50)	Low challenge (n=81)	High challenge (n=85)	Total sample (n=216)	
Age, years					
18	3 (6.0)	20 (24.7)	15 (17.7)	38 (17.6)	.1
19	10 (20.0)	21 (25.9)	21 (24.7)	52 (24.1)	
20	12 (24.0)	13 (16.1)	24 (28.2)	49 (22.7)	
21	9 (18.0)	9 (11.1)	13 (15.3)	31 (14.6)	
≥22	12 (24.0)	11 (13.6)	9 (10.6)	32 (14.8)	
Missing	4 (8.0)	7 (8.6)	3 (3.5)	14 (6.5)	
Gender					
Male	19 (37.3)	37 (45.7)	43 (50.6)	99 (45.6)	.5
Female	27 (54.0)	37 (45.7)	39 (45.9)	104 (47.9)	
Missing	4 (8.0)	7 (8.6)	3 (3.5)	14 (6.5)	
Race/ethnicity					
White/Caucasian	34 (68.0)	50 (61.7)	46 (54.1)	131 (60.2)	.1
Asian	4 (8.0)	15 (18.5)	21 (24.7)	40 (18.5)	
Hispanic/Latino	1 (2.0)	4 (4.9)	3 (3.5)	8 (3.7)	
African American	4 (8.0)	3 (3.7)	10 (11.8)	17 (7.9)	
American Indian/Alaska Native	2 (4.0)	0 (0.0)	1 (1.12)	3 (1.4)	
Missing	5 (10.0)	9 (11.1)	4 (4.7)	18 (8.3)	
Prior cancer screening					
Yes	8 (15.7)	13 (16.1)	11 (12.9)	32 (14.8)	.7
No	38 (74.5)	61 (75.3)	70 (82.4)	169 (77.9)	
Missing	5 (9.8)	7 (8.6)	4 (4.7)	16 (7.4)	
Prior cancer diagnosis					
Yes	1 (2.0)	1 (1.2)	0 (0.0)	2 (14.8)	.6
No	46 (90.2)	73 (74.1)	81 (85.9)	200 (79.6)	
Missing	4 (7.8)	7 (8.6)	4 (4.7)	15 (6.9)	

[a]Percentages may not sum to 100% due to rounding.

[b]Test of association was done from chi-squared test (categorical variables), excluding categories of missing values. No differences were found in demographic characteristics between groups.

Figure 1. CONSORT flow diagram. DNC, participants who did not continue in the study.

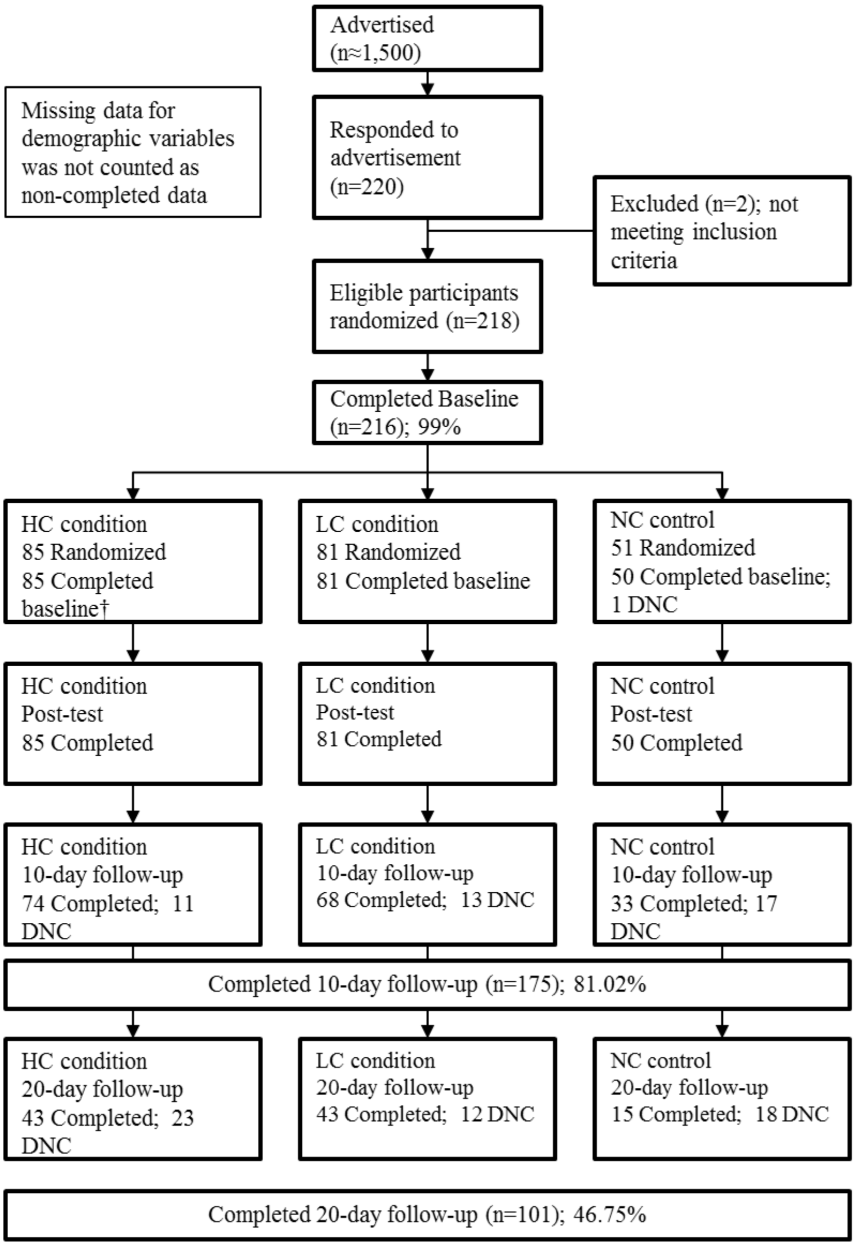


Table 3. Manipulation checks to confirm expected differences and similarities between the conditions.

Variables	Control	Low challenge	High challenge	F [a]	P	η^2
	Mean (SD)	Mean (SD)	Mean (SD)			
Manipulation-dependent personality traits[b] and gaming experience						
General perceived skills in game play	3.77 (1.26)	3.72 (1.37)	3.71 (1.36)	0.01	.98	<.001
General reaction to threat	4.47 (1.00)	4.64 (0.98)	4.70 (1.04)	0.45	.63	.005
General control over stress	4.93 (0.94)	4.98 (1.05)	5.04 (0.94)	0.14	.87	.001
Number of hours of game play per week	3.00 (5.31)	1.70 (3.13)	2.79 (8.50)	0.83	.44	.008
Expected manipulation outcomes[c]						
Attitude toward *Re-Mission*	3.96 (1.61)	4.32(1.65)	4.28 (1.80)	0.74	.48	.006
Perceived control over *Re-Mission*	-	4.42 (1.55)	3.29 (1.21)	27.39	<.001	.14
Perceived challenge in playing *Re-Mission*	-	3.33 (1.47)	4.04 (1.22)	12.06	<.001	.07
Perceived threat from cancer cells	-	4.21 (2.22)	5.18 (2.19)	8.02	.005	.05

[a]Eight separate one-way ANOVAs analyzing the differences between the conditions.

[b]Personality traits that are considered are those that may affect the manipulation of challenge. No differences were found in such traits between groups.

[c]Outcomes that may be affected by the manipulation of challenge. Variables that are specific to *Re-Mission* were not considered in the control group.

Manipulation Checks

A series of ANOVAs were conducted to check for any differences in prior game experience between the conditions, as well as any personality traits that may affect the manipulation (Table 3). There were no significant differences between the conditions with regard to skills in digital game play, general reaction to threat during game play, control over stress when playing games, or prior history of game play.

After the intervention, there was no significant difference between the conditions with respect to attitude toward *Re-Mission*. As expected, players in the HC condition, compared with the LC condition, were less likely to perceive control over game play. In addition, players at HC were more likely to experience positive challenge and perceive threat from cancer cells in the game (Table 3).

At baseline, there were no significant differences between the conditions with respect to perceived severity ($F_{2,110}$=0.49, P=.61) or ISB ($F_{2,139}$=0.49, P=.17). However, perceived susceptibility to cancer was greater among participants in the control group, compared with participants in the HC group ($F_{2,110}$=3.46, P=.03). Subsequent analyses controlled for this baseline group difference by including the baseline variable as an independent variable in the mixed-effect model.

Checking for Confounders

To check for potential demographic confounders, we determined whether intervention effects varied by gender, ethnicity, prior cancer history, or perceived gaming skills. Results generally failed to identify any differential impact as a function of being female (χ^2_2=1.51, P=.47), being White/Caucasian (χ^2_2=5.25, P=.07), having personal or social cancer history (χ^2_2=1.04, P=.59), or reporting having skills in game play ($F_{2,176}$=0.01, P=.98).

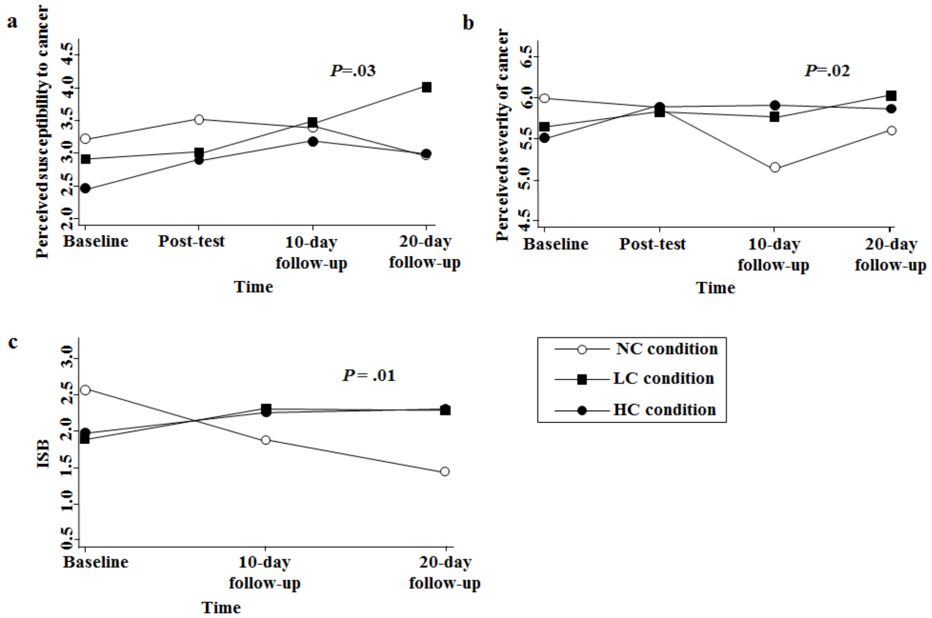
Insuring Power for Main Data Analysis

Post-hoc power analysis for a repeated-measures analysis to test perceived susceptibility or perceived severity revealed that with 3 groups, 4 repeated measures, a constant correlation of 0.5, an alpha of 0.05, and a sample size of 101, there was 90.70% power to detect small-to-moderate overall effect (Cohen's d=0.15). The same power analysis to test ISB revealed that with 3 groups, 3 repeated measures, a constant correlation of 0.5, an alpha of 0.05, and a sample size of 101, there was 84.66% power to detect small-to-moderate overall effect (Cohen's d=0.15).

Perceived Susceptibility to Cancer

Mixed-effect models showed a significant group × time interaction effect on perceived susceptibility to cancer (P=.03; Figure 2 a). For the LC group, there was no significant increase in perceived susceptibility from baseline to posttest (B=0.09, SE 0.16, P=.56). However, a significant increase was observed from posttest to 10-day follow-up (B=0.47, SE 0.16, P=.005) and a marginal increase from 10-day to 20-day follow-up (B=0.56, SE 0.31, P=.07). For the HC group, a significant increase in perceived susceptibility was observed from baseline to posttest (B=0.43, SE 0.14, P=.002), which then exhibited a plateau with no significant change from posttest to 10-day follow-up (B=0.30, SE 0.18, P=.10) and from 10-day to 20-day follow-up (B=−0.19, SE 0.26, P=.46). For the NC control group, no significant change was observed from baseline to 20-day follow-up (B=−0.26, SE 0.38, P=.50; Figure 2 a).

Figure 2. Adjusted predictions of condition-by-time. ISB, information seeking behavior; NC, no challenge; LC, low challenge; HC, high challenge. P-values present significance for the group × time interaction effect.



Perceived Severity of Cancer

A significant group × time interaction effect is also observed for perceived severity (*P*=.02; Figure 2 b). The LC group exhibited no significant changes from baseline to 20-day follow-up (B=0.40, SE 0.33, *P*=.24). On the other hand, the HC group showed a significant increase from baseline to posttest (B=0.39, SE 0.14, *P*=.005), which plateaued from posttest to 10-day follow-up (B=0.005, SE 0.18, *P*=.98) and 10-day to 20-day follow-up (B=−0.007, SE 0.26, *P*=.98). For the NC control group, no significant change is observed from baseline to posttest (B=−0.13, SE 0.18, *P*=.48). However, from posttest to 10-day follow-up, the NC group showed a significant decrease in perceived severity (B=−0.74, SE 0.24, *P*=.002), with no significant change from 10-day to 20-day follow-up (B=0.46, SE 0.32, *P*=.15; Figure 2 b).

Information Seeking Behavior

Mixed-effect analysis showed a significant group × time interaction effect on cancer information seeking by 20-day follow-up (*P*=.01; Figure 2 c). From baseline to 10-day follow-up, participants in the LC group (B=1.09, SE 0.40, *P*=.006) and the HC group (B=0.97, SE 0.38, *P*=.01) were more likely to increase in ISB, compared with participants in the NC group. Also, from baseline to 20-day follow-up, participants in the LC group (B=1.50, SE 0.52, *P*=.004) and the HC group (B=1.42, SE 0.51, *P*=.005) were more likely to increase in ISB, compared with participants in the NC group. Interestingly, ISB significantly decreased for NC from baseline to 20-day follow-up (B=−1.11, SE 0.43, *P*=.01). One-way ANOVA indicated that change in ISB over time (ISB at 20-day follow up – ISB at baseline) was significantly higher for LC compared with NC ($F_{2,42}$=3.60, *P*=.004), and higher for HC compared with NC ($F_{2,42}$= 2.96, *P*=.008).

With ISB as a dichotomous variable, mixed-effect logistic regression indicated a significant group × time interaction effect (*P*=.007). Participants in the LC group were more likely to have sought cancer information at 10-day follow-up (OR 5.10, 95% CI 1.06-24.46, *P*=.04) and 20-day follow-up (OR 121.89, 95% CI 7.05-2105.88, *P*=.001), compared with participants in the NC group. This relationship was also significant for HC compared with NC at 10-day follow-up (OR 6.26, 95% CI 1.40-28.11, *P*=.02) and 20-day follow-up (OR 107.23, 95% CI 7.17-1602.69, *P*=.001).

Relationship Results

Mixed-effect analysis revealed that perceived threat from the virtual cancer cells was significantly related to an increase in perceived severity of cancer (B=0.1, SE 0.03, *P*=.001). However, there was no significant relationship between perceived threat and perceived susceptibility to cancer (B<0.001, SE 0.03, *P*=1.00).

Mixed-effect results indicated that increases in perceived susceptibility were significantly related to increases in ISB from baseline to 10-day follow-up (B=0.21, SE 0.08, *P*=.008). There was a marginal significant relationship between the increase in perceived susceptibility and the increase in ISB by 20-day follow-up (B=0.16, SE 0.09, *P*=.09). On the other hand, increases in perceived severity were not related to increases in ISB by 10-day follow-up (B=−0.07, SE 0.08, *P*=.40), or 20-day follow-up (B=−0.001, SE 0.09, *P*=.99). When considering perceived susceptibility and severity in the same model, perceived susceptibility still exhibits a significant relationship with ISB (B=0.21, SE 0.08, *P*=.007), while perceived severity exhibits no relationship with ISB (B=−0.08, SE 0.08, *P*=.33).

Discussion

Conclusions

While *Re-Mission* has shown success in promoting medication adherence in young cancer patients [17], this randomized controlled study showed that the experience of challenge in *Re-Mission*, when played by young healthy adults, led to an increase in perceived cancer severity and susceptibility, as well as a seeking of cancer-related information. The findings also indicated that the perception of threat during the intervention increased perceived severity of cancer.

This trial is the first to present the potential use of *Re-Mission* for cancer risk communication among healthy young adults. Gaming features in *Re-Mission* increased young adults' cancer risk perception and led to the seeking of cancer-related information. The critical gaming feature manipulated in this study is challenge, which in *Re-Mission* is represented as the conflict between young-adult players and cancer cells. This conflict is characterized by the proliferation of cancer cells and their continuous attack of the player avatar, Roxxi [19].

The results indicate that a high level of challenge in the intervention led to a quick (posttest) change in perceived susceptibility, whereas a low level of challenge was associated with a slower change (10-day follow-up), which marginally increased in the longer term (20-day follow-up). This indicates that the level of challenge may not need to be high to promote enduring change in perceived susceptibility.

Further, the results suggest that a high level of challenge may be needed to change perceived severity of cancer and to maintain that change in the longer term. The HC group exhibited an increase from baseline to posttest, which then plateaued from posttest until 20-day follow-up. On the other hand, the LC group showed no changes from baseline to posttest, 10-day or 20-day follow-up. As hypothesized previously [19], the HC condition, in comparison to the LC condition, exposes players to more aggressive behavior in cancer cells, an experience likely to lead them to perceive cancer as more severe.

The results highlight the limits of mere exposure to health information as opposed to virtual experience of challenge by cancer cells. For the NC control group, no change in perceived severity or perceived susceptibility is observed from baseline to posttest. However, from posttest to 10-day follow-up, the NC group showed a significant decrease in perceived severity. Following the presentation of information that describes cancer cell behavior, young adults did not display any increase in their perception of the severity of cancer, and even showed a decrease in the longer term. Such results are consistent with other research indicating that interactive experience is an important determinant of perceived severity of cancer [39,40].

There was no significant difference between the HC and LC groups in ISB variation over time (Figure 2 c). However, compared with the NC group, the HC and LC groups were more likely to demonstrate an increase in ISB from baseline to 10-day and 20-day follow-up. For this reason, the difference in ISB change between the NC group and the other groups may be driven by factors other than perceived severity.

Our relationship results also attest to this finding. Perceived threat, a direct outcome of the intervention, was related to changes in perceived severity but not susceptibility. However, it is perceived susceptibility that was found to be related to ISB. As a result, the antecedents of perceived susceptibility warrant further investigation in order to understand how ISB occurs following the gaming intervention.

Our previous investigations with *Re-Mission* indicated that threat perception is associated with fear when facing the cancer cells in the game [41]. Effects of perceived threat can be further investigated during future research, in order to understand the role of emotions in driving health outcomes.

Limitations

In this study, young adults participated in 1 session of *Re-Mission* only. Typical use of *Re-Mission* might involve many hours of play over a period of time, and the results found in this study might not be characteristic of more extensive play. The results may also be affected by the tutorial, which can be skipped in a typical setting. However, all *Re-Mission* players were presented with the tutorial to control for its potential effect.

This study ended with a relatively low retention rate (101/216, 46.8%). By the time this study reached 20-day follow-up, college students were at a transition to summer break, and ultimately, several of them were not available to continue in the study. However, this did not stop 101 participants to continue in the study, and keep acceptable power for data analysis. Future work with college students may need to consider a more suitable timing for data collection.

While the results explain short-term effects on information seeking, they do not consider long-term opportunities for actual protective behaviors. Notably, though, this preclinical trial was meant only to test the potential effectiveness of challenge as a moderator of risk perception and ISB. The current study did not inspect specific types of ISB. However, our pilot study indicated relationships between general ISB and players' intentions to obtain information from family members and from doctors during medical visits [19].

Re-Mission is not designed to directly influence actual cancer preventive behaviors among healthy young adults (eg, improvement in healthy eating or prevention of tobacco smoking). However, our results indicate that conflict with cancer cells and the virtual experience of cancer cell behavior could increase young-adults' comprehensive understanding and perception of cancer, which could well act as a driver for acquisition of preventive behaviors.

Implications

The results of this study can be used to inform the design of a novel, game-based intervention for cancer risk communication. Such an intervention might make use of the current findings by providing a balance between LC and HC during conflict with virtual cancer cells. With challenge manipulation, the intervention may create a state of balance between the level of challenge in the game and players' ability to overcome the challenge. Also, the new intervention could allow young adults to discover new cancer information inside the game, and ultimately learn about new ways to protect themselves. With cancer information embedded in the intervention, it may become possible to measure ISB objectively by monitoring the player's accessing of information in the game. The new intervention may boost self-efficacy for self-protection by allowing young adults to enter a virtual environment that facilitates the simulation of engagement in healthy actions. While challenge with cancer cells may drive risk perception, it can also allow young adults to experience the consequences of their actions in the game. In particular, a new intervention may allow young adults to explore the risks of cancer-promoting behaviors (eg, smoking), as well as the benefits of cancer preventive behaviors (eg, healthy eating). Finally, with management options in the game, young adults can create a plan and register electronic reminders to engage in preventive behaviors. For instance, through the game, college students can make Web-based appointments for cancer screening or vaccination at the clinics of their Universities.

Acknowledgments

We would like to thank all research members from the University at Buffalo, the State University of New York (Buffalo, New York) for their contributions to the success of this trial, including Helen Wang, PhD (advisor), Arun Vishwanath, PhD (provider of equipment and laboratory usage), Brian Reynolds and Amanda Damiano, PhD (facilitators of data collection), Wayne Weiai Xu, PhD (research coordinator and assistant), and HopeLab (*Re-Mission* creators, Redwood City, California). Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number R25CA057730 (Principle Investigator: Shine Chang, PhD) and by the Cancer Center Support Grant CA016672 (Principle Investigator: Ronald DePinho, MD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' Contributions

The Principal Investigator, Georges E. Khalil, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A video of the *Re-Mission* game presenting the conflict between Roxxi and the cancer cells.

[MP4 File (MP4 Video), 4MB - games_v4i2e13_app1.mp4]

Multimedia Appendix 2

Screenshots of the *Re-Mission* settings interface that allow the manipulation of the features of challenge.

[PNG File, 49KB - games_v4i2e13_app2.png]

Multimedia Appendix 3

Screenshot of the layout and design of *Re-Mission*; an illustration from the “no challenge” (NC) condition.

[PNG File, 468KB - games_v4i2e13_app3.png]

Multimedia Appendix 4

Checklist for reporting results of Internet e-surveys (CHERRIES).

[PDF File (Adobe PDF File), 39KB - games_v4i2e13_app4.pdf]

References

1. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. Soc Sci Med 2004;59:1795-1806. [doi: 10.1016/j.socscimed.2004.02.020] [Medline: 15312915]
2. National Cancer Institute. Brief 16: Trends in Cancer Information Seeking. In HINTS 2007 database: National Cancer Institute, DCCPS, Behavioral Research Program, Health Communication and Informatics Research Branch; 2007. URL: http://hints.cancer.gov/briefs.aspx [accessed 2016-03-21] [WebCite Cache ID 6gBFGtVO3]
3. Rimal RN. Perceived risk and self efficacy as motivators: understanding individuals' long term use of health information. J Commun 2001;51:633-654. [doi: 10.1111/j.1460-2466.2001.tb02900.x]

4. Shim M, Kelly B, Hornik R. Cancer information scanning and seeking behavior is associated with knowledge, lifestyle choices, and screening. J Health Commun 2006;Suppl 1:157-172. [doi: 10.1080/10810730600637475] [Medline: 16641081]
5. Moorman C, Matulich E. A Model of consumers' preventive health behaviors: The role of health motivation and health ability. J Consum Res 1993;20:208-228. [doi: 10.1086/209344]
6. Rimal RN. Perceived risk and efficacy beliefs as motivators of change. Hum Commun Res 2003;29:370-399. [doi: 10.1093/hcr/29.3.370]
7. Aiken LS, Fenaughty AM, West SG, Johnson JJ, Luckett TL. Perceived determinants of risk for breast cancer and the relations among objective risk, perceived risk, and screening behavior over time. Womens Health 1995;1:27-50. [Medline: 9373372]
8. Katapodi MC, Lee KA, Facione NC, Dodd MJ. Predictors of perceived breast cancer risk and the relation between perceived risk and breast cancer screening: a meta-analytic review. Prev Med 2004;38:388-402. [doi: 10.1016/j.ypmed.2003.11.012] [Medline: 15020172]
9. Champion VL. Revised susceptibility, benefits, and barriers scale for mammography screening. Res Nurs Health 1999;22:341-348. [Medline: 10435551]
10. Prokhorov AV, Yost T, Mullin-Jones M, de Moor C, Ford KH, Marani S, et al. “Look at your health”: outcomes associated with a computer-assisted smoking cessation counseling intervention for community college students. Addict Behav 2008;33:757-771 [FREE Full text] [doi: 10.1016/j.addbeh.2007.12.005] [Medline: 18280668]
11. American CS. Global Cancer Facts and Figures 2nd Edition.: American Cancer Society; 2011. URL: http://www.cancer.org/research/cancerfactsstatistics/global [accessed 2016-03-21] [WebCite Cache ID 6gBFMr5CQ]
12. Wilson T. Information behaviour: an interdisciplinary perspective. Inform Process Manag 1997;33:551-572. [doi: 10.1016/S0306-4573(97)00028-9]
13. Alzougool B, Chang S, Gray K. Towards a Comprehensive understanding of health information needs. Electron J Health Inform 2008;3:15.
14. Pang P, Chang S, Pearce J, Verspoor K. Online health information seeking behaviour: understanding different search approaches. 2014 Presented at: Pacific Asia Conference on Information Systems; 06, 24, 2014; Chengdu, China.
15. HopeLab. Get the Game *Re-Mission*. 2016. URL: http://www.re-mission.net/get-the-game/ [accessed 2016-03-21] [WebCite Cache ID 6gBFQ9edR]
16. Beale IL, Kato PM, Marin-Bowling VM, Guthrie N, Cole SW. Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. J Adolesc Health 2007;41:263-270. [doi: 10.1016/j.jadohealth.2007.04.006] [Medline: 17707296]
17. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008;122:e305-e317 [FREE Full text] [doi: 10.1542/peds.2007-3134] [Medline: 18676516]
18. Cole SW, Yoo DJ, Knutson B. Interactivity and reward-related neural activation during a serious videogame. PLoS One 2012;7:e33909 [FREE Full text] [doi: 10.1371/journal.pone.0033909] [Medline: 22442733]
19. Khalil GE. When losing means winning: the impact of conflict in a digital game on young adults' intentions to get protected from cancer. Games Health J 2012;1:279-286. [doi: 10.1089/g4h.2012.0017] [Medline: 26191631]
20. Rogers R, Prentice-Dunn S. Protection motivation theory. In: Handbook of Health Behavior Research. New York, NY: Plenum Press; 1997:113-132.
21. Rogers RW. A protection motivation theory of fear appeals and attitude change. J Psychol 2010;91:93-114. [doi: 10.1080/00223980.1975.9915803]
22. Ransom S, Jacobsen PB, Schmidt JE, Andrykowski MA. Relationship of problem-focused coping strategies to changes in quality of life following treatment for early stage breast cancer. J Pain Symptom Manage 2005;30:243-253. [doi: 10.1016/j.jpainsymman.2005.03.013] [Medline: 16183008]
23. Rees CE, Bath PA. The psychometric properties of the Miller Behavioural Style Scale with adult daughters of women with early breast cancer: a literature review and empirical study. J Adv Nurs 2000;32:366-374. [Medline: 10964184]
24. Rees CE, Bath PA. Information-seeking behaviors of women with breast cancer. Oncol Nurs Forum 2001;28:899-907. [Medline: 11421149]
25. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007;17:1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
26. Lee TS, Kilbreath SL, Sullivan G, Refshauge KM, Beith JM. The development of an arm activity survey for breast cancer survivors using the Protection Motivation Theory. BMC Cancer 2007;7:75 [FREE Full text] [doi: 10.1186/1471-2407-7-75] [Medline: 17488497]
27. Eysenbach G, CONSORT- EHEALTH. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13:e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]

29. Bedwell WL, Pavlas D, Heyne K, Lazzara EH, Salas E. Toward a taxonomy linking game attributes to learning: an empirical study. Simulat Gaming 2012;43:729-760. [doi: 10.1177/1046878112439444]
30. Crawford C. The Art of Computer Game Design. Berkeley: Mcgraw-Hill Osborne Media; 1984.
31. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6:e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
32. Forest-Boucher A, Brunet J, Fredette M. Brand placement in virtual environments: impact of flow on brand recognition and recall. Les Cahiers du Gerad 2008;75:1-22.
33. Feldner MT, Hekmat H. Perceived control over anxiety-related events as a predictor of pain behaviors in a cold pressor task. J Behav Ther Exp Psychiatry 2001;32:191-202. [Medline: 12102581]
34. Kraft P, Rise J, Sutton S, Røysamb E. Perceived difficulty in the theory of planned behaviour: perceived behavioural control or affective attitude? Br J Soc Psychol 2005;44(Pt 3):479-496. [doi: 10.1348/014466604X17533] [Medline: 16238850]
35. Novak T, Hoffman D. Measuring the flow experience among web users. In: Proceedings of the Interval Research Corporation.: Interval Research Corporation; 1997 Presented at: The Interval Research Corporation; July 31, 1997; Palo Alto, CA.
36. Weinstein ND. Misleading tests of health behavior theories. Ann Behav Med 2007;33:1-10. [doi: 10.1207/s15324796abm3301_1] [Medline: 17291165]
37. Hovick SR, Kahlor L, Liang M. Personal cancer knowledge and information seeking through PRISM: the planned risk information seeking model. J Health Commun 2014;19:511-527. [doi: 10.1080/10810730.2013.821556] [Medline: 24433251]
38. Cliff N. Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich; 1987.
39. Fox J, Bailenson J, Binney J. Virtual experiences, physical behaviors: The effect of presence on imitation of an eating avatar. Presence: Teleoperators and Virtual Environments 2009;18:294-303. [doi: 10.1162/pres.18.4.294]
40. Wang H, Singhal A. Entertainment-education through digital games. In: Ritterfeld U, Cody M, Vorderer P, editors. Serious Games: Mechanisms and Effects. New York, NY: Routledge; 2009:271.
41. Khalil G. Fear and happiness in ?Re-Mission?: teasing out emotional gaming events responsible for cancer risk perception. In: Games for Health. the Netherlands: Springer; 2013:27-44.

Abbreviations

ANOVA: analysis of the variance
CI: confidence interval
HC: high challenge
ISB: information seeking behavior
LC: low challenge
NC: no challenge
OR: odds ratio
PMT: protection motivation theory
SE: standard error

Edited by G Eysenbach; submitted 23.03.16; peer-reviewed by S Perez, PCI Pang, S Koch-Weser; comments to author 28.06.16; revised version received 14.07.16; accepted 18.07.16; published 28.07.16.

Please cite as:
Khalil GE, Beale IL, Chen M, Prokhorov AV
A Video Game Promoting Cancer Risk Perception and Information Seeking Behavior Among Young-Adult College Students: A Randomized Controlled Trial
JMIR Serious Games 2016;4(2):e13
URL: http://games.jmir.org/2016/2/e13/
doi:10.2196/games.5793
PMID:27470927

Original Paper

Gamifying Self-Management of Chronic Illnesses: A Mixed-Methods Study

Alaa AlMarshedi[1], MSc; Gary Wills[1], PhD; Ashok Ranchhod[2], PhD

[1]University of Southampton, School of Electronics and Computer Science, Southampton, United Kingdom
[2]University of Southampton, Winchester School of Art, Winchester, United Kingdom

Corresponding Author:
Alaa AlMarshedi, MSc
University of Southampton
School of Electronics and Computer Science
aaa3g12@soton.ac.uk
Southampton,
United Kingdom
Phone: 44 (0)23 8059 4506
Fax: 44 (0)23 8059 4506
Email: aaa3g12@soton.ac.uk

Abstract

Background: Self-management of chronic illnesses is an ongoing issue in health care research. Gamification is a concept that arose in the field of computer science and has been borrowed by many other disciplines. It is perceived by many that gamification can improve the self-management experience of people with chronic illnesses. This paper discusses the validation of a framework (called The Wheel of Sukr) that was introduced to achieve this goal.

Objective: This research aims to (1) discuss a gamification framework targeting the self-management of chronic illnesses and (2) validate the framework by diabetic patients, medical professionals, and game experts.

Methods: A mixed-method approach was used to validate the framework. Expert interviews (N=8) were conducted in order to validate the themes of the framework. Additionally, diabetic participants completed a questionnaire (N=42) in order to measure their attitudes toward the themes of the framework.

Results: The results provide a validation of the framework. This indicates that gamification might improve the self-management of chronic illnesses, such as diabetes. Namely, the eight themes in the Wheel of Sukr (fun, esteem, socializing, self-management, self-representation, motivation, growth, sustainability) were perceived positively by 71% (30/42) of the participants with *P* value <.001.

Conclusions: In this research, both the interviews and the questionnaire yielded positive results that validate the framework (The Wheel of Sukr). Generally, this study indicates an overall acceptance of the notion of gamification in the self-management of diabetes.

(JMIR Serious Games 2016;4(2):e14) doi:10.2196/games.5943

KEYWORDS
gamification; healthcare; self-management; chronic illnesses; diabetes; motivation; behavioral change

Introduction

The health care industry is experiencing significant changes due to advances in Health 2.0 and mobile technologies [1]. A more user-centered approach is being used to facilitate change in health provision. There has been a focus on utilizing mobile technologies for health behavior interventions. Mobile technologies provide a medium to easily connect to patients and create change. Moreover, apps can be used to track medication, manage illness, and monitor health. Also, online communities can provide patients with the emotional and psychological support they need. However, some Web and mobile health care interventions in their current form lack effective and engaging qualities; they may not appeal to a lot of patients and their effect is temporary.

Gamification could be the solution to effective health care interventions. It is a concept that borrows from game techniques but is not a game by itself. A number of game elements, including badges, levels, leader boards, and progress bars are used in gamification [2]. The principles of engagement, reward, and incentive are incorporated in certain tasks to encourage changes in behavior or motivate users to learn new skills [3]. Gamification employs the innate urge for recognition and instant positive feedback as a way to promote change in behavior and drive user engagement. Moreover, rewards typically leave people feeling happy [4].

Gamification can be thought of as a motivational tool [5]. Generally, motivation can occur in many ways to achieve goals, satisfy personal needs, fulfill leader expectations, and gain rewards or incentives [6]. The health care community is now realizing the power of gamification on motivation [7-10]. Thus, some gamification features have been incorporated in a number of health and fitness apps [11,12]. SuperBetter is one of the successful examples of gamification. It is a tool for self-improvement that provides users with an engaging and interactive experiment to help them reach their health goals [13]. Gamification has been also incorporated in some self-management apps for diabetic patients such as MySugr [14]. MySugr helps users track their blood glucose levels along with other relevant data such as their carbohydrate intake and medication. The app also allows users to keep a photo diary of their meals. Furthermore, it rewards users who are committed to logging their information. However, the app is limited by geographical location as it is available only to users in the United States and Europe. However, the gamified apps for self-management do not follow a specific framework or guideline, and there has been relatively little research on the use of gamification in self-management and adherence to medication.

Chronic illnesses such as diabetes could benefit from the use of gamification [15]. Diabetes is considered the “disease of the 21st century,” and the most common chronic illness in the world [16]. Encouraging adherence to medication and self-management is crucial for the health of a diabetic person. Having diabetes requires a great deal of self-care, such as taking medication, keeping track of food intake, and exercising. It requires self-management skills that are vital in preventing the complications associated with the disease and maintaining a healthy life [17]. This includes the ability to deal with diabetes requirements such as lifestyle changes, medication, and physical and social consequences. Mobile apps can help patients self-manage in a more efficient manner [18].

Gamification has the potential to positively influence patients with chronic illnesses in adhering to medication and self-managing more effectively [15,19]. It can make the tedious and repetitive tasks of managing a chronic illness such as diabetes rewarding and more engaging [12]. Moreover, it can lead to an increase in the adoption of digital health care services, which is generally slow, often because such services are poorly designed and do not meet user needs [20]. In a recent study, 75% of participants showed interest in using digital health services, especially if they provide assistance with routine health tasks [20]. However, the sole reliance on points and badges could damage the longevity effect of gamification and thus diminish the purpose of gamification in the first place. While points and badges are a part of gamification, there are other crucial game techniques that need to be considered. Therefore, to benefit from all the advantages of gamification, one needs to understand the environment to which it is applied, so specific gamification techniques can be tailored and applied to this specific environment.

To address this, we introduced a framework named the Wheel of Sukr [21]. To our knowledge, this is the first framework that targets the use of gamification in the self-management of chronic illnesses (more details of the design rationale are found in [21]). It combines game elements, self-management practices, and behavioral change methods to provide effective and better self-management systems that reinforce healthier behavior. The framework addresses the issues of engagement and effectiveness of self-management applications. It also turns the self-management “tasks” into fun and rewarding activities.

The Wheel of Sukr (Figure 1) contains 28 elements organized under eight different themes:

1. Self-management [22,23]: Basic elements needed to self-monitor blood glucose, including tracking measures of blood glucose, insulin, food intake, and other related information; getting feedback based on the entries; and being notified when blood glucose measures fluctuate.

2. Socializing [24,25]: Being part of a group of people that shares the same situation, which offers social and emotional support and adds to the value of rewards.

3. Self-representation [26,27]: Tailoring the experience to the user to create a bond with the user, thus, increasing engagement and resulting in a meaningful experience.

4. Fun [2,28]: Creating a game-like experience.

5. Esteem [29,30]: Satisfying the fourth level of Maslow’s Hierarchy of Needs, thus, catering to the psychological side of managing diabetes.

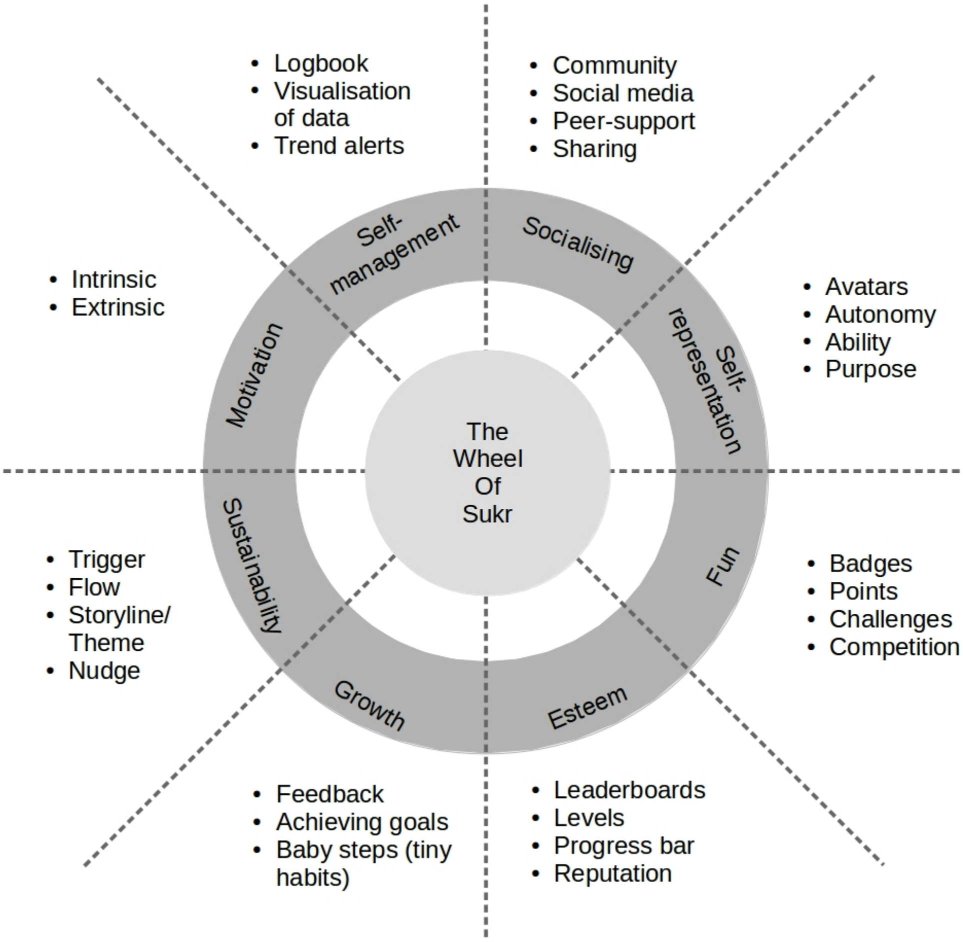
6. Motivation [31]: Appealing to the desire to do things.

7. Sustainability [27,32]: Maintaining the same level of engagement to sustain the desired effect.

8. Growth [33]: Creating a fruitful experience for the user, where gamification in a social and psychological context can result in personal growth in terms of managing diabetes, learning new healthy habits, and understanding the disease better.

In this paper, we present a validation of the Wheel of Sukr framework using a mixed-method approach that includes interviewing experts in the fields of medical practice, psychology, and gamification, as well as a questionnaire for patients with diabetes in Saudi Arabia.

Figure 1. The Wheel of Sukr.



Methods

A mixed-method approach was chosen for validating the framework. In particular, we used expert interviews and patient questionnaires. The interviews were conducted with a group of diabetes doctors and educators, psychiatrists and psychologists, and game experts. The questionnaire was answered by diabetic individuals from Saudi Arabia.

Ethical approvals for both studies were obtained from the Ethics Committee at the university of Southampton prior to conducting the interviews and questionnaires (reference numbers: 14208 and 15296).

Expert Interviews

The interviews were conducted with diabetes doctors, psychologists and psychiatrists, and game experts. The doctors, psychologists, and psychiatrists were selected from public and private hospitals in Riyadh, Saudi Arabia. Those experts were chosen based on their experience with the diabetic community in Saudi Arabia and their expertise in this area. Finally, the game experts were selected from the University of Southampton. The overall total number of experts was eight (Table 1), and no more expert interviews were required since the data from the interviews approached a saturation level.

Table 1. Interview experts.

Expert	Expert #
Diabetic doctors	Experts #1, #3, #4, #8
Psychologists and psychiatrists	Experts #2, #5
Game experts	Experts #6, #7

A thematic analysis approach was used in the analysis of the interviews. The interviews were coded with tags that represent the eight themes of the Wheel of Sukr: self-management, self-representation, fun, growth, sustainability, motivation, esteem, and socializing. After that, the interviews were analyzed based on the coded tags and in relation to the research question and literature.

Questionnaire

The questionnaire consisted of two parts: multiple-choice questions and Likert-scale questions, which had five choices (strongly agree, agree, neutral, disagree, and strongly disagree), each weighted from 5 to 1 respectively. The multiple-choice questions were asked in order to gather information from participants about demographics and self-management habits.

The multiple-choice part of the questionnaire is where the inclusion and exclusion criteria are specified. In particular, the target population is diabetics ranging in age from 18 and 40. The second part of the questionnaire was used to measure the attitude of the participants toward the framework themes. In particular, the questions were designed to capture the overall idea of the framework and its themes instead of focusing on the technical concepts of the elements. This is because gamification may not be a familiar topic to most participants.

The target audience of the questionnaire is diabetics in Saudi Arabia, where diabetes is wide spread (3.4 million in 2015) and has a high prevalence [34,35]. Saudi Arabia is one of the top 10 countries for number of children with type 1 diabetes—it has 16,200 children 15 years old and younger in 2015, which is a quarter of the region's total of 60,700 patients [34].

The sample contains 42 participants. The questionnaire was sent to a number of diabetes accounts on Twitter and Facebook, and participants were asked to answer and share it with other diabetics.

The Likert-scale part of the questionnaire was analyzed using Excel and Mathematica. A normality test was conducted and showed that the data do not follow a normal distribution. Therefore, a non-parametric test was used, specifically the Mann-Whitney test (1-tailed). The reference median (median of the weights) was 3, and the alpha level was specified to alpha=.05. The following hypotheses were used to test the framework themes:

- H_0 (null hypothesis): median ≤3
- H_1 (alternative hypothesis): median 3

To reject the null hypothesis, the *P* value must be *P* ≤.05. Since the questionnaire contains 34 questions, the Bonferroni correction was used [36]. Specifically, alpha is corrected to be .05/34=.0014705. Hence, the *P* value for each question should be less or equal to the new corrected value for the null hypothesis to be rejected. However, it is worth mentioning that the Bonferroni correction can be too conservative.

To ensure the clarity and validity of the questionnaire, we conducted a pilot study in which the questionnaire was given to a number of researchers for feedback. Moreover, the questions were placed randomly (not according to the themes), and a few questions were repeated in different places to ensure validity of the answers. Finally, Cronbach alpha was applied to check the internal consistency of the questionnaire.

Sample

The snowball method was used to select the sample for the patients' questionnaire. Prior to the distribution of the questionnaire, the sample size was estimated using the program G*power, setting alpha=.05, β=0.2, and effect size *d*=0.8 [37]. Based on this, the minimum sample size is 15. However, after distributing the questionnaire, 42 patients participated in the study, which is larger than the minimum sample size but also meets the practical application of the central limit theorem, to represent the mean of the population.

Results

This section is divided into two subsections. The first presents the findings of the expert interviews and the second presents the results of the questionnaire.

Expert Interview Findings

The diabetes doctors and educators provided valuable information based on their immediate experience with diabetics in Saudi Arabia. Similarly, psychologists and psychiatrists provided insight into the psychological issues that diabetics face in Saudi Arabia that can affect their self-management. As for game experts, their input was specifically focused on the elements derived from gamification/game literature. The overall findings of the interviews provided the validation of the Wheel of Sukr.

Fun

The idea of creating an enjoyable experience for diabetic patients was strongly welcomed by experts. Expert #3 said, "Naturally people like to be rewarded. Thus, if this is applied to the self-management of diabetes, it would be very effective." Moreover, Expert #3 added "it will change the view and the experience of self-management of diabetes for the patient." Expert #5 also agreed, saying, "positive reward is enjoyable in whichever form it comes. This will help patients' self-esteem." Furthermore, Expert #8 who regularly participates in events for diabetic patients said, "using games, competition, and fun events has shown a positive effect on diabetic patients." This suggests that creating a game-like experience that entertains the users could be a relevant part to improving the self-management process.

Socializing

The ability to share the same experiences and concerns with other diabetics can offer the social and emotional support that a diabetic patient needs. Moreover, it creates a good environment for gamification where the existence of a social community adds to the value of rewards. As Expert #2 pointed out: "In today's world, the effects of social media on young and early adolescents is very big. In fact, it could leave a stronger impact on the patient than that of the doctor." Therefore, the social and community aspect of the framework can be essential in providing support for diabetics and tying in all the other themes of the framework.

Esteem

Experts agreed that diabetes has a stigma in Saudi Arabia. The parents and family of diabetic individuals enforce this by being overprotective of their diabetic children. This affects their self-management, as Expert #8 stated, "how the patient feels about diabetes has a great effect on their self-management." Therefore, creating competitions between peers and adding leaderboards, levels, and showing progress bars in a self-management tool could be essential in boosting users' self-esteem. This might fulfill the need for recognition and instant positive feedback in human nature. This could result in positive change in behavior regarding the self-management of diabetes.

Self-Management

According to some experts, some diabetics feel stigmatized and may be shy in dealing with their condition when they are participating in social activities. In turn, this could prevent them from maintaining their daily self-management routines, for example, missing taking blood glucose test results. However, the majority of experts (especially psychiatrists) argued that some families play a negative role, which prevents diabetic patients from self-managing diabetes properly. This was supported by Expert #3 who said, "The stigma on diabetes in Saudi Arabia affects the ability of some patients to perform the daily self-management activities in public or around other people." Therefore, a gamified self-management tool could enable patients to self-manage with ease and confidence and without embarrassment or delay. The framework themes collectively could achieve this.

Motivation

The majority of experts agreed that many diabetic patients lack motivation in terms of self-management. Therefore, a gamified self-management tool should take this into consideration. This is why motivation is a significant part of the framework. The diabetes doctors highlighted that there is a lack of motivation in following the right procedures for self-management and caring for oneself among some patients. Both Experts #1 and #3 said that some patients are not motivated to learn about their illness and learn self-management skills. This could be also relevant to the "growth" theme (discussed later) in which feedback and progress are essential. Expert #4 expressed that game elements and rewards could be the solution to the lack of motivation, which supports the "fun" theme discussed earlier.

Growth

The experts agreed that elements of the growth theme such as Feedback are essential to the self-management of diabetes. As mentioned in previous elements, there is a lack of motivation in self-management and a lack of consistency. Therefore, the growth theme is a vital part of the framework, and applying all the themes combined may help in creating new habits in self-management of diabetes and creating consistency.

Self-Representation

Some patients might feel that they are being blamed for not taking care of their condition, for example, if their blood glucose was higher or lower than normal. However, if the patients were in an environment where they are encouraged by other peers, this could help them improve their self-managing skills. Similarly, adolescents might be reluctant to take their doctors' instructions responsibly, possibly because it could undermine their independence, as suggested by Expert #2. This is reinforced by the opinion of Expert #6: "If the user has a sense of control of what they are doing, they will feel that things are not imposed on them and they are the actors." Therefore, a gamified self-management app should provide an environment for diabetic patients where they feel represented and in control. Additionally, the environment should allow them to pursue and achieve their goals regarding self-managing diabetes.

Sustainability

Sustainability is essential to the success of any gamification app. Additionally, maintaining the same level of engagement can result in a positive change in behavior in self-management of diabetes. Experts agreed that the use of triggers and nudge theory, which are the elements of sustainability theme, might direct users into the desired behavior for self-management.

In general, the interview results showed a consensus on the importance of the framework's themes in self-management of Saudi diabetics. In fact, Expert #2 said that the information presented in the framework is enough to start a successful project for the self-management of diabetes. The expert continued to highlight that the first implementation of the framework in a system would show any shortcomings. Next, the system would be enhanced based on user feedback. Moreover, Expert #4 pointed out the importance of such projects for the diabetic people in Saudi, especially because of the conservative nature of the community. Thus, a gamified online system could help immensely in motivating diabetics and keeping their privacy. Additionally, Expert #5 pointed out that using game elements is great for adolescence and young users. The expert continued to highlight that, in Saudi Arabia, there are many young diabetics with overprotective parents (due to their illness). Thus, using gamification and providing them with an enjoyable experience of self-management is important.

Questionnaire Results

Summary

Table 2 shows the results of Part 1 of the questionnaire, which contained multiple-choice questions on demographical questions and habits related to their self-management.

Table 2. Data from Part 1 of the questionnaire.

