JMIR Serious Games

Journal Impact Factor (JIF) (2023): 3.8
Volume 4 (2016), Issue 1 ISSN 2291-9279 Editor-in-Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Original Paper

Behavioral Economics, Wearable Devices, and Cooperative Games: Results From a Population-Based Intervention to Increase Physical Activity

Trevor van Mierlo[1,2*], MScCH, MBA, GEMBA, MSc BMR; Douglas Hyatt[3*], PhD; Andrew T Ching[3*], PhD; Rachel Fournier[1,2*], BSc (Hons); Ron S Dembo[4*], PhD

[1]Evolution Health Systems Inc., Toronto, ON, Canada

[2]Henley Business School, University of Reading, Greenlands, Henley-on-Thames, United Kingdom

[3]Rotman School of Managment, University of Toronto, Toronto, ON, Canada

[4]Zerofootprint, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Trevor van Mierlo, MScCH, MBA, GEMBA, MSc BMR
Evolution Health Systems Inc.
901 King Street West
Suite 400
Toronto, ON, M5V 3H5
Canada
Phone: 1 4166448476
Fax: 1 4166448477
Email: tvanmierlo@evolutionhs.com

Abstract

Background: Health care literature supports the development of accessible interventions that integrate behavioral economics, wearable devices, principles of evidence-based behavior change, and community support. However, there are limited real-world examples of large scale, population-based, member-driven reward platforms. Subsequently, a paucity of outcome data exists and health economic effects remain largely theoretical. To complicate matters, an emerging area of research is defining the role of Superusers, the small percentage of unusually engaged digital health participants who may influence other members.

Objective: The objective of this preliminary study is to analyze descriptive data from GOODcoins, a self-guided, free-to-consumer engagement and rewards platform incentivizing walking, running and cycling. Registered members accessed the GOODcoins platform through PCs, tablets or mobile devices, and had the opportunity to sync wearables to track activity. Following registration, members were encouraged to join gamified group challenges and compare their progress with that of others. As members met challenge targets, they were rewarded with GOODcoins, which could be redeemed for planet- or people-friendly products.

Methods: Outcome data were obtained from the GOODcoins custom SQL database. The reporting period was December 1, 2014 to May 1, 2015. Descriptive self-report data were analyzed using MySQL and MS Excel.

Results: The study period includes data from 1298 users who were connected to an exercise tracking device. Females consisted of 52.6% (n=683) of the study population, 33.7% (n=438) were between the ages of 20-29, and 24.8% (n=322) were between the ages of 30-39. 77.5% (n=1006) of connected and active members met daily-recommended physical activity guidelines of 30 minutes, with a total daily average activity of 107 minutes (95% CI 90, 124). Of all connected and active users, 96.1% (n=1248) listed walking as their primary activity. For members who exchanged GOODcoins, the mean balance was 4,000 (95% CI 3850, 4150) at time of redemption, and 50.4% (n=61) of exchanges were for fitness or outdoor products, while 4.1% (n=5) were for food-related items. Participants were most likely to complete challenges when rewards were between 201-300 GOODcoins.

Conclusions: The purpose of this study is to form a baseline for future research. Overall, results indicate that challenges and incentives may be effective for connected and active members, and may play a role in achieving daily-recommended activity guidelines. Registrants were typically younger, walking was the primary activity, and rewards were mainly exchanged for fitness or outdoor products. Remaining to be determined is whether members were already physically active at time of registration and are representative of healthy adherers, or were previously inactive and were incentivized to change their behavior. As challenges

are gamified, there is an opportunity to investigate the role of superusers and healthy adherers, impacts on behavioral norms, and how cooperative games and incentives can be leveraged across stratified populations. Study limitations and future research agendas are discussed.

(JMIR Serious Games 2016;4(1):e1) doi:10.2196/games.5358

KEYWORDS

behavioral economics, cooperative games, adherence, compliance, serious games, Superusers, health rewards, internet of things, wearable devices

Introduction

Opportunities related to behavioral economics [1-2], wearable devices [3-4], tailored evidence-based behavior change tools [5-7], and community support [8-9] are highlighted in health care literature, but their collective integration into real-world interventions are limited in scope. Subsequently, there is a paucity of outcome data, and health economic effects remain largely theoretical.

While the literature does investigate the relationship between exercise and the use of less sophisticated wearable devices, such as pedometers [10-11], there are little, if any, published outcomes on more sophisticated devices (such as Fitbit, Jawbone, or Apple Watch), the use of tracking apps (such as Moves, Runtastic Pedometer, or Pedometer++), or health-tracking platforms (such as Google Fit, MapMyFitness, or GOODcoins).

However, this will soon change. One source identifies over nearly 300 registered clinical trials that are utilizing devices in their protocols [12]. A recent pharma-focused trade publication notes that although in its infancy, wearables are emerging as a multifaceted solution to typical problems in clinical trials [13], and Google is developing a wristband health tracker specifically for the clinical research industry [14]. In relation to physical activity, tailored prescriptions that leverage personalized algorithms and wearables are feasible [15].

These efforts to leverage digital health tools and behavioral incentives [16-20] have escalated in recent years, perhaps due to the growing health care crisis in North America. Costs of medication and treatment non-adherence are estimated to exceed $300 billion each year [21-22], and policy theorists have identified technology's potential to have an important impact on decreasing costs and increasing intervention efficacy [23-24].

However, researchers are showing concern over high program attrition rates [25], sustainability [26], and the failure of digital health to show impacts at population levels [27-28]. A countermeasure is that increasing amounts of data becoming available, and analysis of specific usage patterns and topologies are becoming more insightful [29-30].

For example, a rule of thumb in digital marketing is the 1% rule, or 90-9-1 principle, which seeks to explain network participatory patterns [31]. The rule states that 90% of network actors observe and do not participate, 9% contribute sparingly, and 1% of actors create the vast majority of new content. This 90%, 9%, and 1% are also known as Lurkers, Contributors, and Superusers, respectively [32].

Since healthy adherers and Superusers tend to be the primary contributors to community-based tools, a concern is that interventions are mainly utilized by healthy adherers, individuals who are already highly engaged in healthy behaviors [33-34]. However, it is also possible that in community-based platforms, Superusers may influence those who are less active.

Objective

The objective of this preliminary study is to analyze descriptive data from GOODcoins, a self-guided, free-to-consumer engagement and rewards platform incentivizing walking, running and cycling.

The Intervention

Registered members accessed the GOODcoins platform through PCs, tablets or mobile devices. Following registration, users were encouraged to sync wearable devices such as Jawbone, Fitbit or the Moves App to their profile.

Zerofootprint Software Inc., the Toronto-based Corporation that manages GOODcoins, is a software company that aggregates data from sensors, databases, medical devices, smart electronics and telematics for the purposes of creating evidence and reward-based behavior change. GOODcoins is a social currency that is being awarded to members for achieving activity goals. There is no cost or membership fee, and any individual can join the GOODcoins platform. Cumulative and anonymized data generated from the program is analyzed by Zerofootprint, program sponsors, or academic partners.

Gamified Group Challenges Measuring Individual Progress

Following registration and the syncing of their wearable device or app, members were encouraged to opt into various gamified group challenges that involved walking, running, or biking.

Each challenge was unique, and had its own reward structure. GOODcoins (or partner organizations that sponsored a challenge) determined the challenge reward structures. Figure 1 illustrates three specific challenges that members could choose to join on November 11, 2015.

Once a member joined a specific challenge, they were able to compare their progress to other GOODcoins members. Figure 2 illustrates the challenge "Walk 30 Minutes". In this challenge, members were rewarded 10 GOODcoins if they walked 30 minutes each day.

The challenge was gamified as it allowed users to measure their daily progress. Normative feedback allowed members to compare their progress to other members who also opted into the challenge. This is illustrated by the chart to the left of the

text “This chart shows how your activity compares to others in this challenge”. The section below the chart outlined the member’s daily progress on an individual level.

Periodically, new, short-term challenges were offered to the community. Figure 3 illustrates the completed challenge “Walk around the Earth”.

In this challenge, cooperative game theory was utilized to encourage members to individually contribute to a single overall goal. The challenge, offered between June 1, 2015 and June 30, 2015, was initiated by the question: “Can the GOODcoins community walk around the earth together this month? Let’s find out! Contribute each day and get 50 GOODcoins each time you meet the daily targets”.

Each member who accepted the challenge was incentivized by being rewarded 50 daily GOODcoins for walking 5 kilometers per day.

Figure 1. GOODcoins opt-in challenges for November 11, 2015.

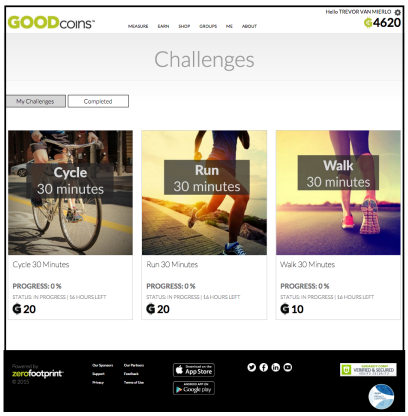
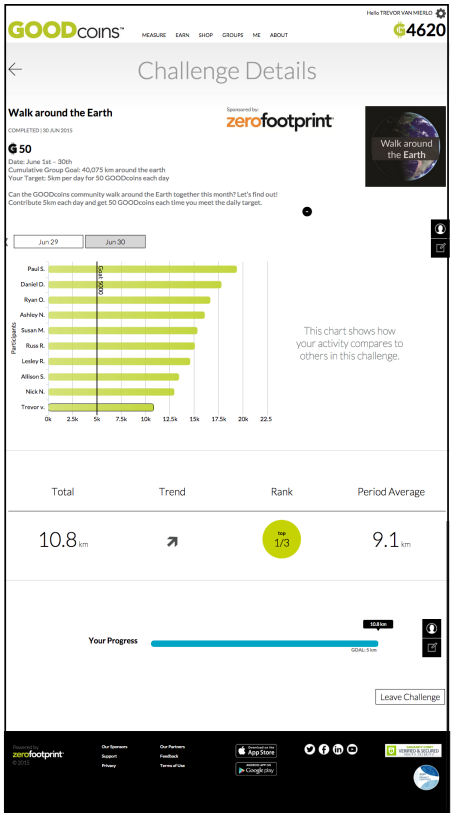


Figure 2. Walk 30 minutes challenge.

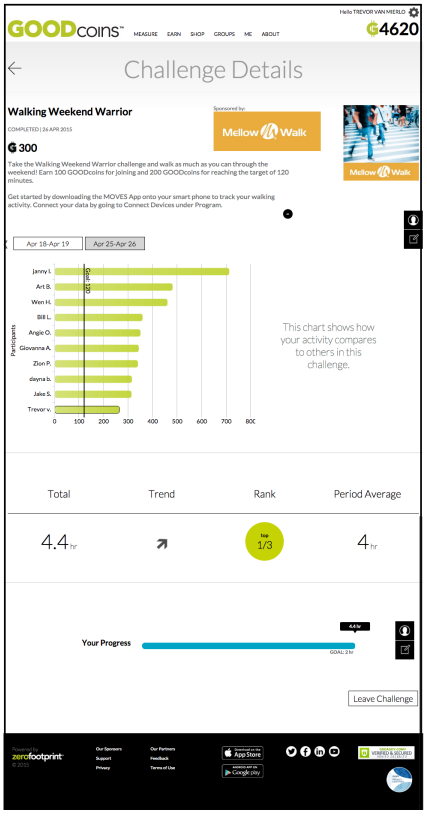
Figure 3. Walk around the earth challenge.



Sponsored Challenges

Challenges are sponsored by Zerofootprint, or partner organizations. Figure 4 is an example of a challenge titled "Walking Weekend Warrior", which was sponsored by Mellow Walk, a Canadian shoe retailer. The challenge was offered over one specific weekend, and members were rewarded with 100 GOODcoins for joining, and 200 GOODcoins for reaching the target of 120 minutes.

Figure 4. Challenge sponsored by Mellow Walk.

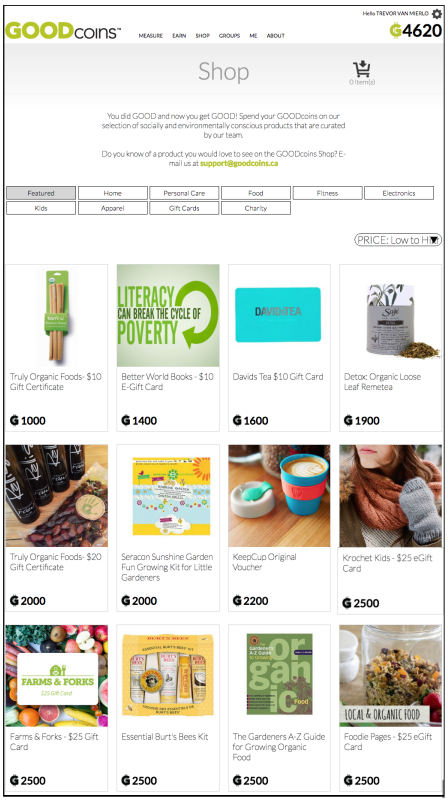


Rewards

As members met challenge targets and accumulated GOODcoins, they had the option of redeeming their GOODcoins for planet- or people-friendly products offered in the GOODcoins shop (Figure 5).

All products were curated by the GOODcoins team, and were deemed to be socially and environmentally conscious.

Figure 5. Sample products in the GOODcoins shop.



Anticheat Measures

Although it is not possible to ensure that members, groups of members, or a single member posing as a group of members do not commit fraud, risk is managed in 4 specific ways. First, statistical techniques can be used to detect whether member movements fall within realistic ranges. Movement that is not within a realistic range is flagged. Second, internal staff reviews challenges and challenge completion rates. Third, redemption rates in the GOODcoins shop are monitored. Finally, members of the GOODcoins community have the opportunity to contact GOODcoins directly if they observe atypical behavior in gamified group challenges.

Methods

The reporting period was December 1, 2014 to May 1, 2015. Descriptive self-report data were analyzed using MySQL and MS Excel. All member data are self-report. Outcome data in this study was obtained from the GOODcoins custom SQL database.

At registration all members consented to the use of their data for research or commercial purposes. Data collection procedures adhered to Canadian privacy guidelines [35].

Prior to analysis, data was scrubbed of test cases, and de-identified. Typical of digital health studies based on retrospective databases which are free of personally identifiable information, the authors deemed the study exempt from formal, ethical review.

Results

Overall Findings

The study period includes data from 1298 users who were connected to an exercise-tracking device. Females consisted of 52.62% (683/1298) of the population, 33.74% (438/1298) were between the ages of 20-29, and 24.81% (322/1298) were between the ages of 30-39. Canadians comprised 89.45% (1161/1298) of the sample (Table 1).

Table 1. Demographic characteristics of GOODcoins members.

General characteristics		n (%)
Total population		4342 (100.0)
Total population using health care challenges		1298 (29.9)
Gender		
	Female	683 (52.6)
	Male	557 (42.9)
	Unknown	58 (4.5)
Age		
	19 and under	70 (5.4)
	20-29	438 (33.7)
	30-39	322 (24.8)
	40-49	254 (19.6)
	50-59	106 (8.2)
	60 and above	30 (2.3)
	Unknown	78 (6.0)
Nationality		
	Canadian	1161 (89.4)
	American	90 (6.9)
	Other	47 (3.6)

Table 2 outlines activity recorded from members' wearable devices. Over 77% of connected and active members (1006/1298) met daily recommended physical activity guidelines of 30 minutes, with a total daily average activity of 107 minutes (95% CI 90-124).

Slightly over 96% of connected and active users (1248/1298) engaged in walking as their primary activity, versus 1.54% who preferred running (20/1298), or 2.31% who preferred cycling (30/1298).

Table 2. Activity.

Primary activity	Percentage of user engagement, n (%)
Walking	1248 (96.14)
Running	20 (1.54)
Cycling	30 (2.31)

Of members who exchanged GOODcoins, the mean balance at time of redemption was 4,000 (equivalent to approximately US $40) (95% CI 3850-4150). Over 50% (61/122) of redemptions were for fitness or outdoor products, while 4.1% (5/122) were for food-related items (Table 3).

Table 3. Redemption categories.

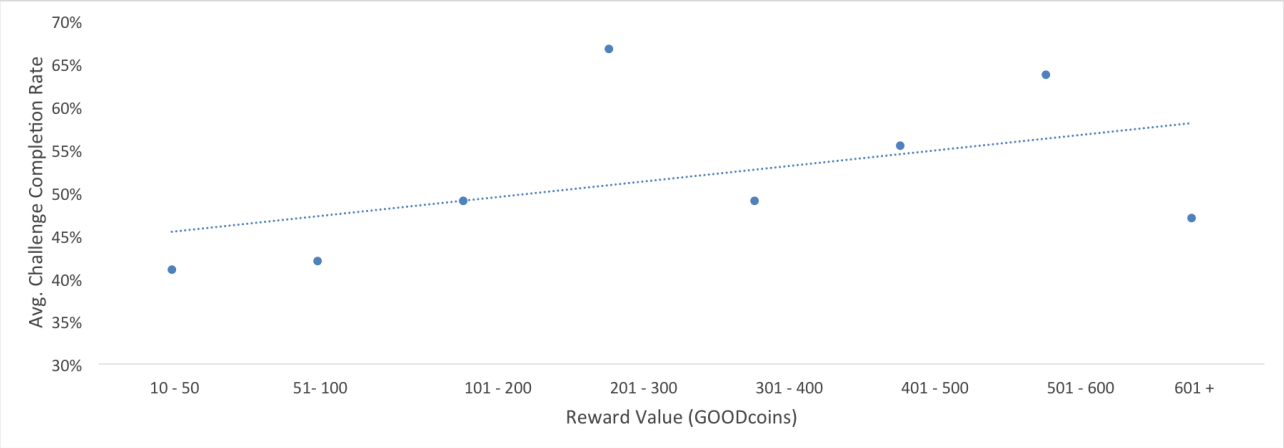
Redemption categories	n (%)
Food	5 (4.1)
Apparel	14 (11.6)
Home and Garden	15 (12.4)
Personal Care	26 (21.5)
Fitness/Outdoor	61 (50.4)

Participants were most likely to complete challenges when rewards were between 201-300 GOODcoins (Table 4 and Figure 6).

Table 4. Challenge completion rate.

GOODcoins value	Average challenge completion rate, %
10-50	41
51-100	42
101-200	49
201-300	67
301-400	49
401-500	55
501-600	64
601+	47

Figure 6. Relationship between reward value & challenge completion rate.



Discussion

Principal Findings

This analysis is observational, and its purpose is to form a baseline for future research in this rapidly emerging field. Results indicate that challenges and incentives may be effective for connected and active members, and may play a role in achieving daily recommended activity guidelines.

Members were generally female (52.62%, 683/1298) and under the age of 40. Walking was the primary activity (96.14%, 1248/1298), and 50.4% of rewards (61/121) were exchanged for fitness or outdoor products. More detailed demographic and psychographic data could assist in the development of profile utilization patterns, motivations, and impact on overall health behaviors.

To achieve health benefits, Canadian adults aged 18-64 should accumulate at last 150 minutes of moderate- to vigorous-intensity aerobic activity per week in bouts of 10 minutes or more [36]. It is encouraging to note that 77.50% (1006/1298) of connected and active members were physically engaged for at least 30 minutes per day, with a total daily average activity of 107 minutes (95% CI 90-124).

Participants were most likely to complete challenges when rewards were between 201-300 GOODcoins (67% completion rate). However the challenge completion rate was 64% when rewards were between 501 and 600 GOODcoins, 55% between 401 and 500 GOODcoins, and 47% when GOODcoin rewards were over 600. This lack of detectable trends warrants investigation into how rewards and incentives are positioned. Researchers may wish to apply models familiar to economics and finance such as hyperbolic discounting, operant conditioning, or matching law.

Yet to be determined is whether members were already physically active at time of registration and are representative of healthy adherers, or were previously inactive and were incentivized to change their behavior.

As outlined in Figure 1, challenges are gamified, and recent health studies have illustrated positive effects from group-based challenges [17-18]. Future research should also investigate the role of Superusers and healthy adherers, and their impact on behavioral norms.

Strengths and Limitations

A strength of this study is that participants belong to a naturalistic, self-seeking population that may be representative of digital patients who seek to participate in rewards programs. However, this same strength may also be interpreted as a weakness as the program may primarily attract healthy adherers. It will also be important for future research to analyze 3, 6 and 12-month trends to determine increases or decreases in individual physical activity levels, and assess if subjects are reaching incentives and goals.

An additional strength is the inclusion of real-time data from wearable devices. Data are continually synced from devices to the GOODcoins platform, so it would be difficult for users to manipulate results.

Absent from this analysis are details examining the potential relationships between number of participants in a specific type of challenge, variance of completion rates, and reward value. Future research should consider the optimization of these relationships through the lens of economic models such as hyperbolic discounting or pooling, or behavior change strategies such as normative feedback or motivational interviewing.

An important limitation is that all connected wearables track walking (steps), however only a few had the capability of calculating movement associated with running or cycling. Therefore, differences between walking, running and cycling should be interpreted with caution.

Conclusions

Challenges and incentives may be effective for connected and active members, and may play a role in achieving daily recommended activity guidelines. Data from rewards-based activity programs can give insights into theoretical constructs related to behavioral incentives, gamification, and strategies associated with cooperative games. Further research examining demographic and psychographic characteristics of rewards-program members, program efficacy rates, and the stratification of member-types is required.

Conflicts of Interest

Author TvM is the CEO and founder of Evolution Health Systems, owner of digital health platforms. Author RF is an employee of Evolution Health Systems. Author RD is CEO and Founder of ZeroFootprint, owner of the GOODcoins platform.

References

1. Volpp KG, Pauly MV, Loewenstein G, Bangsberg D. P4P4P: an agenda for research on pay-for-performance for patients. Health Aff (Millwood) 2009;28(1):206-214. [doi: 10.1377/hlthaff.28.1.206]
2. Wu J. Rewarding healthy behaviors-pay patients for performance. Ann Fam Med 2012;10(3):261-263. [doi: 10.1370/afm.1334]
3. Whalen J. MedTech Boston. 2015. (2015) Healthcare Wearables: Thoughts From the Experts on Challenges and Potential Breakthroughs URL: https://medtechboston.medstro.com/blog/2015/05/13/healthcare-wearables-thoughts-from-the-experts-on-challenges-potential-breakthroughs/ [accessed 2016-01-11] [WebCite Cache ID 6eSgrhu5X]
4. Billinghurst M. Wearable devices: new ways to manage information. Computer 1999;32(1):57-64.
5. Binks M, van Mierlo T. Utilization patterns and user characteristics of an ad libitum Internet weight loss program. J Med Internet Res 2010;12(1):e9. [doi: 10.2196/jmir.1347]
6. Cunningham J, Murphy M, Hendershot CS. Treatment dismantling pilot study to identify the active ingredients in personalized feedback interventions for hazardous alcohol use: randomized controlled trial. Addict Sci Clin Prat 2014;10(1).
7. Kreuter MW, Skinner CS. Tailoring: what's in a name? Health Educ Res 2000 Feb;15(1):1-4 [FREE Full text] [Medline: 10788196]
8. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
9. Selby P, van Mierlo T, Voci SC, Parent D, Cunningham JA. Online social and professional support for smokers trying to quit: an exploration of first time posts from 2562 members. J Med Internet Res 2010;12(3):e34 [FREE Full text] [doi: 10.2196/jmir.1340] [Medline: 20719739]
10. Kolt G, Schofield G, Kerse N, Garrett N, Ashton T, Patel A. Healthy Steps trial: pedometer-based advice and physical activity for low-active older adults. Ann Fam Med 2012;10(3):206-212 [FREE Full text] [doi: 10.1370/afm.1345] [Medline: 22585884]
11. Mendoza L, Horta P, Espinoza J, Aguilera M, Balmaceda N, Castro A, et al. Pedometers to enhance physical activity in COPD: a randomised controlled trial. Eur Respir J 2015 Feb;45(2):347-354 [FREE Full text] [doi: 10.1183/09031936.00084514] [Medline: 25261324]
12. Taylor NP. FierceBiotechIT. 2015. Number of Clinical Trials to Use Wearables Nears 300 URL: http://www.fiercebiotechit.com/story/number-clinical-trials-use-wearables-nears-300/2015-09-16 [accessed 2016-01-11] [WebCite Cache ID 6eSiOjKTd]
13. Robinson R. Pharma Voice. 2015. Wearables in Clinical Trials URL: https://s3.amazonaws.com/aws-mdsol-corporate-website-prod/mHealth_Wearables-Clinical-Trials_20150601_PharmaVOICE_Article.pdf [accessed 2016-01-11] [WebCite Cache ID 6eSiTUSAe]
14. Comstock J. mobihealthnews. 2015. Google X is Developing a Wearable for Clinical Research URL: http://mobihealthnews.com/44622/google-x-is-developing-a-wearable-for-clinical-research/ [accessed 2016-01-11] [WebCite Cache ID 6eSib4LdJ]

15. Coolbaugh CL, Raymond SJ, Hawkins DA. Feasibility of a dynamic web guidance approach for personalized physical activity prescription based on daily information from wearable technology. JMIR Res Protoc 2015;4(2):e67 [FREE Full text] [doi: 10.2196/resprot.3966] [Medline: 26043793]
16. Giles EL, Sniehotta FF, McColl E, Adams J. Acceptability of financial incentives and penalties for encouraging uptake of healthy behaviours: focus groups. BMC Public Health 2015;15:58 [FREE Full text] [doi: 10.1186/s12889-015-1409-y] [Medline: 25636330]
17. Halpern S, French B, Small D, Saulsgiver K, Harhay MO, Audrain-McGovern J, et al. Randomized trial of four financial-incentive programs for smoking cessation. N Engl J Med 2015 May 28;372(22):2108-2117 [FREE Full text] [doi: 10.1056/NEJMoa1414293] [Medline: 25970009]
18. Kullgren J, Troxel A, Loewenstein G, Asch DA, Norton LA, Wesby L, et al. Individual- versus group-based financial incentives for weight loss: a randomized, controlled trial. Ann Intern Med 2013 Apr 2;158(7):505-514 [FREE Full text] [doi: 10.7326/0003-4819-158-7-201304020-00002] [Medline: 23546562]
19. Liu S, Hodgson C, Zbib AM, Payne AYM, Nolan RP. The effectiveness of loyalty rewards to promote the use of an Internet-based heart health program. J Med Internet Res 2014;16(7):e163 [FREE Full text] [doi: 10.2196/jmir.3458] [Medline: 24989982]
20. Mantzari TM, Vogt F, Marteau TM. Financial incentives for increasing uptake of HPV vaccinations: a randomized controlled trial. Health Psychol 2015 Feb;34(2):160-171 [FREE Full text] [doi: 10.1037/hea0000088] [Medline: 25133822]
21. DiMatteo MR. Variations in patients' adherence to medical recommendations: a quantitative review of 50 years of research. Med Care 2004 Mar;42(3):200-209. [Medline: 15076819]
22. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RCM, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 4;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
23. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. Health Aff (Millwood) 2008;27(3):759-769 [FREE Full text] [doi: 10.1377/hlthaff.27.3.759] [Medline: 18474969]
24. Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? Patient Educ Couns 2001 Aug;44(2):119-127. [Medline: 11479052]
25. 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]
26. Itagaki MW, Berlin RB, Schatz BR. MeGenMed. 2002. The Rise and Fall of E-Health: Lessons From the First Generation of Internet Healthcare URL: http://www.medscape.com/viewarticle/431144 [accessed 2016-01-12] [WebCite Cache ID 6eUIYLbAS]
27. Fox S. Pew Internet & American Life Project. 2010. Mobile Health 2010 URL: http://www.pewinternet.org/files/old-media//Files/Reports/2010/PIP_Mobile_Health_2010.pdf [accessed 2016-01-11] [WebCite Cache ID 6eSkDwCiQ]
28. Gill S, Contreras O, Muñoz RF, Leykin Y. Participant retention in an automated online monthly depression rescreening program: patterns and predictors. Internet Interv 2014 Mar;1(1):20-25 [FREE Full text] [doi: 10.1016/j.invent.2014.02.003] [Medline: 25045623]
29. Carron-Arthur B, Cunningham JA, Griffiths KA. Describing the distribution of engagement in an Internet support group by post frequency: A comparison of the 90-9-1 Principle and Zipf's Law. Internet Interventions 2014 Oct;1(4):165-168. [doi: 10.1016/j.invent.2014.09.003]
30. van Mierlo T, Hyatt D, Ching AT. Mapping power law distributions in digital health social networks: methods, interpretations, and practical implications. J Med Internet Res 2015;17(6):e160 [FREE Full text] [doi: 10.2196/jmir.4297] [Medline: 26111790]
31. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res 2014;16(2):e33 [FREE Full text] [doi: 10.2196/jmir.2966] [Medline: 24496109]
32. van Mierlo T, Voci S, Lee S, Fournier R, Selby P. Superusers in social networks for smoking cessation: analysis of demographic characteristics and posting behavior from the Canadian Cancer Society's smokers' helpline online and StopSmokingCenter.net. J Med Internet Res 2012;14(3):e66 [FREE Full text] [doi: 10.2196/jmir.1854] [Medline: 22732103]
33. Dormuth CR, Patrick AR, Shrank WH, Wright JM, Glynn RJ, Sutherland J, et al. Statin adherence and risk of accidents: a cautionary tale. Circulation 2009 Apr 21;119(15):2051-2057 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.824151] [Medline: 19349320]
34. Shrank WH, Patrick AR, Brookhart MA. Healthy user and related biases in observational studies of preventive interventions: a primer for physicians. J Gen Intern Med 2011 May;26(5):546-550 [FREE Full text] [doi: 10.1007/s11606-010-1609-1] [Medline: 21203857]
35. Office of the Privacy Commisioner of Canada. Justice Laws Website. 2000. Personal Information Protection Electronic Documents Act URL: http://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html [accessed 2016-01-11] [WebCite Cache ID 6eSl87nAA]
36. Canadian Society for Exercise and Physiology. CSEP | SCPE. 2015. Canadian Physical Activity Guidelines URL: http://www.csep.ca/CMFiles/Guidelines/CSEP_PAGuidelines_adults_en.pdf [accessed 2016-01-11] [WebCite Cache ID 6eSlINZUU]

Edited by G Eysenbach; submitted 19.11.15; peer-reviewed by S McIntosh, B Carron-Arthur; comments to author 09.12.15; revised version received 11.12.15; accepted 03.01.16; published 28.01.16.

Please cite as:
van Mierlo T, Hyatt D, Ching AT, Fournier R, Dembo RS
Behavioral Economics, Wearable Devices, and Cooperative Games: Results From a Population-Based Intervention to Increase Physical Activity
JMIR Serious Games 2016;4(1):e1
URL: http://games.jmir.org/2016/1/e1/
doi:10.2196/games.5358
PMID:26821955

Original Paper

Feasibility of Applied Gaming During Interdisciplinary Rehabilitation for Patients With Complex Chronic Pain and Fatigue Complaints: A Mixed-Methods Study

Miel A P Vugts[1,2], MSc.; Margot C W Joosen[1], PhD; Alfonsus H M M van Bergen[2], PT; Hubertus J M Vrijhoef[1,3], Dr, Prof

[1]Tranzo Scientific Center for Care and Welfare, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands

[2]Ciran Rehabilitation Centers, Venlo, Netherlands

[3]Saw Swee Hock School of Medicine, National University of Singapore, Singapore, Singapore

Corresponding Author:
Miel A P Vugts, MSc.
Tranzo Scientific Center for Care and Welfare
Tilburg School of Social and Behavioral Sciences
Tilburg University
P.O. Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 13 466 8271
Fax: 31 13 466 8271
Email: m.a.p.vugts@tilburguniversity.edu

Abstract

Background: Applied gaming holds potential as a convenient and engaging means for the delivery of behavioral interventions. For developing and evaluating feasible computer-based interventions, policy makers and designers rely on limited knowledge about what causes variation in usage.

Objective: In this study, we looked closely at why and by whom an applied game (LAKA) is demanded and whether it is feasible (with respect to acceptability, demand, practicality, implementation, and efficacy) and devised a complementary intervention during an interdisciplinary rehabilitation program (IRP) for patients with complex chronic pain and fatigue complaints.

Methods: A mixed-methods design was used. Quantitative process analyses and assessments of feasibility were carried out with patients of a Dutch rehabilitation center who received access to LAKA without professional support during a 16-week interdisciplinary outpatient program. The quantitative data included records of routinely collected baseline variables (t0), additional surveys to measure technology acceptance before (t1) and after 8 weeks of access to LAKA (t2), and automatic log files of usage behavior (frequency, length, and progress). Subsequently, semistructured interviews were held with purposively selected patients. Interview codes triangulated and illustrated explanations of usage and supplemented quantitative findings on other feasibility domains.

Results: Of the 410 eligible patients who started an IRP during the study period, 116 patients participated in additional data collections (108 with problematic fatigue and 47 with moderate or severe pain). Qualitative data verified that hedonic motivation was the most important factor for behavioral intentions to use LAKA (P<.001). Moreover, quotes illustrated a positive association between usage intentions (t1) and baseline level (t0) coping by active engagement (Spearman ρ=0.25; P=.008) and why patients who often respond by seeking social support were represented in a group of 71 patients who accessed the game (P=.034). The median behavioral intention to use LAKA was moderately positive and declined over time. Twenty patients played the game from start to finish. Behavioral change content was recognized and seen as potentially helpful by interview respondents who exposed themselves to the content of LAKA.

Conclusions: Variation in the demand for applied gaming is generally explained by perceived enjoyment and effort and by individual differences in coping resources. An applied game can be offered as a feasible complementary intervention for more patients with complex chronic pain or fatigue complaints by embedding and delivering in alignment with patient experiences. Feasibility, effectiveness, and cost-effectiveness can be evaluated in a full-scale evaluation. New observations elicit areas of further research on the usage of computer-based interventions.

(JMIR Serious Games 2016;4(1):e2) doi:10.2196/games.5088

KEYWORDS

behavioral medicine; therapy; computer-assisted; computer games; mind-body therapies; patient acceptance of health care; feasibility studies; fatigue syndrome, chronic; fibromyalgia; musculoskeletal pain

Introduction

Background

Computer-based interventions (CBIs) can be effective alternatives or complements to face-to-face delivery in psychological treatment and chronic illness management [1-4]. However, systematic reviews on effectiveness of CBIs have concluded that sizable and heterogeneous proportions of patients stop using CBIs before completion [5-7]. Nonusage attrition in CBI studies depends on factors such as therapist involvement, demographics, computer self-efficacy, and health status [6-11]. As a strategy to improve patient engagement, some CBI designs have incorporated interactive features [12,13]. Interactive and visual-enriched designs may support patient demand through perceived personal relevance, social support, and enjoyment [14,15]. Accordingly, computer game technology has been applied to engage people and to promote health behaviors and clinical outcomes [16,17].

Chronic pain and fatigue complaints constitute a major burden for individuals and societies worldwide [18-20]. Functional somatic syndromes (FSS) are diagnosed by medical specialists when bodily functioning is disturbed, somatic symptoms persist longer than a normal healing process, and conditions cannot be fully attributed to a known conventional disease [21]. A high degree of commonality exists between FSS, wherein central sensitization may be a biological substantiation [22]. FSS can be precipitated by profound life events and cultural factors and maintained by psychosocial factors [20]. Evidence supports the effectiveness of various cognitive and behavioral interventions in primary care settings, or within interdisciplinary rehabilitation programs (IRPs), when "unimodal" psychiatric or physiotherapeutic services do not suffice [20]. Nonetheless, patients were often seen by their general practitioners, but seldom accessed specialized behavioral or multi-modal treatment, and often believed that their complaints are inadequately managed (28%-62%) [17].