Question/answer choice	Responses, %
Age group, years	
18-20	21.43
21-25	30.95
26-30	11.90
31-35	14.29
36-40	21.43
Gender	
Female	76.19
Male	23.81
When were you diagnosed with diabetes?	
0-1 years	14.29
2-4 years	26.19
5-8 years	14.29
9+ years	45.24
Do you have friends with diabetes?	
Yes	71.43
No	28.57
Do you keep a log of all your daily test results?	
Yes	40.48
No	59.52
How do you log your daily test results?	
Manually (using a pen and paper)	66.67
Electronically (using mobile apps, computer systems, etc)	33.33

Table 3 shows the results of the second part of the questionnaire, which contained Likert-scale questions on the themes of the framework. Each theme was associated with a number of questions to measure the attitude towards the theme (see Multimedia Appendix 1). For example, the fun theme has 5 questions related to it. The frequencies for each Likert-scale item (eg, strongly agree) were averaged. This result is shown in the first column in Table 3 (39.05%). In addition, the sum of the "strongly agree" and "agree" answers is shown in the "sum" column (75.71%). Similarly, the sum of the "disagree" and "strongly disagree" is shown. The same procedure was done for all the questions.

Table 3. Frequency table (results as percentages).

	Strongly agree	Agree	Sum	Neutral	Disagree	Strongly disagree	Sum	
Fun	39.05	36.67	75.71	19.52	3.81	0.95	4.76	100.00
Social	33.86	34.13	67.99	18.78	11.64	1.59	13.23	100.00
Esteem	34.92	44.44	79.37	11.11	8.73	0.79	9.52	100.00
Self-management	23.81	48.81	72.62	13.10	11.90	2.38	14.29	100.00
Motivation	38.89	38.10	76.98	11.90	8.73	2.38	11.11	100.00
Growth	43.33	44.76	88.10	8.57	2.86	0.48	3.33	100.00
Self-representation	19.05	23.81	42.86	36.51	19.05	1.59	20.63	100.00
Sustainability	27.38	41.07	68.45	20.83	10.71	0.00	10.71	100.00

It is worth mentioning that the Cronbach alpha test was applied to the questionnaire, and the result was 0.91, which indicates an internal consistency of the questionnaire.

Fun

Overall, when we average the frequencies of the answers to each Likert item, we find that 76% (32/42) responded by strongly agree or agree, 19% (8/42) were neutral, and 5% (2/42) strongly disagreed or disagreed. This suggests that the majority of the participants support the importance of the fun theme, since it helps them overcome boredom from the repetitive tasks and provides entertainment and encouragement. It also gives them the opportunity to be appreciated for their efforts in self-management and the opportunity to positively compete with one another.

Socializing

Following the same procedure as described in the fun theme, 68% (28/42) of the participants answered with strongly agree or agree, 19% (8/42) were neutral, and 13% (5/42) supported strongly disagree or disagree. This suggests that diabetics like sharing their positive results with one another and establishing new friendships with their peers. This could help them cope and live positively with their condition.

Esteem

The results of the averaged frequencies of the answers to the Likert scale are as follows: 79% (33/42) strongly agreed or agreed, 11% (5/42) were neutral, and 4/42 (9%) strongly disagreed or disagreed. Clearly, the majority of the participants supported this theme. It is worth recalling that the esteem theme includes progress bars, leaderboards, and reputation. So, by enabling the patients to see each other's scores and progress, they will be encouraged to self-manage. More importantly, the patients will have the chance to support and encourage each other.

Self-Management

Nearly three-quarters (73%, 31/42) of participants strongly agreed or agreed, 13% (5/42) were neutral, and 14% (6/42) strongly disagreed or disagreed. In particular, the majority of participants see that that a self-management system should provide information, tips, and notification. Moreover, when participant were asked about question 18, “I only record my tests to show them to my doctor” (without the conservative Bonferroni correction), 57% (24/42) agreed or strongly agreed with this statement, which could signify a need for developing better self-management habits.

Motivation

Over three-quarters (76%, 32/42) of participants strongly agreed or agreed, 12% (5/42) were neutral, and 12% (5/42) strongly disagreed or disagreed. The participants supported the relevance of the motivation theme. The result indicates that the participants recognize their role in managing their condition, and they are keen to keep their illness in control.

Growth

A majority (88%, 37/42) of participants agreed or strongly agreed, while 3% (1/42) answered disagree or strongly disagree. This theme received much support from the participants. Recall that the growth theme combines feedback, achieving goals, and tiny habits. The participants agreed that receiving feedback regarding inputs (glucose levels, food intake, etc) is important since this enables them to self-manage their condition.

Self-Representation

Fewer than half (43%, 18/42) of participants agreed and strongly agreed, 36% (15/42) were neutral, and 21% (9/42) disagreed or strongly disagreed. If we consider question 27 (It is important to me to keep an eye on my health through improving my self-management skills) only, then 55% (23/42) agreed or strongly agreed, while 14% (6/42) disagreed. This indicates that online self-representation is important to a significant number of the participants, which suggests that the self-representation theme is indeed relevant.

Sustainability

The results showed that 68% (28/42) agreed or strongly agreed, 21% (9/42) were neutral, and 11% (5/42) disagreed or strongly disagreed. A considerably large number of participants agreed that self-management apps should be regularly updated. These updates should keep them encouraged to keep using the app, for instance, by adding more levels or challenges that keep the patients motivated to use the app and therefore continue self-managing their condition in a sustainable manner.

Discussion

Principal Considerations

Gamification has been receiving a great deal of attention in the health care field. It has been pointed out that there is a lack of professional criteria or guidelines to help developers in creating effective apps utilizing gamification and behavioral change theories [11]. This paper presents a validated framework for gamifying the self-management of chronic illnesses to fill a gap in the literature. The validation was carried out by a mixed-method approach, which included expert interviews as well as patient questionnaires. The findings of the interviews and results of the questionnaires support the idea of incorporating gamification in the self-management process of diabetes. Both experts and patients agreed that utilizing the combined themes of the Wheel of Sukr to create a gamified self-management tool might help achieve effective self-management and behavioral change. To our knowledge, the Wheel of Sukr is the first of its kind.

Self-management of chronic illnesses, especially in diabetes, can be turned into an engaging and enjoyable experience by the use of gamification. The results of this study support this notion and indicate that both experts and diabetic patients recognize the potential of gamification in improving self-management of diabetes significantly. In particular, experts highlighted the importance of rewards, competition, and other fun elements in creating an enjoyable and rewarding experience that could lead to positive behavioral change. This, in turn, is reinforced by the findings from the patient questionnaire as shown in the Results section.

Diabetes is a lonely illness and diabetics are more prone to depression [38]. As stated in the Introduction, being part of an online community can provide patients with the emotional and psychological support they need. This is confirmed by the results of the interviews and the questionnaire. In particular, experts emphasized the positive impact of social media and peer support on patients. Additionally, the results of the questionnaire indicate that patients would like to share positive results with their peers and establish friendships with them. This can help them overcome any negative feelings they might encounter. Furthermore, the community aspect of a gamified self-management tool could enhance the value of rewards and other elements of the framework.

Moreover, the results support the notion that creating a fun and enjoyable experience for diabetic patients could help their self-esteem. In particular, the interviews indicate that patient's self-esteem has an effect on self-management, as one expert stated, "how the patient feels about diabetes has a great effect on their self-management". Moreover, the use of leaderboards and creating friendly competition between peers could help in boosting users' self-esteem. The esteem theme and its elements are also supported by the majority of participants in the questionnaire. This is because it enables them to track their progress and compare it to others. Also, it could trigger positive competition between peers in a friendly, non-judgmental environment.

In 2013, a study [39] pointed out that 73% of diabetics do not document their daily glucose tests. Our results are in line with those results since, in general, patients admitted that they only record their test results for their doctors to see. This could be because they are not aware of the importance of self-management or find the self-management process mundane. The results of both the interviews and the questionnaires suggest that a gamified self-management system might help them document daily. Therefore, gamification in self-management is expected to make the self-management experience less mundane.

Furthermore, the results show that patients are willing to learn more about their condition and manage themselves better. However, this could be prevented by the shortage of well-constructed self-management tools (especially in Saudi Arabia where the patients are from). This is supported by the findings of the expert interviews that indicated that many patients are not motivated to learn or self-manage. Gamifying self-management could increase patients' motivation. Yet, many existing gamified applications and services focus only on extrinsic motivation [40]. However, it is known that extrinsic motivation solely does not create a sustainable gamification affect [41]. The Wheel of Sukr considers both types of motivation (intrinsic and extrinsic). This allows it to address some of the issues raised by the experts and enables it to satisfy the patients' expectations.

Providing real-time feedback that is meaningful and relevant to users is an essential part of gamification [21]. The feedback can come in many forms including rewards and graphs of blood tests. The latter will help users learn more about their condition and recognize patterns. This is supported by the majority of participants in the questionnaire who mentioned that they would like to receive feedback regarding their self-managing progress and be notified when their blood glucose fluctuates. Additionally, the importance of being represented and being autonomous were highlighted by the results of the expert interviews. Participants in the questionnaire also supported this. Many of them stressed that they would like their virtual accounts to reflect their personality, which is an integral part of the Wheel of Sukr.

The correlation between gamification and health behavior theories has been discussed [11]. It was mentioned that even though gamification apps for health and fitness do use motivation from the health behavior theory, the use of capacity or triggers is ignored. Patients mentioned the importance of keeping apps updated to sustain their interest in using them. They also noted the need to be encouraged to keep recording daily. This can be done through using triggers, which are an essential part of the framework. Moreover, patients stated that they enjoy being challenged at a level that suits their abilities. The Wheel of Sukr framework answers this by considering the user's ability and employing triggers.

It is worth mentioning that using the mixed-method design provided a clear image of the issue [42]. It created a balance between the weaknesses of qualitative (interviews) and quantitative (questionnaire) methods [42], which allows for a well-rounded representation [43]. Moreover, the data were collected from both experts and patients, using interviews and questionnaires. The results from the interviews and the questionnaires complement each other. In fact, relying on the expert interviews alone would have resulted in a loss of all the important information provided by the patients and vice versa.

Conclusion

This research introduced the Wheel of Sukr, which is a framework that gamifies self-management of chronic illnesses. It establishes the importance of combining gamification, behavioral theories, and standard self-management techniques to create a successful gamified self-management tool. The results are based on the input of medical doctors, psychologist/psychiatrists, and gamification researchers, and the input of patients living with diabetes. The results of the statistical analysis of the questionnaire validated the themes of the framework.

This framework can be used as a guide to help developers in creating better gamified self-managing tools by taking into account all of its themes: self-monitoring, socializing, self-representation, fun, esteem, motivation, sustainability, and growth. Each one of the themes has a number of elements. These themes combined could create the right conditions for a gamified environment to improve the self-management of chronic illnesses and make it an easier and enjoyable process.

Overall, this study suggests a general acceptance of the notion of gamifying self-management of diabetes and that it could be important in improving the experience of patients. It also supports the view that there is a need to change or enhance the current view of self-management of diabetes. The use of gamification in health care and specifically patient self-care is an important research area that needs further investigation. This framework sets the stage for further studies such as creating specific guidelines for gamification (such work is already in progress).

Acknowledgments

AA acknowledges the scholarship provided by the Saudi Government.

Authors' Contributions

AA authored the manuscript, conceived the idea of The Wheel of Sukr and designed it, designed the expert interviews and the patient questionnaire, conducted the interviews, collected, analyzed, and interpreted the data. GW and AR provided guidance, critical feedback, suggestions, supervision, and participated in editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analysis of the questionnaire.

[PDF File (Adobe PDF File), 77KB - games_v4i2e14_app1.pdf]

References

1. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
2. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. New York, NY: ACM; 2011 Presented at: MindTrek '11; Sept. 28-30, 2011; Tampere, Finland p. 9-15. [doi: 10.1145/2181037.2181040]
3. Deterding S. Gamification. interactions 2012 Jul 01;19(4):14. [doi: 10.1145/2212877.2212883]
4. Schultz W. Behavioral theories and the neurophysiology of reward. Annu Rev Psychol 2006;57:87-115. [doi: 10.1146/annurev.psych.56.091103.070229] [Medline: 16318590]
5. Werbach K, Hunter D. For the Win: How Game Thinking Can Revolutionize Your Business. Philadelphia, PA: Wharton Digital Press; Aug 30, 2012.
6. Krippendorff K. Intrinsic motivation and human-centred design. Theoretical Issues in Ergonomics Science 2004 Jan;5(1):43-72. [doi: 10.1080/14639220310000086717]
7. Cugelman B. Gamification: What It Is and Why It Matters to Digital Health Behavior Change Developers. JMIR Serious Games 2013 Dec;1:e3. [doi: 10.2196/games.3139]
8. de Vette F, Tabak M, Dekker-van Weering M, Vollenbroek-Hutten M. Engaging Elderly People in Telemedicine Through Gamification. JMIR Serious Games 2015;3(2):e9 [FREE Full text] [doi: 10.2196/games.4561] [Medline: 26685287]
9. Anderson J, Rainie L. Gamification and the Internet: Experts Expect Game Layers to Expand in the Future, with Positive and Negative Results. Games Health J 2012 Aug;1(4):299-302. [doi: 10.1089/g4h.2012.0027] [Medline: 26191633]
10. Allam A, Kostova Z, Nakamoto K, Schulz P. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]
11. Lister C, West J, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
12. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
13. Roepke A, Jaffee S, Riffle O, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
14. MySugr. URL: https://mysugr.com/?attachment_id=1834 [accessed 2016-07-24] [WebCite Cache ID 6jFVxAxcl]
15. Cafazzo J, Casselman M, Hamming N, Katzman D, Palmert M. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
16. Engelgau M, Narayan KM, Saaddine JB, Vinicor F. Addressing the burden of diabetes in the 21st century: better care and primary prevention. J Am Soc Nephrol 2003 Jul;14(7 Suppl 2):S88-S91 [FREE Full text] [Medline: 12819309]
17. Diabetes Control Complications Trial Research Group (DCCT). The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329(14):977-986. [doi: 10.1056/NEJM199309303291401]

18. Årsand E, Frøisland DH, Skrøvseth SO, Chomutare T, Tatara N, Hartvigsen G, et al. Mobile health applications to assist patients with diabetes: lessons learned and design implications. J Diabetes Sci Technol 2012 Sep;6(5):1197-1206. [Medline: 23063047]
19. McCallum S. Gamification and serious games for personalized health. Stud Health Technol Inform 2012;177:85-96. [Medline: 22942036]
20. Biesdorf S, Niedermann F. Healthcare's digital future. 2014. URL: http://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/healthcares-digital-future [accessed 2016-08-30] [WebCite Cache ID 6k9aa65rw]
21. AlMarshedi A, Wills G, Ranchhod A. The Wheel of Sukr: A Framework for Gamifying Diabetes Self-Management in Saudi Arabia. In: Procedia Computer Science. 2015 Presented at: The 5th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2015); Sept. 27-30, 2015; Berlin, Germany p. 475-480. [doi: 10.1016/j.procs.2015.08.370]
22. Nodhturft V, Schneider J, Hebert P, Bradham DD, Bryant M, Phillips M, et al. Chronic disease self-management: improving health outcomes. Nurs Clin North Am 2000 Jun;35(2):507-518. [Medline: 10873263]
23. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. JAMA 2002 Nov 20;288(19):2469-2475. [Medline: 12435261]
24. van der Ven N. Psychosocial Group Interventions in Diabetes Care. Diabetes Spectrum 2003 Apr 01;16(2):88-95. [doi: 10.2337/diaspect.16.2.88]
25. Zrebiec J. Internet Communities: Do They Improve Coping With Diabetes? The Diabetes Educator 2005;31(6):825-836. [doi: 10.1177/0145721705282162]
26. Reeves B. In: Read J, editor. Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete. Boston, NY: Harvard Business Review Press; 2009.
27. Pink D. Drive: The Surprising Truth About What Motivates Us. London: Canongate Books Ltd; 2010.
28. Antin J, Churchill E. Badges in Social Media: A Social Psychological Perspective. In: CHI 2011 Gamification Workshop Proceedings. 2011 Presented at: CHI 2011; May 7-12, 2011; Vancouver, BC.
29. Deterding S, Sicart M, Lennart N, Kenton O, Dixon D. Gamification: Using game-design elements in non-gaming contexts. In: Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems - CHI. 2011 Presented at: The ACM CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, Canada. [doi: 10.1145/1979742.1979575]
30. Maslow A. A theory of human motivation. Psychological Review 1943;50(4):370-396. [doi: 10.1037/h0054346]
31. Ryan RM, Deci E. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemp Educ Psychol 2000 Jan;25(1):54-67. [doi: 10.1006/ceps.1999.1020] [Medline: 10620381]
32. Csikszentmihalyi M. Finding flow: the psychology of engagement with everyday life. New York, NY: BasicBooks; 1997.
33. Fogg BJ. Behavior Change. Tiny Habits URL: http://tinyhabits.com/ [accessed 2016-07-24] [WebCite Cache ID 6jFWSL8B2]
34. International Diabetes Federation. IDF Diabetes Atlas, 7th ed. Brussels, Belgium: International Diabetes Federation; 2015.
35. International Diabetes Federation- Saudi Arabia. 2015. URL: http://www.idf.org/membership/mena/saudi-arabia [accessed 2016-07-25] [WebCite Cache ID 6jFWYGoUf]
36. Bland JM, Altman D. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170. [doi: 10.1136/bmj.310.6973.170]
37. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L. Erlbaum Associates; 1988.
38. Egede L, Zheng D, Simpson K. Comorbid depression is associated with increased health care use and expenditures in individuals with diabetes. Diabetes Care 2002 Mar;25(3):464-470. [Medline: 11874931]
39. Rose KJ, Koenig M, Wiesbauer F. Evaluating success for behavioral change in diabetes via mHealth and gamification: MySugr's keys to retention and patient engagement. Diabetes Technology & Therapeutics 2013 Feb;15:A114.
40. Seaborn K, Fels D. Gamification in theory and action: A survey. International Journal of Human-Computer Studies 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
41. AlMarshedi A, Wanick V, Wills G, Ranchhod A. SGI: A Framework for Increasing the Sustainability of Gamification Impact. International Journal for Infonomics 2015;8(1):1044-1051. [doi: 10.20533/iji.1742.4712.2015.0123]
42. Jick TD. Mixing Qualitative and Quantitative Methods: Triangulation in Action. Administrative Science Quarterly 1979 Dec;24(4):602. [doi: 10.2307/2392366]
43. Thurmond V. The point of triangulation. J Nurs Scholarsh 2001;33(3):253-258. [Medline: 11552552]

Edited by G Eysenbach; submitted 05.05.16; peer-reviewed by F Oprescu, YY Chao; comments to author 26.05.16; revised version received 25.07.16; accepted 12.08.16; published 09.09.16.

Please cite as:
AlMarshedi A, Wills G, Ranchhod A
Gamifying Self-Management of Chronic Illnesses: A Mixed-Methods Study
JMIR Serious Games 2016;4(2):e14
URL: http://games.jmir.org/2016/2/e14/
doi:10.2196/games.5943
PMID:27612632

Original Paper

Effects of Visual Display on Joint Excursions Used to Play Virtual Dodgeball

James S Thomas[1], PT, PhD; Christopher R France[2], PhD; Megan E Applegate[1], MS; Samuel T Leitkam[1], PhD; Peter E Pidcoe[3], PT, DPT, PhD; Stevan Walkowski[4], DO

[1]Ohio University, School of Rehabilitation and Communication Sciences, Athens, OH, United States

[2]Ohio University, Department of Psychology, Athens, OH, United States

[3]Virginia Commonwealth University, Department of Physical Therapy, Richmond, VA, United States

[4]Ohio University Heritage College of Osteopathic Medicine, Department of Osteopathic Manipulative Medicine, Athens, OH, United States

Corresponding Author:
James S Thomas, PT, PhD
School of Rehabilitation and Communication Sciences
Ohio University
W281 Grover Center
Athens, OH, 45701
United States
Phone: 1 740 593 4178
Fax: 1 740 593 0293
Email: thomasj5@ohio.edu

Abstract

Background: Virtual reality (VR) interventions hold great potential for rehabilitation as commercial systems are becoming more affordable and can be easily applied to both clinical and home settings.

Objective: In this study, we sought to determine how differences in the VR display type can influence motor behavior, cognitive load, and participant engagement.

Methods: Movement patterns of 17 healthy young adults (8 female, 9 male) were examined during games of Virtual Dodgeball presented on a three-dimensional television (3DTV) and a head-mounted display (HMD). The participant's avatar was presented from a third-person perspective on a 3DTV and from a first-person perspective on an HMD.

Results: Examination of motor behavior revealed significantly greater excursions of the knee (P=.003), hip (P<.001), spine (P<.001), shoulder (P=.001), and elbow (P=.026) during HMD versus 3DTV gameplay, resulting in significant differences in forward (P=.003) and downward (P<.001) displacement of the whole-body center of mass. Analyses of cognitive load and engagement revealed that relative to 3DTV, participants indicated that HMD gameplay resulted in greater satisfaction with overall performance and was less frustrating (P<.001). There were no significant differences noted for mental demand.

Conclusions: Differences in visual display type and participant perspective influence how participants perform in Virtual Dodgeball. Because VR use within rehabilitation settings is often designed to help restore movement following orthopedic or neurologic injury, these findings provide an important caveat regarding the need to consider the potential influence of presentation format and perspective on motor behavior.

(JMIR Serious Games 2016;4(2):e16) doi:10.2196/games.6476

KEYWORDS
virtual reality; avatar presentation; joint kinematics

Introduction

Virtual reality (VR) has been used to shape motion in patients with various orthopedic and neurologic impairments (eg, low back pain, cerebral vascular accident) for a number of years [1-4]. As commercial VR systems become increasingly affordable, the feasibility of applying this technology to both clinical and home settings holds great potential for clinical rehabilitation. A particular advantage of VR is that a variety of visual stimuli can be presented and manipulated in real-time to provide insights into neural control of movement and to guide or shape joint motion. However, given the variety of platforms available to present VR stimuli, it is important to better understand how differences in the VR environment can influence motor behavior, cognitive load, and participant engagement.

Levin and colleagues have examined the effects of VR environment on motor behavior in both healthy participants and stroke patients [2,4-6], and in a recent study reported no effect of display type (ie, three-dimensional [3D] image projected on a large screen vs a head-mounted display) on end-effector path straightness, shoulder and elbow joint excursions, or trunk displacements in reaches made to virtual targets from a seated posture [6]. However, differences in vertical and horizontal direction errors were noted between the display types [6]. Others have reported that displacements of the trunk and limbs during reaching tasks were altered by manipulating the viewing angle of the participant's third-person avatar [7]. Specifically, participants reached less when the camera was oriented at 0° (ie, directly behind the avatar) than when it was oriented at angles from 45-77.5° relative to the avatar [7]. Further, the increased segment displacement observed with greater angles was accompanied by a slightly larger displacement of the whole body center of mass (COM). The results are mixed regarding the effects of VR display types on motor performance, but investigations often focused on error of the end-effector [2,4-6,8], effects of the presenting stimulus on distance judgement [9], or the influence of restricted fields of view (FOV) conditions on estimates of distances in virtual environments [9-12]. Further, these investigations have primarily focused on motor tasks that require interception with a static target and have not examined the role of virtual environments on movement strategies adopted during virtual gaming. In fact, to the authors' knowledge no studies to date have examined how VR display types influence apportionment of joint excursions when gameplay requires significant movement of the postural joints. Finally, while some investigations have reported that the sense of actual presence in VR was weakened when the avatar was viewed from a third-person perspective [13,14], there is no evidence in the existing literature that addresses the potential effects of avatar perspective on motor behavior. However, resolving this question is particularly important for developing robust rehabilitation interventions designed to shape motor behavior.

Individuals with back pain and fear of movement due to perceived risk of harm or injury (ie, kinesiophobia) consistently avoid lumbar flexion [15-19]. Indeed, we have shown reduced lumbar flexion among participants with kinesiophobia and experimental low back pain [20], subacute low back pain [16], and chronic low back pain [18], as well as among asymptomatic individuals with kinesiophobia who have recently recovered from an episode of low back pain [17]. To address this problem, we developed a VR intervention, Virtual Dodgeball, to promote lumbar flexion in individuals with chronic low back pain and fear of movement-related injury. We recently completed a Phase I randomized controlled trial to examine feasibility and safety among individuals with chronic low back pain and high levels of pain-related fear [21]. In this initial trial, Virtual Dodgeball was played on a 3D television (3DTV) and the avatar was presented in the third-person perspective. However, because we are interested in enhancing portability of this intervention beyond the laboratory and clinic and into the home environment, this study was designed to compare movement patterns when Virtual Dodgeball is presented on a 3DTV display versus a less expensive and more portable head-mounted display (HMD). Based on existing studies, we predicted that these different display types would not affect joint excursions in full-body reaching tasks.

Methods

Recruitment

We recruited 17 healthy young adults (9 male, 8 female) aged 18-35. Exclusion criteria included a history of low back injury, low back pain within the last 6 months, and any orthopedic, neurological, or visual impairment that would prevent participation. This study was approved by the Institutional Review Board of Ohio University, and written informed consent was obtained at the beginning of the session. Using a within-subjects design, participation consisted of standardized reaches to static targets in the real world (RW) and a round of Virtual Dodgeball using two different visual display types (ie, 3DTV, HMD). Each round of dodgeball consisted of three levels of difficulty. Between each level, the participant had to reach to static virtual targets presented at the same locations as the corresponding reaches performed in RW. This manuscript examines the joint excursions used to intercept the launched virtual balls during Virtual Dodgeball gameplay with two different visual display types.

Instrumentation

Movement of light-reflective marker clusters attached to the head, upper arms, forearms, hands, trunk, pelvis, thighs, shanks, and feet were tracked using a 10-camera Vicon Bonita system sampled at 100 Hz. This optoelectric-based kinematic system can track the 3D coordinates of light reflective marker clusters attached to the participant with a spatial resolution of 0.1 mm.

The time-series joint angle data were derived from the 3D segment coordinate data using an Euler angle sequence of (1) flexion-extension, (2) lateral bending, and (3) axial rotation [22] using MotionMonitor software. Joint excursions were defined as the change in joint angle from initial standing posture to posture at target contact.

Procedures

Participants reached at a comfortable speed holding a regulation dodge ball (24 cm diameter) with both hands. They performed reaches to each of three targets located in the mid-sagittal plane. Target locations were determined for each subject based on their hip height, trunk length, and arm length. The highest target was located such that the subject could, in theory, reach the target by flexing the hips 15° with the shoulder flexed to 90° and the elbow extended. Using the same shoulder and elbow joint positions, the middle and low targets could be reached by flexing the hips 30° and 60°, respectively. Using this individualized method of determining target heights allows for comparison of movement patterns across different individuals [23-25]. We have previously demonstrated that this standardized reaching task challenges participants to produce progressively more lumbar spine flexion, and in doing so is sensitive to individual differences in movement strategies between healthy individuals and those with low back pain [15-19]. In this study, the average lumbar excursions that each participant used to reach the high, middle, and low targets was subsequently used to calculate the intended impact height location of the virtual dodgeballs (described in greater detail below).

In brief, one full game was completed in each visual display type. The order of presentation of the visual display type was randomized and counter balanced such that half the participants played Virtual Dodgeball on the 3DTV first and half played Virtual Dodgeball on the HMD first. For each participant, the impact heights of the virtual balls were identical between the visual display types. During Virtual Dodgeball, participants competed against 4 virtual opponents and the object was to block or avoid virtual balls launched randomly by each of the 4 opponents. Participants earned points and cash rewards by successfully blocking launched virtual balls using a ball that they held in their hands or by avoiding a launched ball by ducking (see Multimedia Appendix 1 for a video of gameplay).

Virtual Environment

Vizard software (WorldViz) was used to develop the virtual environment and control all presented graphics and audio stimuli, including the opposing team's avatars. The six degrees of freedom kinematic data from the clusters of light reflective markers placed on the participant were streamed to the game environment at 100 Hz using Vicon Tracker software to allow for near real-time presentation of the participant's avatar (39 ms latency). The MotionMonitor software was used to control bidirectional communication with Vizard, set game parameters and target locations, and record all kinematic data during the experimental testing session.

In the 3DTV condition, a Samsung 1080p 240 Hz 3D Smart LED TV was paired with 3D shutter glasses providing an effective refresh rate of 60 Hz/eye. The participant viewed their slightly translucent avatar from a third-person perspective from a camera position 1.5 meters directly behind their avatar. The translucent avatar allowed for visibility of objects in front of the avatar. The FOV for gameplay with the 3DTV display was as follows: horizontal=50°, vertical=40°. In the HMD condition, the participant viewed their avatar from a first-person perspective that was projected using an Oculus Rift (Oculus Rift Developers Kit 2). From this perspective, the participant viewed their avatar and the environment from the position of the avatar's eyes. The FOV for the HMD display was as follows: horizontal=100°, vertical=100°, and the refresh rate was fixed at 75 Hz/eye.

Gameplay

The game environment was an indoor basketball arena, with the participant positioned at the free-throw line on one side of the court and the four virtual opponents positioned on the free-throw line on the opposite side of the court. The opposing players moved 3 m fore-aft and 3 m left-right in a random order. Virtual balls were launched every 3.3 ± 0.3 seconds in a randomized order from each of the 4 virtual opponents. The opponent who was about to launch a virtual ball changed color 300 ms prior to launch to alert the participant. If the opponent turned green and the launched ball was yellow, the participant had to attempt to block the ball with the ball held in their hand (co-located with the virtual ball held by the avatar). If the opponent turned red and the launched ball was orange, the participant had to attempt to duck to avoid the ball. A large scoreboard was positioned at the opposite end of the arena (above the opponents) so that participants could track their performance and cash rewards earned. Sound effects were also incorporated, including crowd cheering, buzzers, referee whistles, and a duck quacking sound that occurred whenever an orange ball was launched. An instrumented participant engaged in virtual dodgeball with the HMD is shown in Figure 1.

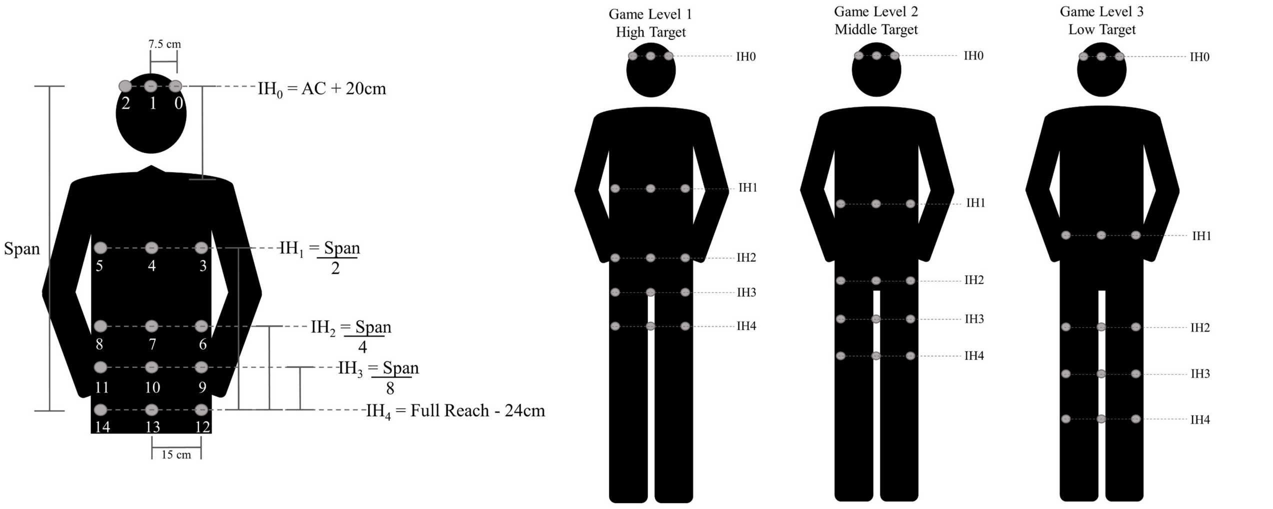
A round of gameplay consisted of a basic practice level to introduce the scoring metrics and three game levels, each lasting approximately 2 minutes. There were two sets of 15 launched balls within each game level. The intended impact locations of the 15 launched balls were distributed to five impact heights (IH) that were determined by the participant's height and the amount of lumbar flexion they used during the baseline standardized reaching tasks. For example, during Level 1 of gameplay, the participant could successfully block the virtual ball launched to IH4 (ie, the lowest impact height) simply by using the identical amount of lumbar flexion used in the standardized reaching task to the high target performed at baseline, whereas during Level 3 of gameplay, the participant could successfully block the virtual ball launched to IH4 (ie, the lowest impact height) simply by using the identical amount of lumbar flexion used in the standardized reaching task to the low target performed at baseline. The five impact heights used in gameplay were scaled to impact between the height of participant's eyes (IH0=highest impact) and approximately their shins (IH4=lowest impact) across the three levels of gameplay (see Figure 2). Three balls were launched at each IH to intersect the participant at their midline, and 20 cm left or right of the midline. It is important to note that the order of the virtual launched balls was permutated at each round of play to make the game exciting, challenging, and to some extent unpredictable. After each set, the participant was presented with a static virtual ball and instructed to reach out and touch the ball with the ball held in their hands. The location of the virtual ball was co-located to individualized target locations used during the standardized reaching task performed in real-world at pretreatment baseline such that the locations of the virtual balls during Levels 1, 2, and 3 were co-located to the real-world location of the high target, middle target, and low target, respectively.

Performance was updated in real-time and displayed on the virtual scoreboard, and the participant was awarded progressively more for each successful block or duck at each level of play (Practice Level=1¢, Level 1=2¢, Level 2=5¢, Level 3=10¢). Successful contact for each highlighted ball presented between each set resulted in a bonus 25¢ reward. Conversely, the participant lost cash rewards for each failure to block or duck. Each player started the game with a cash balance on the scoreboard such that if they failed on every launched or presented ball, their cash balance would be zero. The average gameplay session lasted approximately 15 minutes.

Following each session, the participants rated their overall efforts using the NASA Task Load Index (TLX). The NASA TLX is multidimensional assessment that rates perceived workload across to assess system performance [26] (Multimedia Appendix 2). Specifically, the NASA TLX asked the participants to provide 1 (very low) to 7 (very high) ratings of their experience along six dimensions: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. In the context of assessing display type on Virtual Dodgeball, this measure provides insight into differences in the perceived workload performing nearly identical tasks.

Figure 1. Participant instrumented and engaged in Virtual Dodgeball using the head-mounted display (HMD).

Figure 2. Methods for computing location of the impact heights (IH0-IH4) of the launched virtual balls for a single game level (left). The distribution of launched virtual balls across the 3 levels of gameplay is shown (right). The lowest impact height (IH4) for each gameplay level (1-3) was calculated from the lumbar spine flexion used to reach the high, middle, and low targets during the baseline standardized reaching tasks.



Data Reduction and Analysis

Because the games were played with both hands in fixed locations on the ball and joint excursions were nearly identical for the left and right limbs, analyses are restricted to the right side. First, time-series position vector of the right index fingertip was smoothed using a 41-point fourth-order Savitzky-Golay filter [27]. That is, at each sample time, fourth-order polynomials were fit in the least-squares sense to the data at that point and 20 neighboring samples on each side. The polynomial coefficients were then used to determine velocity. Movement onset was determined from a backwards search from peak velocity and defined as the point where velocity was ≤5% peak velocity. Target contact was defined as the point where velocity was ≤5% peak velocity using a forward search from peak velocity. The change in joint angles (ie, ankle, knee, hip, spine, shoulder, and elbow) and displacement of whole body COM along the anterior-posterior, medio-lateral, and vertical axes was calculated from movement onset to target contact. To determine hand position at target contact, we first calculated the centroid of the hands from the x, y, and z position traces from marker clusters on the left and right hands and adjusted this to the centroid of the left and right ankle joint. We then determined the hand position at target contact for the anterior-posterior, medio-lateral, and vertical axes.

Statistical Power and Analysis

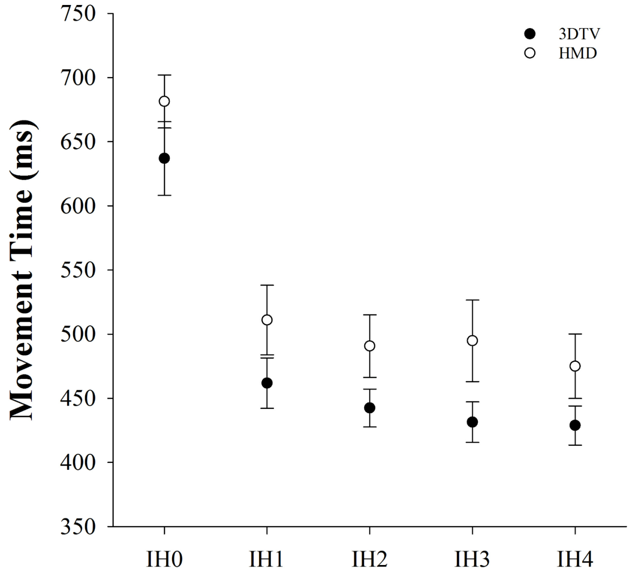
We calculated that we needed 14 participants to determine the within-subject effects of display type with 80% power, assuming alpha=.05, and correlation between measures of .5 and an effect size of *f*=0.4 (large effect) using G*Power 3.19 [28]. The effect size was based on initial pilot testing of the effects of display types on movement strategies. The dependent measures were (1) movement time, (2) hand position at target contact (ie, anterior-posterior [AP], medial-lateral [ML], vertical), (3) joint excursions (ie, right ankle, knee, hip, spine, shoulder, and elbow), and (4) displacement of COM (ie, AP, ML, vertical). Separate 4-way mixed-model multivariate analysis of variance with sex as the between-subjects variable, and environment type (3DTV, HMD), IH (IH0-IH4), and Level (1-3) as the repeated factors were performed on the dependent measures. Posthoc analyses were performed using the method of least significant differences. Interactions were examined using a simple effects model. The NASA TLX data were analyzed used paired *t* tests with a Bonferroni correction for multiple comparisons. All statistical analyses were completed in SPSS 22.

Results

Movement Time

There was a main effect of Display Type on movement time ($F_{1,15}$=6.72, P=.02), with participants moving more quickly in the 3DTV condition. Specifically, mean movement time for reaches made to intercept the launched virtual balls in the 3DTV condition was 480 ms (SD 140) versus 530 ms (SD 210) for the HMD condition. There was also a main effect of IH on movement time ($F_{4,12}$=12.76, P<.001). As illustrated in Figure 3, this is driven primarily by the differences in movement times between the movements to virtual balls launched to IH0 (ie, duck condition) compared virtual balls launched to IH1-4 (ie, block conditions). Specifically, movement times for IH0 averaged 659 ms (SD 220), which was significantly longer than movement times to intercept virtual balls launched to the other targets: IH1=486 ms (SD 200) to IH4=451 ms (SD 180). But posthoc analyses revealed that there were no significant differences in movement times between any of the pairs of IH1-IH4.

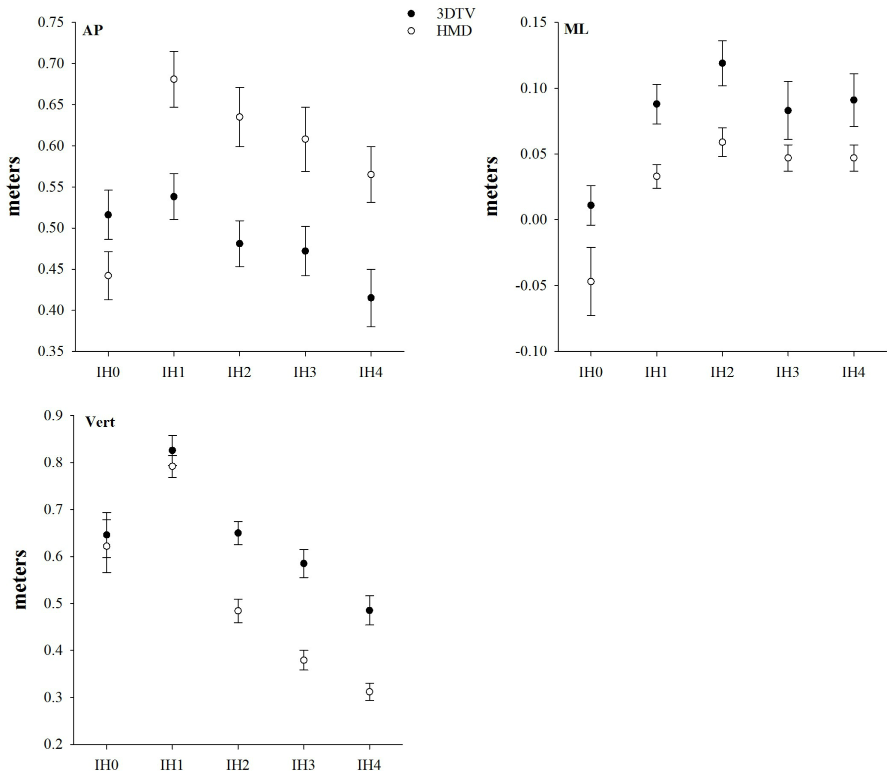
Figure 3. Effect of 3D television (3DTV) versus a head-mounted display (HMD) on movement time for each impact height (IH).



Hand Position at Ball Contact

As illustrated in Figure 4, there was an interaction of Display Type and IH of the launched virtual balls for the AP axis ($F_{4,12}$=8.44, P=.002) and the vertical axis ($F_{4,12}$=9.18, P=.01). As expected, there was also a main effect of IH on hand position at intercept with the launched virtual balls for the AP axis ($F_{1,15}$=29.28, P<.001), the ML axis ($F_{1,15}$=12.11, P<.001), and the vertical axis ($F_{1,15}$=119.66, P<.001). There were also significant effects of Display Type on hand position at intercept with the launched virtual balls for the AP axis ($F_{1,15}$=16.39, P=.001), the ML axis ($F_{1,15}$=17.61, P=.001), and the vertical axis ($F_{1,15}$=54.01, P<.001). Simple analyses of the effects of Display Type revealed that, compared to 3DTV, AP hand position at intercept was approximately 14 cm further forward in gameplay with HMD for IH1-4 (Multimedia Appendix 3). Vertical hand position at intercept was approximately 18 cm higher in gameplay with 3DTV compared to HMD for IH2-4 (Multimedia Appendix 3). There was no effect of Display Type for vertical hand position for IH1-2. Finally, there was a significant effect of Level, but only for the hand position along the vertical axis ($F_{1\text{-}15}$=11.16, P=.001).

Figure 4. Effects of 3D television (3DTV) versus a head-mounted display (HMD) on hand position at target intercept for each impact height (IH) along the anterior-posterior (AP) axis, along the medial-lateral (ML) axis, and along the vertical axis.



Joint Excursions

There were significant interactions of Display Type by IH for on joint excursions of the ankle ($F_{4,12}$=7.43, P=.003), knee ($F_{4,12}$=19.00, P<.001), hip ($F_{4,12}$=8.45, P=.002), spine ($F_{1,15}$=5.26, P=.011), shoulder ($F_{4,12}$=5.76, P=.001), and elbow ($F_{4,12}$=9.95, P<.001). As shown in Figure 5, this interaction is driven primarily by the fact that Display Type had no effect on joint excursions for launched virtual balls to IH0 (ie, balls that required the participant to duck). For the ankle, knee, hip, spine, and shoulder, the joint excursions used to intercept the launched virtual balls to locations IH1-IH4 was significantly greater during gameplay using the HMD compared to gameplay with the 3DTV (Multimedia Appendix 3). The sum of these effects on the apportionment of joint excursions (IH1-IH4) in these full-body reaching tasks is depicted in Figure 6, which is derived from the mean join excursions across target heights, mean participant height (70”), and estimating limb segment lengths from Winter [29].

Figure 5. Interaction of 3D television (3DTV) versus head-mounted display (HMD) by impact height (IH) on the joint excursions of the ankle, knee, hip, spine, shoulder, and elbow.

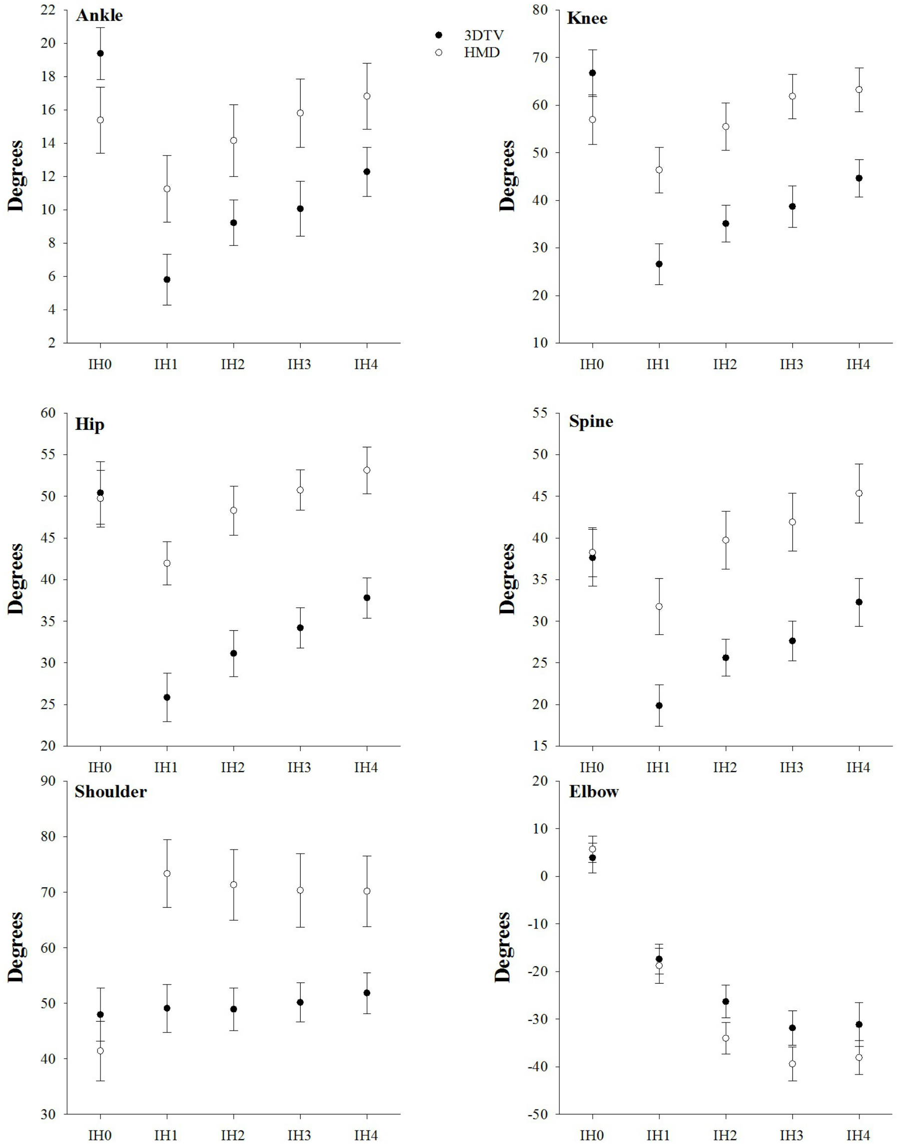
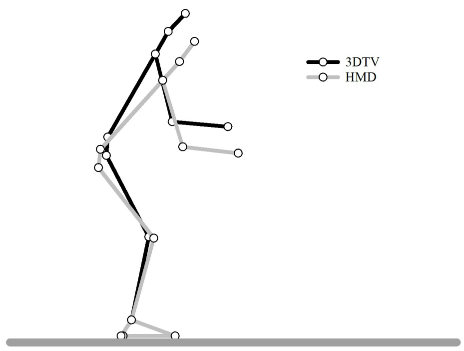


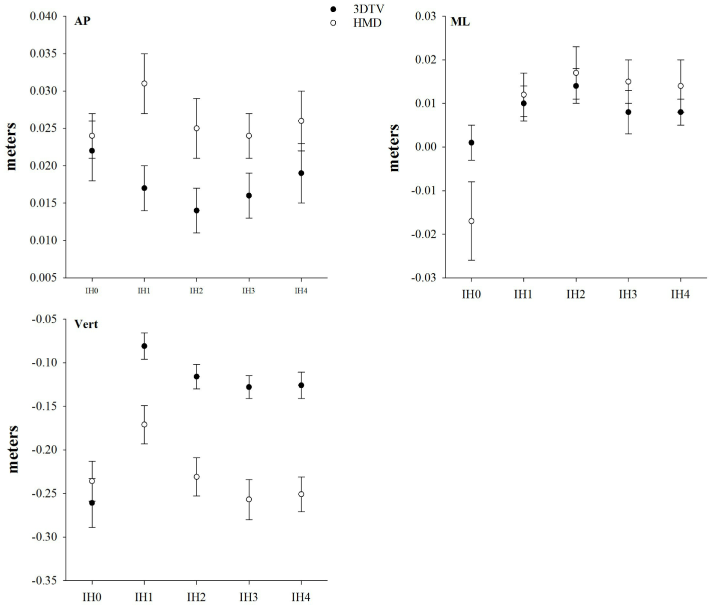
Figure 6. Effects of 3D television (3DTV) versus head-mounted display (HMD) on the posture adopted at target intercept.



Displacement of Center of Mass

There was a significant interaction of Display Type and IH on displacement of COM along the AP axis ($F_{4,12}$=5.63, P=.001), ML axis ($F_{4,12}$=6.03, P=.001), and vertical axis ($F_{4,12}$=9.95, P=.001). As shown in Figure 7, with the exception of IH0, COM displacement along the AP and vertical axes was greater in HMD compared to 3DTV for launched virtual balls launched to IH1-4 (Multimedia Appendix 3). Conversely, the only significant effects of Display Type on COM displacement along the ML axis was for launched virtual balls to IH0. Further, the differences along the ML axis were rather small (ie, 2 cm). There was also a main effect of Display Type on displacement of COM along the AP axis ($F_{1,15}$=12.64, P=.003) and vertical axis ($F_{1,15}$=41.82, P<.001). On average, participants had an 8 cm (SD 0.003) larger forward displacement of the COM along AP axis and an 8.6 cm (SD 0.017) larger downward displacement along the vertical axis during gameplay with the HMD compared to the 3DTV.

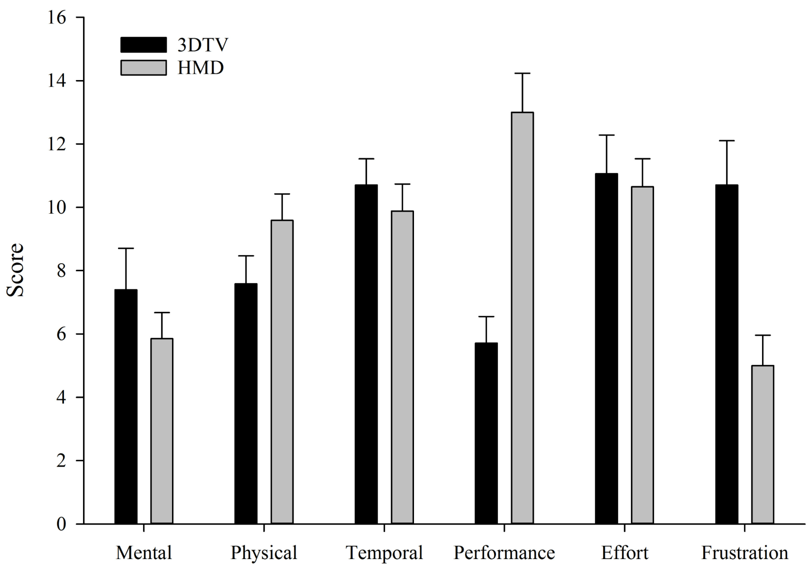
Figure 7. Effects of 3D television (3DTV) versus a head-mounted display (HMD) on displacement of whole-body center-of-mass (COM) for each impact height (IH) along the anterior-posterior (AP) axis, along the medial-lateral (ML) axis, and along the vertical axis.



Task Load Index

Analysis of the Task Load Index data (Figure 8) revealed that, relative to the 3DTV, participants indicated that HMD gameplay resulted in greater satisfaction with overall performance (P<.001) and was less frustrating (P=.001). There were no significant differences noted for physical demand, mental demand, temporal demand (ie, perceived time pressure), or overall effort required. Examination of the effects of display type on success rate during gameplay revealed that participants had a success rate of 38.8% (SD 2.0) for gameplay with the 3DTV compared to 71.2% (SD 2.6) for the HMD ($F_{1,15}$=142.4, P<.001).

Figure 8. Effects of 3D television (3DTV) versus head-mounted display (HMD) on NASA Task Load Index(TLX) scores.



Discussion

Principal Results

The primary goal of this study was to determine the effects of display type on the joint excursions used while playing Virtual Dodgeball. Figure 4 best captures how display type influences motor behavior during gameplay. When playing Virtual Dodgeball with the HMD, participants had larger excursions of the postural joints compared to gameplay on the 3DTV. This appears to reflect an overall shift in how participants respond to the launched virtual balls when using the HMD such that they intercept the virtual balls further in front of their body and from a lower position.

For virtual balls launched to IH1-4, hand position at target intercept contact was about 14 cm forward (AP axis) and 18 cm lower (vertical axis) in gameplay with HMD compared to 3DTV (Figure 2). It has been suggested that distance is underestimated in virtual reality space due, in part, to a restricted FOV [9]. However, the FOV for the 3DTV (horizontal=50°, vertical=40°) is considerably less than the FOV for the HMD (horizontal=100°, vertical=100°). If these results were driven by FOV, then one would expect target intercept with HMD to be less than 3DTV; however, that is clearly not what we found. Thus, FOV does not provide a plausible explanation for the differences in hand position at target intercept between the visual displays. This finding really indicates a difference in strategy, as it is a robust finding across the various IHs of the virtual launched balls. The exception is, of course, for the virtual balls launched to IH0 (ie, required the participant to duck). There was no difference in movement strategies as a function of display type for this portion of Virtual Dodgeball.

While differences in FOV do not provide an explanation for the differences in hand position or joint excursions, perhaps COM displacement can explain the changes in strategy between display types. The vertical displacement for the COM was greater in HMD compared to 3DTV across IH1-4. From an energetics perspective, lowering the height of the COM could result in a fundamentally more stable system. However, the forward displacement of the whole-body COM was also greater in the HMD condition compared to 3DTV. Thus, from the same energetics perspective, the greater forward displacement would not lead to a more stable system. It has been shown that COM displacement is changed in standing reaching tasks performed in virtual reality environments when the viewing angle of the participant's avatar is altered [7]. The change in viewing angle of the avatar could have similar effects on the perception of the task and the evaluation of a participant's location in the virtual environment [13]. Further, while Levin and colleagues found no differences in joint excursions of the trunk, shoulder, and elbow in reaching tasks performed in two visual display types (ie, 3DTV versus HMD) [2], they used a seated reaching task that did not present the same challenges to stability as intercepting dynamic targets performed from a standing position. It does not appear that conservation of displacement of COM is a significant factor in the difference in movement strategies observed in Virtual Dodgeball played with two display types.

It is possible that differences in joint excursions between the visual displays could be driven, in part, by differences in movement time to intercept the virtual launched balls. We have shown that joint excursions of the ankle, knee, and hip increase as movement time to target is reduced by half (ie, when participants move twice as fast to the target) [23,25]. However, we found that movement times were shorter during Virtual Dodgeball played with the 3DTV compared to the HMD (ie, faster movement speeds). Thus, one could expect that joint excursions would be greater in 3DTV compared to HMD. Because we observed just the opposite, movement speed does not appear to explain the differences in joint excursions observed between the visual displays.

Another potential contributor to the differences in observed movement strategies is the difference in refresh rates for the two display types. However, the kinematic input streams to the avatars in both display types is 100 Hz with the 3DTV, paired with the shutter goggles, having an effective update rate of 60 Hz/eye and the HMD having an update rate of 75 Hz. Thus the difference in refresh rates results in an absolute time difference of about 3.3 ms. Further, as both display types use an LED display, there should be no differences in persistence of the displayed images. Finally, according to Ware, the processing time for humans is approximately 166 ms and visual lags effect performance at about 200 ms [30]. Thus, it is highly unlikely that difference in refresh rates was a driving factor of the difference in joint excursions reported.

The difference in movement strategies observed between the two display types could be due to the presentation of the avatar. Some investigations have reported that the sense of actual presence in VR was weakened when the avatar was viewed from a third-person perspective [13,14]. Although differences in sense of presence could influence motor behavior, a recent study found no differences in temporal or spatial performance in a task that required participants to search and walk toward targets in the VR environment [31]. However, Ustinova and colleagues reported that trunk and peripheral joint excursions changed as a function of viewing angle of the avatar in a VR presented on a 3DTV [7], but in that study the avatar was always presented in the third-person perspective. As such, it was not a specific comparison between first- and third-person perspectives as in the current study. In fact, we are unaware of any studies that have examined the effects of avatar perspective on joint excursions. As noted in the methods, Virtual Dodgeball gameplay in 3DTV presented the participant's avatar in the third-person perspective whereas a first-person perspective of the avatar was used in HMD. The visual transformation that must occur while viewing one's avatar, which in the virtual world was located 1.5 meters in front of the participant, can have significant effects on movement control. However, it is possible that display type was the driving factor of the observed differences in joint excursions and not avatar perspective.

Finally, the results of this study provide further support for Virtual Dodgeball as an effective strategy to promote lumbar flexion. Importantly, the current findings also indicate that the clinical utility of Virtual Dodgeball may be enhanced with an HMD because it elicits more lumbar spine flexion, greater participant satisfaction with overall performance, and less frustration.

Limitations

A limitation of this study is that it cannot assign differences in motor performance in these tasks simply to avatar perspective. The differences could also be due to the use of a 3DTV versus the HMD, or to differences in display of the avatar (ie, first-person versus third-person perspective). Accordingly, future studies are needed to carefully isolate the effects of perspective and display type.

Conclusions

The results of this study demonstrate that visual display type influences motor behavior in Virtual Dodgeball. These data are important for the development of virtual reality assessment and treatment tools that are becoming increasingly practical for home and clinic use. Because a primary goal of virtual reality within rehabilitation is often to restore movement following orthopedic or neurologic injury, it is important to understand how presentation of the avatar or, by extension, camera position will affect motor behavior regardless of the display through which it is presented (ie, 3DTV or HMD). Use of home devices such as the Kinect sensor to track and presents an avatar in a third-person perspective may result in very different motor behavior when compared to the same tasks being presented from a first-person perspective.

Acknowledgments

CRF and JST obtained funding for the research; contributed to developing the intervention and study protocols; reviewed, edited, and approved the final version of the paper; and accept full responsibility regarding the conduct of the study and integrity of the data. JST, CRF, MEA, STL, PEP, and SW have read and approved the final manuscript. This study was funded, in part, by the National Institute of Arthritis and Musculoskeletal and Skin Diseases of the National Institutes of Health under Award Number R21AR064430. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Video files.

[PDF File (Adobe PDF File), 13KB - games_v4i2e16_app1.pdf]

Multimedia Appendix 2

The NASA-TLX survey instrument.

[PNG File, 44KB - games_v4i2e16_app2.png]

Multimedia Appendix 3

Supplementary tables.

[PDF File (Adobe PDF File), 14KB - games_v4i2e16_app3.pdf]

References

1. Fung J, Richards CL, Malouin F, McFadyen BJ, Lamontagne A. A treadmill and motion coupled virtual reality system for gait training post-stroke. Cyberpsychol Behav 2006 Apr;9(2):157-162. [doi: 10.1089/cpb.2006.9.157] [Medline: 16640470]
2. Magdalon EC, Michaelsen SM, Quevedo AA, Levin MF. Comparison of grasping movements made by healthy subjects in a 3-dimensional immersive virtual versus physical environment. Acta Psychol (Amst) 2011 Sep;138(1):126-134. [doi: 10.1016/j.actpsy.2011.05.015] [Medline: 21684505]
3. Levin M, Magdalon E, Michaelsen S, Quevedo A. Comparison of Reaching and Grasping Kinematics in Patients with Hemiparesis and in Healthy Controls in Virtual and Physical Environments. 2008 Presented at: Virtual Rehabilitation Conference; September 2008; Vancouver, Canada p. 60. [doi: 10.1109/ICVR.2008.4625123]
4. Subramanian S, Knaut LA, Beaudoin C, McFadyen B, Feldman A, Levin A. Virtual reality environments for post-stroke arm rehabilitation. J Neuroengineering Rehabil 2007 Jun 22;4(20):e31.
5. Subramanian S, Knaut L, Beaudoin C, Levin M. Enhanced feedback during training in virtual versus real world environments. Virtual Rehabilitation 2007:-. [doi: 10.1109/ICVR.2007.4362121]
6. Subramanian SM. Viewing medium affects arm motor performance in 3D virtual environments. J Neuroeng Rehabil 2011;8:1-9 [FREE Full text] [doi: 10.1186/1743-0003-8-36]

7. Ustinova KI, Perkins J, Szostakowski L, Tamkei LS, Leonard WA. Effect of viewing angle on arm reaching while standing in a virtual environment: potential for virtual rehabilitation. Acta Psychol (Amst) 2010 Feb;133(2):180-190. [doi: 10.1016/j.actpsy.2009.11.006] [Medline: 20021998]
8. Knaut LA, Subramanian SK, McFadyen BJ, Bourbonnais D, Levin MF. Kinematics of pointing movements made in a virtual versus a physical 3-dimensional environment in healthy and stroke subjects. Arch Phys Med Rehabil 2009 May;90(5):793-802. [doi: 10.1016/j.apmr.2008.10.030] [Medline: 19406299]
9. Willemsen P, Colton M, Creem-Regehr S, Thompson W. The Effects of Head-Mounted Display Mechanical Properties and Field of View on Distance Judgments in Virtual Environments. Acm Transactions on Applied Perception 2009;6(2). [doi: 10.1145/1498700.1498702]
10. Creem-Regehr SH, Willemsen P, Gooch AA, Thompson WB. The influence of restricted viewing conditions on egocentric distance perception: implications for real and virtual indoor environments. Perception 2005;34(2):191-204. [Medline: 15832569]
11. Thompson W, Willemsen P, Gooch A, Creem-Regehr S, Loomis J, Beall A. Does the quality of the computer graphics matter when judging distances in visually immersive environments? Presence-Teleoperators and Virtual Environments 2004 Oct;13(5):560-571. [doi: 10.1162/1054746042545292]
12. Willemsen P, Gooch A, Thompson W, Creem-Regehr S. Effects of Stereo Viewing Conditions on Distance Perception in Virtual Environments. Presence: Teleoperators and Virtual Environments 2008 Feb;17(1):91-101. [doi: 10.1162/pres.17.1.91]
13. Lenggenhager B, Tadi T, Metzinger T, Blanke O. Video ergo sum: manipulating bodily self-consciousness. Science 2007 Aug 24;317(5841):1096-1099 [FREE Full text] [doi: 10.1126/science.1143439] [Medline: 17717189]
14. Slater M, Spanlang B, Sanchez-Vives MV, Blanke O. First person experience of body transfer in virtual reality. PLoS One 2010;5(5):e10564 [FREE Full text] [doi: 10.1371/journal.pone.0010564] [Medline: 20485681]
15. Thomas JS, France CR. Pain-related fear is associated with avoidance of spinal motion during recovery from low back pain. Spine (Phila Pa 1976) 2007 Jul 15;32(16):E460-E466. [doi: 10.1097/BRS.0b013e3180bc1f7b] [Medline: 17632385]
16. Thomas JS, France CR. The relationship between pain-related fear and lumbar flexion during natural recovery from low back pain. Eur Spine J 2008 Jan;17(1):97-103 [FREE Full text] [doi: 10.1007/s00586-007-0532-6] [Medline: 17972113]
17. Thomas JS, France CR, Lavender SA, Johnson MR. Effects of fear of movement on spine velocity and acceleration after recovery from low back pain. Spine (Phila Pa 1976) 2008 Mar 1;33(5):564-570. [doi: 10.1097/BRS.0b013e3181657f1a] [Medline: 18317203]
18. Thomas JS, France CR, Sha D, Vander WN, Moenter S, Swank K. The effect of chronic low back pain on trunk muscle activations in target reaching movements with various loads. Spine (Phila Pa 1976) 2007 Dec 15;32(26):E801-E808. [doi: 10.1097/BRS.0b013e31815d0003] [Medline: 18091474]
19. Thomas JS, France CR, Sha D, Wiele NV. The influence of pain-related fear on peak muscle activity and force generation during maximal isometric trunk exertions. Spine (Phila Pa 1976) 2008 May 15;33(11):E342-E348. [doi: 10.1097/BRS.0b013e3181719264] [Medline: 18469681]
20. Trost Z, France CR, Sullivan MJ, Thomas JS. Pain-related fear predicts reduced spinal motion following experimental back injury. Pain 2012 May;153(5):1015-1021. [doi: 10.1016/j.pain.2012.02.001] [Medline: 22377437]
21. Thomas JS, France CR, Applegate ME, Leitkam ST, Walkowski S. Feasibility and Safety of a Virtual Reality Dodgeball Intervention for Chronic Low Back Pain: A Randomized Clinical Trial. J Pain 2016 Sep (forthcoming). [doi: 10.1016/j.jpain.2016.08.011]
22. McGill SM, Cholewicki J, Peach JP. Methodological considerations for using inductive sensors (3SPACE ISOTRAK) to monitor 3-D orthopaedic joint motion. Clin Biomech (Bristol, Avon) 1997 Apr;12(3):190-194. [Medline: 11415693]
23. Thomas JS, Corcos DM, Hasan Z. Effect of movement speed on limb segment motions for reaching from a standing position. Exp Brain Res 2003 Feb;148(3):377-387. [doi: 10.1007/s00221-002-1287-2] [Medline: 12541148]
24. Thomas JS, Corcos DM, Hasan Z. Kinematic and kinetic constraints on arm, trunk, and leg segments in target-reaching movements. J Neurophysiol 2005 Jan;93(1):352-364 [FREE Full text] [doi: 10.1152/jn.00582.2004] [Medline: 15342717]
25. Thomas JS, Gibson GE. Coordination and timing of spine and hip joints during full body reaching tasks. Hum Mov Sci 2007;26(1):124-140. [doi: 10.1016/j.humov.2006.08.002] [Medline: 17055098]
26. Hart S. NASA-Task Load Index (NASA-TLX); 20 Years Later. 2006 Presented at: Human Factors and Ergonomics Society Annual Meeting; October 2006; Chicago, IL. [doi: 10.1177/154193120605000909]
27. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical recipes in FORTRAN: the art of scientific computing. Cambridge, UK: Cambridge University Press; 1993.
28. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007;39(2):a-91. [doi: 10.3758/BF03193146]
29. Winter DA. Biomechanics and motor control of human movement. Hoboken, NJ: Wiley; 2009.
30. Ware CR. Target Acquisition in Fish Tank VR - the Effects of Lag and Frame Rate. 1994 Presented at: Graphics Interface; 1994; Banff, Canada p. 1-7.
31. Török A, Nguyen TP, Kolozsvári O, Buchanan RJ, Nadasdy Z. Reference frames in virtual spatial navigation are viewpoint dependent. Front Hum Neurosci 2014;8:646 [FREE Full text] [doi: 10.3389/fnhum.2014.00646] [Medline: 25249956]

Abbreviations

3DTV: three-dimensional television
AP: anterior posterior
COM: center of mass
FOV: field of view
HMD: head-mounted display
ML: medial lateral
VR: virtual reality
TLX: task load index

Edited by G Eysenbach; submitted 09.08.16; peer-reviewed by S Coombes, R Umoren; comments to author 31.08.16; revised version received 02.09.16; accepted 03.09.16; published 15.09.16.

Please cite as:
Thomas JS, France CR, Applegate ME, Leitkam ST, Pidcoe PE, Walkowski S
Effects of Visual Display on Joint Excursions Used to Play Virtual Dodgeball
JMIR Serious Games 2016;4(2):e16
URL: http://games.jmir.org/2016/2/e16/
doi:10.2196/games.6476
PMID:27634561

Original Paper

A Study on the Validity of a Computer-Based Game to Assess Cognitive Processes, Reward Mechanisms, and Time Perception in Children Aged 4-8 Years

Janneke CAW Peijnenborgh[1], MSc; Petra PM Hurks[2], PhD; Albert P Aldenkamp[3,4], PhD; Erik D van der Spek[5], PhD; GWM Rauterberg[5], PhD; Johan SH Vles[6], PhD, MD; Jos GM Hendriksen[1], PhD

[1]Kempenhaeghe, Center for Neurological Learning Disabilities, Heeze, Netherlands
[2]Maastricht University, Faculty of Psychology and Neuroscience, Maastricht, Netherlands
[3]Kempenhaeghe, Department of Research and Development, Heeze, Netherlands
[4]Eindhoven University of Technology, Faculty of Electrical Engineering, Eindhoven, Netherlands
[5]Eindhoven University of Technology, Faculty of Industrial Design, Eindhoven, Netherlands
[6]Maastricht University Medical Center, Department of Neurology, Maastricht, Netherlands

Corresponding Author:
Janneke CAW Peijnenborgh, MSc
Center for Neurological Learning Disabilities
Kempenhaeghe
P.O. Box 61
Heeze, 5590 AB
Netherlands
Phone: 31 (0)40227923
Fax: 31 (0)402260426
Email: peijnenborghj@kempenhaeghe.nl

Abstract

Background: A computer-based game, named Timo’s Adventure, was developed to assess specific cognitive functions (eg, attention, planning, and working memory), time perception, and reward mechanisms in young school-aged children. The game consists of 6 mini-games embedded in a story line and includes fantasy elements to enhance motivation.

Objective: The aim of this study was to investigate the validity of Timo’s Adventure in normally developing children and in children with attention-deficit/hyperactivity disorder (ADHD).

Methods: A total of 96 normally developing children aged 4-8 years and 40 children with ADHD were assessed using the game. Clinical validity was investigated by examining the effects of age on performances within the normally developing children, as well as performance differences between the healthy controls and the ADHD group.

Results: Our analyses in the normally developing children showed developmental effects; that is, older children made fewer inhibition mistakes ($r=-.33$, $P=.001$), had faster (and therefore better) reaction times ($r=-.49$, $P<.001$), and were able to produce time intervals more accurately than younger children ($\rho=.35$, $P<.001$). Discriminant analysis showed that Timo’s Adventure was accurate in most classifications whether a child belonged to the ADHD group or the normally developing group: 78% (76/97) of the children were correctly classified as having ADHD or as being in the normally developing group. The classification results showed that 72% (41/57) children in the control group were correctly classified, and 88% (35/40) of the children in the ADHD group were correctly classified as having ADHD. Sensitivity (0.89) and specificity (0.69) of Timo’s Adventure were satisfying.

Conclusions: Computer-based games seem to be a valid tool to assess specific strengths and weaknesses in young children with ADHD.

(JMIR Serious Games 2016;4(2):e15) doi:10.2196/games.5997

KEYWORDS

experimental games; ADHD; children; neuropsychological test

Introduction

Assessment of children's cognitive strengths and weaknesses is an important focus of clinical child neuropsychological research and clinical care worldwide [1]. Cognitive abilities are quantified traditionally by use of, for example, paper-and-pencil performance tests and more recently with computer-assisted tools [2,3]. Performances on traditional cognitive tests are believed to be influenced significantly by noncognitive functions, such as motivation and perseverance [4]. Therefore, a lower score on, for instance, a working memory test might indicate a memory problem but also, for example, a decreased motivation. This overall performance score therefore only limitedly reflects the underlying "cause" in case of a decreased performance, which makes this score difficult to interpret. In order to test one's cognitive abilities more purely, it would be preferable to (1) optimize motivation in the test situation and (2) assess motivation in a separate test additionally—unfortunately, tests measuring motivation in children are relatively scarcely used in the clinics.

On the basis of the literature, it is known that introducing immediate (vs delayed) rewards and the adaptation of item difficulty levels to the child's abilities are likely to increase motivation in children and help the child to stay focused on the tasks that he or she needs to do [5]. Also, introducing a context (eg, by introducing cognitive tests in the context of a story) may increase motivation, although data on the effects of including a story line on motivation or engagement in games are so far inconclusive [6]. The introduction of a story line improves the child's feeling of being part of a gaming environment [7]. In addition, the use of intrinsic fantasy elements has been found to improve motivation to conduct a specific task [8,9].

The above-mentioned increase in popularity of computer-assisted assessment tools is partly caused by the fact that it is relatively easy to implement these immediate rewards, to adapt difficulty levels, to implement a story, and to use intrinsic fantasy elements. This leads to a situation in which the child does not have the feeling of being assessed but instead thinks that he or she is playing a game [10]. This is especially important in tests that are designed for children because it is known that children can behave differently when they know they are being studied (also known as the Hawthorne effect [11]). For this purpose, we developed "Timo's Adventure," a computer-based game that consists of 6 mini-games. These mini-games aim to assess different cognitive processes, for example, attention, planning, and working memory; delay aversion, as a measure of motivation; and time perception (see method section). The aim of this study was to examine the validity of Timo's Adventure in assessing strengths and weaknesses in the above-mentioned domains of young children, aged 4-8 years. Proof of validity was sought by studying group differences in performance on the mini-games. Two types of relevant group comparisons were made. First, the age of the child is believed to be a relevant variable and was therefore studied in relation to performances on the mini-games: we expected younger children (aged <6 years) to perform less well than older children on all mini-games (in line with studies on, eg, cognitive development [12] and time perception [13,14]). For this purpose, age-related differences in a group of normally developing children (N=96) were examined.

Next, the scores on the mini-games of normally developing children were compared with those of children with attention-deficit/hyperactivity disorder (ADHD; N=40). Attention-deficit/hyperactivity disorder is a developmental disorder that is associated with academic difficulties and social disadvantage [15]. According to the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition; *DSM-V*) [16], 2 main areas of impairment in children with ADHD exist: inattention (eg, difficulty in maintaining attention during a task or problems in dividing attention) and hyperactivity and impulsive behavior (eg, acting out before thinking about the consequences). Previous research has found that cognitive difficulties, more specifically in the domain of working memory and attention, occur in children with ADHD [17]. However, according to the triple-pathway model by Sonuga-Barke et al [18], not all children with ADHD have cognitive weaknesses. In this model 3 distinct patterns of ADHD deficits are distinguished. The first pathway is related to cognitive functions and is called the inhibitory-based executive dysfunction. This pathway views ADHD as a disorder of dysregulation of thought and action associated with diminished inhibitory control (ie, executive functions). In the second pathway, ADHD is explained as a motivational style associated with fundamental alterations in reward mechanisms. Children with ADHD are assumed to prefer small immediate rewards over large delayed rewards, which results in inattentive, overactive, and impulsive behaviors [19]. The third pathway states that deficits in time perception, for instance, deficits in distinguishing between two time intervals, producing time intervals, and estimating time, are another component of ADHD. Indeed, time perception deficits have been reported for children with ADHD [20]. However, the results are not consistent; that is, some authors report no ADHD-related deficits [21]. All 3 pathways are believed to have their own neural substrate [18]. Sonuga-Barke and colleagues [18] found in children with ADHD aged between 6 and 17 years that delay aversion, poor executive functions, and poor time perception are core, but unrelated and independent, characteristics of ADHD. A person with ADHD can have deficits in one of the pathways or a combination of pathways. Neuropsychological measurements (which are used to examine possible deficits in one of the pathways) usually focus primarily on just one of the pathways, whereas one can conclude from the model by Sonuga-Barke et al that it is necessary to gain information on possible deficits in all 3 pathways. Sonuga-Barke and colleagues used several distinctive computerized tasks to collect information about the 3 pathways. These were, however, not connected in a fantasy gaming environment or by a story line. In our game, a story line was included in order to immerse the player in an intrinsic fantasy and possibly improve the reliability of the diagnosis. To our knowledge no computerized diagnostic tools exist in which all 3 pathways are included in combination with all motivation-enhancing elements discussed above (including a story line or fantasy game elements), although some training tools with a story line exist (for instance, Braingame Brian [22]). In Timo's Adventure story line, distracting factors are included to measure real-life distraction. Previous research found that distractors in a computerized continuous performance test resulted in more distractibility in children with ADHD than in their healthy peers [23].

In summary, the aim of this study was to investigate the clinical validity of Timo's Adventure.

Methods

Participants

Normally Developing Children

Parents of all children enrolled in the first 4 grades of 4 Dutch elementary schools were informed by a letter about the study. Informed consent of 102 children was acquired. A total of 4 children were excluded because they were not native speakers and instructions in Timo's Adventure were in the Dutch language. In addition, 2 children were excluded because they had a *DSM-V* diagnosis. The final dataset consisted of 96 children (43 boys), age ranging from 4 to 8 years. An overview of characteristics for this group can be found in Table 1.

All children were tested individually in a private room at their school. Approval for testing this sample was given by the Ethical Review Committee of the Faculty of Psychology and Neuroscience of Maastricht University, the Netherlands.

Children With Attention-Deficit/Hyperactivity Disorder

Parents of patients of the outpatient clinic Center for Neurological Learning Disabilities were asked to participate in this study by their medical specialist. In parallel, parents of children enrolled in a special needs program for children with behavioral problems were informed by a letter about the study and asked to participate, via the children's school. Informed consent of 62 children with a diagnosis of ADHD was acquired. A total of 22 children did not meet the inclusion criteria because they had a comorbid *DSM-V* diagnosis (n=4), because they used medication for attentional problems and hyperactive behavior (stimulants, atomoxetine, tricyclic antidepressants, or clonidine; n=6), or because of a combination of these exclusion criteria (n=12). The final dataset of the ADHD group consisted of 40 children (30 boys), all with a diagnosis of ADHD according to *DSM-V*. These diagnoses were made on the basis of a protocol formulated by Goldman et al [24] that includes (1) extensive history taking, (2) cognitive testing, (3) general physical and neurological examination of the child, and (4) systematic assessment of ADHD characteristics by means of structured questions based on the most recent version of the *DSM* [16]. Age range of the clinical sample was 6-8 years. An overview of characteristics of the ADHD group can be found in Table 1.

Children enrolled in the special needs program for children with behavioral problems were tested individually in a private room at their school. The children who were patients of the Center for Neurological Learning Disabilities were seen for neuropsychological testing as part of clinical care.

Approval for testing this sample was given by the Medical Ethical Board of Kempenhaeghe.

Table 1. Characteristics of all participants.

Characteristics	Typically developing children	ADHD[a] group
Number of participants	96	40
Boys/girls	43/53	30/10
Age range, years	4-8	6-8
Mean age (SD), years	5.85 (1.33)	6.90 (0.74)
Verbal IQ (WPPSI-III-NL[b] Vocabulary), mean (SD)	94.74 (10.79)	87.92 (13.05)

[a]ADHD: attention-deficit/hyperactivity disorder.

[b]WPPSI-III-NL: Dutch version of the Wechsler Preschool and Primary Scale of Intelligence [25].

Materials

Computer-Based Game Timo's Adventure

Timo's Adventure is a single-player game. All tasks are embedded in a story line: the main character in Timo's Adventure is Timo, a friendly alien whose rocket has run out of fuel [26]. He asks the child to go on an adventure together to collect stars that can be used as fuel. To complete all tasks, it takes approximately 20 minutes. The game is categorized as a serious game: a game designed for specific purposes beyond entertainment [27]. The game has a first-person view to simulate the feeling of presence, to make the child feel like he or she is inside the game world.

Development of the game was divided into 3 stages: design, implementation, and evaluation, as based on the iterative software cyclic model and the spiral model [28-30]. Besides the designers of Eindhoven University of Technology, users were invited in the design process. Users in this case were children (who helped us by explaining what they would like and by making drawings) and psychologists from Kempenhaeghe (who participated in the development of the functionalities of the game, the visual graphics, and story line that needed to match the age of the children). In the implementation phase, the functionalities of the game were created (by engineers of Eindhoven University of Technology in close collaboration with the psychologists from Kempenhaeghe). During the evaluation phase (consisting of a user test with a paper prototype technique and an computer-based prototype), the game was played and evaluated by children and by psychologists who were not part of the development team. This feedback was used to improve the game and remove small bugs [31].

The computer-based game consists of 6 different tasks (ie, 6 mini-games), each of which measures different neurocognitive functions and thus gives information about possible deficits. The 6 tasks measure aspects of executive functions, time perception, and reward mechanisms to represent the 3 pathways of Sonuga-Barke's model. All tasks were developed because of the need to measure the specific function and modeled after the theoretical background of the triple-pathway model.

Pathway 1: Executive Functions

All irrelevant mouse clicks are measured and thus give information on impulsivity and hyperactive behavior of the child. Furthermore, the following mini-games are included to assess impairments in this pathway: Dressing up (planning), Sandwich (working memory), Monkey (inhibition), and Magic Land (simple reaction time).

The first task, the mini-game Dressing up (planning), is set in the bedroom of the child. Timo tells the child that he or she needs to get dressed before the adventure can start. Several garments are spread throughout the room, which can be selected by clicking on them (see Figure 1). After clicking, the garment moves toward the reflection of the child in the mirror and the child gets dressed. This task gives information on the planning and organization skills of the child. The order in which the child selects the clothes is assessed to see whether the child is capable of planning his or her actions in the right order. The child can get a total score of 2 points: 1 point for being completely dressed and 1 point for using an executable and correct manner.

The second task, the mini-game Sandwich (working memory), is set in the kitchen (see Figure 2). Timo tells the child that it is necessary to eat something before starting the adventure, and he shows the child pictures of the ingredients. The child needs to remember these ingredients and select them in the same order as presented by Timo. The first sandwich starts with 2 ingredients, adding up to 5 ingredients in the last sandwich. This task is a measurement of the capacity of the visual working memory. According to Craeynest [32], the capacity of the working memory develops from remembering 2 targets when the child is 2.5 years old, to 3 targets when 3 years old, and 5 targets when the child is 7 years old. Martinussen et al [33] found that the capacity of the working memory in children with ADHD is markedly lower. In this task, the child can get a total score of 5 points: one for each correct sandwich.

In the third task, the mini-game Monkey (inhibition), a monkey has thrown banana peels on the road (see Figure 3). To cross the road, the child needs to swipe the banana peels and clear the road. However, if the monkey sees the child swiping banana peels, it will undo the child's actions. The monkey is playing hide-and-seek and appears suddenly. The child needs to wait for the moment the monkey disappears. This is a go or no-go task that gives information on the response inhibition of the child: is the child capable of inhibiting his or her response until the monkey hides? Children with ADHD have deficits in this response inhibition and they can be inclined to react impulsively [19]. The number of failures (ie, when the monkey sees the action of the child) is the outcome variable of this task; the higher this score, the worse the inhibition skills.

In the mini-game Magic Land (simple reaction time), stars shoot upward from magic holes (see Figure 4). The child needs to collect these stars. If the child does not react within 2 seconds, the stars will disappear. The task ends after 50 stars shoot upward. Outcome variables in this task are the number of collected stars (with a maximum of 50) and average reaction time for collected stars. Slower and more variable reaction times have been found to be a characteristic of ADHD [34].

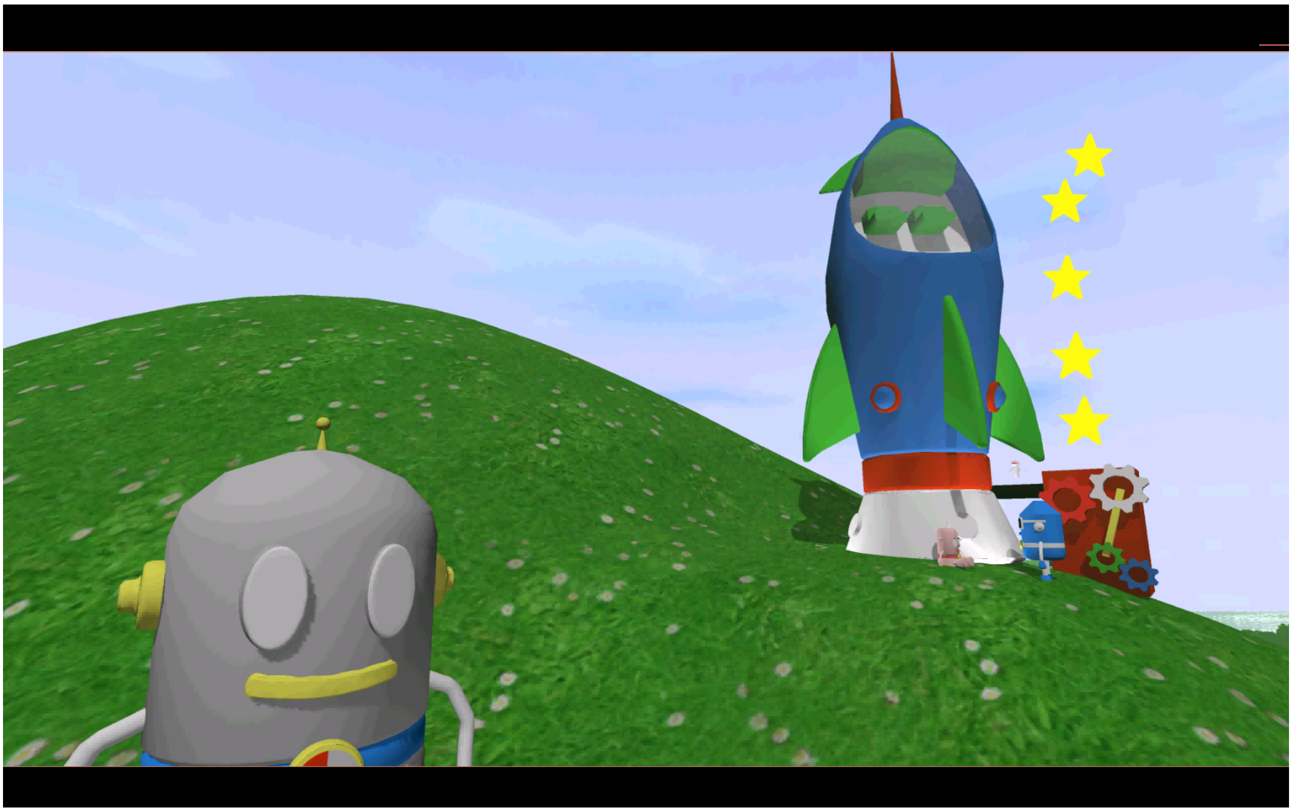
Figure 1. Screenshot of the mini-game Dressing up (planning).

Figure 2. Screenshot of the mini-game Sandwich (working memory).

Figure 3. Screenshot of the mini-game Monkey (inhibition).

Figure 4. Screenshot of the mini-game Magic Land (simple reaction time).

Figure 5. Screenshot of the mini-game Rocket (reward mechanisms).



Pathway 2: Reward Mechanisms

In the task mini-game Rocket (delay aversion; see Figure 5), the child gets a choice between an immediate but small reward (ending of the task) or a delayed but bigger reward (a flight in the rocket, after 2 minutes of waiting). The child can end the task at any moment. Impulsive behavior occurs when responding produces more immediate, relatively smaller rewards at the cost of delayed, larger rewards [35]. Outcome variables are whether the child chooses a small or big reward and how long (in seconds) the child waited.

Pathway 3: Time Perception

The task mini-game Balloon (time production) is set at a river with a broken bridge (see Figure 6). To cross the river, the child needs to inflate a balloon with the balloon machine by producing a time interval of 10 seconds. A produced interval between 9 and 11 seconds results in a perfect balloon. When the produced interval is smaller than 9 seconds the balloon falls into the water, and a produced interval larger than 11 seconds results in a balloon that flies away. The child can make a maximum of 3 perfect balloons, or the task will end after 3 minutes with a perfect balloon (regardless of what the interval is). Barkley et al [36] suggested that the estimation of temporal intervals is atypical in children with ADHD. The number of correct balloons is an outcome variable. Furthermore, the average production interval for the first 3 balloons is measured by subtracting 10 seconds from each of the first 3 balloons, transforming these scores to absolute scores, adding these scores, and then dividing them by 3. The higher this score, the less precise the mean produced intervals are.

Figure 6. Screenshot of the mini-game Balloon (time perception).

Vocabulary

This subscale of the Dutch version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III-NL [25]) was used to estimate verbal intelligence [37]. In this task, the child was asked to give definitions of words such as “umbrella” or “shoe.” The total score can be transformed to IQ scores, with 100 as the mean (SD 15).

Statistics

Data were analyzed with IBM SPSS Statistics version 21.0.0.0. All outliers (scores with z>3.29) within the concerned group (ie, normally developing children and the ADHD group) were replaced by the mean + 3 times its standard deviation as advised by Field [38]. Means and standard deviations of all variables were calculated.

Potential age-related differences on Timo’s Adventure within the sample of normally developing children were examined by conducting Pearson correlation (for scale outcome measurements; eg, number of irrelevant mouse clicks, time used to complete the task) and Spearman correlation analyses (for the ordinal outcome measurements; eg, correct or incorrect, did the child choose the large or small reward). Age was used as a continuous variable in these analyses.

Second, a discriminant analysis was performed to investigate to which level the variables of the game can discriminate between children belonging to the ADHD group and the normally developing children. Sensitivity and specificity were measured. All children aged <6 years in the sample of normally developing children were excluded from these analyses, in order to match with the age of the children in the ADHD group, resulting in a control group of 57 children. An independent samples t test revealed that children in the ADHD group and children in the normally developing group were equal in terms of Vocabulary scores (ADHD group: mean 87.92, SD 13.05; control group: mean 94.74, SD 10.79; t_{67}=1.92, P=.11), therefore it was not necessary to correct for IQ in further analyses.

Finally, it was analyzed on which variables children with ADHD scored significantly different scores from normally developing children. Again, all children aged <6 years in the sample of normally developing children were excluded from these analyses, in order to match with the age of the children in the ADHD group. Three types of analyses were used. For the continuous variables, general linear model univariate analyses were used. In four variables, the assumption of homogeneity in variances was violated; therefore, nonparametric t tests (Mann-Whitney) were used in these variables. In the categorical variables, Pearson chi-square analyses were performed.

Results

Descriptive Statistics of All Variables

Means and standard deviations for all variables of both groups are reported together with the analyses of group differences. All reported results are after correction for outliers.

Developmental Effects

An overview of all correlation analyses between age and variables of Timo’s Adventure can be found in Table 2.

Pathway 1: Executive Functions

Pearson correlation analyses with age as a continuous variable showed significant correlations on 2 tasks: in the inhibition task (Monkey), older children had significantly fewer inhibition failures (*r*=−.33, *P*=.001); and in the reaction time task (Magic Land), older children collected significantly more stars (*r*=.60, *P*<.001) and were faster in collecting these stars (*r*=−.49, *P*<.001).

Pathway 2: Reward Mechanisms

No significant correlations were found in this pathway, indicating that age does not influence reward mechanisms.

Pathway 3: Time Perception

In the time production task (Balloon), older children produced significantly more correct balloons (ρ= .35, *P*<.001) and had more precise time productions than younger children (*r*=−.25, *P*=.01).

Table 2. Correlation between results on Timo's Adventure and age for normally developing children (N=96).

Variables	Age[a]	*P* value
Pathway 1: executive functions		
Dressing up, total score	ρ=−.08	.46
Dressing up, clicks	*r*=−.05	.66
Sandwich, total score	ρ=.21	.05
Sandwich, clicks	*r*=.09	.71
Balloon, clicks	*r*=.14	.18
Monkey, failures	*r*=−.33	.001
Monkey, clicks	*r*=.08	.45
Magic Land, number of collected stars	*r*=.60	<.001
Magic Land, average time for collected stars	*r*=−.49	<.001
Magic Land, clicks	*r*=.16	.14
Pathway 2: reward mechanisms		
Rocket, reward	ρ=.05	.65
Rocket, time waited	*r*=.07	.50
Pathway 3: time perception		
Balloon, correct	ρ=.35	<.001
Balloon, average time for attempts	*r*=−.25	.01

[a]*r*: Pearson correlation; ρ: Spearman correlation.

Table 3. Structure matrix in discriminant analysis.

Output variable	Pooled within-group correlations between discriminating variables and standardized canonical discriminant functions
Magic Land, number of clicks	−.82
Magic Land, number of collected stars	.40
Balloon, number of clicks	.34
Monkey, number of failures	.31
Dressing up, total score	.29
Rocket, time waited	−.27
Balloon, number of correct balloons	−.18
Sandwich, total score	.17
Sandwich, number of clicks	.14
Magic Land, average reaction time for collected stars	.11
Rocket, small (=0) or large (=1) reward	−.10
Monkey, number of clicks	.09
Balloon, average time taken to inflate balloons	−.02
Dressing up, number of clicks	.01

Differences Between Children With ADHD and Healthy Controls

Discriminant Analysis

All variables were included in a discriminant analysis to investigate whether Timo's Adventure can discriminate between the children with ADHD and the healthy controls. A significant difference between the groups was found: Wilks Λ=.51, χ^2_{14}=50.8, *P*<.001. The structure matrix (see Table 3) revealed that especially the number of mouse clicks in several tasks and the mini-games on reaction time (Magic Land), inhibition (Monkey), and planning (Dressing up) were potential predictors. The classification results showed that 72% (41/57) children in the control group were correctly classified, and 88% (35/40) of the children in the ADHD group were correctly classified as having ADHD. Overall, 78% (76/97) of the children were correctly classified as being in the ADHD group or in the control group. Sensitivity of Timo's Adventure was 0.89 and specificity was 0.69.

Group Differences on Individual Variables

Because the combination of all variables was useful in discriminating between children with ADHD and normally developing children, the specific variables for which children with ADHD had a different result from normally developing children were examined. All significant differences between children with ADHD and healthy controls are reported.

Pathway 1: Executive Functions

Results of this pathway can be found in Table 4. There was a significant association between the group (ADHD or control) and the score on the Dressing up task (planning; χ^2_2=11.4, *P*=.003, *V*=.35), indicating that typically developing children had better scores than children with ADHD on a planning task. In the Magic Land task (simple reaction time), children with ADHD used more mouse clicks in collecting stars than children in the control group (*U*=389.50, *P*<.001, *r*=−.55).

Table 4. Results on Timo's Adventure for normally developing children in the control group (N=56) and children in the attention-deficit/hyperactivity disorder group (N=40) in pathway 1, executive functions.

Variable	Group	Mean	SD	Statistic	*P* value	Effect size
Dressing up, total score (0-2)						
	Controls	1.18	0.56	χ^2_2=11.41	.003	*V*[a]=.35
	ADHD[b]	0.93	0.81			
Dressing up, number of clicks (minimum for a satisfying result is 4 clicks)						
	Controls	17.59	14.32	*U*[c]=895.00	.09	*r*[d]=−.17
	ADHD	26.14	22.61			
Sandwich, total score (0-5)						
	Controls	0.89	1.10	$F^f_{1,95}$=0.47	.50	$_p^2$=.01[e]
	ADHD	1.05	0.97			
Sandwich, number of clicks (minimum for 5 correct sandwiches is 22 clicks)						
	Controls	87.49	51.12	*U*=823.00	.25	*r*=−.12
	ADHD	92.57	53.48			
Balloon, number of clicks (minimum for 3 correct balloons is 6)						
	Controls	108.52	93.91	*U*=945.50	.24	*r*=−.12
	ADHD	206.87	227.87			
Monkey, number of failures (minimum is 0)						
	Controls	0.77	1.33	$F_{1,96}$=2.97	.09	$_p^2$=.03
	ADHD	1.26	1.48			
Monkey, number of clicks (minimum to complete the task is 6)						
	Controls	29.72	2.20	$F_{1,96}$=0.82	.37	$_p^2$=.01
	ADHD	33.83	1.62			
Magic Land, number of collected stars (maximum is 50)						
	Controls	37.61	10.87	$F_{1,95}$=2.92	.09	$_p^2$=0.03
	ADHD	41.41	8.33			
Magic Land, average reaction time for collected stars						
	Controls	2.25	0.55	$F_{1,95}$=0.02	.88	$_p^2$=0.00
	ADHD	2.22	0.48			
Magic Land, number of clicks						
	Controls	76.27	19.52	*U*=389.50	<.001	*r*=−.55
	ADHD	155.65	89.68			

[a]*V*: Cramer's *V*.

[b]ADHD: attention-deficit/hyperactivity disorder.

[c]*U*: Mann-Whitney test.

[d]*r*: Pearson correlation coefficient.

[e] $_p^2$: partial variance explained.

Table 5. Results on Timo's Adventure for normally developing children in the control group (N=56) and children in the attention-deficit/hyperactivity disorder group (N=40) in pathway 2, reward mechanisms.

Variable	Group	Mean	SD	Statistic	*P* value	Effect size
Rocket, small (=0) or large (=1) reward						
	Controls	0.55	0.50	χ^2_1=7.3	.01	*V*[a]=.28
	ADHD[b]	0.28	0.46			
Rocket, time waited (minimum is 0 seconds, maximum is 120 seconds)						
	Controls	80.98	50.05	$F_{1,92}$=5.52	.02	$_p^2$=.06[d]
	ADHD	56.26	50.13			

[a]*V*: Cramer's *V*.

[b]ADHD: attention-deficit/hyperactivity disorder.

[c] $_p^2$: partial variance explained.

Pathway 2: Reward Mechanisms

Results of this pathway can be found in Table 5. In the Rocket task, a significant association between the group (ADHD or control) and whether or not a child chose the delayed reward was found (χ^2_1=7.3, *P*=.01, *V*=.28). Also, the total time that the child waited before he or she ended the task was significantly different: children in the control group were able to wait longer for the delayed reward than the children in the ADHD group ($F_{1,92}$ =5.52, *P*=.02, $_p^2$=.06).

Pathway 3: Time Perception

Results of this pathway can be found in Table 6. No significant differences between children with ADHD and normally developing children were found in the time production task.

Table 6. Results on Timo's Adventure for normally developing children in the control group (N=56) and children in the attention-deficit/hyperactivity disorder group (N=40) in pathway 3, time perception.

Variable	Group	Mean	SD	Statistic	*P* value	Effect size
Balloon, number of correct balloons (minimum is 0, maximum is 3)						
	Controls	2.25	1.09	$F_{1,96}$=0.07	.80	$_p^2$=.00[a]
	ADHD[b]	2.21	1.06			
Balloon, average time taken to inflate balloons						
	Controls	4.33	4.05	$F_{1,96}$=1.69	.20	$_p^2$=.02
	ADHD	3.38	2.01			

[b] $_p^2$: partial variance explained.

[c]ADHD: attention-deficit/hyperactivity disorder.

Discussion

Principal Findings

Recently, the development and use of computerized tasks in measuring, for example, neurocognitive abilities is increasing and results in promising effects in the field of interventions. For instance, children with ADHD benefit from game-based training tools on executive functions such as Braingame Brian and Cogmed, as was reported in a review by Peijnenborgh et al [39]. Important elements of these training tools are believed to be the use of fantasy, a story line, adaptation of the degree of difficulty, and the use of immediate rewards. However, most studies focus on training tools, whereas in our research a diagnostic tool was studied. This computerized diagnostic tool, named Timo's Adventure, was developed for young children (between 4 and 8 years old) to investigate the presence of the 3 distinct patterns of possible deficits in ADHD as described in the triple-pathway model by Sonuga-Barke and colleagues. To our knowledge, this is the first computerized tool in which all 3 pathways are assessed. The aim of this study was to investigate the clinical validity of Timo's Adventure.

The first proof of validity was found in the developmental effects of the game. In a population of 96 normally developing children between 4 and 8 years old, we found significant correlations with age in 2 of the 3 pathways. The older the child, the faster he or she is in completing the tasks. Furthermore, it was found that older children are better in inhibiting their response on a go or no-go task. This is in line with previous research (eg, [40,41]). Also, in a reaction time task, older children have better responses on alertness and have better reactions to visually presented stimuli after a visual warning signal. Again, this is in line with previous research (eg, [42]). Furthermore, age-related differences were found in the third pathway (ie, time perception), indicating that the older the child, the better he or she is in producing a predetermined time interval. This is in line with research by, for example, Friedman and Laycock [43] and Pouthas and Jacquet [13] stating that development of time perception skills increases sharply at an early age (ie, before the age of 7 years) and refines in the last grades of elementary school. Interestingly, no developmental effects were found in the second pathway (reward mechanisms), indicating that age does not influence the choice between immediate or delayed rewards. This might be caused by the fact that this aspect might be fully developed before the age of 4 years, as the findings by Mischel et al [44] suggest.

Further proof for clinical validity was found in satisfying results on the discriminant analysis, indicating that Timo's Adventure was correct in most classifications. Sensitivity and specificity of the measurement were satisfying. Our results are similar to, and sometimes even more promising than, other diagnostic measurements. For instance, Williams and colleagues [45] could classify 68% of the children with ADHD correctly when using IntegNeuro. However, only 2 of the 3 pathways are included in IntegNeuro, and it is not suitable for young children.