Literature Review

Efficient use of scarce resources and removal of access barriers are important motives for developing CBIs [2]. Results on the effectiveness of computer-based behavioral interventions are promising, but uncertainties regarding their actual usage certainly applies to FSS patients [2,23]. Virtual reality and gaming technologies have been applied for triggering positive emotions, distraction, or graded exposure in rehabilitation and pain management for improvements in physical functioning, pain symptoms, and daily life activities [24,25]. However, there has been no evaluation of the effectiveness of applied gaming as an independently accessible means for delivering behavioral change messages to patients with FSS [16,17,26]. The actual extent and reasons of patient engagement in applied games will largely determine their effect [16]. A better understanding is needed of why CBIs have not been optimally used by which patients with chronic pain and fatigue symptoms to overcome the treatment barriers they face and why integration of applied gaming can offer a partial solution [2,17-19].

Research Goals

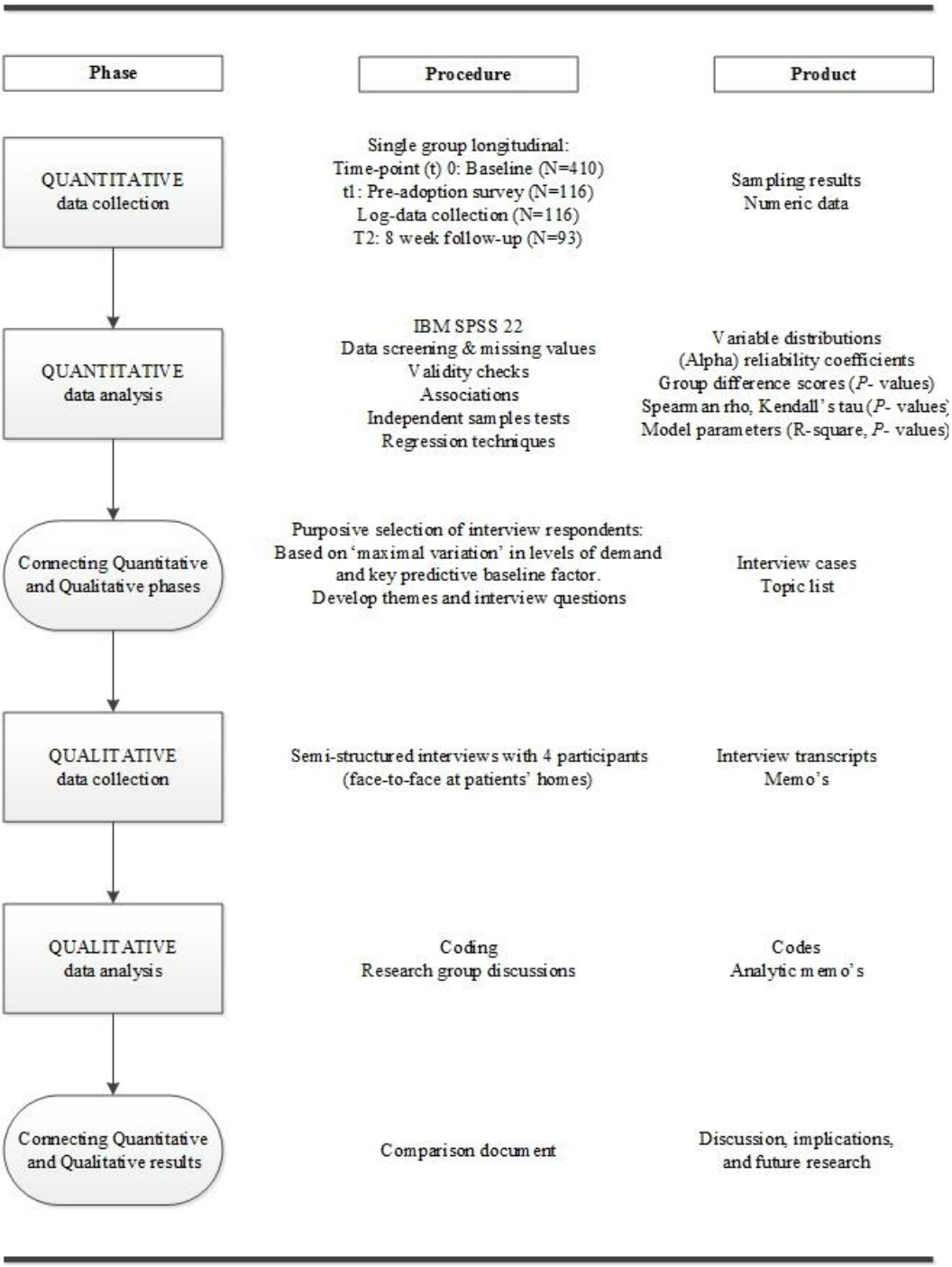
This study aims to explain the usage of applied gaming and provide a comprehensive feasibility description from the perspective of adult patients with chronic pain and fatigue complaints. The opportunity to conduct this study was provided by the planned incorporation of the applied game "LAKA" within a standardized IRP for adult patients with chronic and complex fatigue or pain symptoms in the Netherlands. The primary objective is to explain variation in the demand for applied gaming when offered for voluntary usage during an IRP. Relationships are studied between usage (intentions), behavioral factors, and patient baseline characteristics, including case mix, functional and clinical status, and medical history. In doing so, this study contributes to a better understanding of why applied games are demanded by patients in real health care settings. Second, feasibility was thoroughly described to prepare for a full-scale evaluation in exploring the domains of acceptability, implementation, practicality, and promise for efficacy. Both research goals are reflected in a conceptual framework (see Multimedia Appendix 1) integrating technology acceptance modeling in a feasibility study design [23-27]. Overall, this contribution enables feasible proposals for incorporating and evaluating an applied game for behavioral change within the rehabilitation of patients with complex chronic pain and fatigue complaints.

Methods

Research Design

A mixed-methods design was implemented with sequential quantitative (QN) and qualitative (QL) phases [28] (Figure 1). Owing to the availability of adequate quantitative research instruments, an explanatory sequential mixed-methods design worked well for triangulation, illustration, and complementing QN findings with in-depth QL insights and with practically useful information about feasibility [29,30]. The QN phase was prioritized and set up as a longitudinal single-group study of target patient responses to LAKA when offered for voluntary usage during the first 8 weeks of their IRP. The QL phase provided a complementary inductive approach to both research questions. QN and QL phases were mixed in using QN results for the preparation of QL data collection and again when integrating and documenting QN and QL results.

Figure 1. Mixed-methods design overview.



Recruitment and Data Collection

Ethical approval for the protocol was obtained for this study (at the Psychological Ethical Testing Committee of Tilburg School of Behavioral Sciences, Tilburg University). In total, 410 patients were eligible to start an IRP in 1 of 4 facilities of Ciran, a Dutch rehabilitation center, between 6 January, 2014, and 6 April, 2014 (criteria are listed in Textbox 1) and had given permission to process their diagnostic records for scientific research. Patients in this group were prompted on the day of their first appointment with an email invitation containing information about the nature and consequences of participation in the study, and a link to the "serious gaming page" (see Multimedia Appendix 2). This page guided patients through procedures for software access and additional data collection. Local team leaders were requested to organize face-to-face reminders for patients about the study and the accessibility of LAKA regardless of giving informed consent.

Textbox 1. Inclusion and exclusion criteria for study participants.

Inclusion criteria:

- Is aged between 18 and 67 years
- Lives in the Netherlands and is proficient in the Dutch language
- Reports the presence of pain for more than 6 months, or fatigue complaints, or a musculoskeletal disease for more than 3 months
- Had received previous primary or secondary health care services without a satisfactory result.
- Reported additional problems on at least 2 of the following problem areas: disturbance of participation, individual or environmental factors, psychological distress, and lack of spiritual well-being

Potential exclusion criteria:

- Presence of medical or psychiatric risk factors (nontreated medical condition, drugs- or alcohol addiction, and suicide risk)
- Presence of third party liabilities

Additional QN data were collected from study participants with two surveys and log-file recordings. The first survey was to be completed within 4 weeks after the invitation and before usage (t1). The second survey (t2) was added to a standard midterm evaluation after 8 IRP weeks, and was to be submitted by the 12th week. Participants' actual usage behavior had been logged automatically between the pre and post surveys. In the intermediate phase, the focus was on "demand" as the primary subject of the study. Two extreme cases were selected so that a relationship between the most predictive baseline characteristic for the use of LAKA could be clearly illustrated. Two more cases were selected with demand levels that were poorly explained by this "key" predictor and more likely to provide information about omitted factors or counterfactuals. QN results were also used to set up an interview schedule. Purposefully selected participants were invited with a prescripted telephone call after their IRP was finished. Interviews were held during 1- to 1.5-hour home visits and were tape-recorded and transcribed. Field notes were taken, and full transcripts were sent to respondents by email within a week after the interview.

Intervention

A standardized 16-week IRP was delivered by teams of physiotherapists, psychologists, spiritual counselors, and medical rehabilitation specialists (Textbox 2). A full description of the IRP is provided by Garschagen et al [31].

Textbox 2. Characteristics of the interdisciplinary rehabilitation program (IRP).

Tailored: The program has a modular build-up to match individual care need.

Outpatient, intensive: On average, 100 hours delivered by professionals (both individual and group sessions), and 30 self-directed hours.

Integrated program components:

- 45% exercise therapy, graded activity, graded exposure, and education in physiology
- 15% cognitive behavioral therapy
- 15% counseling and guidance in resuming participation in important life domains, such as work, social activities, and family life
- 25% spiritual education

Target outcomes:

- Primary: improvement of well-being [32], and participation in important life-domains (activity and participation domains 4-9 of the International Classification of Functioning) [33]
- Secondary: reduce pain, fatigue, and emotional distress symptoms

As recommended, functional requirements were specified before the modeling and evaluation of LAKA [34]. LAKA delivers skills training with metaphorical simulation elements (encounters) and guided exercises for focused attention and open awareness [35]. These elements are interspersed with images of real-world environments, immersive mini-games, and in-game debriefings for "transferring" new insights beyond the virtual world (Textbox 3). Basic information on functional specifications and playability feedback are described and illustrated with screenshots and trailers in Multimedia Appendix 3. A Windows version of LAKA was offered for usage wherever and whenever convenient, without support from health professionals. No recommendation for a minimum or maximum amount of usage was given, and no prompts or reminders were sent. On the one hand, it was expected that many target patients would not use LAKA because of this noncommittal mode of delivery. On the other hand, this variation was desired to discover explanations and practical suggestions that generally apply for delivery in open and clinical settings.

Textbox 3. LAKA design.

- **Problem addressed:** In complex cases, suffering associated with FSS has an intrusive impact on patients' existence and their interactions with caregivers, family, or friends [36].
- **Design team:** The design team involved entrepreneurs, researchers, a scriptwriter, game designers, artists, programmers, audio experts, voice-actors, and IT specialists.
- **Stakeholder involvement:** Feedback sessions (on functional specifications, theoretical model, and prototypes) involved experts (in behavioral science, medical technology, and spirituality), and critical users.
- **Genre:** Single-player adventure game.
- **Goal (of the game):** LAKA was designed to provide skills training in "spiritual" practices. Practices are focused on behavioral qualities that are associated negatively with negative emotions and positively with psychological well-being: "generosity," "moral discipline," "patience/forbearance," "enthusiastic energy," and "mental stability." Accordingly, the design includes the delivery of various behavioral change techniques integrated in an immersive simulation environment.
- **Main challenges (in the game):** Identify with a personal Avatar and engage in a quest. The story is about an Avatar, who learns about "the art of living" while traveling the world after a significant deterioration of his/her condition. Tasks primarily entail the consideration and evaluation of response options in virtual "encounters" with nonplaying characters.
- **Application components:** Introduction, 4 training modules (or travel "destinations": London, Turkey, Asia, Africa) with recurrent components, and a celebratory end.
- **Duration:** Completing the game from start to end takes about 2.5 hours.
- **Game controls:** Interaction design and controls (with computer mouse and keyboard) were designed for ease of use. Progression in the game does not depend on gamer performance or skill.
- **Graphics:** Mixed 2D and 3D graphics with comical cut scenes.
- **Sound:** Voice-overs and music convey emotions and atmosphere.
- **Platform:** Personal computer or laptop (MS Windows version).
- **System requirements:** Windows XP or beyond, a 6 gigabytes hard drive, 1 gigabyte memory, and a stable Internet connection.
- **Accessibility:** Via the "serious gaming" Web page by downloading, or by following instructions for picking up a digital versatile disk at local facilities.

Measures

Demand, Demand Factors, and Other Feasibility Domain Outcomes

Objective indicators of demand were based on automatic data logs of participant activities: "frequency" (number of days on which progress was logged), "duration" (sums of time intervals between logins and subsequent data logs), and "progress" (the number of completed encounters). Demand was rated subjectively, before and after usage, by the extent of agreement (1, completely disagree, to 7, completely agree) with 3 statements about their current behavioral intention (BI) to use LAKA during their IRP [24,25]. Seven-point Likert scales were also used to assess behavioral factors, including performance and effort expectancies, social influence, perceived behavioral control, computer anxiety and self-efficacy, hedonic motivation or enjoyment, habit, and trust [23,24,37]. Multimedia Appendix 4 contains details about all survey measures, including variable definitions, items, validity, and reasons why behavioral factors may be relevant [38-46]. Practicality was indicated by counting logins of participants with positive behavioral intentions (BIs at t1 ≥5) as evidence of success in obtaining the software, installation, and running the application. Study and usage attrition were interpreted as indicators of the degree of implementation. Acceptability was operationalized as postusage perceived appropriateness in enjoyment, ease, and knowledge improvement in participants who completed at least the first module of the game.

Baseline Measures

Retrieved baseline variables were categorized into case mix, functional status, clinical status, and previous treatment variables (see Multimedia Appendix 4). Case-mix variables included sex, age, education level, environmental issues, and treatment facility. Preferred coping styles were measured with the Utrecht Coping List (UCL). Functional status variables included the duration and course of health complaints, employment status, absenteeism, and 1-item general subjective health. Pain intensity was assessed with an 11-point Numerical Rating Scale (NRS) [47]. The Checklist Individual Strength was used to assess fatigue dimensions [48]. Clinical status variables included a categorization of the chronic symptom patterns by a rehabilitation specialist (primarily a fatigue or musculoskeletal or other pain condition). Body mass index (BMI) and blood pressure were measured during physical examination. Psychopathology dimensions were assessed with the Dutch 90-item Symptoms Check List (SCL-90) [49]. The Pain Coping and Cognitions List and Tampa Scale of Kinesiophobia were used to measure pain coping and cognitions [50,51]. Finally, patients indicated previous specialized treatments and current medication intake.

Data Analyses

Data Exclusion

Cases were list-wise deleted before analysis if the proportion of missing observations was <5%, or handled by predicting 5

data imputations for each empty cell through regression of all variables in the dataset (using the MCMC algorithm). All full-case QN findings presented as marked results are supported by pooled results.

Participant Statistics

Characteristics of eligible patients, study participants, and participants who logged into the game (players) are described by descriptive statistics and frequencies. Chi square and Mann-Whitney *U* tests were used to compare baseline level characteristics between study participants and participants who logged into LAKA, versus eligible patients that were not included in those groups. Similarly, differences were tested between participants who logged in versus participants who did not log in.

Process Analyses of Demand and Feasibility Descriptions

All feasibility outcomes of applied gaming during the first 8 weeks of the IRP are indicated with descriptive statistics and line graphs. Association measures (Spearman ρ and Kendall τ statistics) between baseline characteristics, behavioral factors, and feasibility outcomes were calculated and tested for significance. Moreover, multiple ordinary least squares regression analyses were performed for the sequential identification of important constituent factors of behavioral intention at t1, to explore whether effects of behavioral factors differed between subgroups of patients (see Multimedia Appendix 5), and to test if marked associations between baseline characteristics and behavioral intentions were mediated by behavioral factors [52].

Qualitative Data Analysis

Interview transcripts were coded by one author (MV) using a software package: MAXQDA 11 (VERBI GmbH) [53]. In the first coding step, all text fragments about the specified interview topics were labeled with short statements that corresponded with contextual meanings. A second author (MJ) independently repeated this first coding step for one interview. These "first order" codes were compared and discussed between MJ and MV to align and refine the coding procedure. In a second coding step, more abstract categories were generated. Throughout this process, first-order codes and emergent categories were constantly compared and hierarchically structured as a means for critical appraisal and to avoid imposing preconceived ideas on the QL data. Finally, categories were related to one another by designating them as context factors, conditions (barriers or facilitators), events or interactions, or consequences.

Mixing Quantitative and Qualitative Results

In connecting QL and QN findings, codes and statistics were provided for comparison for both research questions. QN results were deemed notable for comparison with QL findings if *P* values were below .05. Subsequently, 3 researchers (MV, MJ, and HV) discussed and determined points of convergence, divergence, or complementariness between QN and QL findings. In doing so, observations were summarized to determine which, and to what extent, remarkable and solid QN findings were clearly illustrated and triangulated. Moreover, the point at which qualitative data collection was stopped was determined on the basis of saturation with respect to illustrations of behavioral factors and the role of a key predictive baseline characteristic for usage in early stages.

Results

Participant Statistics

Of the 410 invited eligible patients, 32.2% provided informed consent and completed the first additional survey (Figure 2). The 84 patients who reported why they did not wish to participate mentioned "other obligations" (23), "facilitative problems" (14), "no intention to use the intervention" (14), "not enough energy or concentration" (13), "no interest to participate in the research" (10), "bodily complaints" (8), or "other reasons" (2). One patient withdrew because of a broken computer, and one for experiencing excessive hindrance in attempting to use an unsupported Web browser. The second questionnaire was submitted by 93 participants (80.2%).