When looking more closely at the individual variables that help to differentiate between children with and without ADHD, we found several significant differences between both groups. In the first pathway (executive functions), we found that children with ADHD have significantly more irrelevant mouse clicks on the Magic Land task (reaction time) than healthy peers. This indicates impulsiveness, motor restlessness, and hyperactive behavior, as is also suggested by Hervey and colleagues [46]. Also, children with ADHD had lower planning skills than healthy controls had, what might be expected because planning and organization are affected in children with ADHD [47]. In several tasks we found results that were encouraging but not statistically significant (eg, children with ADHD tend to have more inhibition mistakes in the Monkey task and collect more stars in the Magic Land than typically developing children). This seems promising for the future: maybe, with some adjustments to the tasks, sensitivity and specificity can even increase.

Evidence was also found for differences between the ADHD group and the controls in the second pathway (reward mechanisms): children with ADHD chose the large but delayed reward less often than the control group and did not wait as long as the control group before deciding to end the task. This is in line with previous research, which states that children with ADHD have an aversion for delay [48]. No significant differences were found in the third pathway of Sonuga-Barke's model (time perception). Although timing deficits are known in children with ADHD, it is not uncommon that time production tasks do not result in significant effects [49]. Further research is necessary to gain more information on this aspect of possible ADHD-related deficits.

Finally, we found that user experiences were positive: when asked afterward, 81% of the children said they liked the game very much, and an extra 14% of the children said that they liked the game.

One limitation of this study is that information on reliability cannot be reported at this moment. Because Timo's Adventure consists of several independent functions, analysis of Cronbach alpha would automatically result in low consistency between the items. It would be interesting to test children several times, to collect data for test-retest reliability analyses. Further research is necessary to examine the reliability of this instrument. Another interesting question for future research might be to investigate possible effects of use of medication. In our analysis all children with ADHD who were taking medication (18 in total) were excluded, but it might be possible that medication influences only one (or a combination) of the pathways. Finally, it would be interesting to determine which (or combination of) tasks and corresponding outcome measurements are especially sensitive for the diagnosis of ADHD. It is possible that a total score and normative data can be measured, which can be used to determine a profile of ADHD symptoms. Future research is necessary to determine such a profile or total score.

Conclusions

This is the first time that all 3 pathways of Sonuga-Barke's model are included in one diagnostic computerized tool with a context of rewards and story line. In clinical care, diagnostic instruments on time perception and reward mechanisms are scarce, but it is necessary to gain information on these aspects to complete an analysis of strengths and weaknesses of the child. Proof for validity of Timo's Adventure was found in developmental effects and group differences between normally developing children and children with ADHD, and Timo's Adventure was satisfying accurately when classifying to which group (ie, the ADHD group or the healthy controls) the child belonged. This suggests that Timo's Adventure can be of added value in the diagnosis of ADHD because it helps in formulating a profile of strengths and weaknesses. Further research is necessary to confirm these findings and to examine potential effects of medication.

Acknowledgments

The authors are grateful for the collaboration with Pongpanote Gongsook, PhD, Christian Sallustro, PDEng, and Jun Hu, PhD, from the Eindhoven University of Technology (the Netherlands) in the development of Timo's Adventure.

Conflicts of Interest

None declared.

References

1. Swaab H, Bouma A, Hendriksen J, König C. Klinische Kinderneuropsychologie. Amsterdam: Boom; 2016.

2. Assef EC, Capovilla AG, Capovilla FC. Computerized stroop test to assess selective attention in children with attention deficit hyperactivity disorder. Span J Psychol 2007 May;10(1):33-40. [Medline: 17549876]
3. Messina F, Tiedemann KB, de Andrade ER, Primi R. Assessment of working memory in children with attention-deficit/hyperactivity disorder. J Atten Disord 2006 Aug;10(1):28-35. [doi: 10.1177/108705284299] [Medline: 16840590]
4. Luciana M, Nelson CA. Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated Battery: performance in 4- to 12-year-old children. Dev Neuropsychol 2002;22(3):595-624. [doi: 10.1207/S15326942DN2203_3] [Medline: 12661972]
5. Bioulac S, Arfi L, Bouvard MP. Attention deficit/hyperactivity disorder and video games: a comparative study of hyperactive and control children. Eur Psychiatry 2008 Mar;23(2):134-141. [doi: 10.1016/j.eurpsy.2007.11.002] [Medline: 18206354]
6. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 2013;105(2):249-265. [doi: 10.1037/a0031311]
7. Gorini A, Capideville CS, De LG, Mantovani F, Riva G. The role of immersion and narrative in mediated presence: the virtual hospital experience. Cyberpsychol Behav Soc Netw 2011 Mar;14(3):99-105. [doi: 10.1089/cyber.2010.0100] [Medline: 20649451]
8. Habgood MPJ, Ainsworth SE. Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. Journal of the Learning Sciences 2011 Apr 19;20(2):169-206. [doi: 10.1080/10508406.2010.508029]
9. Van der Spek ED, Sidorenkova T, Porskamp P, Rauterberg M. The effect of familiar and fantasy aesthetics on learning experience of serious games. In: Pisan Y, Sgouros NM, Marsh T, editors. Entertainment Computing – ICEC 2014. Berlin, Heidelberg: Springer; 2014:9783662452110-9783662452138.
10. Gongsook P, Peijnenborgh J, Sallustro C, Van der Spek E, Hu J, Bellotti F, et al. In: De Gloria A, editor. A diagnostic tool on time perception of children with ADHD. Paris: Springer; 2014:9783319121567-9783319121405.
11. McCarney R, Warner J, Iliffe S, van HR, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol 2007;7:30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
12. Diamond A. The early development of executive functions. In: Bialystok E, Craik FIM, editors. Lifespan Cognition - Mechanisms of Change. New York: Oxford University Press; 2006:9780195169539.
13. Pouthas V, Jacquet AY. A developmental study of timing behavior in 4 1/2- and 7-year-old children. J Exp Child Psychol 1987 Apr;43(2):282-299. [Medline: 3585248]
14. Friedman WJ, Lyon TD. Development of temporal-reconstructive abilities. Child Dev 2005;76(6):1202-1216. [doi: 10.1111/j.1467-8624.2005.00845.x] [Medline: 16274435]
15. Swanson J, Sergeant J, Taylor E, Sonuga-Barke E, Jensen P, Cantwell D. Attention-deficit hyperactivity disorder and hyperkinetic disorder. The Lancet 1998 Feb;351(9100):429-433. [doi: 10.1016/S0140-6736(97)11450-7] [Medline: 9482319]
16. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edition. Washington, DC: Author; 2013:A.
17. González-Garrido AA, Gómez-Velázquez FR, Zarabozo D, López-Elizalde R, Ontiveros A, Madera-Carrillo H, et al. Time reproduction disturbances in ADHD children: an ERP study. Int J Neurosci 2008 Jan;118(1):119-135. [doi: 10.1080/00207450601042177] [Medline: 18041610]
18. Sonuga-Barke E, Bitsakou P, Thompson M. Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2010 Apr;49(4):345-355. [Medline: 20410727]
19. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull 1997 Jan;121(1):65-94. [Medline: 9000892]
20. Hurks PP, Hendriksen JG. Retrospective and prospective time deficits in childhood ADHD: The effects of task modality, duration, and symptom dimensions. Child Neuropsychol 2011;17(1):34-50. [doi: 10.1080/09297049.2010.514403] [Medline: 20936546]
21. Bauermeister JJ, Barkley RA, Martinez JV, Cumba E, Ramirez RR, Reina G, et al. Time estimation and performance on reproduction tasks in subtypes of children with attention deficit hyperactivity disorder. J Clin Child Adolesc Psychol 2005 Mar;34(1):151-162. [doi: 10.1207/s15374424jccp3401_14] [Medline: 15677289]
22. Prins P, Ten Brink EL, Dovis S, Ponsioen AJGB, Geurts H, De Groot H, et al. Braingame Brian: Een Executieve Functie training met game-elementen voor kinderen met cognitieve controleproblemen. Amsterdam: Lucertis & Shosho; 2010.
23. Cassuto H, Ben-Simon A, Berger I. Using environmental distractors in the diagnosis of ADHD. Front Hum Neurosci 2013;7:805 [FREE Full text] [doi: 10.3389/fnhum.2013.00805] [Medline: 24319423]
24. Goldman LS. Diagnosis and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. JAMA 1998 Apr 08;279(14):1100. [doi: 10.1001/jama.279.14.1100]
25. Wechsler D. WPPSI-III-NL Nederlandse bewerking: Technische Handleiding 2nd edition. Amsterdam: Netherlands: Pearson Assessment and Information BV; 2010.
26. Sallustro C. Timo&Friends: A diagnostic tool on the time perception of children with ADHD. Eindhoven: University of Technology; 2013:A.

27. Bellotti F, Berta R, De Gloria A. Designing Effective Serious Games: Opportunities and Challenges for Research. Int. J. Emerg. Technol. Learn 2010 Nov 24;5(SI3). [doi: 10.3991/ijet.v5s3.1500]
28. Boehm BW. A spiral model of software development and enhancement. Computer 1988 May;21(5):61-72. [doi: 10.1109/2.59]
29. Rauterberg M, Strohm O, Kirsch C. Benefits of user-oriented software development based on an iterative cyclic process model for simultaneous engineering. International Journal of Industrial Ergonomics 1995 Oct;16(4-6):391-409. [doi: 10.1016/0169-8141(95)00021-8]
30. Rauterberg M. An iterative-cyclic software process model. 1992 Presented at: Proceedings Fourth International Conference on Software Engineering and Knowledge Engineering; 1992; Zurich, Switzerland p. 600-607. [doi: 10.1109/SEKE.1992.227899]
31. Gongsook P. Interactive diagnostic game for time perception. Eindhoven: Technische Universiteit; 2015:978-990.
32. Craeynest P. Psychologie van de levensloop. Leuven: Acco; 2010.
33. Martinussen R, Hayden J, Hogg-Johnson S, Tannock R. A meta-analysis of working memory impairments in children with attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry 2005 Apr;44(4):377-384. [doi: 10.1097/01.chi.0000153228.72591.73] [Medline: 15782085]
34. Gooch D, Snowling MJ, Hulme C. Reaction Time Variability in Children With ADHD Symptoms and/or Dyslexia. Developmental Neuropsychology 2012 Jul;37(5):453-472. [doi: 10.1080/87565641.2011.650809] [Medline: 22799763]
35. Grey I, Healy O, Leader G, Hayes D. Using a Time Timer to increase appropriate waiting behavior in a child with developmental disabilities. Res Dev Disabil 2009;30(2):359-366. [doi: 10.1016/j.ridd.2008.07.001] [Medline: 18926663]
36. Barkley RA, Koplowitz S, Anderson T, McMurray MB. Sense of time in children with ADHD: effects of duration, distraction, and stimulant medication. J Int Neuropsychol Soc 1997 Jul;3(4):359-369. [Medline: 9260445]
37. Alloway T. Can interactive working memory training improve learning? Journal of Interactive Learning Research 2002;23(3):197-207.
38. Field A. Discovering Statistics Using IBM SPSS Statistics 3rd edition. London: Sage Publications Ltd; 2009.
39. Peijnenborgh JC, Hurks PM, Aldenkamp AP, Vles JS, Hendriksen JG. Efficacy of working memory training in children and adolescents with learning disabilities: A review study and meta-analysis. Neuropsychol Rehabil 2015 Apr 17:1-28. [doi: 10.1080/09602011.2015.1026356] [Medline: 25886202]
40. Gathercole SE, Pickering SJ, Ambridge B, Wearing H. The structure of working memory from 4 to 15 years of age. Dev Psychol 2004 Mar;40(2):177-190. [doi: 10.1037/0012-1649.40.2.177] [Medline: 14979759]
41. Davidson MC, Amso D, Anderson LC, Diamond A. Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. Neuropsychologia 2006;44(11):2037-2078 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2006.02.006] [Medline: 16580701]
42. Thomas KM, Nelson CA. Serial reaction time learning in preschool- and school-age children. J Exp Child Psychol 2001 Aug;79(4):364-387. [doi: 10.1006/jecp.2000.2613] [Medline: 11511129]
43. Friedman WJ, Laycock F. Children's Analog and Digital Clock Knowledge. Child Development 1989 Apr;60(2):357. [doi: 10.2307/1130982]
44. Mischel W, Shoda Y, Rodriguez M. Delay of gratification in children. Science 1989 May 26;244(4907):933-938. [doi: 10.1126/science.2658056] [Medline: 2658056]
45. Williams LM, Hermens DF, Thein T, Clark CR, Cooper NJ, Clarke SD, et al. Using brain-based cognitive measures to support clinical decisions in ADHD. Pediatr Neurol 2010 Feb;42(2):118-126. [doi: 10.1016/j.pediatrneurol.2009.08.010] [Medline: 20117748]
46. Hervey AS, Epstein JN, Curry JF, Tonev S, Eugene AL, Keith CC, et al. Reaction time distribution analysis of neuropsychological performance in an ADHD sample. Child Neuropsychol 2006 Apr;12(2):125-140. [doi: 10.1080/09297040500499081] [Medline: 16754533]
47. Gau SS, Chiang H. Association between early attention-deficit/hyperactivity symptoms and current verbal and visuo-spatial short-term memory. Res Dev Disabil 2013 Jan;34(1):710-720. [doi: 10.1016/j.ridd.2012.10.005] [Medline: 23137723]
48. Bitsakou P, Psychogiou L, Thompson M, Sonuga-Barke EJ. Delay Aversion in Attention Deficit/Hyperactivity Disorder: an empirical investigation of the broader phenotype. Neuropsychologia 2009 Jan;47(2):446-456. [doi: 10.1016/j.neuropsychologia.2008.09.015] [Medline: 18929587]
49. Zakay D. The evasive art of subjective time measurement: Some methodological dilemmas. In: Block RA, editor. Cognitive models of psychological time. Hillsdale: Lawrence Erlbaum Associates; 1990:59-84.

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
DSM-V: Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)
WPPSI-III-NL: Wechsler Preschool and Primary Scale of Intelligence, Dutch version

Edited by M Alcañiz; submitted 23.05.16; peer-reviewed by JA Gil Gomez; comments to author 13.07.16; accepted 19.07.16; published 22.09.16.

Please cite as:
Peijnenborgh JCAW, Hurks PPM, Aldenkamp AP, van der Spek ED, Rauterberg GWM, Vles JSH, Hendriksen JGM
A Study on the Validity of a Computer-Based Game to Assess Cognitive Processes, Reward Mechanisms, and Time Perception in Children Aged 4-8 Years
JMIR Serious Games 2016;4(2):e15
URL: http://games.jmir.org/2016/2/e15/
doi:10.2196/games.5997
PMID:27658428

Original Paper

Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study

Abdulrahman Abdulla El-Hilly[1*], BSc (Hons); Sheeraz Syed Iqbal[1*], BSc (Hons); Maroof Ahmed[1*], BSc (Hons); Yusuf Sherwani[1*], BSc (Hons); Mohammed Muntasir[1*], BSc (Hons); Sarim Siddiqui[1*], BSc (Hons); Zaid Al-Fagih[1*], BSc (Hons); Omar Usmani[2], MBBS, PhD, FHEA, FRCP; Andreas B Eisingerich[3], BSc (Hons), MPhil, PhD

[1]Faculty of Medicine, Department of Medicine, Imperial College London, London, United Kingdom

[2]Airway Disease, National Heart and Lung Institute, Imperial College London & Royal Brompton Hospital, London, United Kingdom

[3]Imperial College Business School, Imperial College London, London, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Omar Usmani, MBBS, PhD, FHEA, FRCP
Airway Disease
National Heart and Lung Institute
Imperial College London & Royal Brompton Hospital
Dovehouse Street
London, SW3 6LY
United Kingdom
Phone: 44 (0)20 7351 8051
Fax: 44 (0)20 7351 8051
Email: o.usmani@imperial.ac.uk

Abstract

Background: Finding ways to increase and sustain engagement with mHealth interventions has become a challenge during application development. While gamification shows promise and has proven effective in many fields, critical questions remain concerning how to use gamification to modify health behavior.

Objective: The objective of this study is to investigate how the gamification of mHealth interventions leads to a change in health behavior, specifically with respect to smoking cessation.

Methods: We conducted a qualitative longitudinal study using a sample of 16 smokers divided into 2 cohorts (one used a gamified intervention and the other used a nongamified intervention). Each participant underwent 4 semistructured interviews over a period of 5 weeks. Semistructured interviews were also conducted with 4 experts in gamification, mHealth, and smoking cessation. Interviews were transcribed verbatim and thematic analysis undertaken.

Results: Results indicated perceived behavioral control and intrinsic motivation acted as positive drivers to game engagement and consequently positive health behavior. Importantly, external social influences exerted a negative effect. We identified 3 critical factors, whose presence was necessary for game engagement: purpose (explicit purpose known by the user), user alignment (congruency of game and user objectives), and functional utility (a well-designed game). We summarize these findings in a framework to guide the future development of gamified mHealth interventions.

Conclusions: Gamification holds the potential for a low-cost, highly effective mHealth solution that may replace or supplement the behavioral support component found in current smoking cessation programs. The framework reported here has been built on evidence specific to smoking cessation, however it can be adapted to health interventions in other disease categories. Future research is required to evaluate the generalizability and effectiveness of the framework, directly against current behavioral support therapy interventions in smoking cessation and beyond.

(JMIR Serious Games 2016;4(2):e18) doi:10.2196/games.5678

KEYWORDS

gamification; mhealth; mobile health; smoking cessation; health behavior; health policy; public health; behavioral support

Introduction

Smoking is responsible for 19% of all deaths in the United Kingdom, with a direct cost of £5.2 billion to the National Health Service (NHS) [1]. It is a leading cause of chronic disease [2], and has been declared as the most important cause of preventable morbidity and premature mortality worldwide [3]. However, depressingly, there remains a significant disparity between individuals desiring to quit smoking (~68%), and those actually successfully quitting (~3%) [4]. In 2013/14, the NHS Stop Smoking Services reached only 9% of individuals in the United Kingdom seeking to quit. Alarmingly, this represented a 19% year-on-year reduction [5,6]. As a result, the NHS 5-year forward view pledged ‘hard-hitting national action’ against preventable diseases including smoking, with a new set of smoking cessation services being outlined by Public Health England to promote healthier behavior [7].

A Cochrane review concluded that high-intensity behavioral support combined with pharmacological intervention was the most effective method for smoking cessation [8] with the National Institute for Health and Care Excellence demonstrating a 35% quit rate compared with a 2% background quit rate. However, this approach is expensive with lifetime costs of £7010 per person, where behavioral support contributes to the bulk of the cost [9]. The continuing economic and societal burden created by smoking suggests current smoking cessation techniques are underserving the population and begs the question: are there any other novel approaches we can take to tackle the addiction of smoking?

The rapid technological advancement of mobile phone technologies over the last decade has facilitated a burgeoning market for mHealth apps. However, while thousands of mHealth apps have been released, most have fallen short of their grand expectations owing largely to poor user engagement levels [10]. User engagement was identified as a critical factor to the success of mHealth in an analysis of 945 mHealth apps [11]. Thus, finding ways to increase user engagement with their target audience has become a significant focus of mHealth interventions [12].

Gamification is ‘the use of game design elements in nongame contexts’ [13], making use of the potential motivational ability of games. Gamification empowers users to complete tasks more efficiently, while making them more enjoyable, with the aim of increasing engagement [14]. Cugelman [15] argues that gamification is only effective when used in conjunction with academically grounded behavioral change strategies, and goes on to identify 7 “core ingredients” that can be used as persuasive strategies to promote behavioral change.

The application of gamification in mHealth is an emerging field. *Sparx*, a digital game intervention developed to treat clinical depression in adolescents, represents a successful implementation of gamification. A randomized controlled trial demonstrated noninferiority of the game against traditional face-to-face counseling, along with significantly higher remission rates [16]. From the perspective of health behavior, gamification has shown promising results in encouraging physical activity by turning the ‘work of exercise’ into a game [17]. A recent review published in JMIR Mental Health found no studies had been published explicitly examining the role of gamification features on program adherence with Web-based interventions to manage common mental health disorders [18].

Many health apps have attempted to replicate such success by promoting positive health behavior in a wider context, particularly in relation to smoking, albeit with variable success [12]. A systematic literature review found that the implementation of game elements helps to create motivational affordances that lead to desired psychological outcomes and the consequent behavioral outcome (Figure 1) [19]. However, critical questions remain concerning the mechanism by which gamification exerts its influence, with a particular paucity in research surrounding gamification in the context of health behavior.

While gamification shows promise and has proven effective in many fields, research is required to investigate the beneficial effects on health behavior and disease self-management to warrant the implementation of such interventions [20]. The aim of this study was to take a cognitivist approach, building on the evidence gathered from existing literature and our own data collection, to gain insight into the underlying thought processes and internal rules, which govern the way individuals react to a gamified smoking cessation intervention. Consequently, in this exploratory work, we aim to suggest how gamification might lead to a change in health behavior specifically with respect to smoking cessation.

Figure 1. How motivational affordances lead to behavioral outcomes [19].

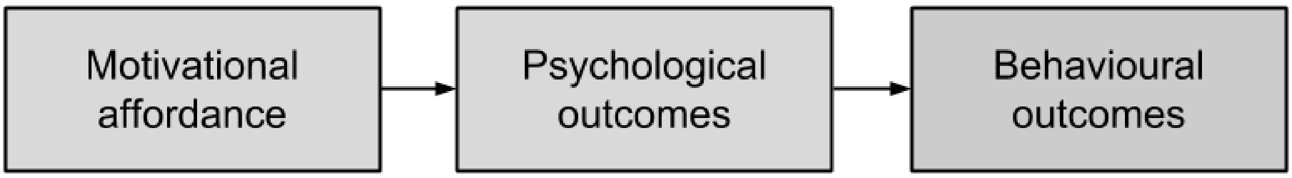


Table 1. Intervention comparison.

Game Components	Kwit 2	Puff Away
Store rating	Apple store - 4.5 /5	Google Play - 4.2 /5
Goal setting	Progress tracking Level system	Nonexistent
Capacity to overcome challenges	Information facilitating growth Learning and development	Nonexistent
Providing feedback on performance	Messages via achievement system	Nonexistent
Reinforcement	Rewards from levels and achievements Avoiding punishment associated with smoking	Nonexistent
Comparing progress	Nonexistent	Nonexistent
Social connectivity	Facebook Twitter	Nonexistent
Fun and playfulness	Minimal	Nonexistent

Methods

Study Approach

We conducted a qualitative longitudinal study with 16 smokers in 2 cohorts. The first cohort used a nongamified mHealth intervention free of any game components, while the second used a gamified mHealth intervention.

Interventions

In order to isolate the game-specific effects we would require 2 identical apps, differing only by the presence of gamified features. To approach this level of distinction, we shortlisted, downloaded, and tested 12 apps to establish which, gamification aside, were the most similar in app mechanics to allow fair comparison. ‘Puff Away’ and ‘Kwit 2’ were chosen because they both used very similar mechanisms to engage the user, focussing on tackling user education and providing progress tracking. The additional game components in Kwit 2 are specified in Table 1 [15].

Participants

All participants met the following 4 criteria: a smoker currently intent on quitting; 18+ years old; English speaker; and owner of a smartphone. We excluded those with smoking-related illnesses. Participants were recruited from local smoking support groups and university campuses in London. Each participant was then randomly allocated to a cohort and asked to install the relevant app onto their own smartphones. Participants were not compensated for their time. Informed verbal and written consent was obtained prior to commencing the study. The study had approval from the Imperial College Research Ethics Committee.

Interview Procedure

We conducted semistructured, one-on-one interviews with participants (30 minutes). Four interviews were conducted with each participant over the course of 5 weeks. The first interview (week 0) assessed their smoking background and demographics. This was accompanied by a standardized set of instructions on how to use their specific app. Subsequent interviews were then conducted at weeks 1, 3, and 5, with changes in participant behavior and emotions being tracked and recorded. Interviewers were instructed to neither encourage nor discourage the participant’s smoking behavior so as to minimize any effect on their behavior. Interview questions were formulated and then discussed with 2 independent, experienced qualitative researchers. The participants were asked about their progress in relation to smoking cessation, their experience using the app, the effect of the app on their behavior and emotions, as well as the specific effects of the game components. The final interview incorporated an exit interview in which participants expressed their overall experience. We conducted an internal pilot to test our methodology with 5 participants. Week 0 and week 1 interviews were conducted with each participant, allowing us to refine the interview questions and confirm the suitability of the 2 apps selected. The methodology employed with the pilot study was deemed satisfactory for the main study and so the results of all 5 pilot participants were included in the full longitudinal study. A semistructured interview guide can be found in the Web-based supplement in Multimedia Appendix 1.

Analysis Procedure

The 6-phase analytic framework was employed in our thematic analysis [21]. Audiotaped interviews were transcribed and read by 3 researchers on 2 separate occasions. These interview scripts were then used to manually generate codes for recurring patterns across participants. The codes were then analyzed to form overarching themes, all of which were defined by the 3 researchers. Ambiguities were resolved in discussion. This was a recursive process, whereby researchers cycled back and forth through the phases to allow for iteration as required. Once the saturation point had been reached and no new themes were emerging, the recruitment of new participants was stopped, conforming to the grounded theory approach [22]. All data regarding theme construction and interpretation was recorded in a reflexivity journal. Once the themes had been completed and defined, the researchers went back to the initial data sample to verify the accuracy of the overarching themes.

Expert Interviews

Semistructured interviews with 4 experts were conducted. The experts were initially shortlisted following our literature search, subsequently looking specifically at the research credentials of candidates. This shortlist was then narrowed to 4 based on their level of expertise within their respective fields of gamification, digital health, and smoking cessation. The experts were: Prof Scott Nicholson, Professor of Game Design and Development, and Director at Because Play Matters game lab, Wilfrid Laurier University; Prof Steven Johnson, Assistant Professor at Fox School of Business, Temple University; Dr Dominic King, Coauthor of the most cited editorial on Gamification and Health Behavior Change; and Dr Omar Usmani, Reader in Respiratory Medicine and Consultant Physician in Respiratory Medicine at the National Heart and Lung Institute, Imperial College London & Royal Brompton Hospital. A semistructured interview was conducted with each expert affording an in-depth multifaceted exploration of both gamification and smoking cessation. The transcripts from these interviews underwent the same manual thematic coding procedure as outlined for the longitudinal participant interviews.

Results

Participant Characteristics

Of the 19 participants initially recruited, 3 participants dropped out after the week 0 interview. Of the 3 dropouts, 2 were in the gamified cohort and 1 was in the nongamified cohort. The reasons given for dropout were refusal of further contact (2) and problems with availability (1). The resulting analysis is of the remaining 16 participants: 9 used the gamified intervention and 7 used the nongamified intervention. All 16 participants reported daily use of their smartphones. All participants expressed a desire to quit smoking prior to recruitment to the study, with 31% (5/16) of the participants attempting to quit for the first time. Additional characteristics are summarized in Table 2

Table 2. Participant characteristics.

Characteristic	Gamified Users (n=9) n (%)	Non-Gamified users (n=7) n (%)
Mean age	26.22 (range, 18-45)	28.1 (range, 20-52)
Cultural split		
South Asian	5 (56)	2 (29)
Arab	2 (22)	1 (14)
Caucasian/British	2 (22)	2 (29)
East Asian	-	2 (29)
Gender		
Female	3 (33)	2 (29)
Male	6 (67)	5 (71)
Occupation		
Student (undergraduate and postgraduate)	6 (67)	6 (86)
Working(part-time or full-time)	4 (44)	2 (29)

Figure 2. Analysis of results.

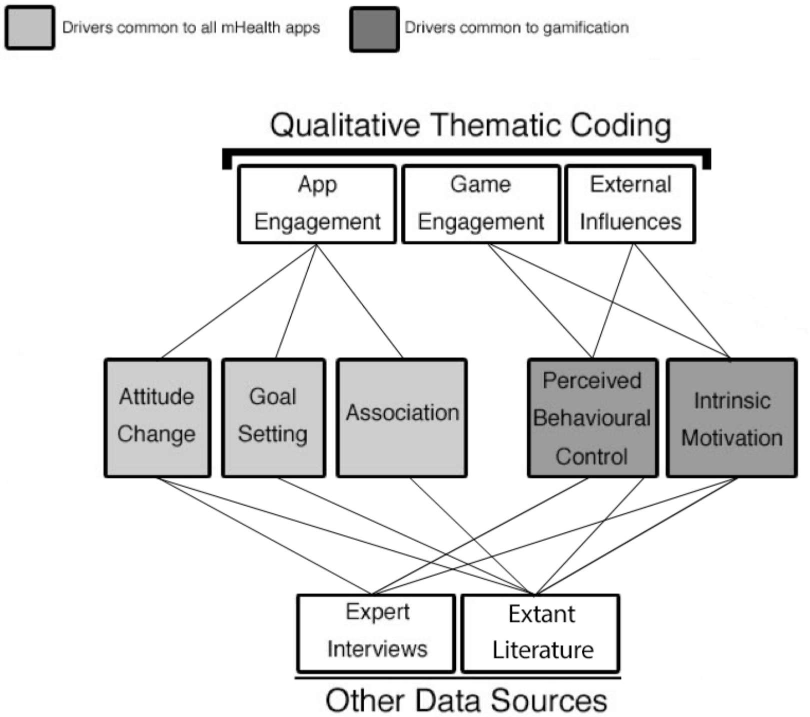
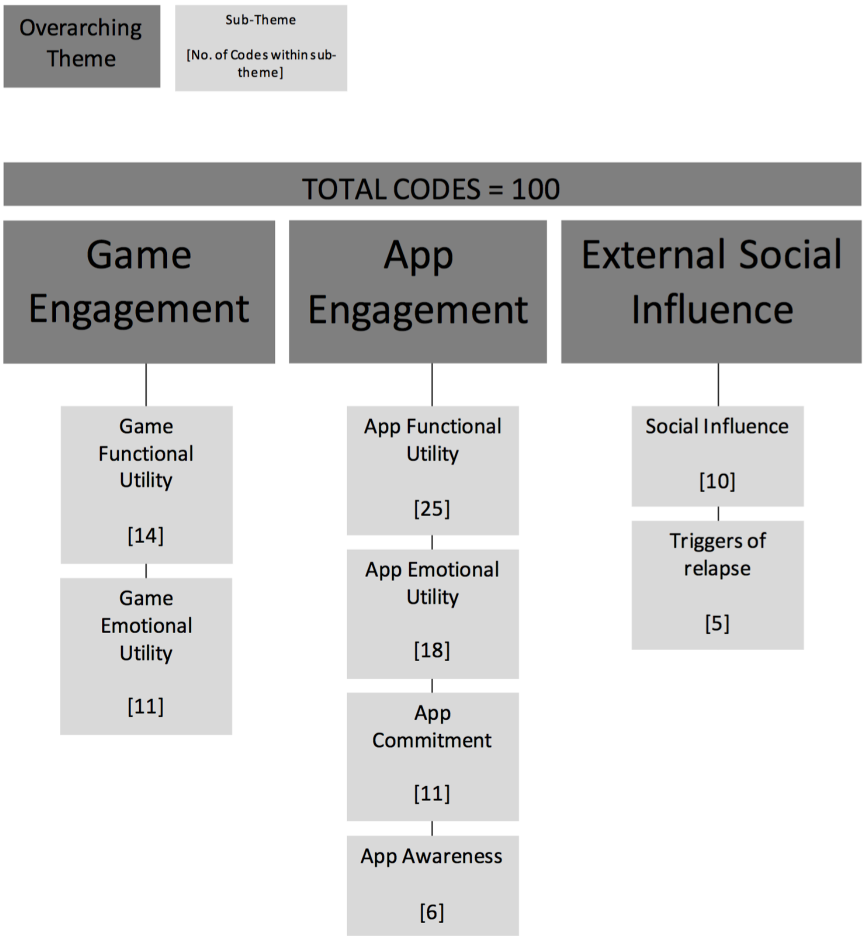


Figure 3. Thematic analysis.



Longitudinal Participant Interviews

A total of 57 interviews were conducted across a 5-week period: 16 interviews at week 0, 16 at week 1, 11 at week 3, and 14 at week 5. Transcripts and audio recordings were available for all the interviews.

Three overarching themes that influenced the impact of the app intervention on health behavior were identified from the longitudinal participant interviews only, with 8 subthemes (Figure 2): app engagement, game engagement, and external influences. App engagement refers to the components common to both gamified and nongamified interventions that helped to create engagement with the user. Game engagement refers to those game components unique to the gamified intervention that helped build engagement with the user. Finally, external influence describes the factors external to both interventions that impacted engagement with the app. The number of codes found in each subtheme can be found in Figure 3.

Expert Interviews

There was a consensus among the experts that technology has the potential to support health care professionals, in providing the behavioral support necessary in certain segments of the population. However, the experts questioned the long-term impact of a gamified intervention, and stated it would be a challenge to maintain long-term user commitment. Experts suggested that gamification can only reinforce desired behaviors, interventions should aim to build intrinsic motivation, rewards should be variable, and that a self-relevant experience is a critical success factor to building engagement with the game.

Analysis of Results

Following a thorough analysis of our findings, we identified drivers and modifiers to health behavior change. Drivers describe the key mechanisms by which behavioral change is produced. Modifiers were identified as those factors whose presence and quality determined the strength of the drivers, and therefore how likely the app was to promote positive health behavior.

We recognized 3 drivers common to the mHealth interventions: attitude change, goal setting, and association (of the mHealth solution with the maladaptive health behavior). Additionally, 2 drivers were proposed as the method by which gamification promotes positive health behavior: perceived behavioral control and intrinsic motivation. Figure 2 illustrates the relationship between the themes from the participant interviews, the other data sources and the drivers to behavioral change.

Drivers and Modifiers Common to the mHealth Interventions

Driver 1 - Attitude Change

A change in attitude toward the maladaptive health behavior described by 1 participant in the following statement:

> *It makes me reconsider if I really need to smoke or not and sometimes that extra few seconds is enough to put my cigarette away… It makes you contemplate and double think 'Do I really need a cigarette right now?'*

This was seen repeatedly with another participant explaining the following:

> *When someone is trying to quit it's like a battle in your head [between smoking and not smoking] … the app helps you in this battle [to not smoke].*

This finding can be explained by the Health Belief Model, which states that positive behavioral change can arise from increasing perceived threat of the negative health behavior and increasing perceived benefit of the positive health behavior [23]. Dr Usmani, with a background in smoking cessation, reinforced this by explaining how highlighting the implications of an individual's actions may result in behavior change:

> *First of all contextualizing the advantage of stopping smoking gives them a scare, a bit of a shock… this is what happens in real life terms of making people want to quit smoking.*

Driver 2 - Goal Setting

The progress tracking mechanism provided a simple visual means for participants to keep track of their efforts with participants appreciating the importance of such features:

> *It gives [me] nice visuals. Sometimes it's hard to visualize exactly what a cigarette means but the bars help you visualize it.*

Others felt the impact even more stating that:

> *it does help [motivate me] … especially with willpower.*

Participants also reported a sense of commitment and duty toward the goals of the app:

> *It feels like I've committed to this, so I am more motivated to try and make it work.*

With some this commitment was often expressed through some sense of guilt when they smoked:

> *When I smoked it's like cheating… you betray yourself when you press the button… Now when I think about it I feel horrendously guilty [when I smoked].*

This commitment was more common among users of the gamified intervention with only 1 participant expressing commitment in the nongamified cohort, but 4 expressing this feeling in the gamified cohort in week 1 interviews. PRIME theory states that in the context of smoking cessation, an intention or commitment is required before an individual can be motivated to change their behavior [24].

Driver 3 - Association

Participants began to associate the act of smoking with the mHealth interventions with a participant explicitly voicing this:

> *The app helped me create an associated between app and smoking.*

Another user went further to explain the change in his habit with the following:

> *My routine has changed now, when I get my cigarette out I automatically get my phone out as well now. The app has become integrated into my smoking.*

Textbox 1. The five modifiers of mHealth interventions.

Extant knowledge: the usefulness of information provided was inversely proportional to the user's existing knowledge. User's persistently commented on meaninglessness of repeated information:

All this information is out there already; I need something with more information [that is not known by me].

Ease of use: perceived simplicity increased engagement. The less convenient and simple the app was to use the lower the level of engagement of the user:

[It's] Impractical, when you reach for a cigarette the last thing you're thinking about is pulling out your phone and update the app… you've made your decision and you're over that dilemma.

Aesthetics: an attractive app design increased engagement with participants specifically mentioning unenticing visuals:

[It is] difficult to engage in the app due to poor presentation and poor visuals.

Initial motivation: the app was not able to change behavior if the user did not already possess the initial motivation to quit. This was a key point mentioned across all data sources as described in the following statement:

If I was on the path of trying to quit and desperately trying [to quit] then the game would help me, but due to not being in that mindset it did not [help me].

Physical distraction: the apps could act only as a supplementary tool in smoking cessation, as it was not able to address the physical side of the smoking habit. As is explained when one user said:

I chew some chewing gum…[it] just gives me something to do with my mouth.

This phenomenon reflects Pavlov's Theory of Classical Conditioning, whereby the instinctive, unconditioned stimulus (the urge to smoke) is paired with a new, conditioned stimulus (drawing for the app) [25]. However, it is important to note that merely forming an association between smoking and the app was in itself insufficient to change health behavior. In order to do so, it should be paired with intrinsic motivation as was emphasised by our expert interviews. Prof Nicholson stated:

> *If the app hasn't built up intrinsic motivation and the user hasn't found their own motivation to continue, then the health behavior will revert if there is no intrinsic motivation.*

Modifiers Common to the mHealth Applications

We identified 5 modifiers of the mHealth interventions as shown in Textbox 1.

Drivers and Modifiers of Gamified mHealth Interventions

Driver 1: Perceived Behavioral Control

Our data indicated that breaking down challenges of changing health behavior into smaller milestones, helped to increase the perceived behavioural control (PBC) of the individual by increasing their control beliefs, illustrated in the following:

> *If the end goal is just to quit smoking it makes it so hard, but if you have a game it enthuses the idea of something to work towards and it can steadily reward or punish you by setting short term goals…*

It has been further suggested that 'Flow' might be involved in shifting the locus of control from external to internal regulation, explaining how gamification might impact control beliefs and subsequently, PBC [26,27]. Achievements and rewards stimulated self-efficacy by providing a feedback mechanism, and thus a form of performance monitoring [28]. In this way, the conditional rewards would reinforce positive health behavior and in turn serve as a conditioned stimulus [29,30]; illustrated with the following users' statements:

> *The (achievements) felt good… achieving something… makes you feel like you can do it.*
>
> *It constantly reminds you, it's like going on a streak, you feel proud of yourself.*
>
> *The game (achievements) showed that I can resist sometimes and proved that I can resist.*

Participants exhibited anxiety at the prospect of going down a level if they were to smoke a concept known as loss aversion [31]:

> *It's so annoying when you go down a level, I want to go up not down. I didn't think much about gaining levels but I really did not want to lose levels.*

It is essential to balance loss aversion against the possibility of negatively impacting self-efficacy by going down a level, and consequently reducing PBC.

PBC was also impacted by external influences, namely local networks (family, friends, and near acquaintances). If they did not support the idea of smoking cessation, it decreased self-efficacy, and thus PBC.

> *He [my husband] actually thinks that it's possibly not the best time to do it [quit] because I have my exams coming in… so not go without any because… you're going back to cigarettes.*

This produced a negative effect as they discouraged the use of the intervention. Therefore, the local network had a profound effect in defining the level of perceived self-efficacy and their involvement should be minimized.

Driver 2: Intrinsic Motivation

Participants using the gamified intervention demonstrated greater levels of motivation and subsequent engagement than the nongamified cohort. Game elements such as rewards and level progression acted as motivational affordances leading to engagement. Participants in the nongamified cohort specifically mentioned game components with statements echoing the following quote: “If you put anything into a game it makes it more fun, and achieving something makes it more fun.”

Self-determination theory defines intrinsic motivation as ‘an activity one does because it is inherently enjoyable’ and extrinsic motivation as ‘doing something because it leads to a separable outcome’ [32]. The motivational effectiveness of extrinsic rewards will reduce over time; in the context of smoking this increases the likelihood of relapse as engagement decreases [33]. However, intrinsic motivation leads to increased frequency of behavior, and therefore increased engagement with the app [34].

Intrinsic motivation was also impacted by external influences. If local networks supported the idea of smoking cessation, participants were more likely to receive words of encouragement regarding their progress. Studies have shown positive feedback to be associated with smoking cessation, with the reverse also being true [35]. However, although words of encouragement from peers can lead to positive behavior change in the short term, the effects are unlikely to be lasting if the encouragement is not self-relevant. Reliance on extrinsic rewards, which create a sense of duty, should be avoided; especially if they are perceived as controlling:

> *The pressure does not help, you don’t want to be told what to do, you want to do it on your own merit and I want to quit when I want to. It feels very parental and ...having people shove their own ideas down my throat as if I am not aware of what I am doing is very patronizing.*

Any such rewards will undermine the game and impact negatively on intrinsic motivation, thereby compromising engagement [32,36].

Modifiers of Gamified mHealth Interventions

There are 7 modifiers that have determined the strength of the drivers specific to gamification: (1) personalization, (2) meaningful framing, (3) challenge-ability balance, (4) unpredictability, (5) user-centered design, (6) fun, and (7) social community.

Personalization

Participants cited that achievements of the gamified intervention lacked self-relevance *:* “Smoking is personal and should not have premade incentives, people should generate their own incentives and the app should empower [them].” The orientation of the individual affects how they will perceive extrinsic rewards; whether they perceive it as controlling (externally oriented), or informational (internally oriented) [32]. An element of personalization can tailor an intervention to the individual, and thus account for an externally oriented user.

Meaningful Framing

Framing a challenge in a meaningful way works synergistically with the gamified reward system to enhance intrinsic motivation [37].

Challenge-Ability Balance

Ensuring a dynamic balance between the participants’ perceived ability and the perceived challenge is a core tenet of flow [36].

Unpredictability

Participants exhibited tedium after using the gamified intervention for some time, which led to disengagement *:* “Achievements became slightly repetitive and need to be more creative.” However, integrating variable rewards, which are informational and unpredictable in nature has been found to increase focus and engagement [38]. In contrast, rewards that are contingent on engagement and performance alone should be minimized as far as possible, where do they not align with the individual’s intrinsic motivation, or they risk undermining it [27].

User-Centered Design

Ensuring that the user’s needs and goals are the primary consideration at every stage of the process [39].

Fun

A common request from participants in the nongamified cohort was to add an element of fun to the game: “If it was a game with milestones and achievements and levels it would definitely be really cool… if it was a game I would definitely do that.” Fun can be defined as a type of intrinsic motivation that may play an important role in achieving a state of flow [40]. It is important to note that ‘fun’ is the product of the relationship between an activity and an individual’s goals, rather than solely as a property of the activity itself.

Social Community

Users were unwilling to share their progress via Facebook and Twitter. With multiple participants sharing the sentiment in the following quote: “I actually think it [sharing on social media] is counterproductive. You do it for yourself, not other people.”

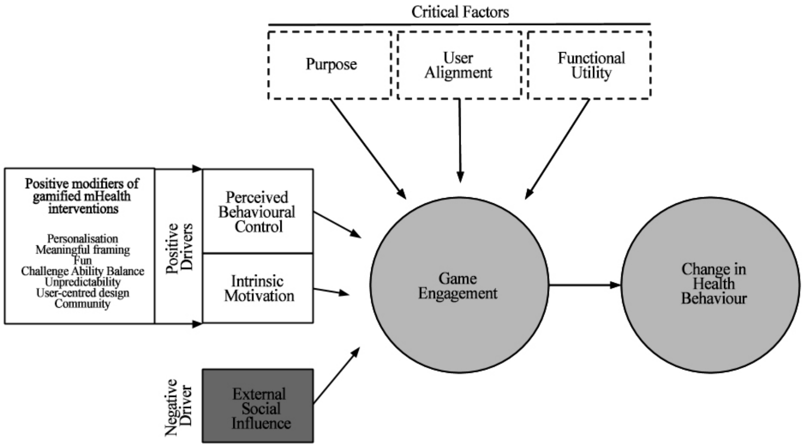
However, they expressed desire to interact with like-minded individuals, with whom they could better relate. Kwon et al [41] reinforced these observations when they found that the motivations for social networking, and the motivations for building up reputation were not mutually exclusive.

Discussion

Framework

In this study, we compared 2 apps, 1 gamified and the other nongamified, in a longitudinal qualitative study and then analyzed our findings in the context of expert opinions and the extant literature. We sought to establish how best to exploit gamification as an effective tool to build and maintain engagement of mHealth apps designed to promote smoking cessation. This work culminated in the development of a framework to isolate the drivers and factors that govern effective gamification (Figure 4). The framework suggests that a change in health behavior is dependent on the degree of engagement with the gamified intervention and that this was influenced by ‘critical factors’ and ‘drivers’ of game engagement.

Figure 4. How positive health behavior can be promoted by the gamification of mHealth apps.



Critical Factors

Critical factors were the 3 components that had to be present in order for users to engage with the game; absence of any one of these critical factors would lead to disengagement. A mHealth app looking to promote positive health behavior change needs a 'purpose' that is made explicit and clear to the user. However, this 'purpose' needs to align with the user's own personal objective ('user alignment'). This 'user alignment' is key to tapping into the user's intrinsic motivation, ensuring sustained engagement with the intervention as explained by both experts and users alike.

The final critical factor is 'functional utility' or the perceived ability of the intervention to fulfil the needs and solve the problems of the participant. This was the most frequently coded code during the thematic analysis, with 39% (39/100) of codes referring to at least 1 of game or app functional utility. We found that when users' perceived functional utility of the intervention was low, they became disengaged. To enhance functional utility, the intervention needs to be easy to use, designed around the user, and integrate a feedback mechanism to allow users to track their progress.

Drivers

We identified 'perceived behavioral control' and 'intrinsic motivation' as positive drivers, which when present, directly led to game engagement. In the context of health behavior, game engagement can be maximized by taking advantage of modifiers that boost self-efficacy and minimize control beliefs. We also observed intrinsic motivation to be a principal driver of game engagement, and should be maximized, by using the modifiers in the presence of the 3 critical factors.

The impact of positive drivers is determined by factors in the framework classed as modifiers. We identified 7 modifiers of gamified mHealth interventions: (1) personalisation: challenges and rewards that are self-relevant, (2) meaningful framing: link the challenge of changing health behavior to an overall self-relevant goal, (3) challenge-ability balance: a dynamic balance must exist between the perceived ability and perceived challenge, (4) unpredictability: unexpected rewards are perceived as least controlling types of rewards, (5) user-centered design: ensuring the user's needs and goals are constantly met, (6) fun: the experience must be innately enjoyable, and (7) social community: create a community of like-minded individuals, where posting accolades will boost an individual's reputation.

We also identified 'external social influence' as a negative driver, which should therefore be minimized to optimally promote positive health behavior. We observed that the presence of an external social influence negatively impacted self-efficacy and consequently decreased the individual's perceived behavioral control and intrinsic motivation. For emphasis, 'external social influence' has been depicted to directly impact game engagement, although it does this by impairing the user's PBC and intrinsic motivation.

Applicability of the Framework

Our aim is for our framework to be used as a guide for health care professionals and app developers in appraising whether a gamified app has the right ingredients to be successful in generating and promoting positive health behavior change. The successful implementation of gamified mHealth interventions will require a multidisciplinary approach, marrying input from clinicians, behavioral scientists, and game designers to build compelling apps [42]. As such, a further application of this framework is to provide a theoretical basis around which the multidisciplinary teams could collaborate.

Limitations

Limitations to our study mainly relate to the infancy of gamification as a field, meaning only a limited number of interventions were available to us. Although the gamified intervention was identified as one of the leading apps implementing game mechanics, it fell short in a number of areas leading to a drop in engagement over time. However, it is unclear whether this was due to shortcomings solely within the app, or gamification itself. In order to fully understand the effect of a gamified intervention, we would have benefited from a more optimal implementation of gamification as well as testing interventions employing a wider range of game elements reflecting the variety of gamification strategies employed by different health apps. Moreover, while we tried to choose 2 apps that employed the same intervention content, bar the presence of gamified features in one and the absence in the other, there may well be some confounding variables responsible for our results that we were unable to identify in our analysis. To combat this, a further study would be required involving the creation, from scratch, of 2 interventions offering the same educational content and differing only by the use of gamified features. In addition, we were unable to examine the long-term impact of gamification beyond the 5-week study period. A more representative analysis of the overall smoking demographic could have been conducted with the inclusion of the diseased population and a larger sample of smokers from different geographical and socioeconomic contexts. A larger sample may also help to better elucidate the scale of the findings we present in the framework, for instance the extent to which external social influences result in a truly negative effect and whether there are instances where they may bolster an individual's intrinsic motivation for example.

Policy Implications of Implanting Gamification in the National Health Service

Offering individuals a gamified mHealth intervention for smoking cessation could be the answer to the inability of current NHS smoking cessation services to serve the population, particularly for millennials who have grown up as 'digital natives' [43]. A gamified mHealth intervention would confer the benefits of evidence-based behavioral therapy, while transforming the expensive interface of patient-doctor consultations, to one between patients and an app. Furthermore, the intervention will always be close at hand to the user helping to provide support when high-risk situations arise.

Gamified mHealth interventions should not be used in isolation, but rather be considered as an additional tool in the delivery of health care. For example, implementation among older, less technologically competent patients will prove challenging, with certain patients still favoring human-human interaction. As such, it will be important to continue to offer conventional behavioral support alongside a new intervention to optimize the effectiveness of the service.

Conclusions

Gamification holds the potential for low-cost, highly effective mHealth solutions that may replace or supplement the behavioral support component found in current smoking cessation programs. Our proposed framework has been built on evidence specific to smoking cessation. We propose that it can also be extended to pave the way toward new methods of public health education, as our findings showed that gamification could be an effective modality for engaging people with the provision of information. However, questions still remain in relation to the long-term effects of gamification. Future research is required to evaluate the effectiveness of the above framework against current behavioral interventions in smoking cessation.

Acknowledgments

Dr Omar S Usmani is supported by the National Institute for Health Research (NIHR) Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London.

Authors' Contributions

MA, YS, MM, SS, AAE, SI, and ZA were all equally involved in designing the study, conducting the research, analysing the data, and writing the paper. ABE helped with the study design, and data analysis. OSU was involved in writing the manuscript.

Conflicts of Interest

Six of the authors (AAE, MA, YS, SSI, MM, and SS) have, following completion of this study in 2015, become shareholders in Digital Therapeutics Ltd; translating the framework from this work into a gamified app () delivering cognitive behavioral therapy for smoking cessation.

Multimedia Appendix 1

Semi-structured interview questions.

[PDF File (Adobe PDF File), 45KB - games_v4i2e18_app1.pdf]

Multimedia Appendix 2

Thematic coding summary with codes and illustrative quotes regarding the expert interviews.