Study participants' average age was 44.4 years (SD 10.8 years; range 21-63 years); 71% were female (Table 1). Sixty-nine participants were completely absent from work. The average duration of absenteeism was 157 days (SD 223.0), with a median slightly more than 100 days. Forty-seven participants (40.5%) reported moderate to severe pain (5-10), and 108 experienced problematic fatigue. Average scores for depressive (42.9, SD 11.4) and anxious (22.2, SD 8.2) symptoms were high. Participants had been regularly surfing the Internet, but only 46 patients (39.7%) had been playing on a computer over the past year. No statistically significant differences between participants and nonparticipants were found for case-mix variables. However, patients with more severe pain symptoms were underrepresented in the sample (Table 2). The group of 71 patients who actually logged in (players) reported relatively higher scores for coping through active engagement and social support seeking, lower scores for pain coping, and fewer environmental issues. The proportion of patients who had received specialist treatment for their current complaint was lower among players than among nonplayers (χ^2_1=4.1; *P*=.042; not in Table 2).

Four interview respondents were selected based on their combination of scores for coping by active engagement and demand (Table 3). Open questions were asked to introduce and focus on topics (see Multimedia Appendix 6). Two topics addressed the primary research question, namely, "initial response" to the digital game offering (topic 1) and patient "experiences" throughout their interactions with LAKA (topic 2). Topic 2 and "suggestions for improvement" (topic 3), served to collect complementary information on feasibility domains. After a first round of mixing, 4 interviews was deemed sufficient to provide clear illustrations of the most notable QN explanations for demand.

Table 1. Characteristics of study participants (N=116).

Characteristic[a]		N (%)
Demographics		
Sex		
	Female	71 (61.2)
Age, years		
	<35	23 (19.9)
	35-45	30 (25.9)
	45-55	41 (35.3)
	55-67	22 (19.0)
Education level ISCED[b]		
	Primary or less	32 (25)
	Lower to postsecondary	44 (37.9)
	Tertiary and posttertiary	40 (34.5)
	Missing	3 (2.6)
Functional status		
Employment in paid work		
	Full-time	49 (42.2)
	Part-time	52 (44.8)
	None	15 (12.9)
Absenteeism		
	Not	15 (26.1)
	Partially	17 (14.8)
	Completely	69 (59.1)
Duration of absenteeism for present somatic symptoms		
	<3 months	31 (26.7)
	0-3 month	41 (35.3)
	3-6 months	22 (19.0)
	6-12 months	14 (12.1)
	1-2 years	6 (5.2)
	>2 years	2 (1.7)
Symptom duration		
	<3 months	3 (2.6)
	3-6 months	11 (9.5)
	6-12 months	30 (25.9)
	1-2 years	27 (23.3)
	>2 years	45 (38.8)
Pain NRS[b]		
	No pain (0)	18 (15.5)
	Mild pain (1-4)	51 (44.0)
	Moderate pain (5-7)	36 (31.0)
	Severe pain (7-10)	11 (9.5)
Fatigue		

Characteristic[a]		N (%)
	No fatigue (NRS[b]=0)	2 (1.7)
	Not problematic (CIS[b] ≤76)	6 (5.3)
	Problematic (CIS >76)	108 (94.7)
CIS subjective fatigue[c]		
	Above average	50 (43.1)
CIS physical activity[c]		
	Below average	67 (58.8)
Clinical status		
Primary diagnosis		
	Chronic musculoskeletal disorder	21 (18.1)
	Chronic pain	17 (14.7)
	Chronic fatigue	78 (67.2)
SCL-90[b] depression[d]		
	Below average (16-31)	24 (20.7)
	Above average (32-35)	7 (6.0)
	High (36-52)	63 (54.3)
	Very high (≥53)	22 (19.0)
SCL-90 anxiety[d]		
	Below average (10-17)	42 (36.2)
	Above average (18)	7 (6.0)
	High (19-28)	39 (33.6)
	Very high (≥29)	28 (24.1)
Previous treatment		
Medical specialist treatment		
	Yes	70 (60.3)
Medication usage		
	Yes	80 (69.0)
	Missing	1 (.9)
Previous use of similar technology		
Habit of frequent Internet usage with a PC or laptop		
	On 6-7 days per week	84 (72.4)
	On 3-5 days per week	22 (19.0)
	On 1-2 days per week	9 (7.8)
	On <1 day per week	1 (0.9)
Experience of digital game play		
	Never played a digital game	37 (31.0)
	More than a year ago	33 (28.4)
	Less than a year ago	14 (12.1)
	Less than a month ago	32 (27.6)
Habit of frequent digital game play		

Characteristic[a]	N (%)
One or more times per month (and less than a month ago)	29 (25)

[a]A selection of individual baseline characteristics is presented to facilitate comparison with previous evaluations of behavioral interventions for FSS patients [54,55].

[b] CIS: Checklist Individual Strength, ISCED: International Standard Classification of Education (according to which highest education levels [Dutch system] were rescaled [low = 0-1, middle = 2-4, high = 5-6]) [56], NRS: Numerical Rating Scale, SCL: Symptom Checklist.

[c]For all participants (2 missing values were ignored; N=114). In comparison with the average in a population of patients with chronic fatigue syndrome [57].

[d]Compared with a population of Dutch patients with chronic pain [49].

Table 2. Overview of independent samples tests.

Variable	Participants (N=116) vs nonparticipants[a]	Players (N=71) vs nonplayers or nonparticipants[a]
	Mean (SD)/frequency (%), *P* value of test statistic[b]	Mean (SD)/frequency (%), *P* value of test statistic[b]
Case mix		
Female (dit.)[b]	71 (61.2), .48	45 (63.4), .92
Age[b]	44.4 (10.8), .91	44.1 (11.3), .88
Education level	3.3 (1.3), .12	3.3 (1.3), .40
UCL[c] active engagement	17.6 (4.0), .53	18.5 (3.9), .02
UCL passive responding	14.3(3.7), .29	14.0 (4.0), .11
UCL social support seeking	13.9 (4.0), .06	14.1 (4.0), .03
UCL comforting thought	12.0 (2.7), .74	12.4 (2.7), .38
Environmental issue (dit.)	61 (53.4), .07	34 (47.9), .02
Location A (dit.)	39 (33.6), .39	18 (25.4), .30
Location B (dit.)	27 (23.3), .41	19 (26.8), .89
Location C (dit.)	29 (25.0), .53	18 (25.4), .59
Location D (dit.)	21 (18.1), .45	16 (22.5), .64
Clinical status and functioning		
Body mass index	27.1 (5.8), .32	27.4 (5,5), .17
Indication for chronic fatigue (dit.)	78 (67.2), .045	54 (76.1), .002
Indication for musculoskeletal disorder (dit.)	21 (18.1), .16	9 (12.7), .03
Indication for chronic pain (dit.)	17 (14.7), .37	8 (11.3), .14
Symptom duration	Median >2 years, .75	Median >2 years, .34
Symptom recurrence (dit.)	74 (63.8) .75	44 (62.0), .68
Symptom deterioration (dit.)	69 (59.5), .04	43 (60.6), .20
Paid work (dit.)	101 (87.1), .12	62 (87.3), .23
SCL-90[c]total	206.9 (50.7), .70	206.3 (51.2), .94
SCL-90 sleeping problems	9.1 (3.3), .18	9.0 (3.4), .23
SCL-90 hostility	11.3 (4.5), .38	11.1 (4.1), .69
SCL-90 interpersonal sensitivity	34.9 (12.7), .58	33.9 (12.4), .63
SCL-90 insufficiency	26.0 (7.0), .73	26.6 (6.9), .27
SCL-90 somatization	30.7 (8.3), .86	31.3 (7.7), .40
SCL-90 depression	42.9(11.4), .69	41.8 (11.7), .55
SCL-90 anxiety	22.2 (8.2), .86	22.4 (8.5), .99
SCL-90 agoraphobia	11.0 (5.4), .60	11.5 (6.0), .73
	N=47[d]	N=27[d]
Pain NRS[c]	6.5 (1.3), .046	6.6 (1.4), .21
PCCL[c] internalization	3.2 (.7), .09	3.3 (.7), .18
PCCL pain coping	2.8 (.8), .06	2.7 (.8), .009
PCCL catastrophizing	3.6 (.8), .04	3.6 (.8), .11
TSK[c]	36.7 (6.8), .80	36.4 (7.1), .64

Variable	Participants (N=116) vs nonparticipants[a]	Players (N=71) vs nonplayers or nonparticipants[a]
	N=108[e]	N=69[e]
CIS[c] subjective fatigue	50.5 (6.0), .32	50.4 (5.6), .85
CIS concentration	26.3 (8.1), .59	26.0 (8.4), .57
CIS motivation	21.3 (6.1), .59	21.4 (6.1), .63
CIS physical (in)activity	17.2 (4.0), .83	16.9 (4.1), .30
CIS total score	110.1 (14.8), .99	109.4 (14.1), .64
	N=101[f]	N=62[f]
Part-time work (dit.)	52 (53.1), .72	33 (46.5), .57
Weekly work hours	31.6 (11.9), .21	29.8 (15.4), .12
Absent (dit.)	86 (74.1), .94	54 (87.1), .60
UBOS-a[c] burnout (dit.)	26 (22.4), .33	16 (22.5), .48
UBOS mental exhaustion	3.9 (1.5), .72	4.0 (1.5), .47
UBOS distancing	2.3 (1.5), .81	2.4 (1.6), .99
UBOS work competence	4.0 (1.2), .20	4.1 (1.1), .08
	N=86[g]	N=54[g]
Partially absent (dit.)	17 (14.7), .31	13 (24.1), .94
Sick leave duration	159.8 (223.4), .91	150.4 (15.4), .36
Previous treatment		
Medication intake (dit.)	70 (60.3), .48	39 (58.2), .06
Previous specialist treatment (dit.)	81 (69.8), .50	47 (66.2), .83

[a]These comparisons were chosen to inform about study sample profiles and how successful the implementation was in recruiting representative subsamples for exploring “within-group” variation in demand. The players versus nonplayers comparison did not yield more remarkable differences.

[b]N (%) and *P* value of chi-square if variable is dichotomous (dit.); median (N) or mean (SD) and *P* value of Mann-Whitney *U* test if variable is an ordinal or a ratio scale value.

[c]CIS: Checklist Individual Strength, PCCL: Pain Coping and Cognitions, SCL: Symptom Checklist, TSK: Tampa Scale of Kinesiophobia, UBOS-a: Utrecht Burnout Scale labor (a) version, UCL: Utrecht Coping List.

[d]Subsample of participants with moderate or severe pain, [e]with problematic fatigue, and [f]with paid work and of those [g]absent from work.

Table 3. Characteristics of interview respondents.

Characteristic	Respondent #1[a]	Respondent #2	Respondent #3	Respondent #4
Usage (session days, encounters)	3, 20	0, 0	1, 1	2, 4
Behavioral intention	6	1	4	7
UCL active engagement[a]	Very high	Very low	Very high	Average
Sex	Male	Male	Female	Female
Age, years	35	57	62	54
Work status	No paid work	Fully absent for 97 days	Fully absent for 287 days	Present at work
CIS fatigue severity[c]	Problematic	Above-average CFS[b]	Above-average CFS	Problematic
Pain NRS	3	1	3	0
SCL[d] anxiety	Average	<Average	High	High

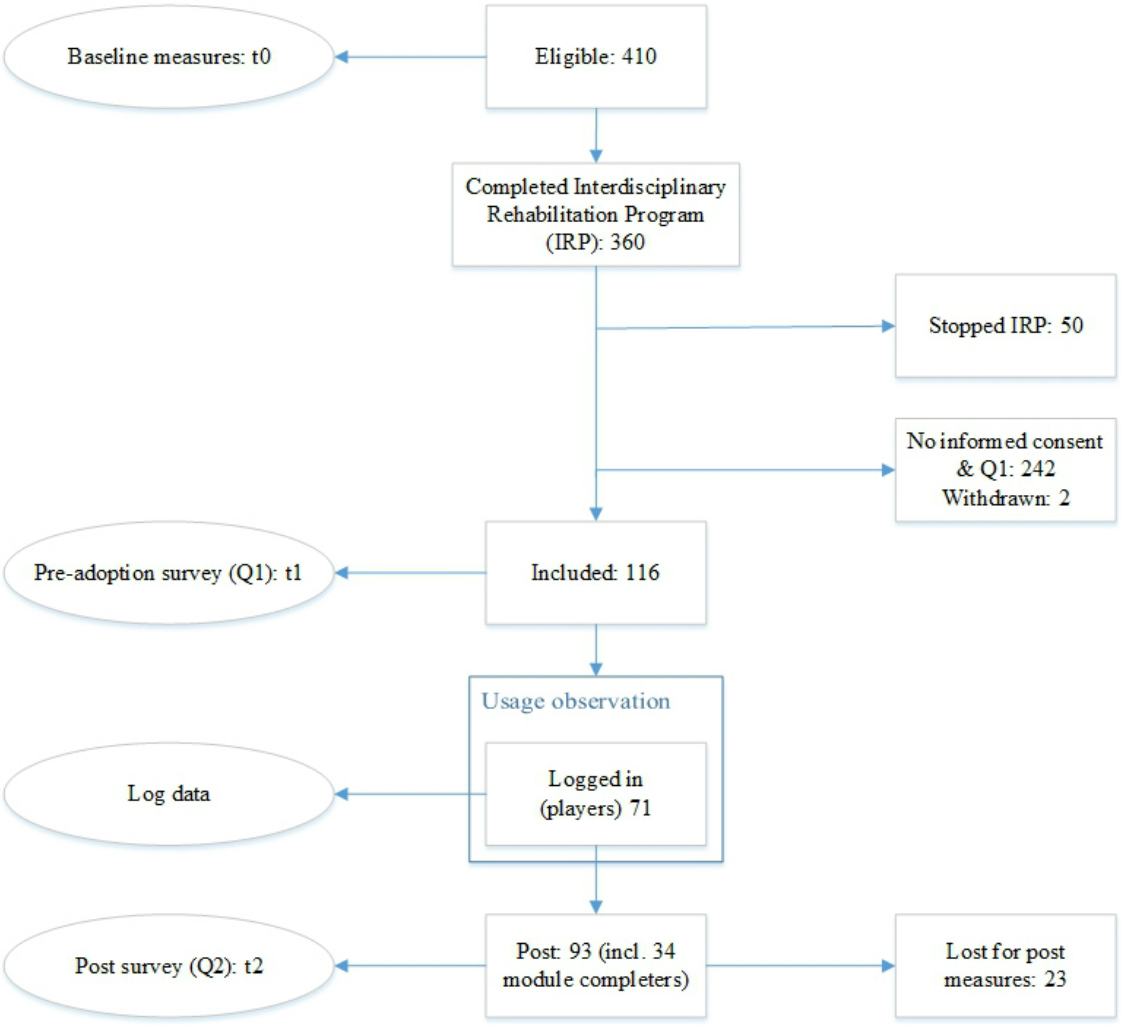
[a]Cases were identified by inspection of a bivariate scatterplot displaying the most predictive individual baseline characteristic on the x-axis; frequency of usage on the y-axis; and marking dots representing negative (<3), neutral (3-5), and positive (>5) behavioral intentions at t1.

[b]Levels of active engagement within the sample are similar to healthy worker population levels. Norm scores are slightly different for males and females.

[c]As compared to average fatigue severity in a sample of patients diagnosed with chronic fatigue syndrome.

[d]CIS: Checklist Individual Strength, NRS: Numerical Rating Scale, SCL: Symptom Checklist, UCL: Utrecht Coping List.

Figure 2. Flow diagram of study participants.



Process of Demand

Direct effects

Actual usage indicators measured at preadoption were associated with behavioral intention at a moderate level at t1 (ρ=.527-.546), and weakly at t2 (ρ=.260-.273). Behavioral intention was associated with various behavioral factors measured at t1 (Table 4). Effects of perceptions on performance, social norms, and knowledge improvement became stronger over time and with exposure to LAKA. Expectancies of effort and performance independently shared a significant amount of variance with behavioral intention at t1 (Table 5). Second step models were superior to the core model if supplemented with hedonic motivation or habit, but not if other factors were added. Inclusion of hedonic motivation weakened the effects of core factors such that performance expectancy was no longer a significant factor for behavioral intention.

If asked openly for an explanation of their own initial response to the offer to play LAKA during their IRP, respondents first explained their affect or attitude at that time, such as their openness to try the intervention. These feelings were substantiated by memories of previous experiences with computer game play. Those who felt positive about using the game did not experience barriers in concerns about privacy, usefulness, opinions of others, or a lack of resources. Nonetheless, respondents desired an explanation about how the intervention could bring about personal benefit.

> *I do like games. I have them myself… I'm just going to see what it is. Not immediately: no, I do not join in… I was just open minded… With two or three people I have talked about it (eds.: study, LAKA), and they said: the only thing for which I look at the computer, is to see what time I should be here. For other things; let me know…* [Respondent #4]

> *I did not recognize a goal… If there was a little more explanation, then I would have probably played ... and especially if it helps.* [Respondent #2]

Various associations between case-mix variables and demand indicators were marked (Table 6). Frequency of coping by active engagement and comforting thought were positively related to demand before exposure. Associations between demand indicators and social support seeking increased by exposure. BI was significantly lower for participants treated in location A, whereas participants treated in location D progressed more within the game. BI measured after 8 weeks was associated negatively with education level and positively with age. Notable differences in demand by functional and clinical status variables were also observed. LAKA was played more frequently by participants who reported partial absenteeism from work and lower pain coping scores. Higher usage was registered for players with higher levels of perceived competence in their job, symptom deterioration, higher pain intensity, lower internalization, and fewer symptoms of anxiety at baseline.

Table 4. Associations between demand indicators and behavioral factors.

Behavioral factor	BF t1 with BI t1[a] N=115[b] (ρ or τ, *P*[c])	BF t1 with BI t2 N=92 (ρ or τ, *P*)	BF t2 with BI t2 N=32 (ρ, *P*)
Performance expectancy	.33, <.001	.19, .08	.59, <.001
Expected ease	.42, <.001	.10, .37	.35, .045
Social influence	0.14, .13	.17, .11	.42, .01
Perceived behavioral control	.33, <.001	.04, .71	.22, .22
Trust	.31, .001	.21, .049	.53, .001
Hedonic motivation	.54, <.001	.43, <.001	.61, <.001
Computer anxiety	−.27, .003	.10, .35	
Computer self-efficacy	.22, .02	.32, .002	
Habit (dichotomous)	.22[c], .007	.06[c], .53	
Perceived knowledge improvement			.77, <.001

[a]BI: behavioral intention, BF: behavioral factor, t: time-point.

[b]Pairwise deletion: one respondent submitted an unfinished web-survey at t1.

[c]Kendall τ (for dichotomous variable) or Spearman ρ (for other variables), *P* value.

Table 5. Parameters and models fit of multiple regression for constituent factors of behavioral intention at preadoption.

Parameters (N=115[a])	Model 1: Core TAM[b] beta (*P* value)	Model 2: UTAUT[b] beta (*P* value)	Model 3: Core+HM[b] beta (*P* value)	Model 4: Core+HB[b] beta (*P* value)
Constant	.84 (.15)	.34 (.63)	.89 (.10)	.69 (.23)
PE[c]	.40 (.002)	.35 (.009)	.04 (.79)	.46 (<.001)
EE[c]	.52 (<.001)	.51 (<.001)	.27 (.02)	.47 (.001)
SI[c]		.15 (.09)		
PBC[c]		.06 (.63)		
HM[c]			.59 (<.001)	
HB[c]				.63 (.02)
	R^2 (*P* of ΔR^2)[c]	R^2 (*P* of ΔR^2)[d]	R^2 (*P* of ΔR^2)[d]	R^2 (*P* of ΔR^2)[d]
Model	.34 (<.001)	.36 (.22)	.43 (<.001)	.38 (.02)

[a]Observations of 1 incomplete case were listwise deleted.

[b]EE: effort expectancy, HB: habit, HM: hedonic motivation, TAM: technology acceptance model, UTAUT: unified theory of acceptance and use of technology, PBC: perceived behavioral control, PE: performance expectancy, SI: social influence.

[c,d]*P* of ΔR^2 is the *P* value of variance explained by the model over [c]a constant-only model, or [d]over model 1.

Table 6. Associations between baseline characteristics and demand indicators.

Baseline variable	BI t1 ρ or τ, *P*[a]	Session days ρ or τ, *P*	Session days ρ or τ, *P*	Time spent ρ or τ, *P*	Progress ρ or τ, *P*	BI t2 ρ or τ, *P*
Case-mix						
	N=116	N=116	N=71	N=71	N=71	N=93
Female (dit)[a]	−.04, .59	−.03, .77	−.10, .34	−.10, .29	−.13, .22	−.08, .38
Age	−.10, .28	−.02, .84	−.05, .69	.03, .78	−.00, .97	.25, .02
Education level (ISCED)[b]	−.05, .60	−.12, .18	−.21, .08	−.24, .048	−.24, .048	.28, .006
UCL[b] active engagement	.25, .008	.30, .001	.23, .06	.13, .29	.13, .28	.11, .31
UCL passive responding	−.09, .32	−.16, .08	−.19, .12	.02, .86	−.02, .89	.02, .82
UCL social support seeking	.09, .37	.19, .045	.23, .052	.20, .08	.24, .04	.08, .43
UCL comforting thought	.20, .03	.19, .04	.11, .36	.09, .44	.09, .48	.05, .62
Environmental issue (dit.)	−.15, .06	−.03, .13	.13, .22	.07, .48	.10, .34	.06, .50
Location A (dit.)	−.17, .03	−.22, .01	−.13, .22	−.18, .06	−.19, .07	−.02, .86
Location B (dit.)	.07, .42	.01, .93	−.15, .15	−.14, .17	−.18, .08	−.091, .31
Location C (dit.)	.07, .37	.06, .47	.15, .18	.12, .23	.14, .18	.030, .74
Location D (dit.)	.06, .46	.19, .03	.15, .17	.21, .03	.24, .02	.084, .35
Clinical and functional status						
	N=116	N=116	N=71	N=71	N=71	N=93
Body mass index	.05, .64	.13, .17	.06, .60	.06, .65	.11, .39	.10, .34
Indication chronic fatigue (dit.)	.06, .48	.07, .43	−.18, .09	−.18, .06	−.16, .12	.15, .10
Indication musculoskeletal (dit.)	−.02, .78	−.05, .56	.16, .15	.10, .30	.10, .31	−.16, .07
Indication chronic pain (dit.)	−.05, .53	−.04, .69	.08, .44	.14, .16	.10, .31	−.02, .83
Pain intensity NRS[b]	−.03, .79	.15, .12	.29, .02	.28, .02	.29, .02	.05, .63
Symptom duration	−.05, .58	−.08, .40	−.01, .92	.01, .94	.03, .79	−.04, .70
Symptom recurrence (dit.)	−.02, .74	−.05, .53	−.08, .49	−.12, .23	−.09, .37	.03, .78
Symptom deterioration (dit.)	.05, .51	.14, .09	.21, .054	.25, .01	.20, .048	.13, .17
Paid work (dit.)	.04, .64	.03, .77	.07, .54	.03, .75	−.00, .98	−.10, .30
Part-time (dit.)	.07, .38	.05, .54	.09, .43	.04, .67	.03, .78	.08, .37
Weekly work hours	.09, .35	.05, .60	.02, .87	.04, .75	.03, .84	−.16, .13
Work absence (dit.)	−.01, .89	.02, 87	−.07, .52	.02, .88	−.01, .97	.05, .64
SCL[b] total	−.08, .41	−.11, .23	−.20, .09	−.06, .63	−.12, .30	−.05, .63
SCL sleeping problems	−.12, .19	−.10, .29	−.09, .48	−.05, .71	−.10, .40	−.01, .96
SCL hostility	−.09, .32	−.01, .94	−.03, .83	.01, .97	−.03, .80	.07, .47
SCL interpersonal sensitivity	−.14, .14	−.13, .17	−.16, .19	−.04, .73	−.06, .61	−.01, .96
SCL insufficiency	.03, .74	.03, .78	−.13, .29	−.02, .86	−.09, .44	−.01, .90
SCL somatization	.01, .94	.09, .36	.02, .89	.07, .56	.01, .92	.02, .88
SCL depression	−.11, .25	−.17, .07	−.18, .13	−.03, .78	−.09, .44	−.11, .28
SCL anxiety	.00, .98	−.13, .18	−.28, .02	−.15, .23	−.21, .08	.04, .74
SCL agoraphobia	.02, .83	−.06, .50	−.29, .02	−.14, .25	−.19, .11	−.07, .53
	N=47	N=47	N=27	N=27	N=27	N=38

Baseline variable	BI t1 ρ or τ, P[a]	Session days ρ or τ, P	Session days ρ or τ, P	Time spent ρ or τ, P	Progress ρ or τ, P	BI t2 ρ or τ, P
PCCL internalization	.14, .37	−.11, .46	−.36, .07	−.42, .03	−.48, .01	−.06, .73
PCCL pain coping	−.03, .87	−.35, .02	−.28, .16	−.25, .21	−.26, .19	−.01, .95
PCCL catastrophizing	.03, .83	.02, .92	.16, .42	.30, .14	.25, .21	.03, .84
TSK kinesiophobia	−.08, .61	−.08, .58	.09, .67	.26, .20	.23, .25	−.08, .64
	N=108	N=108	N=69	N=69	N=69	N=86
CIS[b] subjective fatigue	−.04, .66	−.09, .38	−.07, .55	−.02, .90	−.10, .42	.03, .80
CIS concentration	.07, .49	−.10, .30	−.15, .22	−.05, .67	−.09, .47	.15, .16
CIS motivation	−.08, .44	−.05, .59	−.13, .29	−.04, .72	−.10, .41	.19, .08
CIS physical or inactivity	.02, .81	−.05, .60	.02, .90	.13, .27	.09, .48	−.04, .75
CIS total score	−.02, .82	−.12, .22	−.18, .14	−.06, .64	−.13, .29	.16, .15
	N=101	N=101	N=62	N=62	N=62	N=93
UBOS-a burnout (dit.)	−.01, .95	−.10, .30	−.23, .050	−.17, .10	−.20, .06	.04, 67
UBOS-a mental exhaustion	.03, .77	.01, .90	−.08, .55	−.11, .42	−.11, .38	.11, .35
UBOS-a distancing	.10, .93	−.08, .40	−.19, .15	−.16, .22	−.20, .12	.10, .36
UBOS-a work competence	.15, .14	.19, .052	.29, .02	.27, .03	.32, .01	−.14, .22
	N=86	N=86	N=54	N=54	N=54	N=69
Partially vs. fully absent (dit.)	.06, .55	.26, .01	.27, .03	.18, .11	.20, .09	.09, .40
Sick leave duration	.21, .054	.13, .24	.02, .91	.01, .96	.00, .98	−.10, .41
Previous treatment						
	N=116	N=116	N=71	N=71	N=71	N=93
Medication intake (dit.)	−.09, .44	−.15, .09	−.13, .22	−.10, .31	−.11, .27	.02, .79
Specialist treatment (dit.)	−.03, .72	−.06, .53	.16, .16	.08, .41	.15, .15	.09, .35

[a]ρ: Spearman ρ statistic was calculated when both variables had interval or ratio measurement levels, τ: Kendall τ statistic was calculated for dichotomous level independent variables (dit.) *P*: *P* value of test statistic.

[b]CIS: Checklist Individual Strength, ISCED: International Standard Classification of Education Level, NRS: Numerical Rating Scale, PCCL: Pain Coping and Cognitions, SCL: Symptom Checklist, TSK: Tampa Scale of Kinesiophobia, UBOS-a: Utrecht Burnout Scale labor (a) version, UCL: Utrecht Coping List.

Three interview respondents who exposed themselves to LAKA explained their level of engagement by witnessing that game tasks were welcome challenges in early stages of a rehabilitation process. However, patient users' attention shifted away from gaming tasks toward the pace (slow) and structure of the game when their confidence and engagement in "real-life" roles increased (eg, noticing that selecting preprogrammed alternatives is not as complex as responding in real life, and purposively selecting "bad" responses to explore the "rules" that guide scenarios). Disengagement was also explained by the belief of being incapable to perform a certain task.

> *At the time of the program… I was on sick leave. What could I do? I really had time for the computer, and no energy for anything else... When I stopped, it was enough for me. The game is too slow for me… For my energy that I've built up again... I started working again. I'm going to a sports club. Yes, my life, my rhythm, is different... I have no time.* [Respondent #3]

> *In work, I am constantly adjusting to people. So for me it did not really matter… I have an ADHD problem. So, attention exercises are a disaster for me. I have no patience for that… The first time I went on to see where I got stuck when I was just giving 'wrong' answers… Occasionally, when you had to wait, I was like: come on, hurry.* [Respondent #1]

Moderation and Mediation Effects

Performance expectancy was a significantly stronger predictor of behavioral intention at t1 in patients primarily diagnosed with chronic fatigue rather than a chronic pain condition (beta=.98; P<.001), and high levels of depressive symptoms (beta=.91; P=.006). The relationship between social influence

and behavioral intention was affected negatively by the more than 6-month absenteeism (beta= .61; *P*=.01). Daily Internet usage over the past year strengthened the positive effect of hedonic motivation on behavioral intention at preadoption (beta=.63; *P*=.001). The negative association between computer anxiety and behavioral intention was significantly weaker in participants younger than 45 years (beta=.42; *P*=.009). Mediation analyses showed that perceived behavioral control mediated the effect of active engagement on behavioral intention at preadoption, but did not mediate the effect of active engagement on the presence of a log-in.

Focusing on individual differences in coping with the delivery of LAKA during interviews yielded self-descriptions by patients, which varied between being "curious, a gamer, and capable" to play versus being neither a "games person" nor an "early adopter" and believing that computer games are difficult to play.

> *Anyway, I am someone who games a lot ... Did not doubt about being able to play it. I am someone who wants to follow and keep up with things ... There are buttons, and all the buttons I want to have tried them at least once.* [Respondent #1]
>
> *Most games that happen to PCs, such as Tetris and things like that… That is under time pressure … No, that does not attract me and I cannot do that ... I'm not the pioneer to go on my own.* [Respondent #2]

Feasibility Description

Demand

At the preadoption stage, most participants had a moderately positive intention to use LAKA over the next 4 months in addition to their scheduled IRP activities (Table 7). Nine participants (7.8%) with low initial behavioral intention (2 or lower) were statistical outliers, but were not excluded from further analyses. BI decreased over the course of 8 weeks. On average, players completed 8 encounters, which equals 2 of 4 modules in total. The first module of the game was completed by 40 patients (56.3%). Twenty players (28.2%) completed the game from start to end. A line graph (Figure 3) shows that players were more likely to stop using the game when they headed for a new game, module, or element.

Implementation and Practicality

Of 85 participants, 59 (69.4%) with a positive intention to use (BI ≥5) logged in successfully. At treatment facility A, 12 of the 24 willing participants logged in (50%), which is significantly less (χ^2_1=5.9; *P*=.015) than the proportions of participants at the other 3 locations (70.6%-81.0%). Players who possessed more computer platforms (ie, a tablet, a console, a mobile phone) progressed less within the game (ρ= .39; *P*=.001). Most activity was recorded during the first 4 weeks of participants' IRPs (Figure 4). When playing at home, participants logged in at different times during the day, but mostly after 6 pm (Figure 5).

Table 7. Descriptive results of demand level assessment.

Demand indicator	N	Mean	Median	SD[a]	Min	Max
BI[a] at t1	116	5.1	5	1.4	1	7
BI at t2	93	3.5	4	2.0	1	7
Session days	71	1.8	2	1.4	0	6
Time spend	71	1:14:40	0:52:25	1:07:42	0:00:00	4:22:27
Progression	71	8.1	7	7.3	0	32

[a]BI: behavioral intention, SD: standard deviation.

Figure 3. The proportions of players who stopped using LAKA at certain stages of progress.

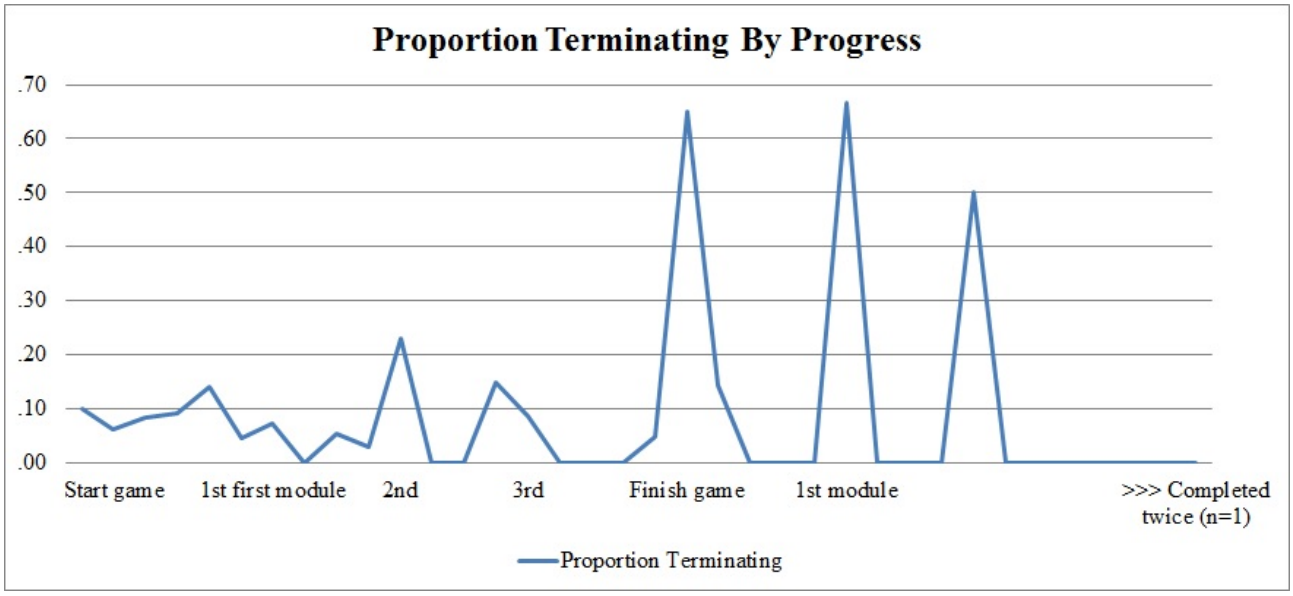


Figure 4. Sums of logged in game activities throughout the first 12 weeks of patients' IRPs.