[PDF File (Adobe PDF File), 34KB - games_v4i2e18_app2.pdf]

Multimedia Appendix 3

Thematic coding summary with codes and illustrative quotes regarding the Drivers for mHealth interventions. Quotes taken from both participant and expert interviews.

[PDF File (Adobe PDF File), 36KB - games_v4i2e18_app3.pdf]

Multimedia Appendix 4

Thematic coding summary with codes and illustrative quotes regarding drivers for gamified mHealth interventions.

[PDF File (Adobe PDF File), 33KB - games_v4i2e18_app4.pdf]

Multimedia Appendix 5

Participant concerns with illustrative quotes regarding gamified mHealth interventions.

[PDF File (Adobe PDF File), 31KB - games_v4i2e18_app5.pdf]

Multimedia Appendix 6

British Thoracic Society Conference presentation December 2015.

[PPTX File, 5MB - games_v4i2e18_app6.pptx]

References

1. Allender S, Balakrishnan R, Scarborough P, Webster P, Rayner M. The burden of smoking-related ill health in the UK. Tob Control 2009;18:262-267. [doi: 10.1136/tc.2008.026294] [Medline: 19509003]
2. Carter BD, Abnet CC, Feskanich D, Freedman ND, Hartge P, Lewis CE, et al. Smoking and mortality--beyond established causes. N Engl J Med 2015;372:631-640. [doi: 10.1056/NEJMsa1407211] [Medline: 25671255]
3. Centers for Disease Control and Prevention. Centers for Disease Control and Prevention. 2015. SmokingTobacco Use; Fact Sheet; Fast Facts - CDC URL: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/ [accessed 2016-02-20] [WebCite Cache ID 6fRL2xvQF]
4. Action on Smoking and Health. 2014. Stopping Smoking; Benefits and Aids to Quitting URL: http://ash.org.uk/files/documents/ASH_116.pdf [accessed 2016-02-20] [WebCite Cache ID 6fRLB6kbd]
5. Lifestyle Statistics Team H and SCIC. 2014. Statistics on Smoking URL: http://www.hscic.gov.uk/catalogue/PUB14988/smok-eng-2014-rep.pdf [accessed 2016-02-20] [WebCite Cache ID 6fRLK9ZX9]
6. Action on Smoking and Health. 2015. Smoking Statistics URL: http://www.ash.org.uk/files/documents/ASH_93.pdf [accessed 2016-02-20] [WebCite Cache ID 6fRLNIGPT]
7. NHS. National Health Service. 2014. Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2016-10-06] [WebCite Cache ID 6l3h8v8si]
8. Stead LF, Koilpillai P, Fanshawe TR, Lancaster T. Combined pharmacotherapy and behavioural interventions for smoking cessation. Cochrane Database Syst Rev 2012;10:CD008286. [doi: 10.1002/14651858.CD008286.pub2] [Medline: 23076944]
9. Flack S, Taylor M, Consultant S, Trueman P. Health Econ. 2007. Cost-Effectiveness of Interventions for Smoking Cessation Final Report URL: https://www.nice.org.uk/guidance/ph10/evidence/costeffectiveness-of-interventions-for-smoking-cessation-mass-media-interventions-369842077 [accessed 2016-02-20] [WebCite Cache ID 6fRLXGhhb]
10. Bacardit A. AT Kearney Insights. 2015. Is Mobile the Cure for the Rising Costs URL: https://www.atkearney.com/documents/10192/422267/EAXIV_1_Is_Mobile_a_Cure_for_the_Rising_Costs_of_Healthcare.pdf/7337673b-bd46-4ae5-aa85-1906f83f6a05 [accessed 2016-02-20] [WebCite Cache ID 6fRLcACdh]
11. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A. Developing a framework for evaluating the patient engagement, quality, and safety of mobile health applications. Commonw Fund Issue Br 2016;5:12.
12. BinDhim NF, McGeechan K, Trevena L. Who uses smoking cessation apps? A feasibility study across three countries via smartphones. JMIR Mhealth Uhealth 2014;2:e4 [FREE Full text] [doi: 10.2196/mhealth.2841] [Medline: 25098439]
13. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness. In: Proceedings of the 15th International Academic MindTrek Conference. New York, New York, USA: ACM Press; 2011 Presented at: International Academic MindTrek Conference; 2011; New York p. 11. [doi: 10.1145/2181037.2181040]
14. Deterding S. Meaningful play. Getting gamification right. In: Google Tech Talk. 2011 Presented at: Google Tech Talk; January 1, 2011; California.
15. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013;1:e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
16. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MFG. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012;344:e2598 [FREE Full text] [Medline: 22517917]
17. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013:76-78. [doi: 10.1177/0141076813480996]
18. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to Web-based mental health interventions: a systematic review. JMIR Ment Health 2016;3:e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]

19. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- A literature review of empirical studies on gamification. In: IEEE Computer Society Digital Library. 2014 Presented at: 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI. [doi: 10.1109/HICSS.2014.377]
20. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10:e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
21. Braun V, Clarke V. Qual Res Psychol Internet. Using thematic analysis in psychology URL: http://eprints.uwe.ac.uk/11735/2/thematic_analysis_revised_-_final.pdf [accessed 2016-09-24] [WebCite Cache ID 6kjoXA9ew]
22. Martin PY. Grounded theory and organizational research. J Appl Behav Sci 1986;22:141-157. [doi: 10.1177/002188638602200207]
23. Carpenter C. A meta-analysis of the effectiveness of health belief model variables in predicting behavior. Health Commun 2010;25:9. [doi: 10.1080/10410236.2010.521906]
24. West R. The European Health Psychologist. 2008. Finding better ways of motivating and assisting smokers to stop URL: http://openhealthpsychology.net/ehp/issues/2008/v10iss3_Sept2008/EHP_Sept_2008_RWest.pdf [accessed 2016-09-23] [WebCite Cache ID 6kjvDPR3n]
25. Pavlov I, Anrep GV. Conditioned Reflexes. Mineola: Dover Publications; 2003.
26. Csikszentmihalyi M. Flow: The Psychology of Optimal Experience. New York: HarperPerennial; 1990.
27. Brühlmann F. Gamification From the Perspective of Self-Determination Theory and Flow [BS thesis]. Basel, Switzerland: University of Basel, Institute of Psychology; 2013.
28. McNamara D, Jackson G, Graesser A. Intelligent tutoring and games (ITaG). 2009 Presented at: Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education; July 6-10, 2009; Brighton, England.
29. Skinner B. The Behaviour of Organisms: An Experimental Analysis. Acton: Copley Publishing Group; 1938.
30. Pavlas D. A Model of Flow and Play in Game-Based Learning: The Impact of Game Characteristics, Player Traits and Player States [PhD thesis]. Orlando, FL: University of Central Florida, Department of Psychology; 2010.
31. Kahneman D, Tversky A. Choices, values, and frames. American Psychologist 1984;39:314-350.
32. Deci EL, Koestner R, Ryan RM. Extrinsic rewards and intrinsic motivation in education: reconsidered once again. Review of Educational Research 2001;71:1-27. [doi: 10.3102/00346543071001001]
33. Lepper M, Sagotsky G, Dafoe J, Greene D. Consequences of superfluous social constraints: effects on young children's social inferences and subsequent intrinsic interest. Journal of Personality and Social Psychology 1982;42:51-65 [FREE Full text] [doi: 10.1037/0022-3514.42.1.51]
34. Deci E, Ryan R. The Handbook of Self-Determination Research. Rochester: University of Rochester Press; 2002.
35. Flora S. The Power of Reinforcement. Albany: State University of New York Press; 2004.
36. Ryan R, Sheldon K, Kasser T, Deci E. All goals are not created equal: an organismic perspective on the nature of goals and their regulation. In: Gollwitzer PM, Bargh JA, editors. The Psychology of Action: Linking Cognition and Motivation to Behavior. New York, NY: Guilford Press; 1996:7-26.
37. Mekler E, Brühlmann F. Disassembling gamification: the effects of points and meaning on user motivation and performance. 2013 Presented at: CHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France URL: http://dl.acm.org/citation.cfm?id=2468559
38. Eyal N. Hooked: How to Build Habit-Forming Products. Ringwood, Vic: Penguin Books Australia; 2013.
39. Norman DA. The Design of Everyday Things. New York: Basic Books; 2013.
40. Draper S. Analysing fun as a candidate software requirement. Pers Ubiquitous Comput 1999;3:22.
41. Kwon KH, Halavais A, Havener S. Tweeting badges: user motivations for displaying achievement in publicly networked environments. Cyberpsychol Behav Soc Netw 2015;18:93-100. [doi: 10.1089/cyber.2014.0438] [Medline: 25684610]
42. King D, Greaves F, Exeter C, Darzi A. 'Gamification': influencing health behaviours with games. J R Soc Med 2013;106:76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
43. Online R, Bennett S, Maton KA, Kervin L, Bennett S, Maton K. The? digital natives? debate: a critical review of the evidence. Br J Educ Technol 2008;39:86. [doi: 10.1111/j.1467-8535.2007.00793.x]

Abbreviations

NHS: National Health Service
PBC: perceived behavioral control

Edited by G Eysenbach; submitted 20.02.16; peer-reviewed by T Fleming, J Lumsden, J McClure; comments to author 06.04.16; revised version received 07.08.16; accepted 22.09.16; published 24.10.16.

Please cite as:
El-Hilly AA, Iqbal SS, Ahmed M, Sherwani Y, Muntasir M, Siddiqui S, Al-Fagih Z, Usmani O, Eisingerich AB
Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study
JMIR Serious Games 2016;4(2):e18
URL: http://games.jmir.org/2016/2/e18/
doi:10.2196/games.5678
PMID:27777216

Original Paper

Quittr: The Design of a Video Game to Support Smoking Cessation

Ivan Bindoff[1], BComp (Hons), PhD; Kristy de Salas[1], BEc, BComp (Hons), PhD; Gregory Peterson[1], PhD; Tristan Ling[1], BComp (Hons), PhD; Ian Lewis[1], BComp (Hons), PhD; Lindsay Wells[1], BComp (Hons); Peter Gee[1], BPharm (Hons), GradDipComp; Stuart G Ferguson[1], PhD

University of Tasmania, Hobart, Australia

Corresponding Author:
Kristy de Salas, BEc, BComp (Hons), PhD
University of Tasmania
Private Bag 87
Hobart, 7001
Australia
Phone: 61 3 62266220
Fax: 61 3 62262900
Email: Kristy.deSalas@utas.edu.au

Abstract

Background: Smoking is recognized as the largest, single, preventable cause of death and disease in the developed world. While the majority of smokers report wanting to quit, and many try each year, smokers find it difficult to maintain long-term abstinence. Behavioral support, such as education, advice, goal-setting, and encouragement, is known to be beneficial in improving the likelihood of succeeding in a quit attempt, but it remains difficult to effectively deliver this behavioral support and keep the patient engaged with the process for a sufficient duration. In an attempt to solve this, there have been numerous mobile apps developed, yet engagement and retention have remained key challenges that limit the potential effectiveness of these interventions. Video games have been clearly linked with the effective delivery of health interventions, due to their capacity to increase motivation and engagement of players.

Objective: The objective of this study is to describe the design and development of a smartphone app that is theory-driven, and which incorporates gaming characteristics in order to promote engagement with content, and thereby help smokers to quit.

Methods: Game design and development was informed by a taxonomy of motivational affordances for meaningful gamified and persuasive technologies. This taxonomy describes a set of design components that is grounded in well-established psychological theories on motivation.

Results: This paper reports on the design and development process of Quittr, a mobile app, describing how game design principles, game mechanics, and game elements can be used to embed education and support content, such that the app actually requires the user to access and engage with relevant educational content. The next stage of this research is to conduct a randomized controlled trial to determine whether the additional incentivization game features offer any value in terms of the key metrics of engagement–how much content users are consuming, how many days users are persisting with using the app, and what proportion of users successfully abstain from smoking for 28 days, based on user-reported data and verified against a biochemical baseline using cotinine tests.

Conclusions: We describe a novel, and theoretically-informed mobile app design approach that has a broad range of potential applications. By using the virtual currency approach, we remove the need for the game to comprehensively integrate the healthy activity as part of its actual play mechanics. This opens up the potential for a wide variety of health problems to be tackled through games where no obvious play mechanic presents itself. The implications of this app are that similar approaches may be of benefit in areas such as managing chronic conditions (diabetes, heart disease, etc), treating substance abuse (alcohol, illicit drugs, etc), diet and exercise, eating disorders (anorexia, bulimia, and binge eating), and various phobias.

(JMIR Serious Games 2016;4(2):e19) doi:10.2196/games.6258

KEYWORDS

smoking cessation; video games; mobile phone; motivation

Introduction

Smoking and Cessation

Smoking is recognized as the largest, single, preventable cause of death and disease in the developed world. It is associated with an increased risk of heart disease, stroke, cancer, emphysema, bronchitis, asthma, renal disease, and eye disease. While the majority of smokers report wanting to quit, and many try each year, smokers find it difficult to maintain long-term abstinence [1-3].

The delivery of education and support to smokers has been the backbone of smoking cessation programs for decades. Such programs were traditionally delivered in face-to-face sessions with trained tobacco cessation counselors, or via self-help booklets and websites, but more recently researchers have turned to technological solutions such as mobile phone apps in order to deliver content to smokers in a way that can be easily scaled up to population levels [4-6]. Regardless of the delivery modality, behavioral support programs have been found to be mildly effective [7], with the educational content involved in such programs (including advice on when and how to use pharmacological support) appearing to have the greatest impact on cessation rates.

Despite significant investment in the development of such programs, a persistent issue is that smokers do not readily engage with the cessation content delivered. Failure to engage with the "active content" of cessation programs is problematic as it limits the potential effectiveness of such programs: the education and support delivered cannot hope to be effective if it is not used by smokers during their quit attempt. Even if users engage initially, it is difficult to keep them interested beyond the first few days (retention). Identifying mechanisms that encourage smokers to engage with smoking cessation content requires ongoing investigation.

Video Games and Health Interventions

Video games are immensely popular, with recent statistics indicating that in the United States 63% of households are home to at least 1 person who plays video games regularly (3 hours or more per week) and 65% of households own a device used to play video games [8]. Rather than video games being seen as simply time-wasting entertainments, they are now recognized as an integral part of our modern society, with games and play increasingly informing other domains of our everyday life.

Research on video games now shows that games have the ability to keep people engaged for a long time, build relationships and communities among players, and cultivate their creative potential. Furthermore, games have been shown to have the power to support knowledge acquisition and even to bring about behavior change [9-11].

Games with a purpose, also known as "serious games," look much like traditional entertainment games, but differ in that they have a defined purpose, outcome, or message that the creators of the game are trying to get across to a particular audience [12]. By employing all the fun and engaging elements found in games and applying them to real-world or productive activities, they have the potential to influence behavior or improve engagement in tasks not generally considered "fun," such as learning, exercising, and practicing.

Within the health domain there have been a wide range of successful serious games initiatives trialed. A 2008 review of 27 publications reporting on 25 video games that promoted health-related behavior change, identified that most of the studies demonstrated positive health-related changes from playing the video games [13].

More recently, a 2012 review included 38 publications regarding video games that provide physical therapy, psychological therapy, improved disease self-management, health education, distraction from discomfort, increased physical activity, and skills training for clinicians [14]. Among the 38 studies, a total of 195 health outcomes were examined. Results indicated that video games improved 69% of psychological therapy outcomes, 59% of physical therapy outcomes, 50% of physical activity outcomes, 46% of clinician skills outcomes, 42% of health education outcomes, 42% of pain distraction outcomes, and 37% of disease self-management outcomes. Further, in 2014, a meta-analysis of 54 serious digital game studies for healthy lifestyle promotion reviewed the overall effectiveness of serious digital games on healthy lifestyle [15], and demonstrated small positive effects on healthy lifestyle promotion outcomes.

Video Games and Smoking Cessation

The concept of using games to promote smoking cessation is not entirely new. There is some existing work that explores the potential of the idea, and even some existing game-based interventions have been developed. Most early attempts at incorporating video game elements have focused on the delivery of a distraction, to help the smoker get through an episode of craving. Crave-Out was an early attempt at this style of game, producing a pattern matching memory game and using the Questionnaire of Smoking Urge-Brief measurement before and after play to determine if it helped to reduce cravings. A small but insignificant reduction in cravings was measured. Critically however, the authors determined during pilot testing that negative reinforcement had a deleterious effect on smokers, as it reminded them of smoking, and this finding is worth bearing in mind in future designs [16]. Subsequently, similar "distraction minigames" have also been incorporated into other mHealth-based smoking cessation apps such as SmokeFree 28 [5].

The most grass-roots evidence to support the idea of game-based smoking cessation interventions comes from Raiff et al [17], who explored the link between video game usage and smoking. They determined that 74.5% of smokers play video games (more than nonsmokers), and that 63.7% of participants believed that a video game–based intervention would motivate smokers to quit [17]. There is some speculation that the "addictive tendencies" of smokers makes them more likely to become frequent video game players.

Hall et al [18] tried to qualitatively determine, from expert opinion, key design characteristics for a successful video game–based smoking cessation intervention. After interviewing experts in game development, tobacco, and health behavior a number of themes emerged. In particular, it was noted that targeting youths through games was likely to be an effective strategy that extensive iterative prototyping was required to create a compelling game experience, and that incorporating behavioral change techniques was likely to result in improved outcomes.

There is also some preliminary evidence exploring Internet implementations. QuitIT is an Internet-based video game that is designed to help smokers practice coping strategies for cravings. Initial evaluation suggests the game was useful in teaching strategies for coping with urges [19], although there is not yet any data to suggest that it improves quit success rates.

Traditionally, serious games in the health behavior change space have focused on directly encouraging the behavior of interest (eg, promoting physical activity by requiring movement as part of the game play). An alternate possible use of games–one that we explore here–is that games could be embedded with education and support content, such that the game play simply acts as the hook that compels the user to access and engage with relevant educational content, in exchange for incentives that further the players goals in the game. Such game play may encourage engagement with proven educationally based intervention strategies, and thereby improve health outcomes.

It has been suggested that games are fun because they appeal to some core drives that motivate us toward certain activities, such as meaning, empowerment, social influence, unpredictability, avoidance, scarcity, ownership, and accomplishment [15]. If we can identify the elements of games that support the acquisition of smoking cessation content, and build a game that provides an engaging environment for players to remain engaged with such content, we may be able to positively influence smoking cessation outcomes.

Theoretical Basis for Design

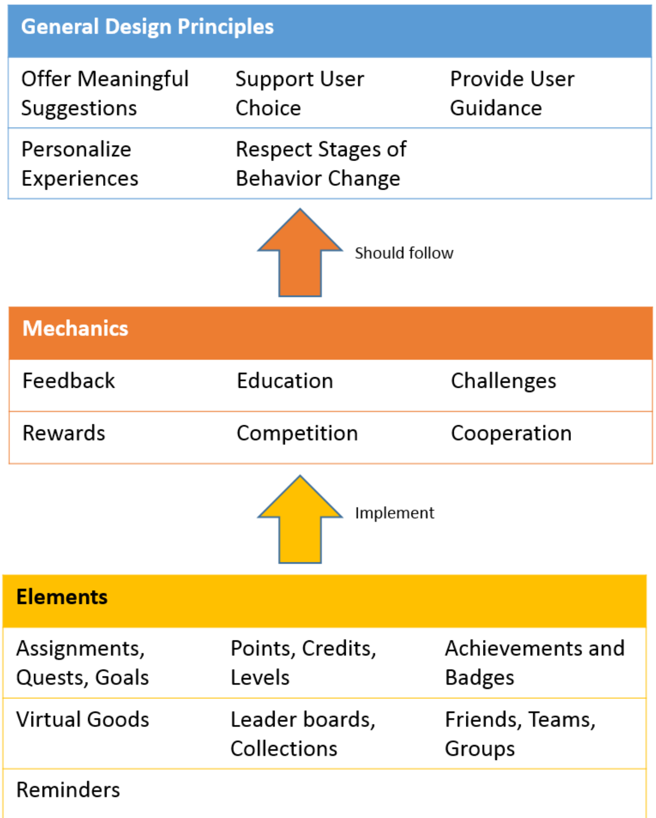
This focus aligns with the assessment that reflective and automatic motivation toward smoking cessation can be enhanced by interventions, such as education, persuasion, intensification, coercion, and modelling [20]. However, the actual content component of the app itself is flexible and has been designed to be easily updatable as new content is developed.

Game Mechanic and Game Elements Design Phase

Within the cycle of design and development, the initial Game Mechanic and Elements design phase was informed by a taxonomy of motivational affordances for meaningful gamified and persuasive technologies (Figure 1) [20]. This taxonomy describes a set of design components that is grounded in well-established psychological theories on motivation [20].

To allow for meaningful gamified and persuasive systems, designers should choose components depending on both the objectives and the users of the system [20]. To guide this activity, the taxonomy outlines both general design principles and mechanics that game designers might benefit from considering, and that might assist in ensuring theoretically grounded serious game development, as is required in this intervention.

Figure 1. Design principles and mechanics for persuasive game design [17].



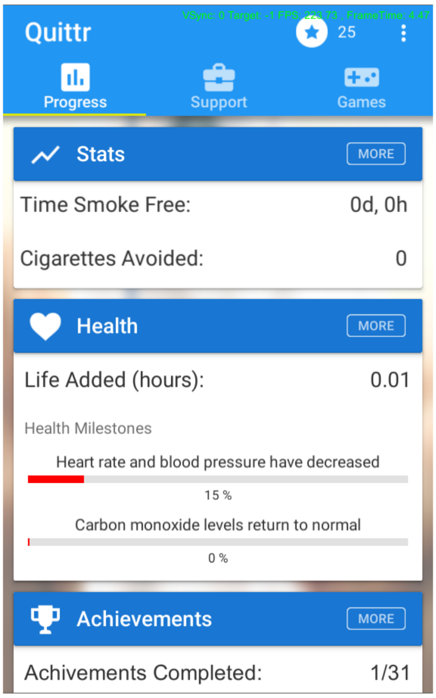
Methods

Overview of the Design

Based on the theoretical underpinnings described above, we designed and developed a mobile phone game called Quittr to support people in their attempt to quit smoking. This design borrowed many elements from existing literature [21,22] and examples of mHealth smoking cessation apps, most notably SmokeFree 28 [5]. The focus was on creating a platform that would enable the delivery of proven effective smoking cessation support content, but with game elements integrated throughout to improve user engagement and retention.

An app was produced compatible with both Android and iOS mobile phones and tablets. Upon opening the app, users are greeted with a dashboard system separated into 3 main pages, which they can swipe between (Figure 2), styled in adherence with the popular Google Material user interface design guidelines [23].

Figure 2. The main dashboard of Quittr.



Data Collection

When the user first opens the app they are prompted to complete an entry survey, which includes questions about the user's smoking status, demographics, health, and financial goals, as well as questions about the user's intentions in using the app. On subsequent openings, the user is asked if they have smoked any cigarettes that haven't been entered into the app. If the user indicates yes to this question, they are asked to indicate how many cigarettes they smoked on each day that has elapsed since their last log. If the user has not opened Quittr for 12 hours they are prompted with a notification on their device asking them to come back and log their cigarettes. This will repeat daily for several days if they still fail to check-in.

At the end of 28 days or when the user indicates their current quit attempt has failed, they are also asked to complete an exit survey. This asks questions about their opinions of the app, the games, and if appropriate, their reasons for giving up their quit attempt early. Having indicated that their quit attempt has failed, the user is able to start a new quit attempt at any time, which will reset all their existing data, triggering a new entry survey, and so on.

All collected data is stored and uploaded to our webserver for analysis and review. For the sake of privacy, we do not collect any personally identifying information, although we do use a unique identifier so that we can track a given user across potentially many quit attempts.

Progress

The first page of the main dashboard is titled "Progress," and includes tracked statistics relating to the financial, health, and social outcomes of the user associated with their current quit attempt. It delivers an easily accessible snapshot of the statistics that are most relevant/important to the user at the current time (based on their current status and data collected during a commencement survey). The user can optionally explore a comprehensive list of every tracked statistic. These progress bars provide meaningful feedback to the user as well as educating them about the benefits of quitting, and have proven to be a valuable feature in a range of successful smoking cessation apps [20,24].

The progress page also includes an achievements section, with a range of predefined goals for the user to strive toward, as well as the ability to set their own financial savings goals to target. The progress page displays the next several achievements that the user is close to achieving, but the user can click on the tile to review the full list of achievements, including achievements they have already obtained. This style of achievement system is a proven staple across the games industry, and is a mainstay in any "gamified" apps. The achievements system ensures that the user always has both short and longer term goals in sight, as well as having a set of accomplishments they can reflect upon to remind themselves how far they've come [20].

Support

The support page includes a range of helpful information and support, including a button, which simply dials the local Quit Support hotline, and the "Quick Tips" button that displays helpful tips about how to use the app. Perhaps the most interesting feature of this section is the "Information Toolbox" (Figure 3).

The Information Toolbox includes a wide variety of educational material, which can support the user on their quit attempt. This includes information about the various therapies and treatment plans available, as well as information on cravings, coping strategies, the benefits of quitting, and the consequences of smoking. The information is neatly structured into 'bite-sized chunks' to make it easy for the user [22]. After reading a section the user is challenged by a multiple choice quiz that asks one of several basic questions that relate to the material they just read.

We have designed the content in this section to be readily customized and expanded upon, with updates being pulled from our webserver as required whenever Internet connectivity is available. The content is also tagged with a set of metadata that enables it to be tailored to the user based on the data collected during the entry survey and the state of their current quit attempt (ie, content about smoking during pregnancy would be more relevant to pregnant women than to men, and content regarding cardiovascular disease would be more important to older smokers).

Games

The games page lists a range of games that the user can engage with (Figure 4). The games are split into 2 categories, those that are intended purely as a distraction from cravings, and those that are used as an incentive to engage with the broader app. Currently, 2 games are available, with plans to include more. By providing a variety of games that will appeal to different users, we can improve the likelihood that a given user will be able to find a game they enjoy playing, and that they find it to be an effective distraction.

Figure 3. The Information Toolbox.

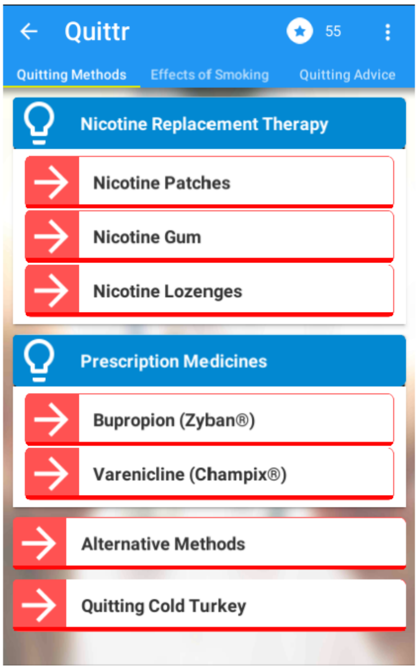


Figure 4. The games currently available in Quittr.

Distraction

The distraction games are simple minigames that can be played in a stand-alone 1- to 5-minute session. They are designed to demand mental focus and both hands, so as to provide a meaningful and effective distraction or an alternative from the act of smoking [5,16,22].

There is currently only one distraction minigame included in Quittr. This is a hidden object style game where the user must search for and tap particular objects in an increasingly cluttered environment, and under time pressure. There are also 2 additional games being prepared for inclusion: a fruit-ninja style game where the user must swipe across missiles, which are launched while avoiding bombs; and an endless-runner style game where the user must navigate a plane through a series of winding caverns without crashing while collecting coins and powerups.

Incentivization

Incentivization games are designed to be played incrementally over the 28-day duration of the quit attempt. These are games that are designed to involve longer term goal setting and planning, as well as frequent monitoring/check-ins by the user. Most importantly, they make use of a premium currency system, which we have called QuitCoins. This system will be familiar to those who have played free-to-play mobile and Facebook games, wherein the user can spend a premium currency (traditionally obtained by spending real money or engaging with video ads) to obtain bonuses in the game. In this case, instead of purchasing QuitCoins using real money, the user is rewarded with QuitCoins in return for engaging with the smoking cessation content of the broader app in a meaningful way (see Textbox 1). This is the key innovation that our approach adds, distinguishing it from earlier attempts in this space.

Currently, we have only created 1 incentivization game. It is a “city builder” style game, which we have called “Tappy Town.” In Tappy Town, the user has an empty plot of land to fill and a large range of structures they can build to improve that land. As the user grows their town the town’s economy will also grow, enabling them to afford increasingly grand and powerful structures. The town will passively collect resources over time, but users are rewarded for frequent check-ins by the intermittent availability of significant bonus resources that require manual collection by tapping the associated buildings. This ensures that every 1 to 2 hours (corresponding to when we might expect the next major craving to hit) the user will have something productive to do in their town, and also provides an additional mechanism by which the user can distract themselves from their cravings.

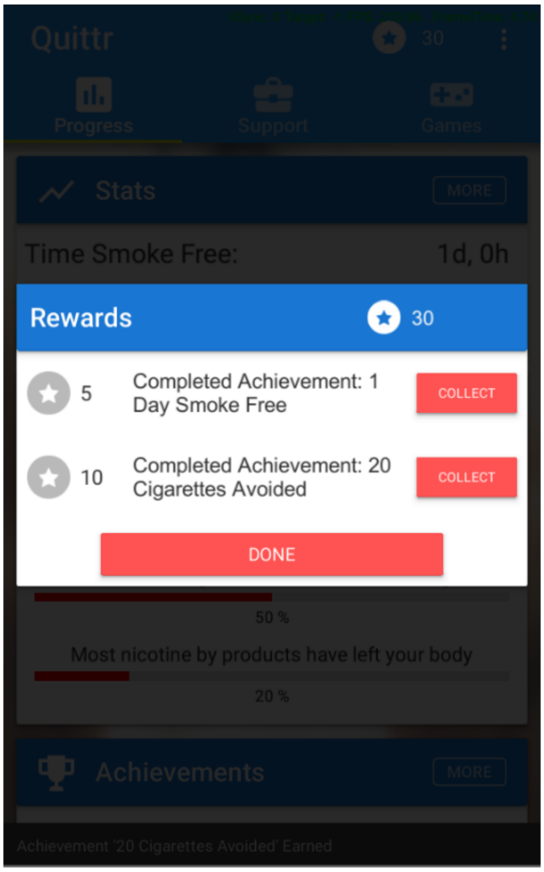
Textbox 1. The ways the user can earn QuitCoins.

- Completing the entry survey
- Opening the app
- Reporting how many cigarettes they’ve smoked
- Reading through their dashboard statistics
- Reading support content
- Successfully completing quizzes
- Playing a distraction minigame
- Earning achievements
- Setting up and achieving personal goals
- Completing the exit survey

However, the QuitCoin rewards for a given piece of content can only be earned periodically. For example, after reading a particular piece of advice the user might be rewarded with a QuitCoin, but they cannot reread the same piece of advice to earn another QuitCoin; they must wait a day before they can earn rewards from reading that particular piece of advice again. QuitCoins are awarded immediately after performing the related action, and this is made visible to the user by an animation and sound-effect that plays in the top right corner of the dashboard. When the user taps the QuitCoin icon they are presented with a list of all the QuitCoin rewards they have been awarded recently. This ensures that the user understands clearly why they earned the reward (Figure 5).

Having earned QuitCoins, the user can redeem these coins in Tappy Town (or any other incentivization game we might add at a later date). The highest performing and most visually impressive structures can only be purchased using QuitCoins. The intention is that the user will become hooked on the reward mechanisms on display in Tappy Town, and will develop a sense of ownership of the town they are gradually building. This will inspire them to engage more with the various contents in the Quittr app so that they can afford to build these impressive showcase structures in their town. These mechanisms are readily apparent to the user, who is regularly notified that they are earning QuitCoins through natural interactions with Quittr, and who are then encouraged to go to Tappy Town to redeem them.

Figure 5. Collecting QuitCoin rewards.



Results

At the time of publication Quittr is largely complete, with only content revision still required. It has been in closed (invite only) beta testing using Apple TestFlight and on the Android Play Store since August 9, 2016, and has at the time of writing (October 4, 2016) been installed by 7 users on iOS and 11 users on Android. The feedback received during this beta testing period has resulted in 6 iterative beta releases during this period to date. At this time, we are confident that Quittr is stable and all features are working as intended. The open beta test is planned to become available for 2017, pending ethics board approval.

Discussion

Further Work

It has yet to be determined whether the innovative game systems we have added into the traditional mHealth smoking cessation recipe provide the additional value we are hoping for in terms of engagement, retention, and ultimately cessation rates. To test the impact of the gaming elements, we are currently planning a research project using a randomized controlled trial design. Users will be randomized into 1 of 2 groups, either the control group that gets access to a "baseline" version of the game without the incentivization game features and the QuitCoin premium currency system, and the intervention group, which receives a copy of the game with the full feature set.

By comparing the results of these 2 groups, we will be able to determine whether the additional incentivization game features we have developed offer any value in terms of the key metrics of engagement–how much content users are consuming; retention–how many days that users are persisting with using the app; and smoking cessation rates–what proportion of users successfully abstain from smoking for given period of time (based on user-reported data and biochemically verified).

Going beyond this, we would ultimately like to compare our approach against a usual care scenario, to see how it compares more directly against current practice. We would also like to do a more specific evaluation in the younger demographic because we believe that our approach may be particularly useful for younger smokers, who are perhaps more willing to engage with a mobile phone–based intervention, and with game-based approaches in general.

Closing Thoughts

If this model proves successful there is a broad range of potential applications for this approach. By using the virtual currency approach in the way we have done here, we remove the need for the game to comprehensively integrate the healthy activity as part of its actual play mechanics. This opens up the potential for a wide variety of health problems to be tackled through games where no obvious play mechanic presents itself. This may be of benefit in areas such as managing chronic conditions (diabetes, heart disease, etc), treating substance abuse (alcohol, illicit drugs, etc), diet and exercise, eating disorders (anorexia, bulimia, and binge eating), and various phobias.

Conflicts of Interest

None declared.

References

1. Burns DM. Epidemiology of smoking-induced cardiovascular disease. Prog Cardiovasc Dis 2003;46:11-29. [Medline: 12920698]
2. Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. JAMA 2003;290:86-97. [doi: 10.1001/jama.290.1.86] [Medline: 12837716]
3. US Department of Health and Human Services. The health consequences of smoking: a report of the Surgeon General. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health 2004:62.
4. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. J Pers Med 2014;4:412-423 [FREE Full text] [doi: 10.3390/jpm4030412] [Medline: 25563359]
5. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015;17:e17 [FREE Full text] [doi: 10.2196/jmir.3479] [Medline: 25596170]
6. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
7. Brown J. A review of the evidence on technology-based interventions for the treatment of tobacco dependence in college health. Worldviews Evid Based Nurs 2013;10:150-162. [doi: 10.1111/wvn.12000] [Medline: 23421669]
8. Entertainment Software Association. Essential facts about the computer and video game industry. 2016 Apr.. URL: http://essentialfacts.theesa.com/Essential-Facts-2016.pdf [accessed 2016-09-21] [WebCite Cache ID 6kfyg6sPI]
9. Gee J. What video games have to teach us about learning and literacy. Computers in Entertainment (CIE) 2003;1(1):20.
10. Granic I, Lobel A, Engels Rutger C M E. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
11. Boot WR, Blakely DP, Simons DJ. Do action video games improve perception and cognition? Front Psychol 2011;2:226 [FREE Full text] [doi: 10.3389/fpsyg.2011.00226] [Medline: 21949513]
12. Susi T, Johannesson M, Backlund P. Serious Games: An Overview. Skövde: University of Skövde; 2007.

13. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008;34:74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
14. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012;42:630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
15. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
16. DeLaughter KL, Sadasivam RS, Kamberi A, English TM, Seward GL, Chan SW, et al. Crave-Out: a distraction/motivation mobile game to assist in smoking cessation. JMIR Serious Games 2016;4:e3 [FREE Full text] [doi: 10.2196/games.4566] [Medline: 27229772]
17. Raiff BR, Jarvis BP, Rapoza D. Prevalence of video game use, cigarette smoking, and acceptability of a video game-based smoking cessation intervention among online adults. Nicotine Tob Res 2012;14:1453-1457. [doi: 10.1093/ntr/nts079] [Medline: 22422929]
18. Hall AK, Mercado R, Anderson-Lewis C, Darville G, Bernhardt JM. How to design tobacco prevention and control games for youth and adolescents: a qualitative analysis of expert interviews. Games Health J 2015;4:488-493. [doi: 10.1089/g4h.2015.0013] [Medline: 26230046]
19. Krebs P, Burkhalter JE, Snow B, Fiske J, Ostroff JS. Development and alpha testing of QuitIT: an interactive video game to enhance skills for coping with smoking urges. JMIR Res Protoc 2013;2:e35 [FREE Full text] [doi: 10.2196/resprot.2416] [Medline: 24025236]
20. Weiser P, Bucher D, Cellina F, De Luca V. A taxonomy of motivational affordances for meaningful gamified and persuasive technologies. Acsr Adv Comput 2015;22:271-280.
21. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10:e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
22. Michie S, Churchill S, West R. Identifying evidence-based competences required to deliver behavioural support for smoking cessation. Ann Behav Med 2011;41:59-70. [doi: 10.1007/s12160-010-9235-z] [Medline: 20936389]
23. Google. Material design. 2016. URL: https://material.google.com/ [accessed 2016-09-21] [WebCite Cache ID 6kfzNNkTW]
24. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

Edited by G Eysenbach; submitted 22.06.16; peer-reviewed by M Businelle, H Ubhi; comments to author 28.07.16; revised version received 05.10.16; accepted 28.10.16; published 01.12.16.

Please cite as:
Bindoff I, de Salas K, Peterson G, Ling T, Lewis I, Wells L, Gee P, Ferguson SG
Quittr: The Design of a Video Game to Support Smoking Cessation
JMIR Serious Games 2016;4(2):e19
URL: http://games.jmir.org/2016/2/e19/
doi:10.2196/games.6258
PMID:27908844

Original Paper

Attentional Bias Modification With Serious Game Elements: Evaluating the Shots Game

Wouter J Boendermaker[1,2], MSc; Soraya Sanchez Maceiras[2], MSc; Marilisa Boffo[2], MSc, PhD; Reinout W Wiers[2], MSc, PhD

[1]Utrecht University, Department of Child and Adolescent Studies, Utrecht, Netherlands

[2]Department of Developmental Psychology, University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Wouter J Boendermaker, MSc
Department of Child and Adolescent Studies
Utrecht University
Heidelberglaan 1
Utrecht, 3584 CS
Netherlands
Phone: 31 302534742
Fax: 31 302534742
Email: w.j.boendermaker@uu.nl

Abstract

Background: Young adults often experiment with heavy use of alcohol, which poses severe health risks and increases the chance of developing addiction problems. In clinical patients, cognitive retraining of automatic appetitive processes, such as selective attention toward alcohol (known as "cognitive bias modification of attention," or CBM-A), has been shown to be a promising add-on to treatment, helping to prevent relapse.

Objective: To prevent escalation of regular use into problematic use in youth, motivation appears to play a pivotal role. As CBM-A is often viewed as long and boring, this paper presents this training with the addition of serious game elements as a novel approach aimed at enhancing motivation to train.

Methods: A total of 96 heavy drinking undergraduate students carried out a regular CBM-A training, a gamified version (called "Shots"), or a placebo training version over 4 training sessions. Measures of motivation to change their behavior, motivation to train, drinking behavior, and attentional bias for alcohol were included before and after training.

Results: Alcohol attentional bias was reduced after training only in the regular training condition. Self-reported drinking behavior was not affected, but motivation to train decreased in all conditions, suggesting that the motivational features of the Shots game were not enough to fully counteract the tiresome nature of the training. Moreover, some of the motivational aspects decreased slightly more in the game condition, which may indicate potential detrimental effects of disappointing gamification.

Conclusions: Gamification is not without its risks. When the motivational value of a training task with serious game elements is less than expected by the adolescent, effects detrimental to their motivation may occur. We therefore advise caution when using gamification, as well as underscore the importance of careful scientific evaluation.

(JMIR Serious Games 2016;4(2):e20) doi:10.2196/games.6464

KEYWORDS

cognitive training; cognitive bias modification (CBM); attentional bias; adolescents; serious games; motivation

Introduction

Heavy alcohol use during adolescence and early adulthood has been related to health problems and academic underperformance [1] and is an important predictor of addictive behaviors later in life [2]. Dual-process models of addiction [3,4] suggest that prolonged use of alcohol, especially when initiated during adolescence [5], can lead to the development of strong automatically triggered reactions toward alcohol, which in turn facilitate the development of addictive behaviors. This is visible in heavy alcohol users' tendency to approach [6] and selectively attend to [7] alcohol-related stimuli more quickly, compared with non-alcohol-related stimuli. Opposite to these strengthened automatic reactions are reflective cognitive processes, including control abilities (eg, working memory [8], inhibition [9]) that can be too weak or too late to moderate the automatically triggered reactions [5]. The resulting imbalance between automatically triggered appetitive processes and reflective control processes may contribute to escalation in problem drinking.

Research has shown that, in both longtime heavy users and clinical patients, both types of processes can successfully be trained or retrained, resulting in less craving and lower relapse rates [7,10-13]. Despite these promising results, application of cognitive training in younger populations has proven more difficult [14], for a number of reasons. First, youngsters tend to perceive stronger positive than negative effects of their alcohol use [15], perhaps because positive effects of alcohol tend to occur sooner than the negative effects [16], making those positive associations stronger. This typically results in lower motivation to change their (drinking) behavior compared with patient populations. Second, the fact that most training paradigms are long and often viewed as tedious and boring [17] adds to the problem. To improve motivation to train, one potential solution could be to make the training sessions more fun to do by adding game elements into the training paradigm. For example, Dovis and colleagues [18] offered children with attention-deficit/hyperactivity disorder (ADHD) a computer game version of a working memory task and observed that they normalized their persistence of performance to the level of children without ADHD. Dennis and O'Toole [19] used a mobile game based on attentional bias modification training as an intervention for stress and anxiety in highly trait-anxious university students and showed a significant reduction in threat bias.

In this paper, we apply similar gamification techniques to a typical cognitive bias modification of attention (CBM-A) training task aimed at training attention away from pictures of alcoholic beverages: the visual probe task (VPT) [20,21]. In the VPT participants are shown pairs of pictures, one of a relevant stimulus (eg, a picture of an alcoholic beverage), the other a visually similar, neutral stimulus (eg, a picture of a nonalcoholic beverage). Next, a probe (eg, an arrow pointing up or down) appears at the location of one of the stimuli, and the instruction is to quickly identify the probe (eg, respond to the direction of the arrow). The contingency between the location of the probe and the stimulus it replaces can be manipulated. To assess attentional bias, the probe appears equally often at the location of both stimulus types; to train attention away from a certain set of stimuli (eg, alcohol), the probe appears more often at the location of the other set of stimuli (eg, nonalcoholic). Schoenmakers and colleagues [21] showed that CBM-A can indeed increase the ability to disengage from alcohol-related cues and found that alcohol-dependent patients who had received the CBM-A training took significantly longer to relapse after training than patients who had not received CBM-A.

It should be noted that although adding game elements to a cognitive training task may help increase participants' motivation to train, they usually also influence the specific task features and parameters and inevitably change to some degree the evidence-based nature of the task [22]. For example, using points as rewards in the gamified training in order to enhance motivation to train may also counterbalance intrinsic motivation to train. As such, we will compare the new gamified VPT training (VPT-G) with both a regular VPT training (VPT-R), to evaluate the added motivational effects of the game elements, as well as with a placebo version of the regular VPT training (VPT-P) to establish whether it has a training effect similar to the VPT-R. This results in the following hypotheses. First, we expect a significant reduction in attentional bias toward alcohol stimuli in both the VPT-R and VPT-G conditions compared with the VPT-P condition (H1). This will be measured using both an assessment version of the VPT and another task that also measures attentional bias but is procedurally different, that is, the visual search task (VST) [23]. Next, we expect the same pattern of results between conditions with regard to decline in actual drinking behavior (H2). Finally, we expect to see that motivation to train is positively affected by the training in the VPT-G condition but not in the VPT-R and VPT-P conditions (H3), while motivation to change is expected to remain unaffected as it is not explicitly targeted by this training (H4).

Methods

Design and Procedure

The training consisted of 4 sessions, at least 1 day apart, over the course of 2 weeks. The first and last training sessions were combined with the assessment tasks in our laboratory; the 2 remaining training sessions were done at home. During the first session, participants were informed about the study's goal, gave digital informed consent, and were randomly assigned to the VPT-P (n=33), VPT-R (n=30), or VPT-G (n=33) condition. They continued with digital versions of the Alcohol Use Disorders Identification Test (AUDIT), a short Readiness to Change Questionnaire (RCQ), the Timeline Followback (TLFB) questionnaire, the Alcohol Use Questionnaire (AUQ), and a Motivation to Train Questionnaire (MTQ). After the questionnaires, they completed the VPT and VST baseline assessment and finished the session with the first VPT training. The following second and third sessions solely consisted of the VPT training task. The last session started with the fourth VPT training, after which they performed the VPT and VST posttraining assessments and the AUQ, TLFB, RCQ, and MTQ questionnaires, supplemented by a brief set of questions about the training itself (EVAL). To evaluate drinking behavior after the training, a follow-up TLFB was filled in via email 2 weeks after session 4.

Participants

A sample of undergraduate students (n=96, mean age 21.2 years, SD 1.8, range 18-28 years, 71%, 68/96 female) was recruited through the university laboratory's website, based on their drinking behavior (≥5 standard glasses of alcohol on average per week for males; ≥4 for females). Participants received study credits or €30 for taking part in the experiment. The study was approved by the Ethics Committee of the University of Amsterdam (Protocol Number: 2015-DP-4215).

Materials

Alcohol use and problems were measured with 3 questionnaires: A TLFB [24,25] was used to measure alcohol consumption per day over the past week and also included a question about the number of binge drinking occasions during the past 30 days (>5 standard glasses of alcohol consumed during one occasion for male participants; >4 for females). An adapted version of the AUQ [26] was used to assess drink-specific alcohol consumption over the past 6 months. For analyses, Mehrabian and Russell's [26] equation 2 was used to calculate the habitual alcohol consumption (HAC), including those items regarding consumption of beer, wine, and liquor, as well as our added items concerning alcohol pops. Alcohol-related problems were measured with the AUDIT [27], which includes 10 multiple-choice questions regarding alcohol consumption and alcohol-related problems. The overall AUDIT score ranges between 0 and 40, with ≥8 indicating an increased risk of alcohol-related problems in normal samples and ≥11 in student samples [28].

Motivation to train was assessed using a self-developed 4-item questionnaire, each item rated on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." *Motivation to change* was assessed using a shortened version of the RCQ [29], consisting of 3 multiple-choice items. The EVAL questions concerned how they rated the training overall. See Multimedia Appendix 1 for an overview of translated questions.

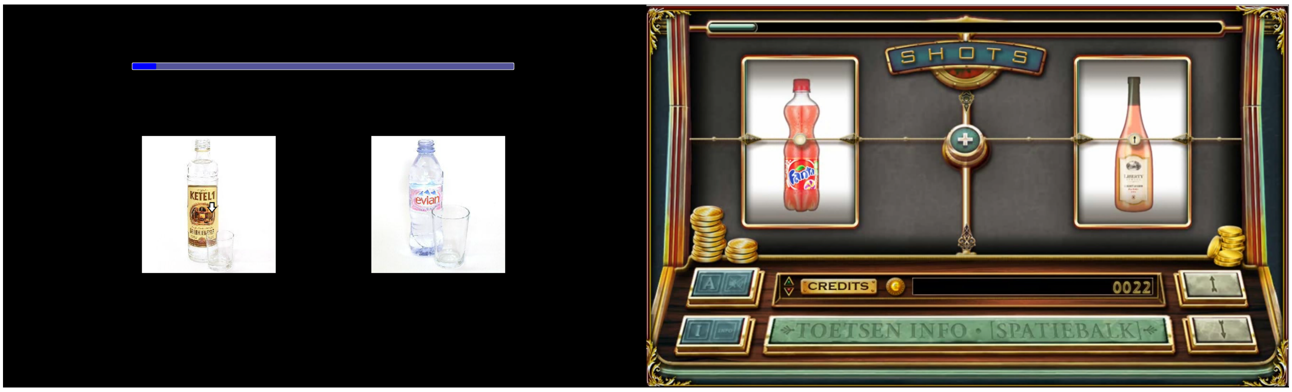
Attentional bias was measured using assessment versions of the VST and VPT paradigms. In the VST [23,30], participants were shown a grid of 4 by 4 pictures of beverages, where only 1 was of a different type: 1 alcoholic among 15 nonalcoholic beverages, or vice versa. The instruction was to find and click the deviant type of beverage as quickly and accurately as possible. To focus visual attention to the center of the grid, each trial started with a fixation cross in the center of the screen the participants had to mouse over in order to start the trial. When an incorrect response was given, feedback was given and the trial had to be redone. The task consisted of 6 blocks of 18 trials, using active (person drinking) or passive (bottle or glass only) pictures of alcoholic and nonalcoholic beverages or neutral pictures of 5- or 7-petaled flowers, following the schedule in Table 1. The order of the blocks containing beverages was counterbalanced over participants. A progress bar indicated the number of trials left in each block. The attentional bias scores for active and passive stimuli were computed by subtracting the respective average reaction times for selecting alcoholic beverages from the average reaction times for selecting nonalcoholic beverages. Given that faster reaction times on alcohol trials suggest an attentional bias toward alcohol, a positive bias score thus indicated a bias toward alcohol.

Table 1. Visual search task block distribution.