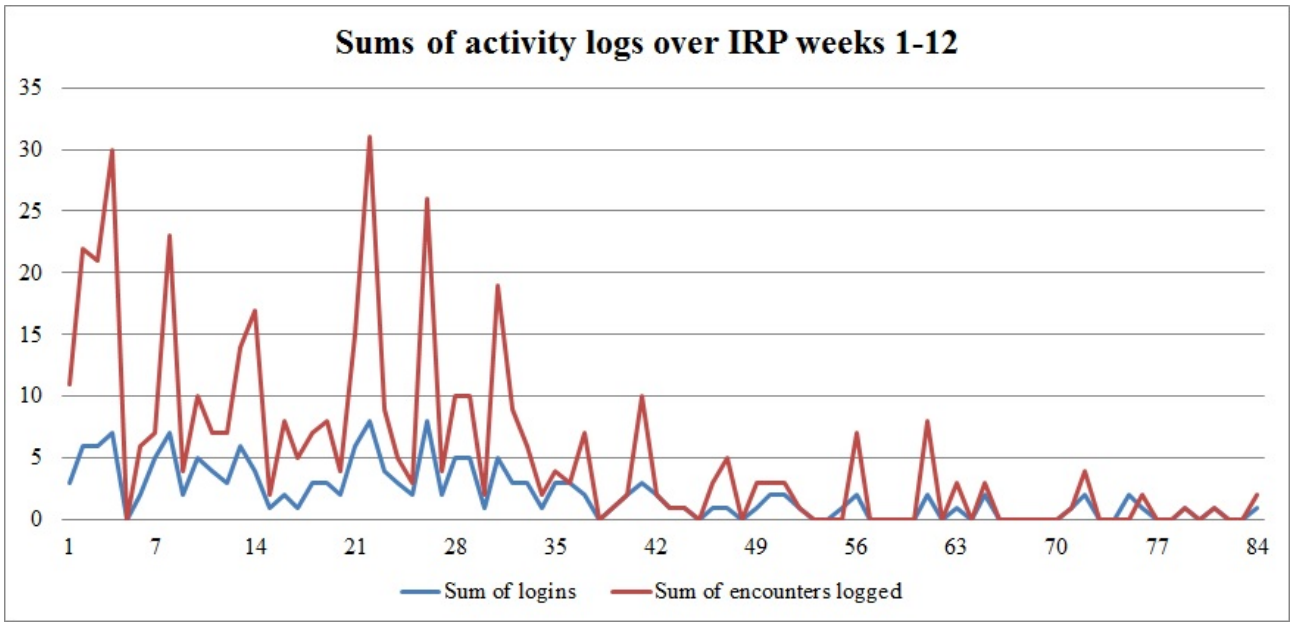
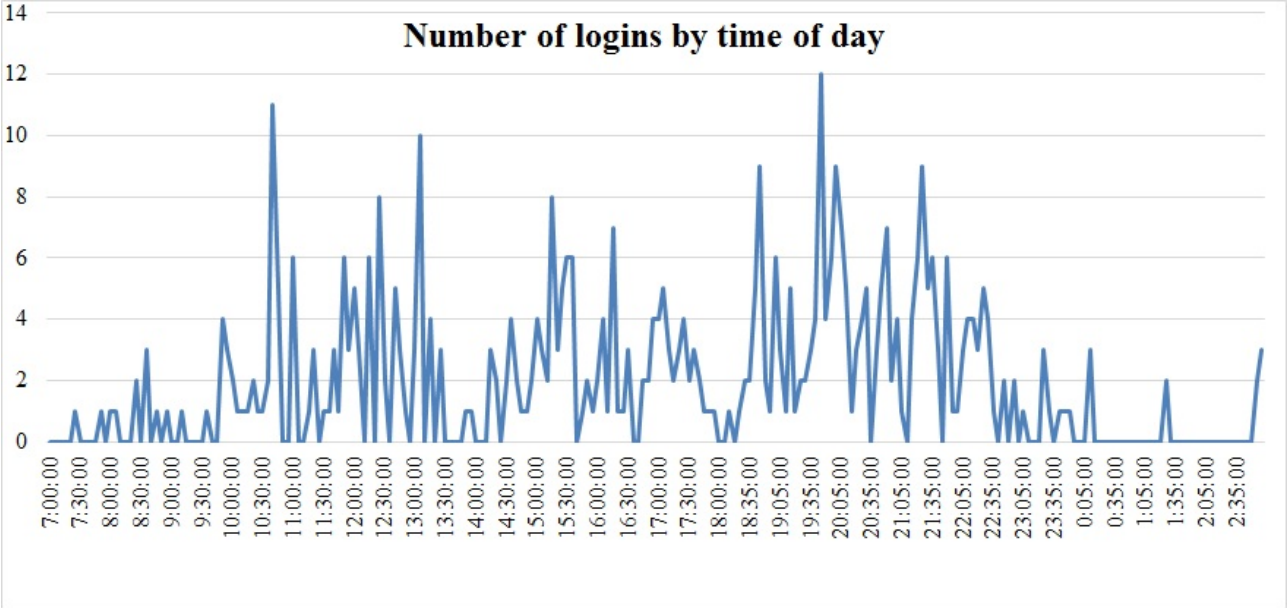


Figure 5. Number of logins by participants by time of the day.



Acceptability and Potential for Efficacy

Postusage perceived ease, rated by those who played the first module of LAKA, was moderately high (mean 5.4/7 (SD 1.54); median 6/7) and varied positively with baseline active engagement (ρ=.45; *P*=.007), motivation (ρ=.57; *P*=.001), and indication of a chronic musculoskeletal disorder (*P*=.008). Perceived enjoyment levels were moderately positive (mean 4.6 (SD 1.7)), and positively associated with baseline concentration problems (ρ=.44; *P*=.01). Postusage perceived knowledge improvement (mean 4.6/7 (SD 1.8)) was lower in participants with a higher BMI (ρ= .42; *P*=.02).

Interview respondents who played LAKA (respondents 1, 3, and 4) believed that it was a suitable program component. A variety of game elements were appointed that were liked. Furthermore, tasks were quickly understood, taken seriously, and experienced as a fit with the approach taken in other program components. “Encounters” were recognized as representations of real-life situations. Respondents who played generally believed that they could select options that corresponded with their intentions. Experienced consequences were acquaintance with meditation, concentration, and reflections on ideal and “healthy” selves, rumination, and adequate ways of responding.

> *I saw pretty quickly where they wanted to go with it. In that respect, it does well with what they do at Ciran.* [Respondent #1]

> *I have a computer, then it is no problem… I could use that game well… Later you find out: oh, it's not just a game. It is something to think about your own situation... Then you're not in the game, but you are in reality… All kinds of possibilities were offered (referring to response options in encounters): what I see as negative, in between, and what I see as 'good'… It was also a bit about … as I was during the illness... I could recognize myself in some situations: Yes, that was the old <patient name>… At the moment however, I react completely differently. I could see that. Maybe that also influenced me: my healing.* [Respondent #3]

> *During the game you were forced to concentrate; not wanting to go through it too fast… Actually you were just forced to use your concentration... It occurs to me that I ruminate long about something, and it also confronts you with that… Those first meditation exercises… I really needed to do it a few times ... That got me meditating…* [Respondent #4]

Suggestions for Improvement

Respondents suggested to integrate LAKA, similar to other program elements patients normally adhere to. Additional support was desired by means of a personalized introduction early in the program, information about how to get something out of the game and about what is achieved afterward, professional feedback on situations in the game, and facilities for gaming at Ciran locations. Suggestions for playability were to match the pace of in-game interaction with skill or health status and to extend software support to multiple platforms.

> *I would certainly continue to offer it. Maybe someone should be designated to introduce it at an early stage... To show it, and to show what you can get out of it for yourself. In a playful way ... If an entire manual should be read, then you put it away quickly… Actually, I do most with the IPad. If that is possible... I rarely use a pc.* [Respondent #4]

> *I would let everyone play. I think if it is offered in <location name>... “If there is a psychologist… that is better. Immediately talk about those reactions taking place at that time… find out faster what kind of situations played a role in becoming so ill, and get better... Maybe you could combine it … a bit slower in the beginning and a little faster at the end.* [Respondent #3]

Discussion

Principal Findings

This study primarily questioned why and by which patients with chronic pain and fatigue complaints is applied gaming for behavior change demanded during an IRP. Voluntary applied gaming (LAKA) was generally explained over time by perceived enjoyment and ease. Coping resources are important to solve delivery issues, get in control, and start using an applied game. Patient environment, health status, and performance expectancies were relevant factors for the usage of an applied game in conjunction with time and exposure.

Second, feasibility was described in preparation for a full-scale evaluation. A substantial number of patients played the game under noncommittal conditions. According to them, LAKA will be acceptable and useful. Active ingredients were recognized and deemed to be relevant in early stages of a rehabilitation process. Patients suggested delivering the game with social support through early and expeditious communication about how and when the game is relevant for their rehabilitation and with extended technical facilities.

Strengths and Limitations

This study provides the first empirical results on a novel applied game for behavioral change in patients with chronic pain and fatigue complaints. Comprehensive information is presented on processes of self-selection, acceptance, and attrition, which provides rare insights into risk factors for bias in CBI evaluations [58]. Mixed methodology strategy worked well to triangulate QN findings with newly collected QL data. Important demand explanations are based on notable and robust statistical results supported by a decent sample size and clear illustrations with QL data. QN results that were not clearly illustrated with QL data, or were based on more selective patient samples, provided practical information and clues for future research. More than a final feasibility assessment, this study contributed with general and utilizable knowledge for the future deployment of applied gaming for FSS patients in practice.

Limitations should be considered when drawing general conclusions about the feasibility of applied gaming for FSS patients. Feasibility was not assessed against a control group level or a reasonable benchmark. Technology acceptance measures are commonly used, but they are often contextually adapted and serve in theory building rather than feasibility assessment. Furthermore, this study builds on pragmatic eligibility criteria and convenience sampling of Dutch patients. It cannot be ruled out that early judgments about the nature of the delivery mode affected results via self-selection. Performing a large number of explorative statistical tests threatens statistical power and internal validity. Caution should be exercised when interpreting causality in relationships between behavioral factors and behavioral intentions because independent and dependent variables were measured at the same point in time, and hedonic motivation was not clearly distinct from performance expectancies and behavioral intentions. Whether the use of technology acceptance questionnaires alone would be an appropriate method for assessing the usage of gaming technology, especially at a time when patients may have difficulty processing information, can be doubted and is not recommended. Finally, advanced statistical techniques such as partial least squares regression [59] or newer process analyses techniques [60] would have been appropriate, but were not used. QN method limitations were addressed by triangulation of key QN findings with QL data, comprehensive sample description, validity checks, residual analysis, and sensitivity analyses (for outlier removal, measures of association, and regression method; see Multimedia Appendix 7).

Comparison With Prior Work

Researchers have stressed that a better understanding of the demand for CBIs is a major concern in overcoming barriers to treatment of patients with chronic pain and fatigue symptoms [2,3]. To our knowledge, this is the first empirical study on applied gaming for the delivery of behavioral intervention for patients with chronic somatic symptoms and functional problems. Findings suggest that voluntary engagement in applied gaming is strongly driven by positive affect. The importance of hedonic motivation for demand is remarkable because this is an often-omitted factor in previous research on the use of information technology in health care [27]. Concerns about utility, demonstrability of results, privacy, or consultation seemed to have a limited effect on demand in this case, when patients had no previous experiences. Ubiquitous interview quotes about “openness” hinted that inclinations to search for meaning or personal growth could partially explain demand for applied gaming [61,62]. Concluding that FSS patients will use an applied game “for the sake of the activity itself” is tentative. Applied gaming interventions are relatively new and barely institutionalized, and limited information was available to patients about the efficacy of LAKA or a similar game. It could also be that patients thought about usefulness and trustworthiness of care before deciding on following an IRP. Moreover, findings on influences of individual differences in coping styles and perceptions of control and ease on the usage of LAKA correspond with those of earlier studies that found a positive effect of internal locus of control on the adherence to a web-based positive psychology intervention [9]. Such results might also reflect differences in executive functioning or capacities for self-control [63].

Other remarkable QN results, which were not clearly illustrated with QL data, are discussed in connection with past research or as areas of future research. Findings on the effect of depressed mood on CBI usage have been heterogeneous [30]. This study pointed in the direction of a negative relationship, but found no statistically significant direct effect. This might be because of the comparatively high levels of psychopathologic symptoms of these FSS patients [64]. A moderation effect is indicated by extremely low BIs that were found exclusively in patients with low to neutral performance expectancies and high levels of depressive symptoms. Furthermore, computer anxiety and experience might explain differences in relationships between age and technology usage found in earlier studies [27]. Moreover, patients with lower scores for pain intensity and those who indicated fatigue as their primary complaint were more likely to self-select as a player, whereas patients with higher pain intensity played more once exposed. Further research on the usage of CBIs could focus on understanding “matches”

between symptom characteristics, "readiness," and demanded delivery mode of behavioral treatment for FSS populations [65]. Another research focus can be patient-environment interactions (ie, coping with issues at home, absenteeism, and return to work) as barriers and facilitators for demand within this target group. For informing design and implementation of computer- and game-based modalities, it is useful to proceed with qualitative research and by formulating and testing theoretically informed hypotheses on how usage varies by patient, program design, behavioral, and context factors [26,61,66-68].

The degree of implementation of LAKA for eligible patients is not satisfactory, as could be expected when a CBI is offered under ad libitum conditions [69]. Besides blending with face-to-face delivery and multiplatform distribution, solutions for additional support can be provided through Web-based features such as tailored messages, prompts, and support via email, chat, or message boards [4]. Acceptability and limited efficacy outcomes should be treated with caution but suggest that LAKA is potentially efficacious and sufficiently engaging to complete once or twice (2-4 hours). Moderately positive enjoyment by users may reflect that the design principle of LAKA was not entirely hedonic, maybe at the expense of "playability" aspects [70]. Eliciting reflective and meditative states, LAKA was pleasant for a patient with concentration problems; however, self-reflections seem to be at the expense of a more satisfactory speed of interaction. Moreover, the game appears to provide opportunity to realize ideal selves, which supports intrinsic motivation [57]. However, LAKA also triggered serious reflections about discrepancy with "actual" selves, which is associated with somatic symptoms and negative emotions [71]. "Slowness" was mentioned as a reason for disengagement, but self-awareness was not. One may also reflect about how self-awareness in virtual reality relates to bodily and behavioral representations of Avatars [72] because extremely low perceived knowledge improvement levels were exclusively reported by patients with high BMI levels at baseline. High-quality and adequately powered studies on the effects of LAKA and similar systems on functional domains are needed to clarify the roles of self-conscious and affective states, learning, and degree of engagement [16,17,73].

Conclusion

Although these first empirical findings support that an applied game is used by FSS patients for enjoyment and convenience, it became very clear that many patients would not be reached with a behavioral intervention of this modality under voluntary conditions. Social factors remain highly important for reaching many patients. LAKA will be feasible as a short and early intervention for patients, with adjustments of social and technical support. A next step in deployment and evaluation of the efficacy and cost-effectiveness of LAKA in a controlled study is recommendable.

Acknowledgments

Ciran, the developer of LAKA, sponsored the research described in this publication and contributed to the study design. Several partners made essential contributions to LAKA. We thank patients for their daily input. Lamas of Tibetan Institute, Yeunten Ling, Jan A.M. Jochijms, and Antonie E. Akkerman, are acknowledged for their contributions to the framing of the concept and operationalization of behavioral change content. Professor Dr. Jac L.A. Geurts guided the formulation of program requirements. Karel Michiels wrote scenarios. Finally, the creation and quality assurance of programming and art were outsourced to Paladin Studios.

Conflicts of Interest

MV is employed by Ciran, and provided time and occasion to conduct independent doctoral research, by way of agreement, at Tranzo, Scientific Center for Care and Welfare. The terms of this arrangement have been reviewed and approved by Tranzo in accordance with its policy on objectivity in research. AB has been serving on the board of directors of Ciran. HV receives compensation for coordinating Ciran's scientific research agenda.

Multimedia Appendix 1

Conceptual framework.

[PDF File (Adobe PDF File), 394KB - games_v4i1e2_app1.pdf]

Multimedia Appendix 2

Serious gaming page.

[PDF File (Adobe PDF File), 654KB - games_v4i1e2_app2.pdf]

Multimedia Appendix 3

LAKA details.

[PDF File (Adobe PDF File), 1MB - games_v4i1e2_app3.pdf]

Multimedia Appendix 4

All measurement details.

[PDF File (Adobe PDF File), 771KB - games_v4i1e2_app4.pdf]

Multimedia Appendix 5

Regression models specified.

[PDF File (Adobe PDF File), 291KB - games_v4i1e2_app5.pdf]

Multimedia Appendix 6

Interview schedule.

[PDF File (Adobe PDF File), 440KB - games_v4i1e2_app6.pdf]

Multimedia Appendix 7

Method sensitivity analyses.