Block	Target picture (1)	Nontarget pictures (15)
1	Active, alcohol related	Active, not alcohol related
2	Passive, alcohol related	Passive, not alcohol related
3 and 4	5-petaled flower	7-petaled flowers
5	Active, not alcohol related	Active, alcohol related
6	Passive, not alcohol related	Passive, alcohol related

In the VPT, participants were shown pairs of alcohol and nonalcoholic beverages followed after 500 milliseconds by a small arrow probe in the location of one of the pictures, pointing upward or downward. The instruction was to press the keyboard's arrow key corresponding to the arrow's direction as quickly and accurately as possible. The task consisted of 168 critical trials with pairs of beverages and 32 filler trials with pairs of neutral objects (office supplies), presented in random order over 3 blocks: a starting block of 10 neutral practice trials, then 2 test blocks of 100 trials (84 critical trials). In one of the two test blocks, the pictures disappeared as the arrow became visible ("Go" block); in the other block the pictures remained visible as the arrow probe was superimposed ("Stay" block). The Stay trials were included as they might better detect difficult disengagement from alcohol cues [31,32], while the standard Go trials may be a better measure of rapid allocation and maintenance of attention on alcohol cues. The order of these blocks was counterbalanced over participants. Location of alcohol picture (left or right) and arrow (at alcohol-related or not-alcohol-related stimulus) was fully counterbalanced. The filler trials were included to slightly mask the contingency between arrow placement and the content of the pictures, as well as maintain participants' attention on the task and avoid anticipatory responses. When an incorrect response was given, the trial had to be redone after feedback was given, and the arrow direction was reapplied randomly. A progress bar indicated the number of trials left in each block. All picture pairs were matched by size and colors (see Figure 1). All stimuli originated from the Amsterdam Beverage Picture Set (ABPS [33]). Attentional bias scores for Stay and Go trials were computed by respectively subtracting the average reaction time for correct trials with the arrow at the location of the alcoholic picture from the average reaction time on correct trials where the arrow was at the location of the nonalcoholic picture. Correction trials following incorrect responses were excluded, as reaction times on those trials were deemed unreliable owing to the stimuli being the same as in the previous trials; reaction times greater than 3000 milliseconds were coded as too late and were also excluded. Given that faster responses on alcohol trials suggest an attentional bias toward alcohol, a positive bias score thus indicated a bias toward alcohol.

Figure 1. The regular visual probe task (left) and the Shots game implementation (right). The arrow size was matched between the tasks but is enhanced slightly in the left panel (visual probe task) for visibility in print.



Training

The VPT training tasks were identical to the assessment version, except that after the practice block there was only 1 training block of 156 critical and 16 filler trials, where Stay and Go trials were presented randomly. Additionally, in the VPT-R and VPT-G the arrow always appeared at the location of the nonalcoholic beverage, thus training attention away from the alcoholic stimuli. In the VPT-G condition, participants trained using the Shots game (S van Schie, unpublished data, 2014). The Shots game was functionally identical to the VPT-R training, while looking like a slot machine game with 2 spinning wheels (see Figure 1; note that although the Shots game looks like a slot machine, it has no gambling elements to it. This was also explained to the participants). The game elements used here constitute an *integrated gamification* of the VPT paradigm as defined by the CBM gamification model by Boendermaker and colleagues [22]. It mainly uses a coin-based reward system (see Step 1 [22]) and nicer looking graphics, animations (eg, the spinning wheels with pictures of beverages), and sound effects (eg, when spinning the wheels or pressing a button; see Step 3 [22]). The participant is rewarded for correct and fast responses (using time bonuses and special bonus trials), requiring a coin in order to spin the wheels (ie, start a new trial) and eventually the possibility of reaching a new level (a new look for the machine). The game used picture stimuli similar to those in the ABPS but slightly edited to fit the graphical style of the game.

Results

Attrition

After the first training session, 2 participants in the VPT-R condition dropped out of the study and were excluded from the training effect analyses. An additional 6 participants failed to do the follow-up assessment and were excluded from the TLFB training effect analysis. Furthermore, baseline data from 2 participants on the RCQ question 3, and 1 participant on the VPT, were missing because of technical problems and therefore excluded from the relevant analyses.

Sample Characteristics

At baseline, participants had consumed an average of 15.09 standard units of alcohol (SD 11.46) during the previous 7 days and binged on average on 6.65 occasions (SD 3.48) during the previous 30 days. The AUQ average HAC score was 230.91 (SD 17.17) and with 94% (90/96) of participants scoring ≥8 and 71% (68/96) scoring ≥11, indicating hazardous drinking in a large proportion of the sample [27,28]. See Table 2 for an overview of baseline characteristics. No significant group differences were detected at baseline.

Training Effects

All dependent variables were screened for univariate extreme outliers based on inspection of stem-and-leaf and box plots. When they were present, or one of the general linear model (GLM) assumptions were violated, a nonparametric method for factorial repeated-measures analysis of variance was used: the Aligned Rank Transform analysis of variance [34], in which data are aligned and then ranked as a preprocessing step, before applying GLM procedures (these results are marked with *).

Attentional Bias Change (H1)

There was a significant reduction in alcohol attentional bias over time on the VPT-Go trials, $F_{1,90}$*=9.407, P=.003, η_p^2=0.095, as well as a significant interaction with condition, $F_{2,90}$*=8.685, P<.001, η_p^2=0.162. Tukey-adjusted contrasts indicated this was due to a significant decrease in bias in the VPT-R condition over time, t_{90}=3.094, P=.031, r=.310, and also confirmed the result presented in Table 2 that the VPT-Go bias score at baseline was significantly higher in the VPT-R condition, compared with the VPT-P condition, $t_{179.05}$=3.055, P=.031, r=.223. The VPT-Stay trials also showed a significant reduction in bias over time, $F_{1,90}$*=10.894, P=.001, η_p^2=0.108, without an interaction with condition. In contrast, no significant changes over time were found on the VST. A significant overall difference was found between the conditions on the VST-ACT trials: $F_{2,91}$=5.480, P=.006, η_p^2=0.107; but post hoc analyses revealed no significant contrasts.

Behavioral Change (H2)

There was no significant reduction in TLFB scores for both binges and total use (P>.05).

Table 2. Baseline characteristics by group.

Characteristics	Placebo (VPT[a]-P)	Regular (VPT-R)	Game (VPT-G)	Total	*P*
Total participants (male)	33 (10)	30 (8)	33 (10)	96 (28)	.936
Age in years, mean (SD)	21.4 (2.1)	21.3 (1.4)	21.0 (2.0)	21.2 (1.8)	.743
AUDIT[b] score, mean (SD)	13.7 (5.5)	13.5 (3.9)	13.6 (4.9)	13.6 (4.8)	.986
TLFB[c] (drinks/7 days), mean (SD)	15.7 (14.2)	13.4 (7.6)	16.0 (11.4)	15.1 (11.5)	.634
TLFB (binges/30 days), mean (SD)	6.9 (4.3)	6.3 (2.8)	6.7 (3.3)	6.7 (3.5)	.809
AUQ[d] (HAC[e]), mean (SD)	229.7 (17.8)	233.9 (20.5)	229.2 (12.4)	230.9 (17.2)	.516
RCQ[f] -1, mean (SD)	2.2 (0.8)	2.3 (0.7)	2.5 (0.9)	2.4 (0.8)	.454
RCQ-2, mean (SD)	3.8 (1.9)	4.0 (1.9)	3.7 (1.9)	3.8 (1.9)	.682
RCQ-3, mean (SD)	1.8 (0.7)	2.0 (0.6)	1.9 (0.7)	1.9 (0.7)	.499
MTQ[g], mean (SD)	15.8 (2.4)	16.7 (1.8)	16.4 (2.0)	16.3 (2.1)	.249
VPT-Go[h] alcohol bias (ms[i]), mean (SD)	−3.7 (20.8)	10.2 (25.1)	0.5 (22.9)	2.1 (23.4)	.053
VPT-Stay[j] alcohol bias (ms), mean (SD)	−0.2 (37.2)	1.9 (26.1)	3.2 (36.0)	1.6 (33.3)	.920
VST-ACT[k] alcohol bias (ms), mean (SD)	64.8 (511.8)	250.9 (596.4)	18.9 (495.8)	107.2 (538.0)	.200
VST-PAS[l] alcohol bias (ms), mean (SD)	116.8 (515.2)	118.3 (536.1)	232.7 (594.4)	157.1 (547.0)	.624

[a]VPT: visual probe task.
[b]AUDIT: Alcohol Use Disorders Identification Test.
[c]TLFB: Timeline Followback.
[d]AUQ: Alcohol Use Questionnaire.
[e]HAC: habitual alcohol consumption.
[f]RCQ: Readiness to Change Questionnaire (items 1, 2, and 3).
[g]MTQ: Motivation to Train Questionnaire.
[h]VPT-Go: VPT trials where the stimulus picture disappeared when the probe appeared.
[i]ms: milliseconds.
[j]VPT-Stay: VPT trials where the stimulus picture remained visible when the probe appeared.
[k]VST-ACT: visual search task with active beverage-related stimuli.
[l]VST-PAS: visual search task with passive beverage-related stimuli.

Motivation to Train (H3)

The MTQ demonstrated sufficient internal consistency, Cronbach alpha=.69. Exploratory principal axis factor analysis indicated a single factor. Therefore, the sum score was analyzed. There was a significant decrease in motivation to train over time, $F_{1,91}$*=54.377, *P*<.001, η_p^2=0.374, with no interaction with condition, indicating that motivation decreased similarly in all conditions. Participants' responses on the EVAL questions only differed between conditions on the question whether they would like to do more training sessions (EVAL-4), H_2=9.987, *P*=.007. Contrasts indicated that the VPT-G conditions scored significantly lower than both the VPT-P condition, *U*=356.0, *P*=.011, *r*=−.313, and the VPT-R condition, *U*=273.0, *P*=.004, *r*=−0.364.

Motivation to Change (H4)

The RCQ showed an overall increase in the degree to which participants planned to drink less after the training (a *lower* score on RCQ-2), $F_{1,91}$*=5.863, *P*=.017, η_p^2=0.061. However, there also was a significant interaction between time and condition, $F_{2,91}$*=3.865, *P*=.024, η_p^2=0.078. Tukey-adjusted contrasts indicated a lower motivation to drink less over time for the VPT-G condition, t_{91}=−2.985, *P*=.041, *r*=.299. The other RCQ items did not show significant effects. See Table 3 for an overview of estimated marginal and interaction means.

Additional Training Analyses

Participants differed in terms of the number of errors made during all training sessions, *H* (2)=9.093, *P*=.011, with participants in the VPT-G condition making more errors (mean 40.48, SD 25.91) than those in the VPT-P condition (mean 23.00, SD 11.96), *U*=781.0, *P*=.002, *r*=.374. The average reaction times over all sessions also differed significantly between all training conditions, *H* (2)=59.421, *P*<.001, with the VPT-P condition (mean 557.75, SD 33.05) being slower than the other conditions and the VPT-G condition (mean 373.71, SD 78.18) being faster than the other conditions (VPT-R, mean 516.49, SD 50.52).

Table 3. Training effects—estimated means.

Training effect	Placebo (VPT[a]-P)	Regular (VPT-R)	Game (VPT-G)	Time
TLFB[b] (drinks/7 days), mean (SE)				
Baseline	124.4 (14.4)	157.5 (14.5)	137.7 (13.7)	134.3 (8.2)
Posttraining	130.7 (14.4)	129.2 (14.5)	120.3 (13.7)	133.2 (8.2)
2-Week follow-up	127.9 (14.4)	127.1 (14.5)	137.8 (13.7)	130.0 (8.2)
Condition	133.9 (11.5)	125.0 (11.6)	138.7 (11.2)	
TLFB (binges/30 days), mean (SE)				
Baseline	128.9 (14.4)	139.3 (14.6)	140.5 (13.9)	136.4 (8.2)
Posttraining	122.2 (14.4)	138.2 (14.6)	133.4 (13.9)	135.6 (8.2)
2-Week follow-up	129.1 (14.4)	121.3 (14.6)	139.5 (13.9)	125.5 (8.2)
Condition	134.5 (12.7)	122.9 (12.7)	140.1 (12.3)	
MTQ[c], mean (SE)				
Baseline	92.9 (9.7)	94.8 (10.1)	100.7 (9.7)	113.6 (5.3)
Posttraining	96.0 (9.7)	90.7 (10.1)	91.9 (9.7)	75.4 (5.3)
Condition	85.0 (8.5)	102.9 (8.75)	95.6 (8.5)	
RCQ[d] -1, mean (SE)				
Baseline	106.3 (9.5)	76.0 (9.9)	91.6 (9.5)	99.0 (5.5)
Posttraining	110.4 (9.5)	89.3 (9.9)	93.5 (9.5)	90.0 (5.5)
Condition	86.5 (8.4)	99.9 (8.7)	97.2 (8.4)	
RCQ-2, mean (SE)				
Baseline	108.2 (9.5)	76.1 (9.8)	86.5 (9.5)	100.4 (5.6)
Posttraining	100.8 (9.5)	85.4 (9.8)	110.1 (9.5)	88.6 (5.6)
Condition	84.3 (8.4)	101.4 (8.7)	97.7 (8.4)	
RCQ-3, mean (SE)				
Baseline	91.6 (9.2)	69.8 (9.6)	88.6 (9.3)	97.5 (5.4)
Posttraining	116.9 (9.2)	83.4 (9.6)	104.7 (9.3)	87.5 (5.4)
Condition	84.1 (7.4)	103.1 (7.6)	90.2 (7.5)	
VPT-Go[e] alcohol bias (ms[f]), mean (SE)				
Baseline	77.1 (9.2)	118.1 (9.7)	80.9 (9.3)	104.8 (5.5)
Posttraining	105.0 (9.2)	76.5 (9.7)	103.3 (9.3)	82.2 (5.5)
Condition	91.7 (7.0)	87.8 (7.2)	101.0 (7.0)	
VPT-Stay[g] alcohol bias (ms), mean (SE)				
Baseline	79.8 (9.5)	103.3 (10.0)	92.6 (9.6)	106.9 (5.5)
Posttraining	105.6 (9.5)	88.6 (10.0)	91.0 (9.6)	80.1 (5.5)
Condition	99.6 (6.5)	88.4 (6.7)	92.4 (6.6)	
VST-ACT[h] alcohol bias (ms), mean (SE)[i]				
Baseline	64.8 (93.5)	226.6 (101.5)	18.9 (93.5)	103.4 (55.6)
Posttraining	50.2 (104.7)	535.1 (113.7)	28.0 (104.7)	204.4 (62.2)
Condition	57.5 (79.7)	380.8 (86.5)	23.5 (79.7)	
VST-PAS[j] alcohol bias (ms), mean (SE)[i]				
Baseline	116.8 (95.8)	153.6 (104.0)	232.7 (95.8)	167.7 (56.9)

Training effect	Placebo (VPT[a]-P)	Regular (VPT-R)	Game (VPT-G)	Time
Posttraining	29.6 (96.8)	178.9 (105.1)	235.5 (96.8)	147.9 (57.5)
Condition	73.2 (75.7)	166.1 (82.2)	234.1 (75.7)	

[a]VPT: visual probe task.

[b]TLFB: Timeline Followback (shows the number of standardized drinks during the week before the pretraining assessment).

[c]MTQ: Motivation to Train Questionnaire.

[d]RCQ: Readiness to Change Questionnaire (questions 1, 2, and 3).

[e]VPT-Go: VPT trials where the stimulus picture disappeared when the probe appeared.

[f]ms: milliseconds.

[g]VPT-Stay: VPT trials where the stimulus picture remained visible when the probe appeared.

[h]VST-ACT: visual search task with active beverage-related stimuli.

[i]Using regular analysis of variance procedure. For the all the other values, the Aligned Rank Transform procedure was used.

[j]VST-PAS: visual search task with passive beverage-related stimuli.

Discussion

Principal Findings

This study aimed to decrease attentional bias toward alcohol and hazardous drinking behavior in young adults by using a CBM-A training with game elements. After training, there was an overall decline in attentional bias on the VPT task, but this effect was primarily driven by the regular VPT training condition, where a stronger bias was observed at baseline. The training effect did not generalize to the VST task, nor was there a decline in alcohol use after the training. Motivation to train decreased equally in all conditions, indicating that the training indeed became boring over time, but also that the motivational elements of the Shots game could not sufficiently counteract this effect. Interestingly, motivation to change, with respect to planning to drink less in the future, increased in the regular and placebo training but decreased in the game training condition. Moreover, participants in the game condition indicated a lower motivation to continue training compared with the other conditions. Participants also differed with regard to overall speed and accuracy of responses during the training, which may be due to the more complex nature of the gamified version of the training or the level of engagement in the training.

These findings regarding the motivational effects of the training may have important implications on the potential risks involved with using certain types of game elements. Although the gamified training arguably looked fancier than the regular training, the game elements in this study merely consisted of upgraded visuals and a coin-based reward system. There was no story line and only limited progression and personalization options were available in the game. As such, it is likely that these rather minimal game elements alone were insufficient to increase motivation to train in our student sample. Moreover, most participants will likely compare any gamified training with what they believe a game experience should look and feel like [22]. If the game training experience then disappoints, motivation could indeed take a dive, even going below the level observed in the regular training conditions, because expectations were higher to begin with. This was perhaps reflected in our finding that participants in the game condition, specifically, were less motivated to continue training, as well as start drinking less, after the training. Finally, these results may also be related to the visual and auditory game elements used, which might have distracted participants from the training elements, rendering it less effective. Indeed, the standard training condition with no game elements did show a small change in attentional bias. Moreover, motivation to change increased only in the nongame conditions. It could be that the exposure to alcohol cues gave participants a push toward a readiness to change but only when there were no distracting game elements surrounding those cues. Although this last point is speculative, it is clear that some game elements may be detrimental not only to motivation to train but also to the training mechanisms themselves. More research is necessary in order to determine which game elements are best suited for cognitive training, and CBM-A in particular. For example, more levels, a background story, or the introduction of character development throughout the multiple sessions of the game could benefit participants' motivation. However, such additional elements would also have to be tested to see what their effect is on cognitive bias and related behavior.

Limitations

Some limitations apply to this study. First, despite hazardous drinking in a substantial part of the sample, the modest training effects in this sample may be partially due to a relatively small alcohol attentional bias at baseline, which is a known moderator of training effects [13,35,36]. However, it should be noted that this notion does not necessarily make CBM-A inappropriate in such samples. For example, Schoenmakers and colleagues [21] detected no bias at baseline in their sample but still found a positive bias after training in their control group and a negative bias in their experimental group. Furthermore, this study included a total of 624 critical training trials divided over 4 sessions. Although this number is similar to that used in other research (eg, [19]), other attentional bias modification studies have used markedly larger numbers (eg, Schoenmakers et al [21], where participants completed 2640 training trials over 5 sessions). Given the very likely dose-response relationship between use and effectiveness of cognitive training paradigms, the amount of training practice may have prevented the training from efficiently changing attentional bias. Finally, a recent meta-analysis [37] concluded that Web-based CBM-A studies usually show smaller effect sizes than laboratory-based studies. Although the assessments took place in the laboratory, it is possible that the option to train at home had a negative effect on participants' motivation, for example, by making participants take the training less seriously.

Conclusions

In sum, the novel game-like approach used in this study proved insufficient to motivate young adults to train, in comparison with a regular CBM-A training. In fact, some aspects of motivation appeared to deteriorate rather than improve, suggesting that gamification can have drawbacks if not done optimally. It could be concluded from this study that a point-based reward system in combination with fancy graphics does not satisfy participants' expectations of what constitutes a game. Because one expects a game to be fun, this may have detrimental effects on motivation. Moreover, when those game elements distract participants from the training elements, they may actually impair performance. A second notion that can be taken from this study is that the observed attentional biases toward alcohol as measured with the VPT in this heavy drinking student sample were remarkably low. Whether this has implications for the presence of attentional bias in adolescent samples in general or merely pertains to the VPT paradigm as a valid assessment measure of attentional bias remains to be determined by future research. If nothing else, however, these results underscore the importance of careful scientific evaluation before serious games are used as interventions.

Acknowledgments

The authors thank the students involved in the data collection for this study: Aida Alai, Guillaume Rensink, and especially Sylvia van Schie, who also provided the game's graphics. WJB is supported by the National Initiative Brain and Cognition grant 433-11-1510, MB by the NextLevel project grant 314-99-102, and RWW by the VICI grant 453-08-001, all funded by the Netherlands Organisation for Scientific Research (NWO).

Authors' Contributions

SSM was involved as a master's student.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Motivational questions.

[PDF File (Adobe PDF File), 20KB - games_v4i2e20_app1.pdf]

References

1. Singleton RA. Collegiate alcohol consumption and academic performance. J Stud Alcohol Drugs 2007 Jul;68(4):548-555. [Medline: 17568960]
2. Kuntsche E, Rossow I, Simons-Morton B, Ter Bogt T, Kokkevi A, Godeau E. Not early drinking but early drunkenness is a risk factor for problem behaviors among adolescents from 38 European and North American countries. Alcohol Clin Exp Res 2013 Feb;37(2):308-314 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01895.x] [Medline: 23240610]
3. Deutsch R, Strack F. Reflective and impulsive determinants of addictive behaviors. In: Wiers RW, Stacy AW, editors. Handbook of Implicit Cognition and Addiction. Thousand Oaks, CA: Sage; 2006:45-57.
4. Wiers RW, Bartholow BD, van den Wildenberg E, Thush C, Engels RC, Sher KJ, et al. Automatic and controlled processes and the development of addictive behaviors in adolescents: a review and a model. Pharmacol Biochem Behav 2007 Feb;86(2):263-283. [doi: 10.1016/j.pbb.2006.09.021] [Medline: 17116324]
5. Gladwin TE, Figner B, Crone EA, Wiers RW. Addiction, adolescence, and the integration of control and motivation. Dev Cogn Neurosci 2011 Oct;1(4):364-376 [FREE Full text] [doi: 10.1016/j.dcn.2011.06.008] [Medline: 22436562]
6. Wiers RW, Rinck M, Dictus M, van den Wildenberg E. Relatively strong automatic appetitive action-tendencies in male carriers of the OPRM1 G-allele. Genes Brain Behav 2009 Feb;8(1):101-106 [FREE Full text] [doi: 10.1111/j.1601-183X.2008.00454.x] [Medline: 19016889]
7. Field M, Duka T, Eastwood B, Child R, Santarcangelo M, Gayton M. Experimental manipulation of attentional biases in heavy drinkers: Do the effects generalise? Psychopharmacology (Berl) 2007 Jul;192(4):593-608. [doi: 10.1007/s00213-007-0760-9] [Medline: 17361393]
8. Grenard JL, Ames SL, Wiers RW, Thush C, Sussman S, Stacy AW. Working memory capacity moderates the predictive effects of drug-related associations on substance use. Psychol Addict Behav 2008 Sep;22(3):426-432 [FREE Full text] [doi: 10.1037/0893-164X.22.3.426] [Medline: 18778136]
9. Houben K, Wiers RW. Response inhibition moderates the relationship between implicit associations and drinking behavior. Alcohol Clin Exp Res 2009 Apr;33(4):626-633. [doi: 10.1111/j.1530-0277.2008.00877.x] [Medline: 19183132]

10. Bickel WK, Christensen DR, Marsch LA. A review of computer-based interventions used in the assessment, treatment, and research of drug addiction. Subst Use Misuse 2011;46(1):4-9 [FREE Full text] [doi: 10.3109/10826084.2011.521066] [Medline: 21190401]
11. Houben K, Wiers RW, Jansen A. Getting a grip on drinking behavior: training working memory to reduce alcohol abuse. Psychol Sci 2011 Jul;22(7):968-975. [doi: 10.1177/0956797611412392] [Medline: 21685380]
12. Wiers RW, Gladwin TE, Hofmann W, Salemink E, Ridderinkhof KR. Cognitive Bias Modification and Cognitive Control Training in Addiction and Related Psychopathology: Mechanisms, Clinical Perspectives, and Ways Forward. Clin Psychol Sci 2013 Jan 09;1(2):192-212. [doi: 10.1177/2167702612466547]
13. Eberl C, Wiers RW, Pawelczack S, Rinck M, Becker ES, Lindenmeyer J. Approach bias modification in alcohol dependence: do clinical effects replicate and for whom does it work best? Dev Cogn Neurosci 2013 Apr;4:38-51 [FREE Full text] [doi: 10.1016/j.dcn.2012.11.002] [Medline: 23218805]
14. Lindgren KP, Wiers RW, Teachman BA, Gasser ML, Westgate EC, Cousijn J, et al. Attempted Training of Alcohol Approach and Drinking Identity Associations in US Undergraduate Drinkers: Null Results from Two Studies. PLoS One 2015;10(8):e0134642 [FREE Full text] [doi: 10.1371/journal.pone.0134642] [Medline: 26241316]
15. National Institute on Alcohol Abuse and Alcoholism. The effects of alcohol on physiological processes and biological development. Alcohol Res Health 2005;28:125-131.
16. Goldman MS. Risk for substance abuse: Memory as final common etiological pathway. Psychol Sci 1999;10:196-198.
17. Beard C, Weisberg RB, Primack J. Socially anxious primary care patients' attitudes toward cognitive bias modification (CBM): a qualitative study. Behav Cogn Psychother 2012 Oct;40(5):618-633 [FREE Full text] [doi: 10.1017/S1352465811000671] [Medline: 22127022]
18. Dovis S, Van der Oord S, Wiers RW, Prins PJ. Can motivation normalize working memory and task persistence in children with attention-deficit/hyperactivity disorder? The effects of money and computer-gaming. J Abnorm Child Psychol 2012 Jul;40(5):669-681 [FREE Full text] [doi: 10.1007/s10802-011-9601-8] [Medline: 22187093]
19. Dennis TA, O'Toole L. Mental Health on the Go: Effects of a Gamified Attention Bias Modification Mobile Application in Trait Anxious Adults. Clin Psychol Sci 2014 Sep 1;2(5):576-590 [FREE Full text] [doi: 10.1177/2167702614522228] [Medline: 26029490]
20. MacLeod C, Rutherford E, Campbell L, Ebsworthy G, Holker L. Selective attention and emotional vulnerability: Assessing the causal basis of their association through the experimental manipulation of attentional bias. J Abnorm Psychol 2002 Feb;111(1):107-123. [Medline: 11866165]
21. Schoenmakers TM, de Bruin M, Lux IF, Goertz AG, van Kerkhof DH, Wiers RW. Clinical effectiveness of attentional bias modification training in abstinent alcoholic patients. Drug Alcohol Depend 2010 Jun 1;109(1-3):30-36. [doi: 10.1016/j.drugalcdep.2009.11.022] [Medline: 20064698]
22. Boendermaker WJ, Prins PJ, Wiers RW. Cognitive Bias Modification for adolescents with substance use problems - Can serious games help? J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):13-20 [FREE Full text] [doi: 10.1016/j.jbtep.2015.03.008] [Medline: 25843611]
23. Dandeneau SD, Baldwin MW. The inhibition of socially rejecting information among people with high versus low self-esteem: The role of attentional bias and the effects of bias reduction training. J Soc Clin Psychol 2004 Aug;23(4):584-603. [doi: 10.1521/jscp.23.4.584.40306]
24. Sobell MB, Sobell LC, Klajner F, Pavan D, Basian E. The reliability of a timeline method for assessing normal drinker college students' recent drinking history: utility for alcohol research. Addict Behav 1986;11(2):149-161. [Medline: 3739800]
25. Wiers RW, Hoogeveen KJ, Sergeant JA, Boudewijn Gunning BW. High- and low-dose alcohol-related expectancies and the differential associations with drinking in male and female adolescents and young adults. Addiction 1997 Jul;92(7):871-888. [Medline: 9293046]
26. Mehrabian A, Russell JA. A questionnaire measure of habitual alcohol use. Psychol Rep 1978 Dec;43(3 Pt 1):803-806. [doi: 10.2466/pr0.1978.43.3.803] [Medline: 740823]
27. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption--II. Addiction 1993 Jun;88(6):791-804. [Medline: 8329970]
28. Fleming MF, Barry KL, MacDonald R. The alcohol use disorders identification test (AUDIT) in a college sample. Int J Addict 1991 Nov;26(11):1173-1185. [Medline: 1743817]
29. Rollnick S, Heather N, Gold R, Hall W. Development of a short 'readiness to change' questionnaire for use in brief, opportunistic interventions among excessive drinkers. Addiction 1992 May;87(5):743-754. [doi: 10.1111/j.1360-0443.1992.tb02720.x]
30. De Voogd EL, Wiers RW, Prins PJ, Salemink E. Visual search attentional bias modification reduced social phobia in adolescents. J Behav Ther Exp Psychiatry 2014 Jun;45(2):252-259. [doi: 10.1016/j.jbtep.2013.11.006] [Medline: 24361543]
31. Staugaard SR. Reliability of two versions of the dot-probe task using photographic faces. Psychol Sci Q 2009;51:339-350.
32. Sportel BE, de Hullu E, de Jong PJ, Nauta MH. Cognitive bias modification versus CBT in reducing adolescent social anxiety: a randomized controlled trial. PLoS One 2013;8(5):e64355 [FREE Full text] [doi: 10.1371/journal.pone.0064355] [Medline: 23691203]

33. Pronk T, van Deursen DS, Beraha EM, Larsen H, Wiers RW. Validation of the Amsterdam Beverage Picture Set: A Controlled Picture Set for Cognitive Bias Measurement and Modification Paradigms. Alcohol Clin Exp Res 2015 Oct;39(10):2047-2055. [doi: 10.1111/acer.12853] [Medline: 26431117]
34. Wobbrock J, Findlater L, Gergle D, Higgins J. The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. 2011 Presented at: ACM Conference on Human Factors in Computing Systems (CHI '11); May 7-12, 2011; Vancouver, British Columbia p. 143-146.
35. Boettcher J, Leek L, Matson L, Holmes EA, Browning M, MacLeod C, et al. Internet-based attention bias modification for social anxiety: a randomised controlled comparison of training towards negative and training towards positive cues. PLoS One 2013 Sep;8(9):e71760 [FREE Full text] [doi: 10.1371/journal.pone.0071760] [Medline: 24098630]
36. Kuckertz JM, Gildebrant E, Liliequist B, Karlström P, Väppling C, Bodlund O, et al. Moderation and mediation of the effect of attention training in social anxiety disorder. Behav Res Ther 2014 Feb;53:30-40 [FREE Full text] [doi: 10.1016/j.brat.2013.12.003] [Medline: 24373984]
37. Mogoaşe C, David D, Koster EH. Clinical efficacy of attentional bias modification procedures: an updated meta-analysis. J Clin Psychol 2014 Dec;70(12):1133-1157. [doi: 10.1002/jclp.22081] [Medline: 24652823]

Abbreviations

ABPS: Amsterdam Beverage Picture Set
ADHD: attention-deficit/hyperactivity disorder
AUDIT: Alcohol Use Disorders Identification Test
AUQ: Alcohol Use Questionnaire
CBM-A: cognitive bias modification of attention
GLM: general linear model
HAC: habitual alcohol consumption
MTQ: Motivation to Train Questionnaire
RCQ: Readiness to Change Questionnaire
TLFB: Timeline Followback
VPT: visual probe task
VPT-G: gamified VPT training
VPT-P: placebo VPT training
VPT-R: regular VPT training
VST: visual search task

Edited by G Eysenbach; submitted 08.08.16; peer-reviewed by C Beard, A Taly; comments to author 07.09.16; revised version received 21.09.16; accepted 21.09.16; published 06.12.16.

Please cite as:
Boendermaker WJ, Sanchez Maceiras S, Boffo M, Wiers RW
Attentional Bias Modification With Serious Game Elements: Evaluating the Shots Game
JMIR Serious Games 2016;4(2):e20
URL: http://games.jmir.org/2016/2/e20/
doi:10.2196/games.6464
PMID:27923780

Original Paper

Cardiopulmonary Resuscitation Training by Avatars: A Qualitative Study of Medical Students' Experiences Using a Multiplayer Virtual World

Johan Creutzfeldt[1,2], MD, PhD; Leif Hedman[1,3], PhD; Li Felländer-Tsai[1,2], MD, PhD

[1]Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, Stockholm, Sweden
[2]Center for Advanced Medical Simulation and Training, Karolinska University Hospital, Stockholm, Sweden
[3]Department of Psychology, Umeå University, Umeå, Sweden

Corresponding Author:
Johan Creutzfeldt, MD, PhD
Department of Clinical Science, Intervention and Technology (CLINTEC)
Karolinska Institutet
K32
Karolinska University Hospital
Stockholm, 14186
Sweden
Phone: 46 8 585 82102
Fax: 46 8 585 82224
Email: johan.creutzfeldt@ki.se

Abstract

Background: Emergency medical practices are often team efforts. Training for various tasks and collaborations may be carried out in virtual environments. Although promising results exist from studies of serious games, little is known about the subjective reactions of learners when using multiplayer virtual world (MVW) training in medicine.

Objective: The objective of this study was to reach a better understanding of the learners' reactions and experiences when using an MVW for team training of cardiopulmonary resuscitation (CPR).

Methods: Twelve Swedish medical students participated in semistructured focus group discussions after CPR training in an MVW with partially preset options. The students' perceptions and feelings related to use of this educational tool were investigated. Using qualitative methodology, discussions were analyzed by a phenomenological data-driven approach. Quality measures included negotiations, back-and-forth reading, triangulation, and validation with the informants.

Results: Four categories characterizing the students' experiences could be defined: (1) Focused Mental Training, (2) Interface Diverting Focus From Training, (3) Benefits of Practicing in a Group, and (4) Easy Loss of Focus When Passive. We interpreted the results, compared them to findings of others, and propose advantages and risks of using virtual worlds for learning.

Conclusions: Beneficial aspects of learning CPR in a virtual world were confirmed. To achieve high participant engagement and create good conditions for training, well-established procedures should be practiced. Furthermore, students should be kept in an active mode and frequent feedback should be utilized. It cannot be completely ruled out that the use of virtual training may contribute to erroneous self-beliefs that can affect later clinical performance.

(JMIR Serious Games 2016;4(2):e22) doi:10.2196/games.6448

KEYWORDS

avatars; cardiopulmonary resuscitation; educational technology; medical students; experiences; multiplayer virtual worlds; patient simulation; virtual learning environments

Introduction

New generations of students, having grown up playing videogames and computer games, embrace these technologies for learning [1,2]. It has been shown that videogames used for entertainment may possess a transferable effect to medical skills [3]. Although several studies within diverse educational areas find serious games to be of benefit, more critical analyses of the effectiveness of computer-based serious games have questioned this by pointing out ambiguities and a need for further research [4,5]. In a review of serious games designed for health care professionals, Wang et al found positive training effects [6]. However, effects were hard to compare and heterogeneities and methodological difficulties were noted [6].

Several reports on benefits of this learning tool exist in the medical field [7-10]. Furthermore, skills and procedures taking place in group settings (as is often the case in the medical field) may potentially utilize training in multiplayer virtual worlds (MVWs) [11,12]. In contrast to so called *virtual patients*, MVWs enable training for teamwork skills, such as team coordination and team communication.

Traditionally, cardiopulmonary resuscitation (CPR) training is performed on mannequins under instructor supervision, focusing on single-person psychomotor skills, but in the last two decades alternative forms of basic CPR training have been introduced [13-15], including CPR team training using MVWs [16].

Studies of serious games would benefit from moving the focus away from the acquisition of CPR knowledge and skills, and instead focus on how students think and feel during MVW-CPR training. Furthermore, this approach may also indicate potential uses of MVWs and give rise to new questions about the use of MVWs for medical education. By using qualitative research methodology, new knowledge will be generated about this phenomenon, contributing to a novel and rapidly developing educational field. The aim of this study was to explore medical students' experiences using MVW-CPR in groups via broad qualitative analyses.

Methods

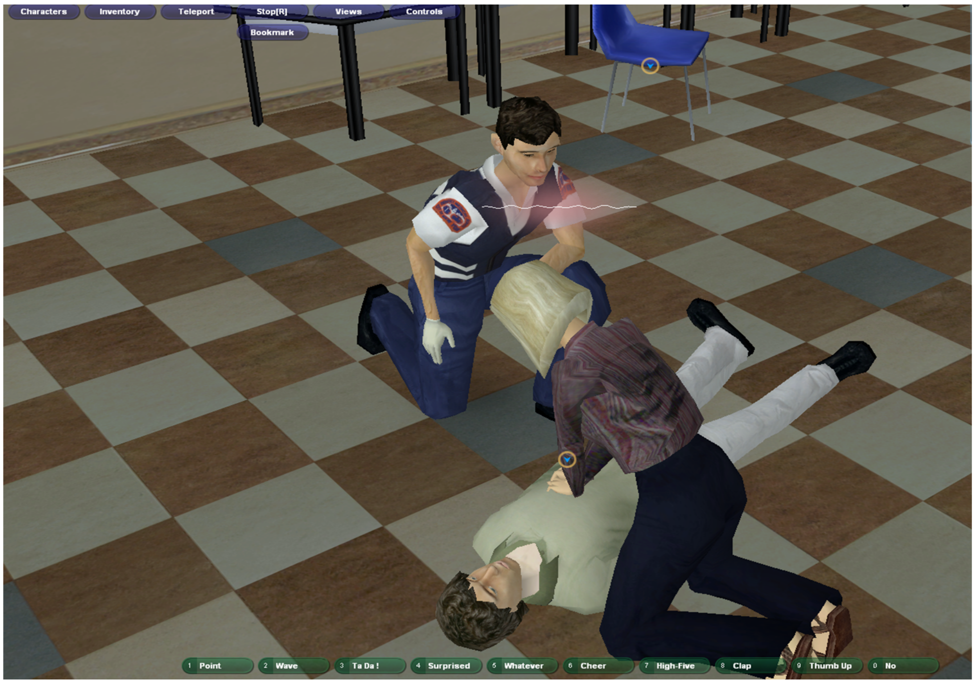
After regional research ethics committee approval, 12 first year medical students at Karolinska Institutet (Stockholm, Sweden) were recruited to the study. Inclusion criteria included elementary computer knowledge and previous CPR training in medical school. Exclusion criteria were restricted to any students who had previously used a virtual world-based serious game. To elucidate how virtual world training in groups was conceived, the students were invited to share their thoughts and experiences in focus group discussions.

The study was an after-training follow-up in which the trainees served as informants. Convenience sampling was used for this study. All subjects were enrolled by answering an invitation distributed by email to all first semester medical students at Karolinska Institutet. Upon enrollment, written consent was obtained and confidentiality was guaranteed by the authors.

The virtual environment used during training (On-Line Interactive Virtual Environment) was developed in conjunction by the authors, coresearchers at Stanford University Medical Media and Information Technologies, and game developers at Forterra, Incorporated (San Mateo, CA). The MVW included a school building (interior and exterior) and a parking lot (Figure 1), and was accessed by standard personal computers connected to a server on the Internet. The subjects interacted by use of their avatars and a headset (voice over Internet protocol). The avatars' movements and actions were controlled by a keyboard and computer mouse. Conventional gaming commands were used. Some essential commands for examination and treatment were accessible on *action tab-lists* defined by situation. During MVW training, the computers were isolated to prevent users from overhearing each other.

Before engaging in the training, participants were introduced to the virtual world by a virtual world instructor (avatar) that aided the trainees with navigational and procedural software commands, and taught the participants how to communicate with others. Altogether, familiarization lasted for approximately 15 minutes, and was followed by four scenarios in which a team of three trainees were instructed to act upon the need of the situation. In all scenarios, there was a victim-avatar (controlled by the instructor) that collapsed. The first two scenarios took place in a classroom, in which a teacher collapsed due to a cardiac arrest. The students (trainees) witnessed the event and had to take action: circulatory arrest was to be diagnosed and CPR started in accordance with existing bystander-CPR guidelines [17]. These guidelines require the rescuers to quickly start CPR when appropriate, call for help after circulatory arrest has been confirmed, and to relieve the rescuers in order to maintain the effectiveness of chest compressions. The third and fourth scenarios occurred outside the school building in a parking lot, where a person close to of a group of students (trainees) collapsed. The participants had to perform in the same manner as described above. In addition to taking care of the victim, the participants were expected to run to a phone to call 911, and to guide the paramedics to the victim and give a brief report. The scenarios lasted for 5-7 minutes each and ended when an avatar-paramedic entered the scene. After each scenario the participants were reassembled to receive real-world oral feedback from the instructor. The feedback focused on adherence to the CPR algorithm and how to coordinate activities in the resuscitation team.

Figure 1. Interior of the multiplayer virtual world classroom, with a student (trainee) taking care of a victim as a paramedic arrives.



As part of the MVW-CPR training program, a second identical training session was undertaken after 6 months. A semistructured group discussion approach was used, in which the aim of the discussion was to explore how the subjects had experienced the virtual world training. The discussants were the three participants who had just trained in the MVW, and the moderator was the instructor (JC). The discussion started 10-to-15 minutes after the end of the second training session. Data was collected by use of an audio recorder, and the discussions lasted for 55-to-65 minutes. Additional notes were collected by the moderator.

An interview guide was used (Multimedia Appendix 1), focusing on the central question of experiences, but the spontaneous emergence of topics was allowed. Participants were encouraged to speak freely and at length, and the moderator's role was to pose open questions, highlight inconsistencies, and follow up on ambiguities. A verbatim transcription of the audio material was performed, and validation of the transcriptions was made by comparing parts of transcripts and recordings (JC, LH). By examining the degree of novelty of information in the following discussions, saturation was reached after 4 focus discussions with 12 participants.

A qualitative methodology described by Malterud was used [18]. This approach aims to describe a phenomenon by generating descriptions, categories, models, or theories. The aim of the analysis was to get a broad picture of the experience of using MVW for CPR training among medical students. A phenomenological data-driven approach was used, including analytical reduction, distillation, and aggregation [18]. Transcriptions were read back-and-forth independently by two researchers (JC, LH) who were blinded to the identity of the subjects. Themes characterizing the transcripts were negotiated, and a matrix was created in which the identified meaning bearing units from the informants were arranged according to theme. To enable further decontextualizing, the themes were replaced by codes that were refined, split, and combined as the process of connecting the analysis to the transcriptions and notes continued. Categories evolved as end products. To assert quality in the evolving process of analysis, grouping, and structuring of the decontextualized material, discussions and negotiations took place between two authors (JC, LH). All discussions were in Swedish and quotes were translated to English by the authors. Seven informants representing different gaming backgrounds, sexes, and experiences during the training were consulted at a late stage of analysis to verify and comment on the results.

Results

The participants are characterized in Table 1.

Quantitative Data on Self-Efficacy, Mental Strain, and Concentration

Following the same protocol as a previous study, before-training and after-training assessments of the subjects' level of self-efficacy (using a 5-item validated instrument) was performed, and the levels of mental strain and concentration were assessed during the training using questions from validated instruments [16]. These assessments revealed that the level of mental strain in general was low and stable (mean 22/100, standard deviation [SD] 19 during the first scenario; mean 20/100, SD 14 during the last scenario) and concentration was moderate (mean 60/100, SD 13 during the first scenario; mean 65/100, SD 18 during the last scenario). Self-efficacy ratings were high before training (mean 5.8/7, SD 0.8) and increased further after training (mean 6.3/7, SD 0.6; $P<.001$, Wilcoxon signed rank test). A theoretical triangulation against these quantitative data was carried out after the qualitative analysis.

Table 1. Background demographics of the subjects.

Characteristic	Total (N=12)	Male (n=6)	Female (n=6)
Age, mean (standard deviation)	22.7 (2.6)	23.3 (3.6)	22.2 (1.2)
Number with access to computer at home	12	6	6
Level of computer experience[a], median (range)	2 (1-3)	2 (2-3)	2 (1-2)
Videogame and computer game use[b], median (range)	1 (0-3)	2 (1-3)	0.5 (0-1)

[a]Level of computer experience was graded on a 0-to-3 Likert-type scale (0=none, 1=beginner, 2=experienced, and 3=very high or expert).

[b]The use of computer and video games was graded on a 0-to-5 Likert-type scale (0=none, 1=less than once a month, 2=once every second week, 3=once a week, 4=several times every week, and 5=every day).

Qualitative Analysis

The following four categories evolved as end products: (1) Focused Mental Training, (2) Interface Diverting Focus From Training, (3) Benefits of Practicing in a Group, and (4) Easy Loss of Focus When Passive.

Focused Mental Training

The participants generally enjoyed the training. In particular they mentioned that MVW was a reasonable and good way to go through the procedure of CPR in their minds. Their focus was on when and how to perform CPR. Their perceptions of it being a cognitive training are illustrated by the following representative quotes:

> *…it makes you improve each time you repeat… it feels like a very good complement - you really want to do more times.* [Woman, 21, little previous gaming experience]

> *Yes sometimes you think about that, you know, thinking ahead in the game and well to be a step ahead, kind of…* [Man, 30, moderate previous gaming experience]

However, frequent comments indicated that realism was lacking in the scenarios. There was lack of physical realism (virtual versus real-world) and participants were unable to perform all parts of the CPR procedure hands-on. The participants also indicated that the level of mental stress, although not completely absent, was much lower than would be expected in a real-world CPR event, exemplified by:

> *Alright, it feels like I remember what we have done, but it doesn't feel like training because it is so unrealistic to me.* [Woman, 22, little previous gaming experience]

> *… it is not, no emotional tension at all, possibly there is some mental challenge, you have to think about, at least a bit about, what there is to do, but it´s no, it's no stress in that sense.* [Man, 19, moderate previous gaming experience]

Interface Diverting Focus From Training

All participating medical students could relate to videogames and computer games, but the level of gaming experience varied. Despite this variability, the interface to the virtual world and the quality of the virtual world received a great deal of attention, substantiated in the following discussion. Students with less videogame and computer game experience commented on this issue, and indicated that they were unfamiliar with how they should interact in the virtual environment and control the avatar. This lack of familiarity had several consequences: it made it difficult to control and navigate the avatar, resulting in negative reactions; and trainees unaccustomed to this technology felt more distant to what was going on in the virtual setting.

> *Because, you know, if you haven't played computer games it would take like a week before you felt comfortable in these movements and where to look and how it works…* [Man, 21, moderate previous gaming experience]

> *I think I was a bit afraid to press the wrong button so I stood somewhat passive, so to test - nope, it didn't work.* [Woman, 20, little previous gaming experience]

Conversely, students with more videogame and computer game experience demonstrated a tendency to criticize and compare the interface and the environment with previous experiences:

> *…exactly, I think that it can help when you can choose yourself, you know, build [an avatar] yourself a bit, it is quite fun.* [Man, 20, large previous gaming experience]

> *[An improvement would be] maybe just monitoring [on the computer screen] on how, how well or bad this person [the victim] feels, kind of.* [Man, 22, large previous gaming experience]

To find further support, a post hoc quantitative analysis of meaning-bearing units was performed, focusing on technical difficulties versus sex. In this analysis, female participants addressed their own experienced difficulties five times, whereas males never did. Males mentioned the importance of gaming skills, in general terms, to act in the virtual world a total of seven times. This concept was never mentioned by females.

Benefits of Practicing in a Group

MVW technology enables trainees to learn in team constellations. In all virtual world scenarios, the participants trained together in groups of three. The strengths with this concept arose repeatedly during the discussions, and could be broken down into two subgroups: Practicing a Team-Based Activity in a Group, and Training in a Group is More Engaging.

Practicing a Team-Based Activity in a Group

CPR is often carried out in team settings. In the virtual world, the trainees could perform group activities that resembled what could be expected in real life; they were able to communicate to inform, seek support, and make decisions. Some team aspects of the CPR guidelines could also be simulated (eg, starting resuscitation while someone else was calling for help, relieving each other during CPR):

> *It was really good because otherwise you would have sat down yourself and done everything, but now I have to think about who does what; should I do that now or… and that was much better. You learn to cooperate in a completely different way than when you just do it yourself, and then when you actually join other persons too which do the same thing you get stunned, [what] should I do now? But you learn how to divide the tasks, I think, in quite short time.* [Woman, 21, little previous gaming experience]

> *Yes cooperate, and it's actually where the biggest problems lie… Moderator: And here the training contributed? Yes, maybe it can't be trained in any other way than just like this.* [Man, 22, large previous gaming experience]

Training in a Group is More Engaging

Some participants declared that performing the training together with others increased their engagement and made the training feel more fruitful (eg, making it easier to suspend feelings of unreality, and prosper from feedback and support from peers during training):

> *When, when I was coached kind of, like what they told me, I liked it, [it] was good…Then, besides, I guess it was much more fun [working together].* [Man, 21, moderate previous gaming experience]

> *Well, you get a little more focused if there are two others watching and three others looking who know it [what to do].* [Woman, 22, little previous gaming experience]

Easy Loss of Focus When Passive

A large category of conversation focused on feeling engaged in the task during the virtual training. Direct statements, as well as many ideas on how to increase the sense of directedness, arose during the discussions. In general, the students felt most challenged in the beginning of the training. With the addition of similar scenarios, a common experience indicated that the training became repetitive and less demanding. When there was a demand for action the participants were engaged, but when they were less active, they easily lost focus and got bored. It was pointed out that the trainee having the most mentally-demanding tasks was more focused, whereas others (eg, awaiting the arrival of paramedics) quite easily lost the sense of engagement in the endeavor, exemplified by the following quotes:

> *…and before, when it is tedious, you think about completely different things that, well, have to do with life outside.* [Man, 30, moderate previous gaming experience]

> *I think it is boring when you get the task to wait for the ambulance because then you really notice how long time this takes.* [Woman, 23, very little previous gaming experience]

Discussion

Principal Findings

Using serious games for CPR training was a novel experience for all participants. Several benefits of this training tool proposed in the literature (eg, the focus of the cognitive part of the training and the added value of group practice) have also reoccurred among our four categories. However, our results also draw attention to other important properties of scenario-based MVW training when used for medical education.

The most important finding was the close relationship between the level of activation and degree of difficulty, and the level of engagement reported by the subjects. One of the commonly assumed strengths using serious games is the capacity to engage the *player* [11,19]. Our data support this, as all participants were positive about the experience and gave examples of their engagement in the virtual world training. This engagement could be an effect of the training tool itself, but might also be related to the seriousness and importance of the subject. Conversely, the inherent characteristic of this technology to give rise to engagement cannot be taken for granted, as pointed out by Choi and Baek [20]. Conceptual similarities exist between the construct of *engagement* and that of *flow*, as described by Csikszentmihalyi [21]. In his model of optimal experience, Csikszentmihalyi predicts a flow experience when the level of challenge and the level of personal skills are matched [21]. Based on our results, this model seems appropriate. The CPR scenarios were quite repetitive and the level of difficulty did not increase, so the subjects (presumably getting more skilled) tended to experience less challenge, and hence moved from a sense of anxiety and arousal towards one of control, relaxation, and boredom. This finding underscores the concept of *leveling* in gaming practice (ie, a progressive increase in difficulty to maintain challenge and motivation) that should also be considered in serious games.

Focusing on our psychometric data, mean concentration (a conceptually important component of *flow*) displayed a tendency to increase after the first training scenario. Triangulating these data with the students' experiences could imply that part of the increase is to be ascribed to the increased proficiency of maneuvering in the virtual environment and overcoming technical difficulties.

The seemingly close relationship between activation and challenge with engagement might be more pronounced than that experienced in real-life learning. According to constructivist learning theories, as the learner gets more accustomed and proficient, he or she moves into more complex reasoning and deeper understanding [22]. Accordingly, this development might not be as pronounced in a virtual world, and engagement in a virtual world might be more dependent on novelty and challenge. The assumption that a virtual world offers a less rich environment, in which the altered perception in the learning space affects our ways of learning, should be further elucidated. Furthermore, the use of serious games for tasks that may be monotonous can be questioned, since one of the key features of this educational tool (*engagement*) may be lost.