[PDF File (Adobe PDF File), 322KB - games_v4i1e2_app7.pdf]

References

1. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: A meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
2. Eccleston C, Fisher E, Craig L, Duggan GB, Rosser BA, Keogh E. Psychological therapies (Internet-delivered) for the management of chronic pain in adults. Cochrane Database Syst Rev 2014;2:CD010152. [doi: 10.1002/14651858.CD010152.pub2] [Medline: 24574082]
3. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: A systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
4. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
5. Beatty L, Lambert S. A systematic review of internet-based self-help therapeutic interventions to improve distress and disease-control among adults with chronic health conditions. Clin Psychol Rev 2013 Jun;33(4):609-622. [doi: 10.1016/j.cpr.2013.03.004] [Medline: 23603521]
6. Bleakley CM, Charles D, Porter-Armstrong A, McNeill MD, McDonough SM, McCormack B. Gaming for health: A systematic review of the physical and cognitive effects of interactive computer games in older adults. J Appl Gerontol 2015 Apr;34(3):NP166-NP189. [doi: 10.1177/0733464812470747] [Medline: 24652863]
7. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
8. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: A systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
9. Geraghty AW, Wood AM, Hyland ME. Attrition from self-directed interventions: Investigating the relationship between psychological predictors, intervention content and dropout from a body dissatisfaction intervention. Soc Sci Med 2010 Jul;71(1):30-37. [doi: 10.1016/j.socscimed.2010.03.007] [Medline: 20400220]
10. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: A review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
11. Brouwer W, Kroeze W, Crutzen R, de Nooijer J, de Vries NK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
12. Rosser BA, Vowles KE, Keogh E, Eccleston C, Mountain GA. Technologically-assisted behaviour change: A systematic review of studies of novel technologies for the management of chronic illness. J Telemed Telecare 2009;15(7):327-338. [doi: 10.1258/jtt.2009.090116] [Medline: 19815901]
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 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]

14. Camerini L, Schulz PJ. Effects of functional interactivity on patients' knowledge, empowerment, and health outcomes: An experimental model-driven evaluation of a web-based intervention. J Med Internet Res 2012;14(4):e105 [FREE Full text] [doi: 10.2196/jmir.1953] [Medline: 22810046]
15. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
16. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
17. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: A systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
18. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. Eur J Pain 2006 May;10(4):287-333. [doi: 10.1016/j.ejpain.2005.06.009] [Medline: 16095934]
19. Kroenke K. Patients presenting with somatic complaints: Epidemiology, psychiatric comorbidity and management. Int J Methods Psychiatr Res 2003;12(1):34-43. [Medline: 12830308]
20. Henningsen P, Zipfel S, Herzog W. Management of functional somatic syndromes. Lancet 2007 Mar 17;369(9565):946-955. [doi: 10.1016/S0140-6736(07)60159-7] [Medline: 17368156]
21. Harvey SB, Wessely S. How should functional somatic syndromes be diagnosed? Int J Behav Med 2013 Jun;20(2):239-241. [doi: 10.1007/s12529-013-9300-7] [Medline: 23479340]
22. Bourke JH, Langford RM, White PD. The common link between functional somatic syndromes may be central sensitisation. J Psychosom Res 2015 Mar;78(3):228-236. [doi: 10.1016/j.jpsychores.2015.01.003] [Medline: 25598410]
23. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. MIS Quarterly 2012;36(1):157-178.
24. Venkatesh V, Thong JYL, Chan FKY, Hu PJ, Brown SA. Extending the two-stage information systems continuance model: Incorporating UTAUT predictors and the role of context. Information Systems Journal 2011;21(6):527-555.
25. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
26. Kim J, Park HA. Development of a health information technology acceptance model using consumers' health behavior intention. J Med Internet Res 2012;14(5):e133 [FREE Full text] [doi: 10.2196/jmir.2143] [Medline: 23026508]
27. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Aug;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
28. Ivankova NV, Creswell JW, Stick SL. Using mixed-methods sequential explanatory design: From theory to practice. Field Methods 2006 Feb 01;18(1):3-20. [doi: 10.1177/1525822X05282260]
29. Creswell DJW, Clark VLP. Designing and Conducting Mixed Methods Research. 2nd edition. USA: Sage Publications, Inc; 2007.
30. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: A mixed method study. J Med Internet Res 2013;15(10):e223 [FREE Full text] [doi: 10.2196/jmir.2742] [Medline: 24132044]
31. Garschagen A, Steegers MA, van Bergen AH, Jochijms JA, Skrabanja TL, Vrijhoef HJ, et al. Is there a need for including spiritual care in interdisciplinary rehabilitation of chronic pain patients? Investigating an innovative strategy. Pain Pract 2015 Sep 17;15(7):671-687. [doi: 10.1111/papr.12234] [Medline: 25229884]
32. Ryff CD, Singer B. Psychological well-being: Meaning, measurement, and implications for psychotherapy research. Psychother Psychosom 1996;65(1):14-23. [Medline: 8838692]
33. Magasi S, Post MW. A comparative review of contemporary participation measures' psychometric properties and content coverage. Arch Phys Med Rehabil 2010 Sep;91(9 Suppl):S17-S28. [doi: 10.1016/j.apmr.2010.07.011] [Medline: 20801275]
34. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
35. Lutz A, Dunne JD, Davidson RJ. Meditation and the neuroscience of consciousness: An introduction. In: Zelazo PD, Moscovitch M, Thompson E, editors. The Cambridge Handbook of Consciousness. Cambridge, UK: Cambridge University Press; 2007:497-549.
36. Jackson J. Stigma, liminality, and chronic pain: Mind-body borderlands. American Ethnologist 2005;32(3):332-353.
37. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: Toward a unified view. MIS quarterly 2003:425-478.
38. Turner H, Bryant-Waugh R, Peveler R, Bucks RS. A psychometric evaluation of an English version of the Utrecht Coping List. Eur Eat Disord Rev 2012 Jul;20(4):339-342. [doi: 10.1002/erv.2173] [Medline: 22419565]
39. Nunnally J, Bernstein IH. Psychometric methods. In: Psychometric Theory. New York: McGraw-Hill; 1994.

40. Or CK, Karsh B, Severtson DJ, Burke LJ, Brown RL, Brennan PF. Factors affecting home care patients' acceptance of a web-based interactive self-management technology. J Am Med Inform Assoc 2011 Feb;18(1):51-59 [FREE Full text] [doi: 10.1136/jamia.2010.007336] [Medline: 21131605]
41. Van der Heijden H. User acceptance of hedonic information systems. MIS Quarterly 2004:695-704.
42. Vallejo MA, Jordán CM, Díaz MI, Comeche MI, Ortega J. Psychological assessment via the internet: A reliability and validity study of online (vs paper-and-pencil) versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). J Med Internet Res 2007;9(1):e2 [FREE Full text] [doi: 10.2196/jmir.9.1.e2] [Medline: 17478411]
43. Brenninkmeijer V, VanYperen N. How to conduct research on burnout: Advantages and disadvantages of a unidimensional approach in burnout research. Occup Environ Med 2003 Jun;60 Suppl 1:i16-i20 [FREE Full text] [Medline: 12782742]
44. Fu FL, Su RC, Yu SC. EGameFlow: A scale to measure learners' enjoyment of e-learning games. Computers & Education 2009;52(1):101-112.
45. Oshlyansky L, Cairns P, Thimbleby H. Validating the Unified Theory of Acceptance and Use of Technology (UTAUT) tool cross-culturally. In: Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI but not as we know it. 2007 Presented at: 21st British HCI Group Annual Conference on People and Computers; September 3-7; Lancaster, United Kingdom.
46. Davis FD, Venkatesh V. A critical assessment of potential measurement biases in the technology acceptance model: Three experiments. International Journal of Human-Computer Studies 1996;45(1):19-45.
47. Price DD, Bush FM, Long S, Harkins SW. A comparison of pain measurement characteristics of mechanical visual analogue and simple numerical rating scales. Pain 1994 Feb;56(2):217-226. [Medline: 8008411]
48. Dittner AJ, Wessely SC, Brown RG. The assessment of fatigue: A practical guide for clinicians and researchers. J Psychosom Res 2004 Feb;56(2):157-170. [doi: 10.1016/S0022-3999(03)00371-4] [Medline: 15016573]
49. Arindell WA, Ettema JHM. SCL-90, Handleiding bij een Multidimensionele Psychopathologie-Indicator. Amsterdam, Netherlands: Swets & Zeitlinger B.V; 1986.
50. Stomp-van der Berg S, Vlaeyen J, Ter KM, Spinhoven P, van BG, Kole-Snijders A. Meetinstrumenten chronische pijn deel 2 Pijn Coping en Cognitie Lijst (PCCL). Maastricht: Pijn Kennis Centrum Academisch Ziekenhuis Maastricht 2001.
51. Roelofs J, Goubert L, Peters ML, Vlaeyen JW, Crombez G. The Tampa Scale for Kinesiophobia: further examination of psychometric properties in patients with chronic low back pain and fibromyalgia. Eur J Pain 2004 Oct;8(5):495-502. [doi: 10.1016/j.ejpain.2003.11.016] [Medline: 15324781]
52. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986 Dec;51(6):1173-1182. [Medline: 3806354]
53. VERBI GmbH. 1995. MAXQDA 11 URL: http://www.maxqda.com/products/maxqda11 [accessed 2015-04-16] [WebCite Cache ID 6XpOq7xmD]
54. Morley S, Williams A, Hussain S. Estimating the clinical effectiveness of cognitive behavioural therapy in the clinic: Evaluation of a CBT informed pain management programme. Pain 2008 Jul 31;137(3):670-680. [doi: 10.1016/j.pain.2008.02.025] [Medline: 18394806]
55. Prins JB, Bleijenberg G, Bazelmans E, Elving LD, de Boo TM, Severens JL, et al. Cognitive behaviour therapy for chronic fatigue syndrome: A multicentre randomised controlled trial. Lancet 2001 Mar 17;357(9259):841-847. [doi: 10.1016/S0140-6736(00)04198-2] [Medline: 11265953]
56. UNESCO. 1997. International Standard Classification of Education URL: http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm [WebCite Cache ID 6VcqPwizW]
57. Vercoulen JH, Hommes OR, Swanink CM, Jongen PJ, Fennis JF, Galama JM, et al. The measurement of fatigue in patients with multiple sclerosis. A multidimensional comparison with patients with chronic fatigue syndrome and healthy subjects. Arch Neurol 1996 Jul;53(7):642-649. [Medline: 8929171]
58. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Sampling bias in an internet treatment trial for depression. Transl Psychiatry 2012;2:e174 [FREE Full text] [doi: 10.1038/tp.2012.100] [Medline: 23092978]
59. Geladi P, Kowalski BR. Partial least-squares regression: A tutorial. Analytica Chimica Acta 1986;185:1-17. [doi: 10.1016/0003-2670(86)80028-9]
60. Hayes AF. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. Communication Monographs 2009 Dec;76(4):408-420. [doi: 10.1080/03637750903310360]
61. Szalma JL. On the application of motivation theory to human factors/ergonomics: Motivational design principles for human-technology interaction. Hum Factors 2014 Dec;56(8):1453-1471. [Medline: 25509824]
62. Przybylski AK, Weinstein N, Murayama K, Lynch MF, Ryan RM. The ideal self at play: The appeal of video games that let you be all you can be. Psychol Sci 2012 Jan 1;23(1):69-76. [doi: 10.1177/0956797611418676] [Medline: 22173739]
63. Nes L, Roach A, Segerstrom S. Executive functions, self-regulation, and chronic pain: A review. Annals of Behavioral Medicine 2009;37(2):173-183.
64. Arrindell WA, Barelds DP, Janssen I, Buwalda FM, van der Ende J. Invariance of SCL-90-R dimensions of symptom distress in patients with peri partum pelvic pain (PPPP) syndrome. Br J Clin Psychol 2006 Sep;45(Pt 3):377-391. [Medline: 17147103]

65. Ritterband LM, Cox DJ, Gordon TL, Borowitz SM, Kovatchev BP, Walker LS, et al. Examining the added value of audio, graphics, and interactivity in an Internet intervention for pediatric encopresis. Children's Health Care 2006;35(1):47-59.
66. Ahadzadeh AS, Pahlevan SS, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: An investigation of health-related internet use. J Med Internet Res 2015;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
67. Beaudry A, Pinsonneault A. The other side of acceptance: Studying the direct and indirect effects of emotions on information technology use. MIS Quarterly 2010;34(4):689-710.
68. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: A systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
69. Botella C, Garcia-Palacios A, Vizcaíno Y, Herrero R, Baños RM, Belmonte MA. Virtual reality in the treatment of fibromyalgia: A pilot study. Cyberpsychol Behav Soc Netw 2013 Mar;16(3):215-223. [doi: 10.1089/cyber.2012.1572] [Medline: 23496678]
70. Sánchez JLG, Vela FLG, Simarro FM, Padilla-Zea N. Playability: Analysing user experience in video games. Behaviour & Information Technology 2012 Mar;31(10):1033-1054. [doi: 10.1016/j.jpainsymman.2003.07.001]
71. Waters SJ, Keefe FJ, Strauman TJ. Self-discrepancy in chronic low back pain: Relation to pain, depression, and psychological distress. J Pain Symptom Manage 2004 Mar;27(3):251-259. [doi: 10.1016/j.jpainsymman.2003.07.001] [Medline: 15010103]
72. Kim Y, Sundar SS. Visualizing ideal self vs actual self through avatars: Impact on preventive health outcomes. Computers in Human Behavior 2012;28(4):1356-1364.
73. Glegg SM, Tatla SK, Holsti L. The GestureTek virtual reality system in rehabilitation: A scoping review. Disabil Rehabil Assist Technol 2014 Mar;9(2):89-111. [doi: 10.3109/17483107.2013.799236] [Medline: 23713408]

Abbreviations

BI: behavioral intention
CBI: computer-based interventions
FSS: functional somatic syndromes
IRP: interdisciplinary rehabilitation program
QL: qualitative
QN: quantitative

Edited by G Eysenbach; submitted 29.08.15; peer-reviewed by D Bossen, FS Ong, P Schulz, WC Su; comments to author 04.11.15; revised version received 24.12.15; accepted 19.01.16; published 01.04.16.

Please cite as:
Vugts MAP, Joosen MCW, van Bergen AHMM, Vrijhoef HJM
Feasibility of Applied Gaming During Interdisciplinary Rehabilitation for Patients With Complex Chronic Pain and Fatigue Complaints: A Mixed-Methods Study
JMIR Serious Games 2016;4(1):e2
URL: http://games.jmir.org/2016/1/e2/
doi:10.2196/games.5088
PMID:27036094

Original Paper

The Relationship Between Engagement and Neurophysiological Measures of Attention in Motion-Controlled Video Games: A Randomized Controlled Trial

Amber M Leiker[1*], MEd; Matthew Miller[1*], PhD; Lauren Brewer[1], MEd; Monica Nelson[1]; Maria Siow[1]; Keith Lohse[2], PhD

[1]Auburn University, Auburn, AL, United States

[2]Rehabilitation Informatics Lab, School of Kinesiolgy, Auburn University, Auburn, AL, United States

[*]these authors contributed equally

Corresponding Author:
Keith Lohse, PhD
Rehabilitation Informatics Lab
School of Kinesiolgy
Auburn University
301 Wire Road
Auburn, AL,
United States
Phone: 1 334 844 1982
Fax: 1 334 844 4483
Email: rehabinformatics@gmail.com

Abstract

Background: Video games and virtual environments continue to be the subject of research in health sciences for their capacity to augment practice through user engagement. Creating game mechanics that increase user engagement may have indirect benefits on learning (ie, engaged learners are likely to practice more) and may also have direct benefits on learning (ie, for a fixed amount of practice, engaged learners show superior retention of information or skills).

Objective: To manipulate engagement through the aesthetic features of a motion-controlled video game and measure engagement's influence on learning.

Methods: A group of 40 right-handed participants played the game under two different conditions (game condition or sterile condition). The mechanics of the game and the amount of practice were constant. During practice, event-related potentials (ERPs) to task-irrelevant probe tones were recorded during practice as an index of participants' attentional reserve. Participants returned for retention and transfer testing one week later.

Results: Although both groups improved in the task, there was no difference in the amount of learning between the game and sterile groups, countering previous research. A new finding was a statistically significant relationship between self-reported engagement and the amplitude of the early-P3a (eP3a) component of the ERP waveform, such that participants who reported higher levels of engagement showed a smaller eP3a (beta=−.08, *P*=.02).

Conclusions: This finding provides physiological data showing that engagement elicits increased information processing (reducing attentional reserve), which yields new insight into engagement and its underlying neurophysiological properties. Future studies may objectively index engagement by quantifying ERPs (specifically the eP3a) to task-irrelevant probes.

(JMIR Serious Games 2016;4(1):e4) doi:10.2196/games.5460

KEYWORDS

games; engagement; motivation; eP3a; Kinect; exergame

Introduction

Motion-controlled video game systems such as Microsoft Kinect, Playstation Move, and Nintendo Wii have seen a surfeit of research in recent years because these kinds of virtual environments provide the opportunity for motivating interactions and full-bodied movements and providing additional feedback beyond the body's intrinsic sensory systems [1-3]. One of the major areas of interest in this research is the potential for certain game mechanics to reliably increase participant engagement [4,5]. Engagement has been defined as an affective quality or experience of a participant in a task that emerges from focused attention, aesthetic pleasures, perceptions of novelty, perceptions of usability, and the extent to which the participant feels involved in the task (ie, choices in the game have meaningful consequences [6,7]). Viscerally pleasing stimuli, choice, clear mechanics/feedback, novelty/exploration, and adaptive difficulty are game mechanics thought to contribute to engagement [5,8,9]. Engagement is thus related to, but distinct from, motivation. Participants could be motivated to play a game, but if the game no longer offers adequate challenge, they may not be engaged by the game, potentially reducing future motivation.

Major reasons for using games in rehabilitation are the indirect effects that engagement might have on learning. That is, increased engagement might be beneficial for skill learning and rehabilitation because a participant will be inclined to practice more (ie, have greater compliance with therapy). Beyond this indirect benefit of engagement, recent evidence suggests that increased engagement during practice might also have a direct effect on learning [10]. In that study, Lohse et al manipulated the aesthetics of a gaming environment while keeping the amount of practice and the mechanics of the game constant. The group of participants who trained in the *game* group (complex, space-themed graphics with ambient and task-relevant sound) showed statistically superior retention and transfer performance compared to participants in the *sterile* group (simple, geometric graphics with no sounds), although the groups did not differ during practice (ie, gamification specifically enhanced learning). Furthermore, the game group self-reported statistically higher levels of engagement than the sterile group (using a language-adapted version of a user-engagement scale [6]). However, even though increased engagement was observed coincident to improved learning in the game group, individual engagement scores were not correlated with participant posttest performance, raising questions about the relationship between engagement and observed learning benefits.

This potential for engagement during practice to augment the learning of a novel motor skill was the impetus for the current experiment. Adapting the methods of Lohse et al [10], we conducted an electroencephalography (EEG) study in which participants practiced in either game or sterile conditions while task-irrelevant auditory probes were played at random intervals. Measuring event-related potentials (ERPs) in response to complex tones is a common research paradigm. In particular, we chose to focus on the amplitude of the early P3a (eP3a) component of the ERP waveform. The eP3a in response to auditory stimuli has been shown to be a reliable index of attentional reserve [11-13]. That is, when more attention is being paid to the primary task (ie, more information being processed), the magnitude of the eP3a in response to an irrelevant tone will be lower as a consequence of fewer attentional resources being available to process the tone. Thus, we hypothesized that participants in the game-training group would show a reduced eP3a compared to the sterile group, suggesting that more attentional resources are absorbed by the task in the game condition than in the sterile condition. Consistent with the results of Lohse et al [10], we also hypothesized that the game group would show superior learning (ie, better performance on retention and transfer tests) compared to the sterile group. The experiment was powered specifically to detect these effects, but in order to follow up these a priori hypotheses, we also conducted exploratory analyses of the relationships between posttest performance, self-reported engagement, and eP3a amplitude.