During training, feedback on performance was mainly given by peers in the virtual world, and by an instructor in the real-world immediately after each scenario. Overall, this feedback was appreciated, but the participants asked for more direct feedback within the virtual world. The virtual environment provides possibilities that tantalize the user to test and vary (eg, actions, appearances, and surroundings). In particular, with such characteristics incorporated into the virtual world, it is not surprising that the participants, whether previously accustomed to computer games or not, seek immediate feedback. Serious games lacking such features might render less active experimentation, and possibly less engagement [20].

The training was characterized as being mainly cognitive. Participants were accustomed to traditional CPR training, and several trainees noted the lack of psychomotor skills training as an important insufficiency. During training, the level of mental strain was low to moderate; this was confirmed in the discussions, and lack of stress was seen as another unrealistic feature of the virtual world training. The low level of stress has the potential to create a more structured and optimal learning situation during the initial phase of training. However, without training in a stressful and complex environment, transfer to real-world CPR situations may be hampered.

Discussions with the participants demonstrated an attitude of high capability in bystander-CPR. This theme was reflected by an attitude of mastery and a demand for more complex and difficult scenarios; there were many creative suggestions for variation and increased difficulty. In a previous study, we found that medical students reported that a weaknesses of MVW-CPR was in the area of, “tasks too easy, more options wanted” [16]. These finding agree with the increase in self-efficacy beliefs. It can be argued that shifting from a clinical high-stakes environment to a simulated (at least in this case) low stakes training environment might alter self-efficacy beliefs. Such an effect would be predicted by Bandura’s ideas about cost of failure, in which increased social cost tends to lower self-efficacy beliefs more if performance does not meet the needs of the situation [23]. If situations shift from partly decontextualized learning and training to high-stakes clinical reality, problems could arise in terms of misconceived self-beliefs and attitudes. Based on this issue, some positive aspects of using virtual worlds for learning and training (eg, engagement and experimentation) might have a downside of inadequate attitudes about situational demands and individual capabilities, as discussed by Wang et al [6]. Taking such risks into consideration, using virtual worlds for training in medicine (and probably other high risk professions) should be addressed with caution, not only by the trainees, but also by teachers, instructors, managers, and developers.

Serious games have been mentioned as an attractive alternative for digital natives [19], although this argument has also been disputed [24,25]. In this study, all participants were between 20 and 30 years of age, and the level of previous videogame and computer game experience varied. There was no clear correlation between age and experiences in the virtual training environment. The participants that were most clear about the lack of realism in the MVW were among the youngest. Therefore, the common belief that younger people would ask for more (and more easily accept) learning in virtual environments, could be questioned. These attitudes are likely attributable to more factors than simply age alone.

One category highlighted in this study was related to the influences of the interface. Most participants held strong opinions about its effects. Sweden is a highly computer technology-developed country, and students are accustomed to using computers; the Internet is regularly accessed for private and educational purposes [26]. Despite this general trend, students that were less experienced with computer games often mentioned difficulties with navigation in the MVW, along with other technical problems, whereas students with more experience tended to be forgiving towards software deficiencies. Representatives of the latter group tended to compare the MVW-CPR training with previous gaming experiences, and provided suggestions about how to make the interface and software more compelling, immersive, and enjoyable. To some extent, the opinions of the less-experienced and more-experienced participants are contradictory. Making use of high-end technological solutions that create more options for interactions, with the goal of making the experience more realistic, complex, and unpredictable, could render a virtual world that takes longer to get familiarized with. Interest in technology and computer games will vary among different users in medical fields, so a balance between technical solutions and usability must always be considered [27]. Furthermore, in the partly-experimental setting of this study, technical support was readily accessible, and interest from the instructors was high. Using this kind of training tool in an average educational setting also emphasizes the need to keep such technologies manageable for the users.

General discussions about videogames and computer games indicate that sex differences is a common issue; females and males seem to be attracted to different game genres [28]. Males and females also spend different amounts of time in different computer-related activities. It has also been reported that videogames and computer games have different meanings to men and women [29,30], and that men and women are attracted by different features in virtual worlds [31]. Kron et al discovered sex differences concerning attitudes towards video games among medical students [1], although a simplistic gender view is questioned by others [32]. Addressing this issue was not an aim of this study, making it hard to draw any further conclusions from present work on this issue. However, there seem to be some differences between males and females in our findings. Among the subjects, more men had a history of using videogames or computer games: this likely explains the tendency for female participants to focus more on technical difficulties in the MVW. Conversely, this trend does not explain why the importance of the training (versus importance of the interface itself) was raised more often by the females. Male participants appeared to display a more relaxed attitude towards the seriousness of the topic, and showed an interest in discussing how the tasks could be made more challenging, and the technical solutions more appealing. This difference in the importance of

CPR competence among future physicians is mirrored by our previous results that demonstrated differences in self-efficacy beliefs before virtual world training among males and females [16].

In medicine, insufficient teamwork skills have been identified as a common cause for suboptimal performance and harm [33-35]. This problem also seems to be true during CPR [36,37]. Our findings indicate that the participants appreciated the team focus, and the creation of an atmosphere of shared tasks and responsibilities seemed to be of importance for engagement in the scenarios. Not only was there a common belief in the strength of practicing a team endeavor in a team setting, but the students also considered it to be more fun, inspiring, and rewarding to train together. Inferring theories of situated learning [38], it is likely that positive experiences might occur due to the students learning with their peers. This matter, investigated by deNoyelles and Seo, might be of great importance for virtual world learning [39].

The participants trained in four short virtual world scenarios on two occasions. Several participants were aware of the rapid deterioration of CPR skills, and suggested MVW training as a good way to repeat and retrain. Distributed training using self-directed methods, such as a serious game technology, seems like an attractive option for retraining CPR.

Our group of informants consisted of a very homogenous group of early Swedish medical students, making it difficult to generalize our results. However, we believe that our findings can be applicable to other virtual world scenario-based training programs in medicine, in which clinically inexperienced (but at least moderately computer-experienced) users can train while reflective feedback is given. The strengths of this study lie in its novelty and ability to highlight certain characteristics of medical team training in virtual worlds, and how these MVWs are experienced. Methodologically we have also triangulated our results with psychometric process variables and previous results to reach better credibility [16].

Further investigations should involve studies of more heterogeneous groups, and include outcome measures focusing on behaviors and performance after training (*transfer*). Studies focusing on when and where virtual world training is particularly effective would also be of interest. The perceptions and experiences of instructors, teachers, and other stakeholders involved in this educational technology are also warranted. In our data there were many suggestions about how to create better and more challenging scenarios, and make use of more exciting technical solutions. Adding stressful elements, such as a real-time timing and scoring that reflect chances for successful resuscitation, unforeseen interruptions, or multiple emergency medical conditions might increase the experienced levels of stress, and could be included when the learner has completed the basic procedural and team-oriented steps. Technology is advancing quickly in this area, and it would also be interesting to understand which features would facilitate learning. There is a definite risk of being captured by the possibilities of this learning tool, and how it is used and varied among game developers. Instead, it is important to utilize experience-based and scientifically grounded knowledge in the field of learning psychology.

There are several limitations of this study. Our participants were 12 first year Swedish medical students that actively answered a call for participants. Adding more informants could have added more experiences and unraveled additional categories. The scenarios were all CPR-related and the virtual training environment was partly developed for this study, and lacked many features of real-world scenarios (eg, noises, bystanders). Although our study was designed for a rich input of data, the aim of understanding how virtual world team CPR training is experienced cannot be fully reached. Experiences depend on individuals' subjectivity, how the teams are arranged, and the learning situation in which the study is performed.

Conclusions

Four categories—*Focused Mental Training, Interface Diverting Focus From Training, Benefits of Practicing in a Group, and Easy Loss of Focus When Passive*—illustrate the phenomenon of virtual world team CPR training among medical students. In order to be successful, we suggest that the use of scenario-based virtual world team training should address how to actively engage users in the training of shorter, well-established behaviors, and focus on procedures that contain a high degree of group member interaction and feedback. Learning by use of MVW training occurs on several levels, and pedagogic validity should be examined when changing from traditional real-world training to that of a virtual world.

Acknowledgments

We would like to express our sincere gratitude to Christopher Medin, BS, Department of Biomedical Technology at Karolinska University Hospital, for technical assistance and acting during virtual world training, and Marie Eliasson, secretary, Department of Anaesthesiology and Intensive Care at Karolinska University Hospital, for professional transcriptions of audio recordings. This study was fully financed by research grants from the Wallenberg Global Learning Network, Stockholm County Council and EU Research Programme Goal 1, Northern Sweden, and the Stockholm County Council.

Authors' Contributions

All authors participated in the conception and overall design of the study. JC and LH designed the interview guide and chose qualitative methods for interpretations and analysis of the data. JC collected data and moderated the group discussions. JC and LH carried out the analysis, interpretation, and triangulation of the data. JC drafted the manuscript, and LH and LFT contributed with critical revisions. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview guide for focus group discussions.

[PDF File (Adobe PDF File), 17KB - games_v4i2e22_app1.pdf]

References

1. Kron FW, Gjerde CL, Sen A, Fetters MD. Medical student attitudes toward video games and related new media technologies in medical education. BMC Med Educ 2010 Jun 24;10:50. [doi: 10.1186/1472-6920-10-50] [Medline: 20576125]
2. Schwarz D, Štourač P, Komenda M, Harazim H, Kosinová M, Gregor J, et al. Interactive algorithms for teaching and learning acute medicine in the network of medical faculties MEFANET. J Med Internet Res 2013 Jul 08;15(7):e135 [FREE Full text] [doi: 10.2196/jmir.2590] [Medline: 23835586]
3. Kolga SM, Hedman L, Enochsson L, Kjellin A, Felländer-Tsai L. Transfer of systematic computer game training in surgical novices on performance in virtual reality image guided surgical simulators. Stud Health Technol Inform 2008;132:210-215. [Medline: 18391288]
4. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, et al. The effect of educational games on medical students' learning outcomes: a systematic review: BEME Guide No 14. Med Teach 2010 Jan;32(1):16-27. [doi: 10.3109/01421590903473969] [Medline: 20095770]
5. Pasquier P, Mérat S, Malgras B, Petit L, Queran X, Bay C, et al. A serious game for massive training and assessment of French soldiers involved in forward combat casualty care (3D-SC1): development and deployment. JMIR Serious Games 2016 May 18;4(1):e5 [FREE Full text] [doi: 10.2196/games.5340] [Medline: 27194369]
6. Wang TJ. Educating avatars: on virtual worlds and pedagogical intent. Teaching in Higher Education 2011 Dec;16(6):617-628. [doi: 10.1080/13562517.2011.570433]
7. Andrade AD, Bagri A, Zaw K, Roos BA, Ruiz JG. Avatar-mediated training in the delivery of bad news in a virtual world. J Palliat Med 2010 Dec;13(12):1415-1419. [doi: 10.1089/jpm.2010.0108] [Medline: 21091407]
8. LeFlore JL, Anderson M, Zielke MA, Nelson KA, Thomas PE, Hardee G, et al. Can a virtual patient trainer teach student nurses how to save lives--teaching nursing students about pediatric respiratory diseases. Simul Healthc 2012 Feb;7(1):10-17. [doi: 10.1097/SIH.0b013e31823652de] [Medline: 22228285]
9. Patel V, Aggarwal R, Osinibi E, Taylor D, Arora S, Darzi A. Operating room introduction for the novice. Am J Surg 2012 Feb;203(2):266-275. [doi: 10.1016/j.amjsurg.2011.03.003] [Medline: 21703594]
10. Wiecha J, Heyden R, Sternthal E, Merialdi M. Learning in a virtual world: experience with using second life for medical education. J Med Internet Res 2010 Jan 23;12(1):e1 [FREE Full text] [doi: 10.2196/jmir.1337] [Medline: 20097652]
11. Oblinger D. The Next Generation of Educational Engagement. J Interactive Media Educ 2004 May 21;2004(1):10-20. [doi: 10.5334/2004-8-oblinger] [Medline: 9178381]
12. Yee N. The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. PRESENCE Teleoperators Virtual Environments 2006 Jun;15(3):309-329. [doi: 10.1162/pres.15.3.309]
13. Braslow A, Brennan RT, Newman MM, Bircher NG, Batcheller AM, Kaye W. CPR training without an instructor: development and evaluation of a video self-instructional system for effective performance of cardiopulmonary resuscitation. Resuscitation 1997 Jun;34(3):207-220. [Medline: 9178381]
14. Owen H, Mugford B, Follows V, Plummer JL. Comparison of three simulation-based training methods for management of medical emergencies. Resuscitation 2006 Nov;71(2):204-211. [doi: 10.1016/j.resuscitation.2006.04.007] [Medline: 16987587]
15. Wik L, Myklebust H, Auestad BH, Steen PA. Retention of basic life support skills 6 months after training with an automated voice advisory manikin system without instructor involvement. Resuscitation 2002 Mar;52(3):273-279. [Medline: 11886733]
16. Creutzfeldt J, Hedman L, Medin C, Heinrichs WL, Felländer-Tsai L. Exploring virtual worlds for scenario-based repeated team training of cardiopulmonary resuscitation in medical students. J Med Internet Res 2010 Sep 03;12(3):e38 [FREE Full text] [doi: 10.2196/jmir.1426] [Medline: 20813717]
17. International Liaison Committee on Resuscitation. 2005 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. Part 2: adult basic life support. Resuscitation 2005;67(2-3):187-201. [doi: 10.1016/j.resuscitation.2005.09.016] [Medline: 16324988]
18. Malterud K. Qualitative methods in medical research: an introduction. Lund, Sweden: Studentlitteratur AB; 2009.
19. Prensky M. Digital game-based learning. New York: McGraw-Hill; 2001.
20. Choi B, Baek Y. Exploring factors of media characteristic influencing flow in learning through virtual worlds. Comput Educ 2011 Dec;57(4):2382-2394. [doi: 10.1016/j.compedu.2011.06.019]
21. Csikszentihalyi M. Flow: The Psychology of Optimal Experience (P.S.). New York: Harper Perennial Modern Classics; 2008.

22. Vygotsky L, Cole M. Mind in society: the development of higher psychological processes, 14 edition. Cambridge: Harvard University Press; 1978.
23. Bandura A. Self-efficacy: the exercise of control. New York: W.H. Freeman; 1997.
24. Margaryan A, Littlejohn A, Vojt G. Are digital natives a myth or reality? University students' use of digital technologies. Comput Educ 2011 Feb;56(2):429-440. [doi: 10.1016/j.compedu.2010.09.004]
25. Sanchez J, Salinas A, Contreras D, Meyer E. Does the new digital generation of learners exist? Brit J Educ Technol 2011;42(4):543-556 [FREE Full text] [doi: 10.1111/j.1467-8535.2010.01069.x]
26. World Wide Web Foundation. Webindex 2015. World Wide Web Foundation; 2015. URL: http://www.internetlivestats.com/internet-users-by-country/ [accessed 2016-11-16] [WebCite Cache ID 6m2rv9HlV]
27. Dodds TJ, Mohler BJ, Bülthoff HH. Talk to the virtual hands: self-animated avatars improve communication in head-mounted display virtual environments. PLoS One 2011;6(10):e25759 [FREE Full text] [doi: 10.1371/journal.pone.0025759] [Medline: 22022442]
28. Carr D. Contexts, gaming pleasures, and gendered preferences. Simul Games 2005 Dec 01;36(4):464-482. [doi: 10.1177/1046878105282160]
29. Broos A. Gender and information and communication technologies (ICT) anxiety: male self-assurance and female hesitation. Cyberpsychol Behav 2005 Feb;8(1):21-31. [doi: 10.1089/cpb.2005.8.21] [Medline: 15738690]
30. Hoeft F, Watson CL, Kesler SR, Bettinger KE, Reiss AL. Gender differences in the mesocorticolimbic system during computer game-play. J Psychiatr Res 2008 Mar;42(4):253-258. [doi: 10.1016/j.jpsychires.2007.11.010] [Medline: 18194807]
31. Choi G, Chung H, Kim Y. Are stereotypes relative to gender usage applicable to virtual worlds? Int J Hum Comput Interact 2012 Jun;28(6):399-405. [doi: 10.1080/10447318.2011.601973]
32. Jenson J, de Castell S. Gender, simulation, and gaming: research review and redirections. Simul Games 2010 Jan 26;41(1):51-71. [doi: 10.1177/1046878109353473]
33. Mills P, Neily J, Dunn E. Teamwork and communication in surgical teams: implications for patient safety. J Am Coll Surg 2008 Jan;206(1):107-112. [doi: 10.1016/j.jamcollsurg.2007.06.281] [Medline: 18155575]
34. Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S, et al. Surgical team behaviors and patient outcomes. Am J Surg 2009 May;197(5):678-685. [doi: 10.1016/j.amjsurg.2008.03.002] [Medline: 18789425]
35. Neily J, Mills PD, Young-Xu Y, Carney BT, West P, Berger DH, et al. Association between implementation of a medical team training program and surgical mortality. JAMA 2010 Oct 20;304(15):1693-1700. [doi: 10.1001/jama.2010.1506] [Medline: 20959579]
36. Hunziker S, Johansson AC, Tschan F, Semmer NK, Rock L, Howell MD, et al. Teamwork and leadership in cardiopulmonary resuscitation. J Am Coll Cardiol 2011 Jun 14;57(24):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2011.03.017] [Medline: 21658557]
37. Mäkinen M, Aune S, Niemi-Murola L, Herlitz J, Varpula T, Nurmi J, et al. Assessment of CPR-D skills of nurses in Göteborg, Sweden and Espoo, Finland: teaching leadership makes a difference. Resuscitation 2007 Feb;72(2):264-269. [doi: 10.1016/j.resuscitation.2006.06.032] [Medline: 17113208]
38. Lave J, Wenger E. Legitimate peripheral participation in communities of practice. In: Harrison R, Reeve F, Hanson A, Clarke J, editors. Supporting lifelong learning. London: Routledge Falmer; 2002:111-126.
39. deNoyelles A, Seo KJ. Inspiring equal contribution and opportunity in a 3d multi-user virtual environment: bringing together men gamers and women non-gamers in Second Life. Comput Educ 2012 Jan;58(1):21-29. [doi: 10.1016/j.compedu.2011.07.007]

Abbreviations

CPR: cardiopulmonary resuscitation
MVW: multiplayer virtual worlds
SD: standard deviation

Edited by G Eysenbach; submitted 04.08.16; peer-reviewed by P Stourac, J Adjedj, A Richardson-Hatcher; comments to author 13.10.16; revised version received 15.11.16; accepted 23.11.16; published 16.12.16.

Please cite as:
Creutzfeldt J, Hedman L, Felländer-Tsai L
Cardiopulmonary Resuscitation Training by Avatars: A Qualitative Study of Medical Students' Experiences Using a Multiplayer Virtual World
JMIR Serious Games 2016;4(2):e22
URL: http://games.jmir.org/2016/2/e22/
doi:10.2196/games.6448
PMID:27986645

Original Paper

Design of a Serious Game for Handling Obstetrical Emergencies

Estelle Jean dit Gautier[1], MD; Virginie Bot-Robin[2], MD; Aurélien Libessart[3], MD; Guillaume Doucède[1], MD; Michel Cosson[1], PhD; Chrystèle Rubod[1], PhD

[1]Department of Gynecology Surgery, Hopital Jeanne de Flandre, University of Lille, Lille cedex, France
[2]Department of Gynecology, Centre hospitalier de Seclin, Seclin, France
[3]3DDUO, Tourcoing, France

Corresponding Author:
Estelle Jean dit Gautier, MD
Department of Gynecology Surgery
Hopital Jeanne de Flandre
University of Lille
Rue Eugéne Avinée
Lille cedex, 59037
France
Phone: 33 624634623
Fax: 33 320446561
Email: estelle.jdg@gmail.com

Abstract

Background: The emergence of new technologies in the obstetrical field should lead to the development of learning applications, specifically for obstetrical emergencies. Many childbirth simulations have been recently developed. However, to date none of them have been integrated into a serious game.

Objective: Our objective was to design a new type of immersive serious game, using virtual glasses to facilitate the learning of pregnancy and childbirth pathologies. We have elaborated a new game engine, placing the student in some maternity emergency situations and delivery room simulations.

Methods: A gynecologist initially wrote a scenario based on a real clinical situation. He also designed, along with an educational engineer, a tree diagram, which served as a guide for dialogues and actions. A game engine, especially developed for this case, enabled us to connect actions to the graphic universe (fully 3D modeled and based on photographic references). We used the Oculus Rift in order to immerse the player in virtual reality. Each action in the game was linked to a certain number of score points, which could either be positive or negative.

Results: Different pathological pregnancy situations have been targeted and are as follows: care of spontaneous miscarriage, threat of preterm birth, forceps operative delivery for fetal abnormal heart rate, and reduction of a shoulder dystocia. The first phase immerses the learner into an action scene, as a doctor. The second phase ask the student to make a diagnosis. Once the diagnosis is made, different treatments are suggested.

Conclusions: Our serious game offers a new perspective for obstetrical emergency management trainings and provides students with active learning by immersing them into an environment, which recreates all or part of the real obstetrical world of emergency. It is consistent with the latest recommendations, which clarify the importance of simulation in teaching and in ongoing professional development.

(JMIR Serious Games 2016;4(2):e21) doi:10.2196/games.5526

KEYWORDS

serious game; obstetric emergencies; gynecology

Introduction

In most cases, pregnancy and childbirth are conducted without complications. However, when pathology occurs, sometimes severe and urgent, a fast and efficient care led by a perfectly trained team is needed. In these cases, the training is mostly performed on real patients under the supervision of a senior. However, emergency situation does not facilitate the learning process.

XSL•FO
RenderX

The occurrence of certain situations is therefore random and depends on the maternity services with which the student is affiliated.

The emergence of new technologies in the obstetrical field should lead to the development of learning applications, specifically for obstetrical emergencies. Indeed, according to the latest 2014 Haute Autorité de Santé recommendations, related to birth care quality and safety, simulation exercises are integrated into the team training for obstetrical emergencies [1]. Many trainings using simulation were designed to teach medical students and to develop their capacity to perform surgeries in the future. Effectiveness of this form of learning is now well established [2,3]. Furthermore, using virtual glasses-headphones coupling allows total immersion of the player in a 3D environment and affects both hearing and sight. This immersion is more complete as compared with a situation in which the player is just in front of a computer screen with keyboard and mouse control devices [4].

In the obstetrical field, many childbirth simulations have been recently designed for educational purposes [5]. However, to date, none of them have been integrated into a serious game, allowing a global "virtual patient" approach.

Our goal is to develop a new type of immersive, virtual reality serious game using virtual glasses. We describe the design of a new game engine, placing the student in some maternity emergency situations and delivery room simulations. In-game decision-making should lead, in the second step, to an accurate obstetrical gesture realization on a physical manikin.

Methods

For several years, our obstetricians' team has been working in partnership with an educational engineer.

Different pathological situations during pregnancy that have been targeted are as follows: threat of early spontaneous miscarriage, threat of preterm birth, forceps operative delivery for abnormal heart rate, and reduction of a shoulder dystocia.

This serious game targets different categories of learners.

Indeed, "spontaneous miscarriage" and "possible preterm birth" scenarios deal with 2 topics that are part of the French educational program of the medicine study second cycle [6]. They also form a part of the midwife educational program. Thus, these scenarios target a wide audience, composed of medicine students, obstetrical and gynecological internes and residents, and apprentice midwives. The "forceps operative delivery for abnormal fetal heart rate" scenario will permit the gynecology and obstetrics residents training.

Obstetricians and midwives must know the "shoulder dystocia" diagnosis and the related reduction gestures. The last 2 scenarios could be integrated as part of an ongoing medical training.

An obstetrician initially wrote a scenario based on a real clinical situation. All scenarios were designed following the same method. The first phase put the learner in an action scene, as a doctor. The second phase lead to making the diagnosis of the pathology. An actions diagram tree referring to all the possible actions enables the students to progress following a clinical reasoning. Players can have access to various additional clinical elements, for example, the possibility to perform an ultrasound, a biological assessment, consult medical records, or fetal monitoring. Once the diagnosis is made, different treatments are offered to learners.

We want to reach the most real-like situation by offering students the maximum dialog choices and additional tests.

In order to offer a certain freedom in action, a complex actions diagram tree lead in-game actions and allows students to make their own choices in the game.

Both an educational engineer and a doctor have designed this actions tree. It contains all the possible actions, which are given to players in the game. Thanks to this method, the scenario moves away from a linear path, and allows students to learn through attempts and errors. All these actions diagram trees are designed with a diagramming software, such as Ed Graph Editor, (see Figure 1).

This actions diagram has been evaluated for its relevance by doctors with different knowledge levels. Thus, medicine students, obstetrical residents, and hospital practitioners tested it. It was aimed to highlight possible misunderstandings or inconsistencies, which may persist in the scenario realization. This will also enable us to expand the range of options offered to players. Some answers may not have been initially envisaged and might be relevant in diagnostic or therapeutic procedures.

The graphic universe has been fully 3D modeled and is based on hospital photographic references. Different views have been taken in some strategic locations such as obstetrical emergencies unit, guardrooms, and delivery rooms (see Figure 2). So far, 2 locations have been 3D modeled in order to compose the environment of the 4 scenarios. "Spontaneous miscarriage" and "possible preterm birth" scenarios take place in the obstetrical emergencies unit. "Forceps operative delivery for fetal abnormal heart rate" and "reduction of a shoulder dystocia" scenarios take place in a delivery room. These different locations are 3D modeled using "Autodesk 3ds Max" software.

Game development was carried out following 2 distinct phases.

First, it was necessary to develop a game engine. Indeed, once the actions tree design was completed, each actions branch had to be linked to in-game actions. Each action contains, via the game engine, the action description, the previous actions (leading to this action), and the following actions (unlocked by this action). Some other information can also be set up, such as corresponding interactive objects, code names, and so on. Second, indicated actions are connected with the graphic universe. For example, ultrasound actions type appeared by clicking on the ultrasound equipment, respecting the timeline and the causes or consequences relationship established in the actions tree diagram.

Players are immersed in virtual reality thanks to the Oculus Rift technology, a virtual reality device designed by the Oculus VR Company. The device looks like a mask covering eyes and can be strapped to the face at the rear of the head. A digital screen is placed a few centimeters in front of each eye, perpendicular to the sight line. This screen displays a stereoscopic picture, digitally distorted by 2 lenses located in front of each eye, in order to inverse the optical distortion. It expands the visual field and the definition in front of the fovea. The screen is placed in the focal plane of these lenses. The created virtual picture is projected to infinity. Various sensors detect user head movements, which make a real-time picture adaptation on the screen possible and produce a total immersion into the rendered scene.

A score, which can be either positive or negative, is set for every possible in-game action. A good action is positively rewarded, and conversely, bad actions are negatively rewarded. Therefore, certain choices would be rewarded, and others would be penalized. Score points setting depend on the student diagnostic process and its relevance. For example, if the student immediately led a biological assessment before interrogating or examining the patient, the student will be rewarded less score points than a student who follows a correct approach, with a medical logic (interrogation, followed by clinical tests, and additional tests). In delivery room scenarios, a concept of penalty depending on the student's decision time is added.

All these actions are then summarized at the end of the game in a score table (see Figure 3). The score table is divided into 5 sections. On the top, the global player score is provided that the player earned on the entire scenario. The first column reminds each action he or she performed, the second shows the actions category, the third shows the score associated with this particular action, and the last column often leads to additional content (videos, articles, and courses).

Every time a session ends, a personal and downloadable assessment of the player is edited through a spreadsheet software (Excel-like). Thus, the player can see his mistakes and successes at the end of the game. A scoring system permits the player assessment. The learning assessment is a two-step process. The first step is made through the game based on the choices the player virtually made and for which he or she has been given points (positive and negative) assigning the player a final grade. The second step occurs through the actual debrief at the end of the learning session, with the teacher who will go through the blunt assessment given by the game.

Figure 1. Orange: environment changes and specific content display; blue: player decision making; purple: interaction with object; green: game automatic feedback; black: checkpoint.

Figure 2. Elaboration of a graphic environment: (a) 3D modeled graphic universe, based on hospital photographic references: guard room and (b) obstetrical emergencies unit.

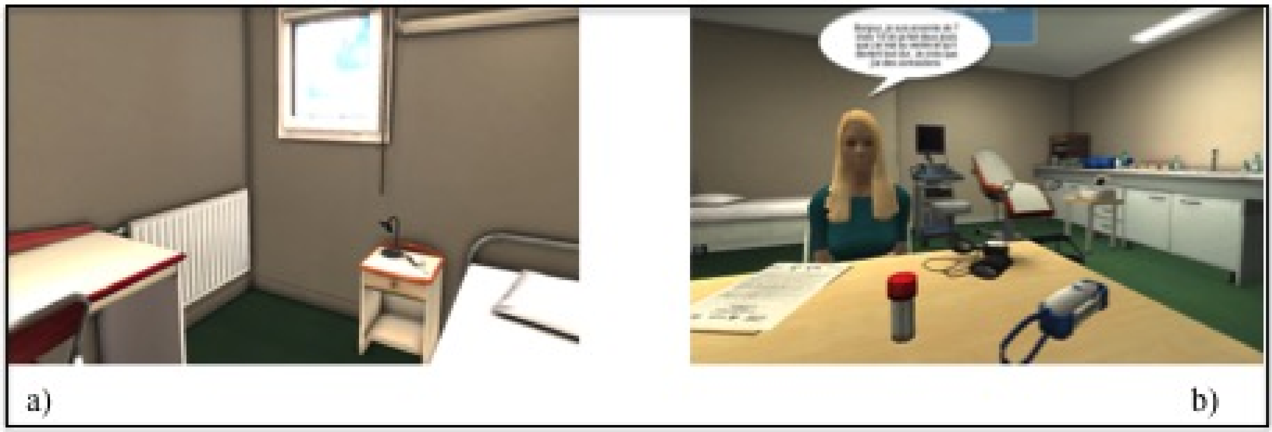


Figure 3. Score table: personal and downloadable score table of the player.

Results

All the scenarios mentioned above have been developed.

Two scenarios take place in obstetrical emergencies unit. The 2 selected pathologies are frequent motives for consultations: spontaneous miscarriage and possible preterm birth [7].

The player is in an emergency room. He or she has access to all the necessary equipment in order to solve the clinical case: personal medical records, cardiotocography, lookup table, ultrasound device, tensiometer, thermometer, tubes for biological samples, or even urine sample bottles for urinalysis strips uses. Each element has a specific role in the scenario.

The threat of preterm birth scenario is a simple situation in which a patient comes to obstetrical emergencies unit for pelvic pain at 30 weeks of gestation. We have created a virtual dialogue between the patient and the learner or player in order to provide him with the necessary information on diagnosis. The learner can also consult the personal medical records in order to access various data such as patient history, allergies, and pregnancy monitoring. If the player wishes, he or she can perform an ultrasound, or consult the fetal heart rate and tocography. If so, he or she will have an access to ultrasound images we have included in the scenario (cervical length and estimated fetal weight). The purpose of these additional elements is to guide the learner toward a possible preterm birth diagnosis. However, only the player can choose if he or she realizes (or not) a medical interrogation and additional examinations. He or she can then choose among several therapeutic treatments. To some extent, the player can constantly access the ongoing diagnosis before choosing the appropriate treatment and modify it if he or she wants. However, there are points of no return: Once he or she reaches one, the player cannot go back would he or she want to obtain additional information and modify his diagnosis. The diagnostic process is estimated at 20 minutes, which is the time amount allotted to the player to solve the case. With this countdown system, we want to recreate the stressful conditions of emergency services (average consulting time).

The "spontaneous miscarriage scenario" takes place in the same environment and is based on the same game action engine, with a similar diagnostic approach. A patient, at the first trimester of pregnancy, comes for metrorrhagia. The player can realize detailed examinations, ultrasound, and biological check-up (pregnancy blood test, complete blood count). He or she should eventually make a spontaneous miscarriage diagnosis and choose the adapted therapy. Regarding the scenarios taking place in the delivery room, we needed to create a new graphic environment.

An obstetrical unit midwife calls the player and asks him to come into the delivery room. He or she can consult the personal medical records, the monitoring (showing an abnormal fetal heart rhythm and expulsive efforts), ultrasound, and has access to various obstetrical tools.

Regarding "forceps operative delivery through forceps for abnormal heart rate scenario," the player can lead a patient and a midwife interrogation, perform a vaginal examination, an ultrasound, and collect some information about fetal head engagement and orientation. The goal is to achieve as soon as possible the abnormal fetal heart rate diagnosis requiring forceps operative delivery. The decision must be taken within 10 minutes in order to represent the urgency of this type of situation.

The "reduction of a shoulder dystocia scenario" takes place in the same way. The same additional examination possibilities are offered to the player. Most of them are useless, and even have a negative impact and result in a waste of time because the shoulder dystocia diagnosis is purely clinical. The player goal is to reach this diagnosis as soon as possible and choose among various proposed gestures.

Discussion

Principal Findings

We designed a new type of immersive serious game, using a new game engine. It supports the Oculus Rift technology, allows the integration of new scenarios with minimal effort, especially in the gynecologic field, and could be linked to the medical training manikin. It has been designed so that it is sensor-equipped to allow gesture realization.

The use of new simulation technologies has been widespread in the obstetrical field for the last ten years. However, to our knowledge, there is no publication about serious games used as pedagogic tools for learning in obstetrical emergency situations.

Nowadays, learning through simulation mostly concerns trainings related to breast and pelvic clinical examination as well as to postpartum hemorrhage management [8]. This learning form is commonly linked to the technical skills reinforcement, especially with postexercise debriefing. Through the game, our educational tool can help stimulate the motivation of medicine students, obstetrical interns, residents, and midwifery students by offering the possibility to face frequent emergency situations, which require quick management answers and are based on essential knowledge.

Some commercial or nonprofit platforms already exist but have not been totally adapted to the medical training needs. For instance, these game engines do not support the Oculus Rift technology. Based on our training needs, we created our own specific tool. It allows the incorporation of a complex diagram tree, based on a real clinical situation, which offers a multiple action combination and leads to a realistic and immersive game.

Moreover, the project must lead to the integration of a medical training manikin, be sensor-equipped, and link to the Serious Game, to allow obstetrical gesture training with real-time feedbacks. Our game engine should therefore be custom designed in regard with this evolution. Serious games developments in the medical field are often very specific to each project. Few of them use a game engine, which allows creating new scenarios with minimal efforts, especially in the gynecologic field. Our work led to the development of a new game engine, which allows integrating, quickly and easily, new scenarios, both in the gynecologic and obstetric fields, as well as in other medical fields, without requiring the intervention of a programmer. Each scenario can be integrated into a new graphic environment, corresponding to various hospital locations, using already created graphic assets. Moreover, few serious games offer a complete immersion in the medical field currently. Immersion, through the Oculus Rift, can increase the realism of a clinical situation and the involvement of the player [9].

Learning by playing seems to be a solid method to gain better appropriation for the learner [10]. Serious games, mostly developed for surgical skills, have been tested to prove their validity [11-13]. Serious games seem to better enable the learner to feel immersed, to improve their confidence, and to enhance their clinical skills [12]. Serious games offer an innovative approach and seem more attractive than the "old fashioned way" of learning. Nevertheless, in order to do so, it is important to cooperate in designing and validating a serious game for a specific educational problem [14,15]. That is why we combined both the work and the skills of a pedagogic engineer, a game designer, a medical doctor, and a medical professor.

Our educational method aims at learning various common gynecological pathologies through realistic virtual situations. The game provides the learner with the opportunity to think and to follow a patterned diagnosis approach. Thanks to the postgame score analysis, the learner can evaluate himself and get debriefed about the mistakes he or she committed during the game session.

Naturally, this tool has to be properly tested on a student's sample in order to validate its performances and demonstrate its pedagogic potential, and also evaluate the good learning acquisition. Thanks to the game and to the learning session with the debrief-time.

Even today, *serious game* is not a well-defined concept. So far, it has many definitions, which differ from author to author [16-20]. Our work matched with the definition of Julian Alvarez that defines serious games as any "computer application whose initial intention is to combine, with consistency, both serious aspects with fun spring from video game" [18]. We have described a system that meets this definition by using scoring system, action and decision tree, stressful and immersive graphic environment narrative systems, and game mechanics based on flow and game design theory (especially from point and click adventure game type). This Serious Game is developed as a part of a continuing medical training, funded by Lille2 University. It is used within the framework of learning sessions, managed by a professor. A medical teacher carries out the briefing and the debriefing. In this context, the game is only a support, used to initiate a dialogue involving the player's own knowledge. However, this game can be also used as a "standalone game." The final report screen allows nonetheless a basic assessment, even without learning sessions. The scoring system, initially designed by a doctor, allows the player to visualize his mistakes, and to get the "correct answer" related to the clinical situation he or she just experienced. Prior to any training, we will use these sessions as an opportunity to make A-B testing (with and without the game) and evaluate its usability, through a form filled by the student at the end of each session.

So far, delivery room scenarios end at the obstetrical gesture choice. Thereafter, we would allow the learner to realize these technical gestures on a sensor-equipped obstetrical anatomic manikin, which tracks each learner's gestures.

We are working on a virtual simulator of the pregnant woman's pelvic system in order to allow gestures on a digital model. The game action engine of our serious game can also integrate new future scenarios such as ectopic pregnancy, operative delivery for obstructed labor, and/or vacuum extraction.

Conclusion

Simulation teaching offers active learning, conducted by the learner, immersed in an environment recreating all or part of the real world, promoting knowledge integration and technical and behavioral skills in a short time.

Our serious game is part of this type of educational training, and offers a new perspective for obstetrical emergency learning. We need to test it on a student sample to validate its pedagogic potential in order to justify its integration into an obstetrical learning program.

We want to develop it by creating some new scenario in the gynecologic and obstetrical fields to extend its pedagogical impact.

Acknowledgments

The authors would like to thank the University of Lille for the financial support.

Conflicts of Interest

None declared.

References

1. Granry JC, Moll MC. Rapport de mission État de l'art (national et international) en matiere de pratiques de simulation dans le domaine de la sante: Dans le cadre du developpement professionel continu et de la prevention des risques associes aux soins. Haute Autorité de Santé. 2012 Jan. URL: http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-01/simulation_en_sante_-_rapport.pdf [accessed 2016-08-30] [WebCite Cache ID 6k9DApsIO]
2. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, et al. Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 2011 Sep 7;306(9):978-988. [doi: 10.1001/jama.2011.1234] [Medline: 21900138]
3. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg 2012 Oct;99(10):1322-1330. [doi: 10.1002/bjs.8819] [Medline: 22961509]
4. Starner T. Fundamentals of wearable computers and augmented reality. In: Wearable Comput Meetting Chall. France: CRC press; 2015.
5. Dupuis O, Moreau R, Pham MT, Redarce T. Assessment of forceps blade orientations during their placement using an instrumented childbirth simulator. BJOG 2009 Jan;116(2):327-32; discussion 332 [FREE Full text] [doi: 10.1111/j.1471-0528.2008.02004.x] [Medline: 19076965]
6. Ministère de l'éducation nationale et ministère de l'enseignement supérieur et de la recherche. Deuxième partie du deuxième cycle des études médicales. Bulletin officiel. 2007 Jun 07. URL: http://www.education.gouv.fr/bo/2007/22/MENS0753287A.htm [accessed 2016-08-30] [WebCite Cache ID 6k9DfNadw]
7. Alouini S, Mesnard L, Coly S, Dolique M, Lemaire B. Urgences gynécologiques : nature et degré de gravité. J Gynécologie Obstétrique Biol Reprod 2012;41(1):48-54.
8. Daniels K, Parness A. Development and use of mechanical devices for simulation of seizure and hemorrhage in obstetrical team training. Simul Healthc 2008;3(1):42-46. [doi: 10.1097/01.SIH.0000290632.83361.4b] [Medline: 19088641]
9. Csikszentmihalyi M. Toward a Psychology of Optimal Experience. In: Flow and the Foundations of Positive Psychology. Netherlands: Springer; 2014:209-226.

10. Al-Smadi M. Assessment in serious games: An enhanced approach for integrated assessment forms and feedback to support guided learning. 2012 Presented at: 2012 15th International Conference on Interactive Collaborative Learning (ICL); September 26-28, 2012; Austria. [doi: 10.1109/ICL.2012.6402089]
11. Youngblood P, Harter P, Srivastava S, Moffett S, Heinrichs W, Dev P. Design, development, and evaluation of an online virtual emergency department for training trauma teams. Simul Healthc 2008;3(3):146-153. [doi: 10.1097/SIH.0b013e31817bedf7] [Medline: 19088658]
12. Heinrichs W, Youngblood P, Harter P, Kusumoto L, Dev P. Training Healthcare Personnel for Mass-Casualty Incidents in a Virtual Emergency Department: VED II. Prehosp Disaster Med 2012 Jun 28;25(05):424-432. [doi: 10.1017/S1049023X00008505]
13. LeRoy HW, Youngblood P, Harter P, Dev P. Simulation for team training and assessment: case studies of online training with virtual worlds. World J Surg 2008 Feb;32(2):161-170. [doi: 10.1007/s00268-007-9354-2] [Medline: 18188640]
14. Schijven M, Jakimowicz J. Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005;14(4):244-246. [Medline: 16754170]
15. Gallagher A, Ritter E, Satava R. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003 Oct;17(10):1525-1529. [doi: 10.1007/s00464-003-0035-4] [Medline: 14502403]
16. Abt CC. Serious games. Lanham, MD: University Press of America; 1987.
17. Michael D, Chen S. Serious Games: Games That Educate, Train, and Inform. Boston, MA: Course Technology PTR; 2005.
18. Alvarez J. Du jeu video au serious game: approches culturelle, pragmatique et formelle. Games. URL: http://ja.games.free.fr/These_SeriousGames/TheseSeriousGames.pdf [accessed 2016-08-25] [WebCite Cache ID 6k1jmaIyV]
19. Sawyer B. The “Serious Games” Landscape. Biu. 2007. URL: http://u.cs.biu.ac.il/~ariel/download/mm664/resources/digital_games/Serious%20Games%20Landscape.ppt [accessed 2016-08-30] [WebCite Cache ID 6k9FTgirk]
20. Djaouti D. Serious Game Design: considerations theoriques et techniques sur la creation de jeux video a vocation utilitaire. Ludoscience. URL: http://www.ludoscience.com/files/these_djaouti.pdf [accessed 2016-08-25] [WebCite Cache ID 6k1kClCEs]

Edited by C Boyer; submitted 14.01.16; peer-reviewed by P Staccini, D Katz, C Boyer; comments to author 30.03.16; revised version received 25.05.16; accepted 08.08.16; published 21.12.16.

Please cite as:
Jean dit Gautier E, Bot-Robin V, Libessart A, Doucède G, Cosson M, Rubod C
Design of a Serious Game for Handling Obstetrical Emergencies
JMIR Serious Games 2016;4(2):e21
URL: http://games.jmir.org/2016/2/e21/
doi:10.2196/games.5526
PMID:28003175

Review

Usability Evaluation Methods for Gesture-Based Games: A Systematic Review

Fernando Winckler Simor[1], MS; Manoela Rogofski Brum[1]; Jaison Dairon Ebertz Schmidt[1], BS; Rafael Rieder[1], PhD; Ana Carolina Bertoletti De Marchi[1], PhD

Universidade de Passo Fundo, Passo Fundo, RS, Brazil

Corresponding Author:
Rafael Rieder, PhD
Universidade de Passo Fundo
BR 285 KM 292.7
Passo Fundo, RS, 99052-900
Brazil
Phone: 55 5433168354
Fax: 55 5433168354
Email: rieder@upf.br

Abstract

Background: Gestural interaction systems are increasingly being used, mainly in games, expanding the idea of entertainment and providing experiences with the purpose of promoting better physical and/or mental health. Therefore, it is necessary to establish mechanisms for evaluating the usability of these interfaces, which make gestures the basis of interaction, to achieve a balance between functionality and ease of use.

Objective: This study aims to present the results of a systematic review focused on usability evaluation methods for gesture-based games, considering devices with motion-sensing capability. We considered the usability methods used, the common interface issues, and the strategies adopted to build good gesture-based games.

Methods: The research was centered on four electronic databases: IEEE, Association for Computing Machinery (ACM), Springer, and Science Direct from September 4 to 21, 2015. Within 1427 studies evaluated, 10 matched the eligibility criteria. As a requirement, we considered studies about gesture-based games, Kinect and/or Wii as devices, and the use of a usability method to evaluate the user interface.

Results: In the 10 studies found, there was no standardization in the methods because they considered diverse analysis variables. Heterogeneously, authors used different instruments to evaluate gesture-based interfaces and no default approach was proposed. Questionnaires were the most used instruments (70%, 7/10), followed by interviews (30%, 3/10), and observation and video recording (20%, 2/10). Moreover, 60% (6/10) of the studies used gesture-based serious games to evaluate the performance of elderly participants in rehabilitation tasks. This highlights the need for creating an evaluation protocol for older adults to provide a user-friendly interface according to the user's age and limitations.

Conclusions: Through this study, we conclude this field is in need of a usability evaluation method for serious games, especially games for older adults, and that the definition of a methodology and a test protocol may offer the user more comfort, welfare, and confidence.

(JMIR Serious Games 2016;4(2):e17) doi:10.2196/games.5860

KEYWORDS

usability testing; evaluation; computer games; gestural input; usability evaluation; method; gesture-based games

Introduction

Interactive systems can only be considered useful and practical if they have good usability. According to Karray et al [1], usability is the variety and the degree to which system features can be used efficiently so that the user can accomplish tasks effectively and intuitively. The balance between functionality and usability allows achieving the system effectiveness. Among the usability characteristics defined by Nielsen [2] are ease in performing basic tasks, efficiency when performing these tasks, facility by reusing resource, reestablishment of services when mistakes occur, and satisfaction with use.

Researchers in the area of human-computer interaction have been developing several usability evaluation methods in order to determine whether a system or interactive device is usable or not. According to Cockton [3], usability evaluation is essential to establish a relationship between the quality of an interactive system and interaction quality. The author mentions that when a usability evaluation shows that an application or device can be used, methods and metrics can determine the extent to which a system is easy and pleasant to use.

The constant development of usability evaluation mechanisms occurs due to the high supply of interactive systems on the market, constantly bringing to the user new ways of interacting. Gestural interactions are among the styles that have evolved in more recent years and used in largely in entertainment applications, such as virtual reality environments and games.

According to Morelli and Folmer [4], gesture-based games typically simulate real physical activities because they use whole-body gestures. These kinds of games are intuitive to play, they have successfully attracted users, and they provide different social forms of gaming—especially because they allow natural interaction and immersion. This interaction style is present through different motion-sensing input devices [5]. Microsoft Kinect and Nintendo Wii are examples of devices that use gestures as an interaction method. They apply unimodal and multimodal resources, meaning they can combine audio, video, and gestures to emulate interactive environments [1]. According to Karam and Schraefel [6], depending on the application type, it is possible to use more than one input device in the interaction base, which will all allow the same action.

Given the market demand for systems with gestural interaction, it is necessary to establish procedures to evaluate the usability of these interfaces in order to minimize interaction problems. In this perspective, Keskinen et al [7] proposed a method to evaluate user experience in interactive systems. Rautaray and Pandey [8] presented comparative studies that characterized gestural interaction elements and Maidi and Preda [9] organized how gestures could be evaluated. However, it is difficult to establish a consensus between these and other studies regarding what should or not be evaluated, especially when it comes to evaluating the usability of gestural interaction applied to games.

Different tools are being used now to assist the usability evaluation process: some present qualitative results and others quantitative, some focus on perception or acceptance, some consider physiological measures, and so on. The big gap is the lack of a validated approach that makes it possible to ensure consistent assessment results and increase its credibility. Still, usability patterns could be defined from this approach along time to use as benchmarks in evaluations.

In this sense, the aim of this study is to present a systematic review about the usability evaluation methods applied to games with gestural interaction, considering devices with motion-sensing capability. To reach this goal, specific objectives have been set: (1) identify and analyze techniques applied to usability analysis of games interfaces for gesture-based devices, (2) identify common problems found in gesture-based games interfaces, and (3) relate strategies and technologies that have been used to solve user interface problems in gesture-based games.

Methods

This study is a systematic review, explicit and rigorous research that identifies, critically evaluates, and synthesizes relevant studies about a specific subject [10].

Eligibility Criteria

The eligibility criteria to identify studies in the primary phase included were (1) games for gestural interaction devices, (2) Kinect and/or Wii as the gestural interaction device, and (3) description of a usability evaluation technique for analysis of user interfaces.

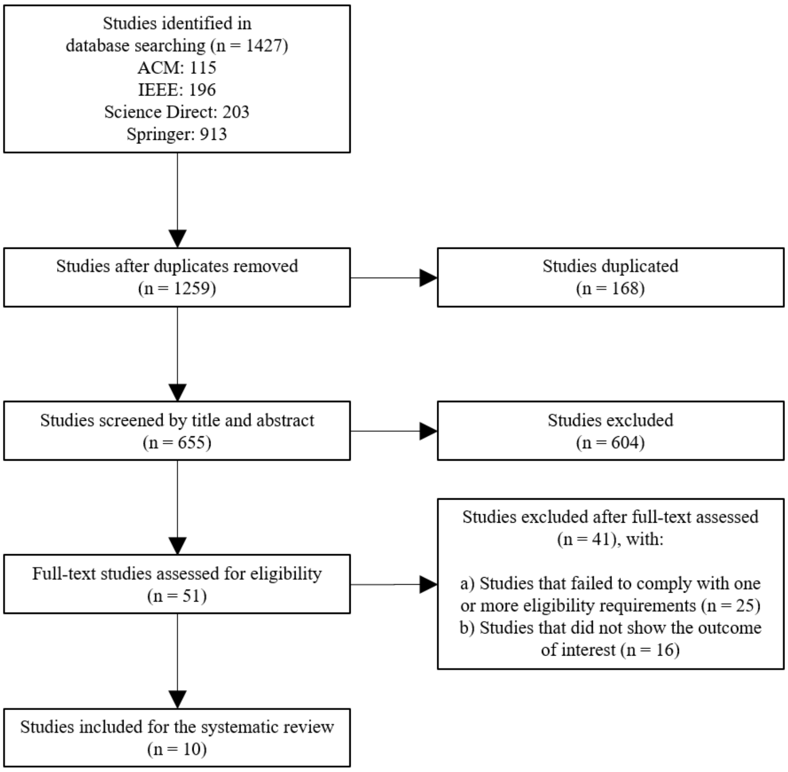
Search Strategy for Primary Studies

The research was concentrated into four electronic databases: IEEE Xplore Digital Library, Association for Computing Machinery (ACM), Springer International Publisher Science, and Science Direct. This work collected studies published in English between September 4 and 21, 2015, not limited by date, and using the following expression: games AND usability AND evaluation AND (“Kinect” OR “Wii”).

Flowchart of Identified Studies

Figure 1 shows the flowchart of studies identified by the search strategy used in this study. The selection process and reading of these materials involved at least two researchers.

Figure 1. Flowchart of identified studies. ACM: Association for Computing Machinery.



Results

Table 1 presents the 10 studies that met the eligibility criteria. These studies showed no standardization in their data or their methods, which made it impossible to conduct a statistical analysis.

Each study will be presented considering the following items:

1. Evaluation aim: approach and focus of each study;

2. Gesture-based devices used: equipment used during the experiments;

3. Evaluator's profile: professionals and researchers team;

4. Profile and number of participants: characterize the sample;

5. Evaluation method and its application steps: the author's methodology and how it was applied;

6. User's tasks: experiment tasks;

7. Type of interface (2D or 3D) and software used: the applications used during the experiments;

8. Time required for the user experience and evaluation: the period for each user evaluation; and

9. Results: the analysis and conclusions of the authors of each study.

Komlódi Et Al

Komlódi et al [11] aimed to test the Wiimote device in basic navigation, object manipulation, and menu gestural interaction tasks in a virtual environment. The users perform pointing and turning gestures using their hands and the Wii controllers. During the interaction process, they also pushed the buttons of the controllers to operate different navigation modes.

Researchers from the University of Maryland in the United States and the Budapest University of Technology and Economics in Hungary conducted the evaluation. A group of 14 Hungarian undergraduate students (seven men and seven women) participated in the evaluation; the mean age of participants was 24 years. Half of these users had some experience with computer game simulators. Four had used the Wiimote previously, but none had significant experience.

Although this study was not a game, it was included because, in the demographic questionnaire, the authors considered the user's previous game experience important. According to Bowman et al [21], this has a strong impact on the ability to interact with virtual environments. Moreover, virtual environments can simulate or represent a game in 3D space. In this study specifically, one of the user tasks involved the handling of dominoes using the Wiimote.

During the pilot test, three participants performed the tasks in a virtual environment and answered questionnaires. The procedure had several changes throughout the testing period: extension of the training period, review of tasks, and changes in the interview questions.

The study combined qualitative and quantitative methods to explore the utility of interaction methods in a 3D environment. After a brief training session, 14 participants executed two immersive tasks in three pilot tests. A video camera recorded the entire process and the researchers interviewed participants about their experience immediately after the interaction.

After the orientation and training session, participants read instructions for tasks on a poster. This poster displayed the following guidance: (1) walk around the room, find the dominoes, stack the dominoes one on top of another, and dismantle them; and (2) use the KUKA robot to move the gray black balls on the table and, when finished, tell the session coordinator.

Participants answered a questionnaire containing 19 questions about their experience using the game. Some questions were issues, such as ease of interaction, first impressions, reactions during use, satisfaction when using, and suggestions for future game improvements. Only two questions used a scale of 0 to 100; the others were dissertated.

In addition to this questionnaire, participants were also requested to complete a sociodemographic questionnaire and the Myers-Briggs Type Indicator (MBTI) personality type questionnaire. This questionnaire identifies users' characteristics of extroversion/introversion, sensing/intuition, thinking/feeling, and judgment/perception. Participants also performed the VZ-2 paper folding test to measure cognitive ability of spatial visualization, which can influence the user's ability to navigate in 3D space and manipulate objects in space, and the Reading the Mind in the Eyes Test.

For effective usability studies, the authors found the need for more training time and practice with games as well as several sessions of interaction involving experts and novices with games and devices. In addition, reducing the memory load for users by including tasks and feedback functions, as well as help in the environment also improves the game usability.

Legouverneur Et Al

Legouverneur et al [12] aimed to conduct a usability study of two sports games for the Wii. The purpose was to determine whether the elderly with cognitive impairment could learn to play and control their movements with the wireless controller (Wiimote). A secondary objective was to examine how specific neuropsychological deficits may modulate the game usability.

Broca Hospital professionals from Paris, France, conducted the evaluation with two groups of users recruited from a health care center. The first group consisted of elderly people with mild-to-moderate Alzheimer disease as the diagnostic criteria. The second group consisted of elderly people with mild cognitive impairment. A third group consisted of healthy elderly individuals. All users were aged between 75 and 90 years.

The test protocol included an introductory session and four test sessions, with mean duration of 1 hour per week. The introductory session included a neuropsychological evaluation. During this session, participants also created their own avatar as a way to learn to use the Wiimote. In the test sessions, the participants interacted with two bowling games and two tennis games alternately. Each test session had two cameras to record the game screen and the player simultaneously.

The use of the Wiimote was to mimic the actions of swinging a racket (tennis) or rolling a ball down an alley (bowling). According to the authors, the movement performed by each user was analyzed considering the approach of games (no specific gesture was evaluated).

In addition to the game console and cameras, other equipment involved in the experiment was a 46-inch plasma TV and behavioral analysis software. In order to collect performance and behavior data, the authors used the video recordings. A questionnaire using a five-point Likert scale evaluated user preferences. The authors used the questionnaire at the end of the first test session and after session 4 to evaluate whether familiarity with the games influenced user preferences.

The results showed that all participants, regardless of their cognitive status, could use the wireless controller and learn to play both games. A positive experiment result, according to the authors, was the improvement of skills with games throughout the sessions on performance measures observed for most participants. The study also confirmed the importance of usability testing with end users before introducing traditional technology to older adults who have cognitive dysfunction. Multiple sessions allowed users with cognitive impairment to be comfortable with technological devices in order to learn how to use them and have a positive experience with them. This experience also confirmed the role of motivation and a socially supportive environment on how a person learns to use new technologies.

Francese Et Al

In the Francese et al study [13], the aim was to evaluate two games developed for 3D interaction with the use of navigation maps from Bing Maps. The games used were Wing for the Wii and King for Kinect. The main objective was to evaluate the gestural interaction by controlling user navigation in Bing Maps through the devices previously mentioned.

With the Wiimote, the gestures were inspired by the motorcycle metaphor: roll/rotation of the Wiimote acts as a motorcycle throttle command connected to navigation forward and backward movements; the turning gestures resemble the turning handlebar of an imaginary motorcycle. With the Nunchuk, the airplane cloche metaphor was used to control altitude: its tilting direction determined the vertical variations of the navigation.

Using Kinect, the bird (or airplane) metaphor was used, with natural gestures associated to the various commands. The idea was to mimic the bird's wing movements, when possible, with arm gestures. For example, the user would move their aligned arms downward to the left as the bird or airplane did onscreen. A virtual paper plane was presented to the user to give a feedback about their movement.

The evaluation process involved 24 volunteers (16 men and 8 women), who were staff and students from the University of Salerno in Salerno, Italy. Ages ranged between 18 and 41 years, with a mean of 24 years.

Before beginning the experiment, the skills of participants in the games were evaluated. Eight participants mentioned they played at least once a week, three played Wii, and only two played Xbox and Kinect. Each participant answered 12 questions, using a seven-point Likert scale, and three factors were evaluated: involvement, distraction, and control. The study was performed in a research laboratory of the University of Salerno.

For the experiment, participants were quickly introduced to gestural interfaces and performed two navigation tasks. After being instructed on how to use both games, Wing and King, users were asked to navigate in two geographical routes involving well-known Italian cities: MAR (Cagliari-Naples-Palermo) and TERRENO (Genoa-Rome-Venice).

Both tasks were compatible in terms of distance and difficulty in locating the destination cities. In order to avoid bias in evaluation tasks, the approach defined two user groups in which each member of the same group started the experiment with the same system. After each task, all participants filled-in After-Scenario Questionnaires (ASQ) to evaluate the time spent, the ease of completion, and the adequacy of support information.

Authors used the Computer System Usability Questionnaire (CSUQ), consisting of 19 questions, to evaluate user satisfaction with the 3D maps game for four factors: general evaluation, system utility, information quality, and interface quality. According to the authors, evaluation results conducted through questionnaires confirmed that, if the interface is more natural, the user will be as satisfied and engaged in the navigation experience.

Table 1. Studies included in the systematic review.

Study ID	Paper title	Device
Komlódi et al [11]	Empirical Usability Evaluation of the Wii Controller As an Input Device for the VirCA Immersive Virtual Space	Wii
Legouverneur et al [12]	Wii Sports, a Usability Study with MCI and Alzheimer's Patients	Wii
Francese et al [13]	Wiimote and Kinect: Gestural User Interfaces Add a Natural Third Dimension to HCI	Wii and Kinect
Norouzi-Gheidari et al [14]	Interactive Virtual Reality Game-Based Rehabilitation for Stroke Patients	Kinect
Liu et al [15]	An Approach of Indoor Exercise: Kinect-Based Video Game for Elderly People	Kinect
Shin et al [16]	A Task-Specific Interactive Game-Based Virtual Reality Rehabilitation System for Patients with Stroke: a Usability Test and Two Clinical Experiments	Kinect
Fang et al [17]	Interactive Physical Games: Improving Balance in Older Adults	Kinect
Harrington et al [18]	Assessing Older Adults' Usability Challenges Using Kinect-Based Exergames	Kinect
Nakai et al [19]	Investigating the Effects of Motion-Based Kinect Game System on User Cognition	Kinect
Sheu et al [20]	User-Centered Design of Interactive Gesture-Based Fitness Video Game for Elderly	Kinect

Norouzi-Gheidari Et Al

Norouzi-Gheidari et al [14] developed a study to use a research protocol to validate a virtual reality system for rehabilitation. This system used five games for the motor recovery of the upper limb in stroke survivors. The gestural interaction device used was the Kinect. The evaluation involved eight health professionals, with at least 1 year of experience in neurological work, who evaluated 24 stroke patients with different skill levels.

The games' activities used arm movements (unilateral and bilateral) and torso control in the sitting position. For this evaluation, the active range of motion of the arm was only considered within the context of reaching for each patient to determine target placement for the activities.

For the experiment, the patients were divided into four groups of six patients based on level of motor ability. Each patient participated in three 20-minute game sessions of over 10 days. During each session, patients should sit in a chair in front of the Kinect camera at a fixed distance of 1.5 meters as per device specifications. The calibration at this distance was important because the active range of arm movement within the game's reach space determined the target position for the activities.

The patient interacted with all five games at least once during the 20-minute session. The doctor defined the level of difficulty of each activity (eg, required speed, destination number, repetitions), which could be adjusted during the session.

After interacting with each game, the patient had access to a score of correct answers (according to their level of performance) as a way to encourage him/her to continue. At the end of each session, the doctor had access to a global performance report. In the final session, doctors and patients should complete a questionnaire based on the Technology Acceptance Model (TAM) of medical information to assess their opinions about the game system. Additionally, the Fugl-Meier Assessment of sensorimotor recovery after stroke test evaluated the upper limbs. For each patient, session, and activity, they measured the success rate, the performance in tests of success (medium speed and precision), and highest difficulty level reached within the session.

However, this article showed only preliminary results, specifically compiling and identifying the advantages and limitations perceived by clinicians and patients with stroke in rehab games. Success rates, performance scores, and difficulty levels have not been studied yet.

Liu Et Al

Liu et al [15] evaluated the usability of a game they had developed, which used the Kinect device. The game allowed users to accumulate points and stimulated competition among friends. The interaction tasks were to select bubbles that fell from the top of the screen area.

The authors randomly chose six volunteers in a park (two men and four women), who were aged between 50 and 88 years.

The players were able to move one of their hands to select any button options to configure the game. During the interaction process, hands were used freely to select the game objects.

The evaluation experiment consisted of three stages: a brief introduction to the game, testing activity (use the game), and a follow-up interview to collect their comments. First, the evaluator presented the game and the user could ask questions. After that, the user played the game prototype for 1 minute then answered a list of open questions and provided any suggestions for the game design. The activities performed by the users when using the game consisted of indoor exercises, like jump, hit, and grab onto it.

As for results, the authors emphasized some issues. First, the video game presents less danger to the player. Second, it offers more entertainment and, therefore, motivation. The mechanism of obtaining points and prizes by exercising helped the players to increase the amount of exercise unconsciously. The event of winning served as a strong motivator, while also helping to maintain long-term exercise habits for the elderly. Third, through the online gaming platform, players could still have fun together with their friends as if being together somewhere. All current devices work individually and keep the elderly away from their companions, which could have a negative impact on their social life. However, online platforms offer access to players to share their scores and comments with their friends. This not only increased the game's entertainment, but also added competition among friends, helping with motivation and having a positive effect during the exercises. Although all users showed interest in the game and provided a good evaluation, the evaluation questionnaires showed some directions to consider improving the game. The game should have different difficulty levels and diverse tasks. Game messages should provide clearer instructions and not just information. In addition, it must provide kinds of exercises that can help exercise different body parts, yet must be simple for easy learning and understanding by the elderly.

Shin Et Al

Shin et al [16] aimed to combine gestural rehabilitation exercises with game elements using PrimeSense technology, which are 3D-depth camera sensor chips part of Microsoft's Kinect motion-sensing system. Researchers at the University of Hanyang in Seoul, South Korea, conducted the evaluation. The organization of two user groups was as follows: stroke patients and health professionals (occupational therapists and physiatrists) who were involved in the software design of RehabMaster, a game-based virtual reality rehabilitation system developed by the authors. Two clinical studies were performed; the first with seven patients and the second with 16 patients, all diagnosed with stroke. The first was an observational study in which seven patients with chronic stroke received the RehabMaster intervention for 30 minutes per day for 2 weeks. The second was a randomized controlled study of 16 patients with acute or subacute stroke, who received 10 sessions of conventional occupational therapy and plus 20 minutes of the RehabMaster intervention.

The authors evaluated patients' routine tasks individually and focal group studies were performed once a week for approximately 6 months. The software categorized user feedback during the development process.

Regarding gestures and motions, the interventions aimed to stimulate patients through tasks using arm and trunk movements. The motions were intended to promote incremental improvement in range of motion and endurance, strength, and deviation from synergistic motion patterns. RehabMaster provided games to train the patient's forearm movement and eye-hand coordination; upper extremity control, endurance, speed, accuracy, and range of motion; and to increase the control, speed, and accuracy of extremity control and trunk movements.

The same users also participated in the usability study later on. The objective was to evaluate RehabMaster from the perspective of each group. Meetings held with stroke patients took 20 minutes at regular intervals, twice a week for two weeks, under supervision of occupational therapists and physiatrists. Each of the three groups answered a different questionnaire using a five-point Likert scale, so the authors could collect diverse viewpoints. The Fugl-Meyer Assessment and the modified Barthel Index also were used during the evaluation.

Patient involvement was a key point of the RehabMaster intervention. With the stroke patients, the authors wanted to evaluate RehabMaster's ability to provide strong motivation, pleasure, and an optimal flow experience. With the secondary user group (occupational therapists and physiatrists), they wanted to assess the usability of RehabMaster for improving upper limb function and the ability to provide adequate challenge levels for all different patients in the stroke group.

To diagnose if the game provided stroke patients with a desirable rehabilitation level, the study considered three factors in their game experience: attention maintenance, ability, and motivation. These factors were identified through six questions asked of participants. Generally, it found that participants had serious attention and a pleasant experience (immersion), even considering the users' motor limitations.

Tests showed the viability of using RehabMaster in stroke patients with different levels of severity within a safe virtual environment. However, their results were inconsistent due to the different experimental protocols using different intervention times in both experiments.

The authors emphasized the need for a new study. One reason was because cognitive function, motivation, and depression, which are common in stroke patients, were not considered. Another factor was that the usability evaluation did not compare the perspectives of each group.

Fang Et Al

Fang et al [17] developed an interactive prototype motion-based game called Evergreen Fitness System (EFS), in order to train balance in older adults. Health care experts carefully selected the exercises for the users.

The EFS recognizes body gestures and body motions using Kinect. Gestures were used to select the menu (hand movements), whereas body motions were required as part of the exercises available on the game system. Because the goal was to improve senior's balance performance, the focus of the exercises was on lower body strength. Tasks developed consisted of specific exercises for balance training and strengthening of the lower limbs. Six exercises designed for improving balance explored knee marching, side hip raise, lunges, partial squats, wide squats, and standing knee flexion.

Thirteen participants were involved in the study (2 men and 11 women) aged between 60 and 80 years. The study used six exercises, specifically designed by experts to increase lower body strength in older adults, and integrated the elderly in games. Before the test, users were asked to complete a Physical Activity Enjoyment Scale (PAES) questionnaire to evaluate if they were physically capable of performing the test. After the exercises, participants answered a second questionnaire (Physical Activity Readiness Questionnaire, PARQ) that measured the degree of pleasure performing the activity [22].

This study showed that elderly participants approved the exercises based on games, culminating in a positive experience with the EFS. They provided feedback on improving the system design, on the appropriateness of the six exercises, system operation, game design, and demonstrated intention of using the game. It also verified the system should include a navigation requiring less learning, corrective feedback, and warnings while idle.

Harrington Et Al

Harrington et al [18] stated that few studies had examined the usability challenges faced by the elderly using exergames. Thus, it is necessary to identify these challenges and how they translate into guidelines to provide user-friendly exergames for seniors. The objective of this study was to identify the challenges of usability based on Kinect exergames for seniors. Particularly, it aimed to identify which were the most difficult assimilation aspects for the elderly. To do so, 10 people aged between 60 and 69 years (five male and five female) and 10 people (five male and five female) aged between 70 and 79 years, recruited from Georgia Institute of Technology, participated in the study. Pretrials ensured that participants would be able to perform the expected actions; none had experience with the Microsoft Xbox 360 or any device that used Kinect.

The proposed activities were two games that encouraged physical activity: "Body and Brain Connection" and "Your Shape Fitness Evolved." Both games used body motion, providing different activities as participants used their hands or feet to select objects, used their arms in balance challenges, and did torso exercises. Hand gestures were used to select a particular activity in each game, without any restriction.

The evaluation was developed as follows: before participating, each participant completed a health questionnaire, a demographic questionnaire, a technology experience survey, and a game experience questionnaire. Health and demographic questionnaires evaluated the health of the participants and collected basic information, including age, sex, race, education, and limitations (vision, hearing, or mobility). The technology experience questionnaire evaluated the use and familiarity of participants with various technologies. The game experience questionnaire [23] evaluated the participants' levels of familiarity with games and their playing habits.

Additional questionnaires were filled out after each individual session to evaluate user satisfaction and performance. A questionnaire with five items measured satisfaction with the motion controls and gestures for navigation. A seven-item questionnaire assessed satisfaction with the activity developed in the program. Both questionnaires used a scale ranging from 1 (strongly disagree) to 7 (strongly agree). The game experience questionnaire was adapted from Boot et al [23]; information about the other questionnaires was not detailed.

After completing the questionnaires, participants interacted with the Kinect device, including training and definition of the participant's starting position. Test sessions began with researchers giving details of what would be required from the participant. Researchers informed the participants that they could stop the test at any time if they felt they could not complete an activity. After that, an interview assessed participant behaviors and opinions about the programs and the experience. During the interviews, the participants described what they liked and disliked about each program and their line of thought. In addition, participants also answered if they used some kind of help or additional instruction throughout the program. The purpose of these interviews was to determine what made participants feel more frustrated and what types of assistance would be most beneficial.

After completing both exergame programs, each participant completed three questionnaires. The first questionnaire evaluated usability and included the following propositions measured on a Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree): clear and understandable system interaction, useful system for daily life, daily use of game to make one more physically active, and it improves well-being. The second questionnaire assessed the ease of use including the following items measured on the same Likert scale: easy to use, flexibility to interact, increase of skill, clear tasks, and learn to use. The first and second questionnaires were adapted from Davis et al [24]. The third questionnaire, the System Usability Scale (SUS), was adapted from Brooke [25], and was used to give a global view of subjective assessments of usability. Sessions lasted between 1.5 and 3 hours for each participant.

The study showed that older people realize the benefits of exergames, believing it to be a useful means to exercise. Regarding ease of use, the responses were diverse. Most participants in the 60 to 69 year group agreed that the interface was friendly, whereas most in the 70 to 79 year group disagreed with ease of use.

Nakai Et Al

Nakai et al [19] conducted a study to evaluate the usability of a game using the Kansei engineering method or "feelings engineering." Kansei engineering is a method to develop or improve products and services by translating the customer's psychological feelings and needs into the domain of product design. It explores the emotions between a user and a computer system. The study involved 12 users who performed the tasks and system usability testing. They had 10 minutes to play different levels in a game prototype called "The Glider."

In this game, the user controls a virtual glider using body motions and rotations, such as front-back movements to change speed and pitch axis, left-right movements to change direction and roll axis, and torso rotations to control yaw axis. A Kinect device was used to capture the movements of users.

This approach divided the evaluation into five steps as follows: questionnaire, behavioral observation, speech watching, game testing, and analysis. First, the authors developed a gaming environment, which was tested preliminarily by three participants. There were two game preparation sessions. Then, a pretest questionnaire collected basic information about the topics. After the observations, they used a posttest questionnaire to collect participants' impressions.

During these observations, speech and behavioral data from participants were collected based on the think-aloud method. In this method, players are invited to express aloud what they are thinking, doing, and feeling. Therefore, the researcher must take care to explain the experiment purpose to participants, making it clear that is not to test the player's playing skills, but the product itself. It is important to clarify the aim so that users know what is being analyzed so the quality of the experiment is guaranteed. Users performed the task and reported their feelings about the product whenever they failed to complete any of the tasks. In addition, users reported their impressions and thoughts; the analysis of these data depended on the task observation, recorded in a time sequence.

The results demonstrated that the first four levels had good playability and the necessity for players to receive the largest possible amount of information about the game. Motivation was the key point in the game because while they were motivated the game flow looked promising: the game attracted the players' attention and players showed eagerness to learn new things.

Sheu Et Al

Sheu et al [20] aimed to address issues on how to design a gesture-based system that allows older people to play in a secure, convenient, and enjoyable way. This study used two gesture-based games (EG I and EG II) developed by researchers for Kinect. The EG II is an optimized version of the EG I based on the feedback obtained from usability tests performed in the first game.

The exercises used gestures for selections (eg, swing right arm to the right to make cursor move one step to the right and use left arm for moving cursor to the left).

Seven users participated in the experiment (four men and three women), who were aged between 60 and 77 years. The selection criteria for the participants were not detailed. The test had three stages: (1) pretest questionnaires and basic living information for user selection; (2) procedure introduction (game), signing the consent form, and using game for the test procedures (tasks); and (3) posttest questionnaires and interview. Tasks performed consisted of selecting operations in the program interface.

Users were able to complete all tasks. On average, users performed the tasks more quickly on the EG II interface, suggesting that EG II was more usable than EG I. Furthermore, the subjective score for EG II was higher than for EG I. For task selection, it was suggested that, in terms of effectiveness and efficiency, vertical selection works better than horizontal selection because moving the right arm to the right to move the cursor to the right and the left arm to the left to move the cursor to the left can be exhausting. If this movement is vertically oriented, the system interaction becomes less tiring.

To compile the results from all studies, Tables 2 and 3 present the main characteristics of each study and their results to correlate the differences and provide support for the Discussion section.

Table 2. Summarization of the studies included for the systematic review.

Study ID and publication year	Evaluation methods	Focus and devices	Evaluation stages	Participants
Komlódi et al [11] (2011)	Sociodemographic and health questionnaire, observation, video recording, MBTI[a], Folding Test, Eyes Test	Test device for navigation and task manipulation using gestures (Wii)	(1) Questionnaires and tests; (2) verbal guidance and reading; (3) tasks in a virtual environment; (4) evaluation	N=14 (7 men, 7 women); age: mean 24 years
Legouverneur et al [12] (2011)	Author's questionnaire, video recording	Conduct a usability study for 2 sports games (Wii)	(1) Neuropsychological evaluation; (2) sessions tests	N=undefined; age: range 75-90 years
Francese et al [13] (2012)	ASQ[b], CSUQ[c], Presence Questionnaire	Evaluate two 3D interaction games in navigation tasks (Kinect and Wii)	(1) Questionnaire; (2) instructions; (3) tests; (4) ASQ and CSUQ questionnaires	N=24 (16 men, 8 women); age: range 18-41 years
Norouzi-Gheidari et al [14] (2013)	User performance report, author's questionnaire, TAM, Fugl-Meyer	Using a protocol for evaluating a virtual reality system as motor rehabilitation tool of upper limb (Kinect)	(1) System interaction; (2) questionnaires and evaluation	N=24 with stroke; age: undefined
Liu et al [15] (2014)	Interview	Evaluate a game usability for select objects with top-down movements (Kinect)	(1) Game introduction; (2) game activities; (3) interview	N=6 (2 men, 4 women); age: range 50-88 years
Shin et al [16] (2014)	Author's questionnaire, Observation, Fugl-Meyer, Barthel	Combining rehabilitation exercises with game elements (Kinect)	Different experimental protocols	Group 1: n=7, group 2: n=16; age: undefined
Fang et al [17] (2015)	Interview, PARQ[e], PAES[f]	Check the user's experience; train the equilibrium in elderly with upper limb (Kinect)	(1) Physical evaluation; (2) exercises; (3) satisfaction evaluation	N=13 (2 men, 11 women); age: range 60-80 years
Harrington et al [18] (2015)	Sociodemographic and health questionnaire, author's questionnaire, Interview, technology experience and videogame experience questionnaires, TAM, SUS[g]	Identify usability challenges based on exergames for seniors (Kinect)	(1) Questionnaires; (2) training; (3) test; (4) interview; (5) satisfaction questionnaires and usability	Group 1: n=10 (5 men, 5 women), age: range 60-69 years; group 2: n=10 (5 men, 5 women), age: range 70-79 years
Nakai et al [19] (2015)	Sociodemographic and health questionnaires, author's questionnaire, video recording, think-aloud protocol	Evaluate a game usability using evaluation methods based on Kansei Engineering (Kinect)	(1) Questionnaire; (2) behavioral observation; (3) think-aloud; (4) game test; (5) analysis	N=12; age: "seniors"
Sheu et al [20] (2015)	Sociodemographic and health questionnaire, Interview, PARQ, PAES, SUS	To list design issues of a gesture-based system that allows seniors to interact naturally in selection tasks (Kinect)	(1) Questionnaires; (2) game introduction; (3) test procedure; (4) questionnaires posttest; (5) interview	N=7 (4 men, 3 women); age: range 60-77 years

Table 3. Summarization of results of included studies.

Study ID and publication year	Results
Komlódi et al [11] (2011)	For effective usability studies, it is necessary to provide more training time and practice with games. In addition, reducing users memory load, including tasks and feedback functions, and environment helps also improve games usability.
Legouverneur et al [12] (2011)	All participants, regardless of their cognitive status, could use the wireless controller and learn to play both games.
Francese et al [13] (2012)	If the interface is more natural, the user will be as satisfied and engaged in the navigation experience as you want.
Norouzi-Gheidari et al [14] (2013)	This review presented preliminary results. It aimed to compile and identify benefits and limitations perceived by clinicians and patients with stroke in rehab games.
Liu et al [15] (2014)	The game offers more entertainment and less physical risks than physical activity, and may motivate seniors to increase the practice of exercises to get more points.
Shin et al [16] (2014)	It is necessary to define a standard assessment protocol and a time of intervention in order to evaluate the usability of the game.
Fang et al [17] (2015)	Seniors like exercises based on games and showed a positive experience using EFS.
Harrington et al [18] (2015)	Regarding ease of use, most participants in 60-69 year group agreed that the interface was friendly, whereas most in 70-79 year group disagreed with ease of use.
Nakai et al [19] (2015)	It is necessary to provide a game training session and initial guidance to the participants. The motivation proved to be a key point.
Sheu et al [20] (2015)	The way of selecting elements in the program interface must be vertically.

[a]MBTI: Myers-Briggs Type Indicator.
[b]ASQ: After-Scenario Questionnaire.
[c]CSUQ: Computer System Usability Questionnaire.
[d]TAM: Technology Acceptance Model.
[e]PARQ: Physical Activity Readiness Questionnaire.
[e]PAES: Physical Activity Enjoyment Scale.
[f]SUS: System Usability Scale.

Discussion

Overview of Selected Studies

Of the 10 selected studies, seven used the Kinect device for interaction, two used the Wii device, and one used both devices. From this observation, it is possible to propose a study about the reason for this difference because the interaction device may interfere positively or negatively in a usability evaluation. This research could evaluate, for example, if there is really a usability difference between both devices or if the increased use of Kinect is due to its popularity and its complete controller-free gaming experience.

In relation to usability evaluation, nine studies used games in their experiments. Of these, six studies were directed toward the elderly (60%), showing that there are several efforts in serious games for this population. Thus, it is evident there is a need for creating a usability evaluation protocol for serious games for seniors, capable of generating qualitative and quantitative results, because there were no standard serious game evaluation testing protocols found in this research. This trend is supported by the requirements needed to adapt the interface according to age including, for example, the sensitivity of effort and having enough time to do the tasks. In addition, there can be evaluated potential differences in serious game evaluations of 2D or 3D gaming interfaces for the elderly.

Of the nine studies that used games to assess usability, eight studies (89%) used games developed by researchers and only one [12] used a commercial game, in this case for the Wii. This result promotes questions about what caused this situation. It can evaluated as there are no games on the market that meet the objectives of the proposed studies or as the existing games would not be adequate for testing for some reason.

According to Harrington et al [18], inside the elderly population there is further fragmentation that results in groups with special needs for good interface usability. In their study, the majority of participants in the 60 to 69 year group agreed that the interface was friendly, whereas most of the 70 to 79 year group disagreed with ease of use. It is necessary to identify these challenges and apply them into the development process of recommendations for the project in order to provide user-friendly systems to the elderly population and its subgroups.

On the other hand, the study from Legouverneur et al [12] showed that all participants, independently of their cognitive status, were able to use the wireless controller and learn to play both proposed games. They also argued that seniors could improve their game skills throughout the sessions based on their collected performance measures. Yet, several usability sessions allowed users with cognitive impairment to become familiar with technological devices and learn how to use them and have a positive experience.

Nakai et al [19] determined that it is necessary that players receive guidance about the game. This can be an alternative for the same prototype to be applied to many elderly groups even if they have different characteristics.

Most of the studies identified motivation as a major incentive for the elderly to use games to practice physical activities. Zhao et al [26] confirmed this condition, showing that physical exercises are the main intervention instrument in the preventive health and rehabilitation area. In this context, Cary et al [27] and Göbel et al [28] emphasized that serious games are alternative tools and aids for disease prevention. They can stimulate the practice of beneficial activities to human body and increase the patient's interest for his treatment, which is often slow and painful [29]. Another factor that can encourage motivation in users is the use of movement track devices; the more natural the interaction process, the more the user is satisfied and motivated to explore the game resources [13].

Concerning evaluations, questionnaires were the most used instruments (70%, 7/10 studies), followed by interviews (30%, 3/10 studies) and observation and video recording (20%, 2/10 studies). There was also use of think-aloud methods and the Folding Test, but both were applied only in one study.

All studies that used questionnaires were evaluating the interface exclusively. Only two studies used sociodemographic questionnaires [11,18]. All approaches used a questionnaire to evaluate user experiences with natural interaction devices. Four studies that used questionnaires used a Likert scale for responses. Half of them were composed of five variation degrees [12,16] and the other half with seven variation degrees [13,18]. Furthermore, the use of questionnaires was heterogeneous in relation to discussed studies. This shows that there is no standardization. It is possible that evaluations that used Likert scales should vary according to the application under study, but currently there is not specific research about this topic.

No author proposed new approaches for evaluation methods. Some authors suggested and used some instruments, such as the CSUQ questionnaire, MBTI questionnaire, Folding Tests, Eyes Test, Fugl-Meyer Assessment, PAES, the think-aloud method, and modified questionnaires from Boot et al [23] and Brooke [25]. However, it is clear that there is still no protocol for interface usability evaluation for serious games, especially for specific populations such as elderly.

Usability Evaluation Methods Used in Selected Studies

Regarding usability evaluation methods of this review, we observed a variety of techniques applied by authors in different approaches, with low adherence between selected studies. Table 4 summarizes these methods and their intended use in related work.

This heterogeneity shows that researchers are concerned with particularities of their samples and their experiments, avoiding biases. This situation also shows a lack of standardization, at least partially, of a protocol or tools for evaluating games based on gestures and/or movements. This makes the choice of best methods or techniques difficult for a unified approach in future work.

The benefits of this diversity are a good number of approaches that used evaluation instruments are well established in the literature. A group of authors, for instance, applied tools to evaluate cognitive, emotional, and motor skills in experiments. Another group applied methods and techniques to evaluate user satisfaction, user perception, and user performance.

It is important to keep it in mind during the creation of a standardization process to evaluate usability because there are consolidated techniques and scales for measuring gains according to the specificity of each approach. For example, comparison of games for people with upper limb impairment can use the Fugl-Meyer test as part of the evaluation process. In a rehabilitation context, it can support decisions and appropriately choose and validate a game according to the user's profile.

Considering the evaluation instruments used by selected studies, it is possible to think about advantages and disadvantages to define the basis of a standardized procedure to evaluate usability of gesture-based games regardless of the user experience.

In reference to sociodemographic and health questionnaires, it is important to verify the age and previous game experience (software and hardware) of the user. This can help divide the user groups and the test conduction. Preliminarily, physical and cognitive limitations can also be identified to avoid health hazards and contribute to a better gaming experience.

Interviews are recommended when the number of participants is small, given that the collection data are qualitative and demands a time-consuming analysis. For example, in preliminary assessments, groups are smaller and interviews are useful to understand the reasoning of the user facing a problem. On the other hand, the subjective nature of the interviews leads to different interpretations by the evaluators in groups with many users.

Observations also require time for analysis because of the large amount of data acquired. However, it may provide a different perspective to the evaluator that other techniques do not provide, such as the moment when a problem occurred. Used together with video recording, observations can facilitate the review process of user actions and enrich the usability evaluation.

Think-aloud protocols affect user performance because they force the participant to do more than one task at a time, resulting in the loss of focus on game tasks or in unintended actions. Another problem is that motion-based systems use sensors for speech recognition as an interaction technique (because the feature is available on motion-sensing input devices), making it impossible in the use of this protocol. Therefore, we not recommend it in a usability evaluation for motion-based games.

User performance reporting is an interesting instrument because the game software can collect the measures during the interaction process and it is useful for collating with other assessment tools. For example, you can make a relationship between time spent executing the task with acceptance of the technology in order to see whether the user liked the game or not. However, if analyzed without comparison with another instrument, it is essential to instruct users in a very specific way about how they should perform the task in order to obtain balanced results. This is important because there is an implicit relationship between task performance metrics, such as speed and accuracy. The participant may be faster, but be less accurate, or the participant can increase accuracy, but decrease the speed [21].

Table 4. Purposes of each method in selected studies.

Method and study ID	Used to...
Sociodemographic and health questionnaire	
Komlódi et al [11]	Identify the users' profiles
Harrington et al [18]	Identify the users' profiles and their physical limitations
Nakai et al [19]	Identify the users' profiles
Sheu et al [20]	Obtain personal information
Author's questionnaires	
Legouverneur et al [12]	Get the user satisfaction and verify if familiarization with the games have influence on user preference measures
Norouzi-Gheidari et al [14]	Evaluate the acceptance of virtual reality technology for games
Shin et al [16]	Test the usability of the game from expert perspective
Harrington et al [18]	Evaluate the user task performance and user satisfaction
Nakai et al [19]	Get the user feedback about the game
Observation	
Komlódi et al [11]	Verify the users' behaviors and issues during a session
Shin et al [16]	Assess the usability and the negative effects of the game
Video recording	
Komlódi et al [11]	Verify the task time and get usability issues
Legouverneur et al [12]	Used to elicit users' behavior and performance
Nakai et al [19]	Verify the users' behaviors during a session
User performance report	
Norouzi-Gheidari et al [14]	Evaluate success rate, speed, and accuracy during the tasks
Think-aloud protocol	
Nakai et al [19]	Verify the users' behaviors during a session
Interview	
Liu et al [15]	Collect suggestions about the game in open questions (qualitative data analysis)
Harrington et al [18]	Identify what users liked or not in each game, the reason of the answers, their frustration, and what form of aid is the most beneficial
Sheu et al [20]	Verify the user experience and get doubts
Fang et al [17]	Evaluate the combined use of exercises and verify the user experience
Other relevant tools	
Komlódi et al [11]	Verify cognitive, motor, and emotional aspects (MBTI[a], Folding Test, Eyes Test)
Francese et al [13]	Measure the user satisfaction during interaction process, the usability and the quality of the system, and the presence and immersion (ASQ[b], CSUQ[c], Presence Questionnaire)
Norouzi-Gheidari et al [14]	Assess of sensorimotor function of upper limbs (Fugl-Meyer Assessment)
Shin et al [16]	Assess of sensorimotor function of upper limbs, and functional capacity (Fugl-Meyer Assessment, Barthel)
Fang et al [17]	Ensure that participants were physically and mentally ready to perform to play, and check the degree of pleasure during the tasks (PARQ[d], PAES[e])
Harrington et al [18]	Evaluate gameplay experience, check the acceptance of technology, the familiarity with technology, and measure the usability (effectiveness, efficiency and satisfaction) (Technology Experience and Videogame Experience questionnaires, TAM[f], SUS[g])
Sheu et al [20]	Ensure that participants were physically and mentally ready to perform the tasks, check the degree of pleasure during the activities, and measure the usability (effectiveness, efficiency and satisfaction) (PARQ, PAES, SUS)

[a]MBTI: Myers-Briggs Type Indicator.

[b]ASQ: After-Scenario Questionnaire.

[c]CSUQ: Computer System Usability Questionnaire.

[d]PARQ: Physical Activity Readiness Questionnaire.

[e] PAES: Physical Activity Enjoyment Scale.

[f]TAM: Technology Acceptance Model.

[g]SUS: System Usability Scale.

Finally, the use of posttest questionnaires without validation is not a good practice in evaluations because it is not certain that the listed issues are relevant and appropriate inside the application context. Moreover, it is necessary to manage specific questionnaires at the end of experimental session because they can tire the participants leading to them not adequately indicating their impressions about the system. On the other hand, it is important to create specific questionnaires to evaluate the usability of motion-based games. In this case, the questionnaires must be validated in preliminary studies to prove the quality of the results. Preferably, they must be applied with other posttest tools in preliminary assessments in order to define specific time for the experiment.

Suggested Guidelines for a Usability Evaluation Approach

In view of our impressions, a suggested usability evaluation approach for gesture-based and motion-based games can follow some guidelines within the pretest, test, and posttest stages.

In the pretest stage, we recommend the use of instruments capable of distinguishing users with previous experience considering the resources under evaluation and capable of indicating user limitations that may interfere during the experiment. We suggest the use of a questionnaire to characterize the sample. This instrument must track cognitive and physical aspects to avoid biases. For example, it can identify a problem that affects the understanding of the game activities during the sessions, or even a limitation that affects the movement required in gestural interaction. Among the instruments listed in the studies included in this systematic review were PARQ and PAES.

During the test, we suggest collecting performance and physiological user data using software. User performance reporting is useful if the purpose is to verify the user's progress during the interaction process, usually from objective measures, such as speed and accuracy. However, it is interesting to use tracker video software to study the evolution of movements. Physiological measures can also be useful to identify emotions that may affect the user's interaction with the game tasks. Heart rate, for example, can show evidence of stress during the interaction process. Observations can also record the evaluator's perception during the session. We also recommend testing different versions of the same game to identify relevant issues, such as a version that uses a specific guideline and another without using it.

In the posttest stage, one can opt for a qualitative or quantitative analysis, depending on the purpose of the evaluation. If the option is for qualitative results, we recommend an interview to collect the user's perception of the game. On the other hand, in a quantitative analysis, the use of instruments such as TAM and SUS, for example, may be useful for evaluating user acceptance and user satisfaction. In this case, we recommend scales that allow for statistical analysis, such as a Likert scale.

Conclusions

Results show that there is no standardization in evaluation methods because they use different analysis variables. The definition of usability in games, especially in relation to gestural interaction, and of who should be evaluated and how it should be assessed, were not evident in the selected studies. Some studies evaluated users and others experts, using qualitative and/or quantitative methods.

We observed that the studies in this systematic review do not use the same methods in the user selection process or similar criteria in pilot tests or protocols for usability evaluations. There was also not similarity between the questionnaires and answer options.

We suspect that the lack of studies and methods (as well as theoretical foundation) that indicate the appropriate interaction techniques for each experience or application is the reason for the lack of standardization.

For this reason, the use of a standard usability evaluation process for gesture-based games and definition of criteria to enable a quantitative analysis in evaluations can be significant to this area. Therefore, understanding the different types of goals in usability evaluation and its implementation becomes relevant. Once the different types of goals have been established, it becomes possible, for example, to create a usability evaluation tool for rehabilitation of people with some type of motor or cognitive impairment, or even for unimpaired people with a focus on entertainment or training.

An evaluation for gesture-based games may also consider other factors in order to contribute to the usability applications. Physiological metrics, user anxiety levels, and stress issues are measures that could be collected during the evaluation. It is essential to consider these questions to advance the study in this area, defining at least one basic usability method to guide user assessments.

With this in mind, it is important to consider the adoption of a standard for usability evaluations for two reasons: to guide the validation of gesture-based and/or motion-based systems in future case studies and to check whether a particular hardware or software is suitable for their intended use. This valuation approach is substantial because the market regularly offers new interaction methods without consideration for how the public will use these solutions. As an example, the elderly, a group that is increasing globally, are using interaction systems more and more every day.

Future Work

Our proposal is to validate the suggested approach in a usability evaluation protocol using, as a case study, two versions of a serious game based on gestures and movements (2D and 3D) for the elderly. We will regard three stages suggested in this study. We believe that this validation can be the basis to consolidate a standardized usability evaluation approach for gesture-based games based on our experiences and the identified studies in this review.

The focus for the elderly is justified by the fact that the game industry has been investing in this age group due to the growth of this population and increased life expectancy in most countries. In this systematic review, it was observed that some studies also used this population [12,17,18,20], with serious games for physical exercise and rehabilitation processes.

Adults today have greater contact with technologies and, therefore, will tend to use them more in the future. Easy access to a larger set of gestural interaction devices and general information systems will also contribute to the use of these technologies by a greater number of users.

Acknowledgments

The authors are grateful to Brazilian National Council for Scientific and Technological Development for their support through the public notice MCTI/CNPQ/Universal 14/2014.

Conflicts of Interest

None declared.

References

1. Karray F, Alemzadeh M, Saleh JA, Arab MN. Human-computer interaction: overview on state of the art. Int J Smart Sens Intell Sys 2008 Mar;1(1):137-159 [FREE Full text]
2. Nielsen J. Nielsen Norman Group. 2012. Usability 101: introduction to usability URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/ [accessed 2016-04-09] [WebCite Cache ID 6gd0Zfdkr]
3. Cockton G. Usability evaluation. In: Soegaard M, Friis R, editors. The Encyclopedia of Human-Computer Interaction. Aarhus, Denmark: The Interaction Design Foundation; 2012.
4. Morelli T, Folmer E. Real-time sensory substitution to enable players who are blind to play video games using whole body gestures. Entertain Comp 2014;5(1):83-90. [doi: 10.1145/2159365.2159385]
5. Ribeiro A, Duarte C. Move and speak: interaction with multiple surfaces using gestures and voice. 2012 Presented at: First International Symposium on Pervasive Displays; Jun 4-5, 2012; Porto, Portugal.
6. Karam M, Schraefel MC. A Taxonomy of Gestures in Human Computer Interactions [monograph]. Southampton, UK: University of Southampton; 2005. URL: http://eprints.soton.ac.uk/261149/ [accessed 2016-09-02] [WebCite Cache ID 6kEYcwBT2]
7. Keskinen T, Hakulinen J, Heimonen T, Turunen M, Sharma S, Miettinen T, et al. Evaluating the experiential user experience of public display applications in the wild. In: MUM '13. 2013 Presented at: 12th International Conference on Mobile and Ubiquitous Multimedia; Dec 2-5, 2013; Lulea, Sweden p. 7. [doi: 10.1145/2541831.2541840]
8. Rautaray S, Pandey M. Single and multiple hand gesture recognition systems: a comparative analysis. IJISA 2014 Oct 08;6(11):57-65. [doi: 10.5815/ijisa.2014.11.08]
9. Maidi M, Preda M. Interactive media control using natural interaction-based Kinect. 2013 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing; May 26-31, 2013; Vancouver, BC p. 1812-1815. [doi: 10.1109/ICASSP.2013.6637965]
10. Dyba T, Dingsoyr T, Hanssen GK. Applying systematic reviews to diverse study types: An experience report. 2007 Presented at: First International Symposium on Empirical Software Engineering and Measurement; Sep 20-21, 2007; Madrid, Spain p. 225-234. [doi: 10.1109/ESEM.2007.59]
11. Komlódi A, Józsa E, Hercegfi K, Kucsora S, Borics D. Empirical usability evaluation of the Wii controller as an input device for the VirCA immersive virtual space. In: CogInfoCom. 2011 Presented at: 2nd International Conference on Cognitive Infocommunications; Jul 7-9, 2011; Budapest p. 1-6.
12. Legouverneur G, Pino M, Boulay M, Rigaud A. Wii sports, a usability study with MCI and Alzheimer's patients. Alzheimer Dement 2011 Jul;7(4):S500-S501. [doi: 10.1016/j.jalz.2011.05.2398]
13. Francese R, Passero I, Tortora G. Wiimote and Kinect: gestural user interfaces add a natural third dimension to HCI. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. 2012 Presented at: International Working Conference on Advanced Visual Interfaces; May 22-25, 2012; Capri Island (Naples), Italy p. 116-123. [doi: 10.1145/2254556.2254580]
14. Norouzi-Gheidari N, Levin M, Fung J, Archambault P. Interactive virtual reality game-based rehabilitation for stroke patients. In: Proceedings of the ICVR. 2013 Presented at: International Conference on Virtual Rehabilitation; Aug 26-29, 2013; Philadelphia, PA p. 220-221.

15. Liu Z, Liao C, Choe P. An approach of indoor exercise: Kinect-based video game for elderly people. In: Rau PL, editor. Cross-Cultural Design: Proceedings 6th International Conference, CCD 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014. Cham, Switzerland: Springer International Publishing; 2014:193-200.
16. Shin J, Ryu H, Jang SH. A task-specific interactive game-based virtual reality rehabilitation system for patients with stroke: a usability test and two clinical experiments. J Neuroeng Rehabil 2014;11:32 [FREE Full text] [doi: 10.1186/1743-0003-11-32] [Medline: 24597650]
17. Fang W, Sheu F, Lin Y, Lee Y, Chen N. Interactive physical games: improving balance in older adults. In: Chang M, Li Y, editors. Smart Learning Environments. Berlin: Springer Berlin Heidelberg; 2015:159-174.
18. Harrington CN, Hartley JQ, Mitzner TL, Rogers WA. Assessing older adults' usability challenges using Kinect-based exergames. In: Zhou J, Salvendy G, editors. Human Aspects of IT for the Aged Population. Design for Everyday Life: First International Conference, ITAP 2015, Held as Part of HCI International 2015. Cham, Switzerland: Springer International Publishing; 2015:488-499.
19. Nakai A, Pyae A, Luimula M, Hongo S, Vuola H, Smed J. Investigating the effects of motion-based Kinect game system on user cognition. J Multimodal User Interface 2015;9(4):403-411. [doi: 10.1007/s12193-015-0197-0]
20. Sheu F, Lee Y, Yang S, Chen N. User-centered design of interactive gesture-based fitness video game for elderly. In: Chen G, Kumar V, Kinshuk , Huang R, Kong CS, editors. Emerging Issues in Smart Learning. Berlin: Springer Berlin Heidelberg; 2015:393-397.
21. Bowman DA, Kruijff E, LaViola JJ, Poupyrev I. 3D User Interfaces: Theory and Practice. Boston, MA: Addison-Wesley; 2005.
22. Raedeke T. The relationship between enjoyment and affective responses to exercise. J Applied Sport Psychol 2007;19(1):105-115. [doi: 10.1080/10413200601113638,]
23. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. Acta Psychol (Amst) 2008 Nov;129(3):387-398. [doi: 10.1016/j.actpsy.2008.09.005] [Medline: 18929349]
24. Davis F, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 1989 Aug;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
25. Brooke J. SUS: A quick and dirty usability scale. Usability Eval Ind 1996;189(194):4-7.
26. Zhao W, Feng H, Lun R, Espy DD, Reinthal M. A Kinect-based rehabilitation exercise monitoring and guidance system. In: Proceedings of ICSESS. 2014 Jun Presented at: 5th IEEE International Conference on Software Engineering and Service Science; Jun 27-29, 2014; Beijing, China p. 762-765. [doi: 10.1109/ICSESS.2014.6933678]
27. Cary F, Postolache O, Girão PS. Kinect based system and serious game motivating approach for physiotherapy assessment and remote session monitoring. In: Proceedings of ICST. 2014 Sep Presented at: International Conference on Sensing Technology; Sep 2-4, 2014; Liverpool, UK p. 474-479.
28. Göbel S, Hardy S, Wendel V, Mehm F, Steinmetz R. Serious games for health: personalized exergames. In: Proceedings of ACM Multimedia 2010. 2010 Presented at: 18th ACM international conference on Multimedia; Oct 25-29, 2010; Firenze, Italy p. 1663-1666. [doi: 10.1145/1873951.1874316]
29. Broeren J, Bjorkdahl A, Claesson L, Goude D, Lundgren-Nilsson A, Samuelsson H, et al. Virtual rehabilitation after stroke. In: Andersen SK, Klein MC, Schulz S, Aarts J, Mazzoleni MC, editors. Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2008:77-82.

Abbreviations

ACM: Association for Computing Machinery
ASQ: After-Scenario Questionnaire
CSUQ: Computer System Usability Questionnaire
EFS: Evergreen Fitness System
MBTI: Myers-Briggs Type Indicator
PAES: Physical Activity Enjoyment Scale
PARQ: Physical Activity Readiness Questionnaire
SUS: System Usability Scale
TAM: Technology Acceptance Model

Edited by G Eysenbach; submitted 09.04.16; peer-reviewed by M Joselli, E Taranta; comments to author 28.06.16; revised version received 31.07.16; accepted 22.08.16; published 04.10.16.

Please cite as:
Simor FW, Brum MR, Schmidt JDE, Rieder R, De Marchi ACB
Usability Evaluation Methods for Gesture-Based Games: A Systematic Review
JMIR Serious Games 2016;4(2):e17
URL: http://games.jmir.org/2016/2/e17/
doi:10.2196/games.5860
PMID:27702737

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org