Methods

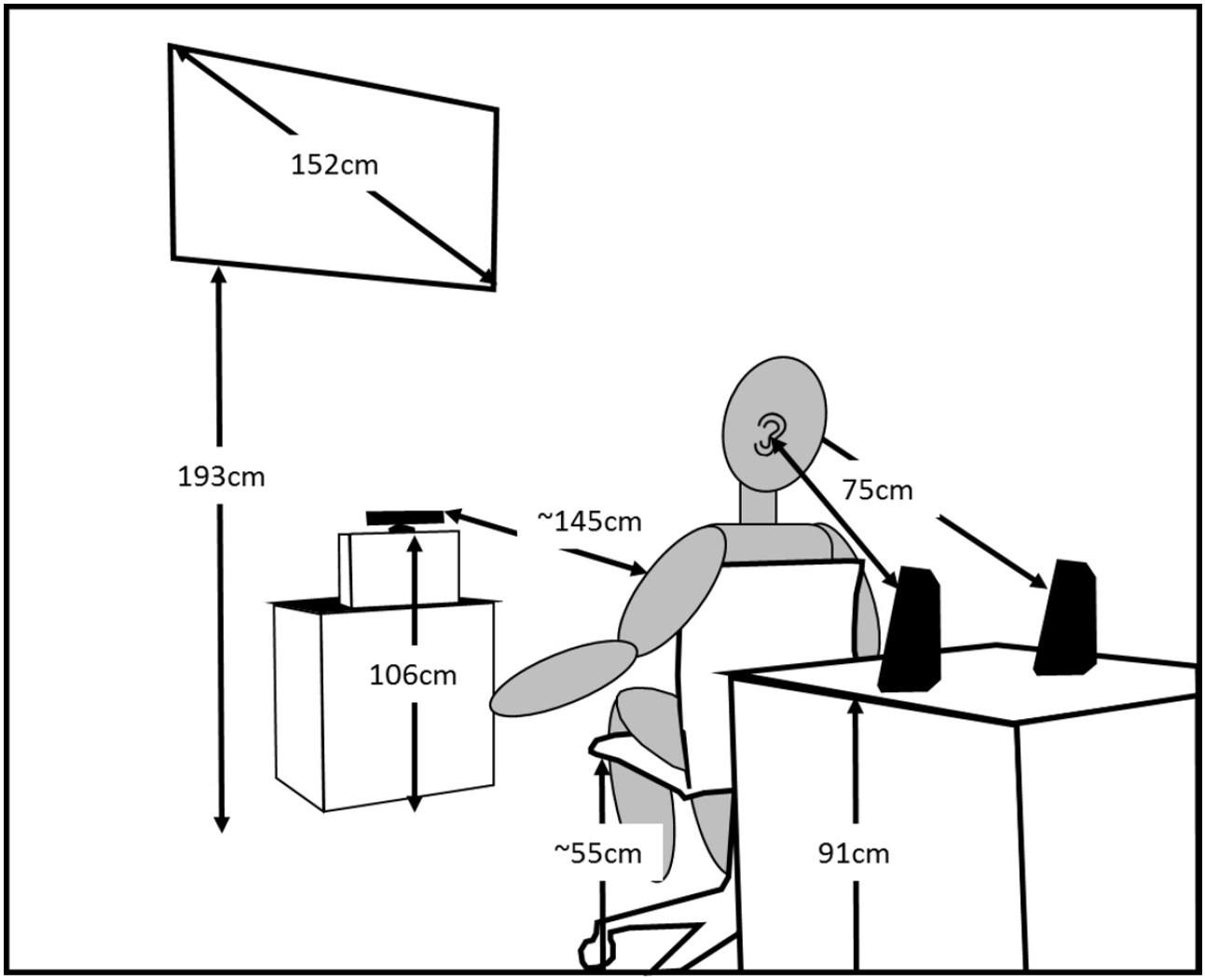
Participants

A group of 40 right-handed participants was recruited through classes, flyers, and an online advertisement at Auburn University (17 male, 23 female). The average age of the participants was 22.6 (SD 3.15) years. Six participants indicated that they had used the Kinect system at some point in the past, but none of the participants had played in the last 3 months or regularly played before that (self-reported frequency 0 (SD 0.0) days/week). Many participants (n=31) indicated that they played some other form of motion-controlled game (mostly Nintendo Wii), with an average frequency of 0.2 (SD 0.61) days/week, and 36 participants indicated that they played games in some other medium (most commonly a mobile phone) with an average frequency of 1.34 (SD 1.59) days/week. Participants were randomly assigned to either the game group or the sterile group using blocked random assignment within sex to balance the groups. Participants self-reported no musculoskeletal or neurological impairments that would affect their performance, and all had normal or corrected-to-normal vision.

Game Apparatus

Participants played a custom-built computer game written in Visual Studio 2013 using XNA Game Studio 4.0 and the Kinect SDK 1.8 using the Microsoft Kinect. The game was displayed on a 152 cm Samsung HDTV that was 193 cm above the ground (see Figure 1). The Kinect camera was placed 106 cm above the ground and approximately 145 cm away from the participant (who could move forward or back to improve tracking).

Figure 1. A schematic of experimental setup. The Kinect camera and speakers are shown in black.



Auditory Probes

Speakers for presenting the auditory probes were placed on a table 91 cm above the ground. The center of each speaker was set to an initial radial distance of 75 cm to the center of the ear of each participant. Probes consisted of 30 novel complex sounds (eg, door knock, dog bark, whistle) employed in previous studies using task-irrelevant auditory stimuli to index attentional reserve [11,14,15]. Probes were presented in random order at 75-95 dB SPL with interstimulus intervals varying randomly between 10 and 50 s.

Procedures

All procedures were approved by the Internal Review Board of Auburn University (14-502 EP 1411). On day 1, participants provided written informed consent and completed an initial survey measuring handedness and past experience with video games. Next, participants were prepared for EEG recording while probe tones were played in order to habituate participants to the tones. Participants were told that these tones would be playing in the background during the experiment but they had nothing to do with the game.

The Kinect system was then calibrated to track the nondominant left hand, and all participants were given standardized instructions on how to play the game. In the game/sterile conditions, participants controlled the motion of a spaceship/cursor on the screen in order to catch asteroids/circles and throw them into yellow targets that would appear at the top, bottom, or sides of the screen. The two graphic types are shown in Figure 2; other than this difference, the game conditions were mechanically identical. All participants were instructed to catch the objects as quickly as possible and hit as many targets as they could. This combined speed-accuracy constraint was reinforced by participant in-game scores. Participants lost a single point for every 10 frames (approximately 167 ms) that they had not yet hit the target and scored 100 points for every target hit.

Following the standardized instructions, all participants completed a 20-trial pretest in both the same condition they would practice in and the opposite condition (40 trials total). The order of the pretest was counterbalanced across participants. The pretest was given under both test conditions to detect any potential baseline differences in the difficulty of the two conditions. No tones were played and no EEG data were recorded. Following the pretest, participants completed 200 practice trials in their given condition (game or sterile). EEG data were continuously recorded during this time period and probe tones were played. After 200 practice trials, participants were given the opportunity to rest. When ready, they began the

second round of 200 practice trials in the same condition (400 practice trials total) with probe tones playing. Because the length of time required to complete the trials varied by how quickly participants caught and threw objects, the number of probe tones played varied as well (ie, faster players would hear fewer tones). The number of probe tones played ranged from 49 to 60 (mean 54.9, SD 2.86).

Approximately one week later (5-9 days with a median of 7 days), participants returned for retention (same condition as practice) and transfer (opposite condition) posttests. As with the pretest, the order of the posttests was counterbalanced across participants, but participants always completed the posttests in the same order they completed the pretest. Each posttest consisted of 20 trials (40 trials total). No tones were played, and no EEG data were collected during the posttest.

Figure 2. Sterile condition (left) and game condition (right).

Electroencephalograph Processing and Measures

Scalp EEG was collected from 20 channels of an EEG cap housing a 64-channel BrainVision actiCAP system (Brain Products GmbH) labeled in accord with an extended international 10-20 system [16]. EEG data were online-referenced to the left earlobe, and a common ground was employed at the FPz electrode site. Electrode impedances were maintained below 25 kΩ throughout the study, and a high-pass filter was set at 0.016 Hz with a sampling rate of 250 Hz. The EEG signal was amplified and digitized with a BrainAmp DC amplifier (Brain Products GmbH) linked to BrainVision Recorder software (Brain Products GmbH).

EEG data processing was conducted with BrainVision Analyzer 2.1 software (BrainProducts GmbH). Data were rereferenced to an averaged ears montage, band-pass filtered between 0.1 and 30 Hz with 24-dB rolloffs with a 60 Hz notch employing a zero phase shift Butterworth filter. Next, eye blinks were reduced employing the independent component analysis (ICA)-based ocular artifact rejection function within the BrainVision Analyzer software (electrode FP2 served as the vertical electrooculogram channel; BrainProducts, 2013). This function searches for an ocular artifact template in channel FP2 and then finds ICA-derived components that account for a user-specified (70%) amount of variance in the template-matched portion of the signal from FP2. These components were removed from the EEG signal, which was then reconstructed for further processing. ERPs were obtained by extracting the epoch of 200 ms prior to probe onset through 800 ms postprobe, then baseline-corrected with reference to the preprobe interval. Next, ERPs containing changes of more than 100 μV within a moving 200-ms window were excluded from subsequent analysis. This resulted in 1.3% of trials being rejected. The remaining trials were then averaged. Visual inspection of participants' average waveforms revealed substantial interindividual differences in component latencies. Accordingly, the adaptive mean amplitude quantification technique was employed [17] to quantify the N1, eP3a, and late P3a (lP3a) components, although the primary component of interest was the eP3a. For each participant, a 40-ms time window was centered on the peaks within the 200-300 ms and 350-450 ms time ranges for the eP3a and lP3a components, respectively, and a 20-ms time window was centered on the peak within the 100-200 ms range for the N1 component. Mean amplitude was calculated for each component within the time window at the electrode where the component was maximal when averaged across all participants. This resulted in the N1 and eP3a being calculated at Cz and the lP3a being calculated at Fz.

Survey Measures

Following the end of practice on day 1, participants completed a posttraining survey that included a language-adapted version of a user-engagement scale developed in the human-computer interaction literature [6] and a language-adapted version of the Intrinsic Motivation Inventory (IMI) [18] edited to include only the interest/enjoyment, perceived competence, effort, and pressure/tension subscales.

Statistical Power and Analyses

All statistical analyses were conducted using SPSS version 22.0 (IBM Corp). The experiment was designed to test two a priori hypotheses: (a) the game group would show superior learning relative to the sterile group as measured by in-game performance on the retention and transfer tests and (b) the game group would show a decreased eP3a relative to the sterile group.

We operationally defined the learning effect as the interaction of test (pre- vs post-) and training condition (game vs sterile) in a mixed-factorial analysis of variance (ANOVA). Assuming an alpha of .05, Cohen's *f* of .25 (a medium effect), and a positive correlation between the pretest and posttest (*r*=.50), a total sample size of N=40 was needed to achieve approximately 80% power. For the eP3a, this sample size would also give us approximately 80% power to detect a Cohen's *f* of .45 (a large effect), assuming alpha is .05, operationally defined as the main effect of group (game vs sterile) in an independent samples *t* test. Power calculations used G*Power 3.1 [19].

Points scored in-game were analyzed in blocks of five trials (maximum 500 points per block). Learning was measured by points per block using a mixed-factorial ANOVA with a between-subjects factor of training condition (game vs sterile) and within-subject factors of test (pre- vs post-) and testing condition (game vs sterile).

N1, eP3a, and lP3a mean amplitudes were assessed by separate 1-way ANOVAs with a between-subjects factor of group.

For the engagement scale, we first conducted a reliability analysis of the questions for each subscale. The minimum Cronbach's alpha was .78, which allowed us to collapse across questions. Similarly, among average subscale scores the Cronbach's alpha was .83, allowing us to collapse across subscales into a single engagement score. Between-group differences in composite engagement were measured using independent samples *t* tests. Reliability for the subscales of the IMI was also quite good with a minimum Cronbach's alpha of .83, but among the average subscale scores the Cronbach's alpha was −.08, preventing us from collapsing across subscales into a composite IMI score. Between-group differences in IMI subscales were measured using independent samples *t* tests.

Results

No Differences in Learning Between Groups

As shown in Table 1, participants in both groups improved from pretest to posttest, which was confirmed by the main effect of test, $F_{1,38}$=37.92, P<.001, η_p^2=0.50. However, there was no main effect of training condition, $F_{1,38}$<1, and no test by training condition interaction, $F_{1,38}$<1. The main effect of testing condition was not significant, $F_{1,38}$=3.28, P=.08, η_p^2=0.08, although participants scored fewer points per block on average during the game test than during the sterile test (350.55 [SD 46.54] vs 363.99 [SD 38.77], respectively). None of the other interactions was statistically significant, with the largest being the 3-way interaction of test, training condition, and testing condition, $F_{1,38}$=1.47, P=.23, η_p^2=0.04.

Table 1. Means (SDs) for performance variables and electrophysiological variables as a function of group (game, n=19; sterile, n=21).

	Pretest score[a]	Posttest score[a]	N1[b]	eP3a[b]	lp3a[b]
Game, mean (SD)	330.18	385.27	−8.59	7.92	3.47
	(58.32)	(34.07)	(5.76)	(3.90)	(3.82)
Sterile, mean (SD)	322.14	391.50	−9.99	10.30	3.18
	(56.61)	(37.79)	(4.00)	(5.30)	(2.24)

[a]Pretest and posttest scores are in points per block (maximum of 500) and refer to average performance across the two different test-types.

[b]N1, eP3a and lP3a are in µV.

No Differences in Engagement Between Groups

As shown in Table 2, there was no difference between groups on the engagement scale overall, t_{38}=−0.72, P=.48. Similarly, there were no significant differences in the subscales of focused attention, endurability, novelty, or perceived involvement. The difference in the usability subscale was not statistically significant (t_{38}=−1.86, P=.07), and although the difference in the aesthetics subscale was statistically significant (t_{38}=−2.07, P=.04), neither of these differences was significant after correcting for multiple comparisons (Bonferroni correction). On the IMI subscales, Table 3, the only statistically significant difference was in competence (t_{38}=−2.45, P=.02), but this difference was not significant following a correction for multiple comparisons (Bonferroni correction).

Table 2. Means (SDs) for the overall engagement scale and subscales as a function of group (game, n=19; sterile, n=21). The maximum on the engagement and IMI subscales is 7.

Engagement	Overall	FA[a]	US[b]	AS[c]	EN[d]	NO[e]	IN[f]
Game, mean (SD)	4.08 (1.03)	4.32 (1.40)	5.13 (1.23)	3.99 (1.46)	3.96 (1.18)	3.05 (1.61)	4.04 (1.45)
Sterile, mean (SD)	3.87 (0.84)	4.16 (1.02)	4.49 (0.96)	3.12 (1.19)	3.67 (1.13)	3.30 (1.25)	4.48 (1.23)

[a]FA: focused attention.

[b]US: usability.

[c]AS: aesthetics.

[d]EN: endurability.

[e]NO: novelty.

[f]IN: perceived involvement.

Table 3. Means (SDs) for the subscales of the IMI as a function of group (game, n=19; sterile, n=21). The maximum on the engagement and IMI subscales is 7.

IMI subscales	I/E[a]	CM[b]	EF[c]	P/T[d]
Game, mean (SD)	3.68 (1.61)	5.30 (0.96)	4.98 (1.58)	5.18 (1.21)
Sterile, mean (SD)	4.02 (1.19)	4.52 (1.04)	5.37 (1.24)	4.74 (1.35)

[a]I/E: interest/enjoyment.

[b]CM: competence.

[c]EF: effort.

[d]P/T: pressure/tension (reverse-coded so higher numbers mean less pressure).

A Trend for a Difference Between Groups in the eP3a

The left panel of Figure 3 displays grand average ERPs for the game and sterile groups at Fz, Cz, and Pz electrodes, and the N1, eP3a, and lP3a components are indicated. The topographies of the components collapsed across groups are displayed in the right panel. For EEG measures, there was no statistically significant difference in the N1 (t_{37}=−0.88, P=.38), eP3a (t_{37}=1.60, P=.12), or lP3a (t_{37} =.29, P=.77). As shown in Table 1, however, the eP3a effect was in the predicted direction with a moderately large effect size (Cohen's d=.51).

eP3a Is Correlated With Self-Reported Engagement, Controlling for Group

Although the a priori group difference in the eP3a was not supported, we were interested in exploring individual variability in the eP3a and how those differences related to engagement (overall engagement score) and learning (defined as posttest performance, given equivalent baselines). For these exploratory analyses, a step-up series of regression models was tested in which training condition, eP3a, and their interaction were regressed onto performance on the posttest (in points per block) and overall engagement scale scores (on a 1- to 7-point scale). Results of best fitting regressions for each series are shown in Multimedia Appendix 1. The predictions of the regression equations are shown in Figure 4. There were no reliable relationships observed between the eP3a and posttest performance even when controlling for training condition and the interaction. There was, however, a statistically significant negative relationship between the eP3a and self-reported levels of engagement. There was no evidence that this negative relationship changed as a function of group because the training condition by eP3a interaction was not significant. (Note that the relationship between eP3a and engagement was statistically significant even without controlling for group. However, given the difference between groups in the eP3a, the most appropriate analysis is the multivariable regression controlling for group.)

Figure 3. Grand average ERPs for the sterile and game groups (left). Topographies of the N1, eP3a, and lP3a components collapsed across groups (right).

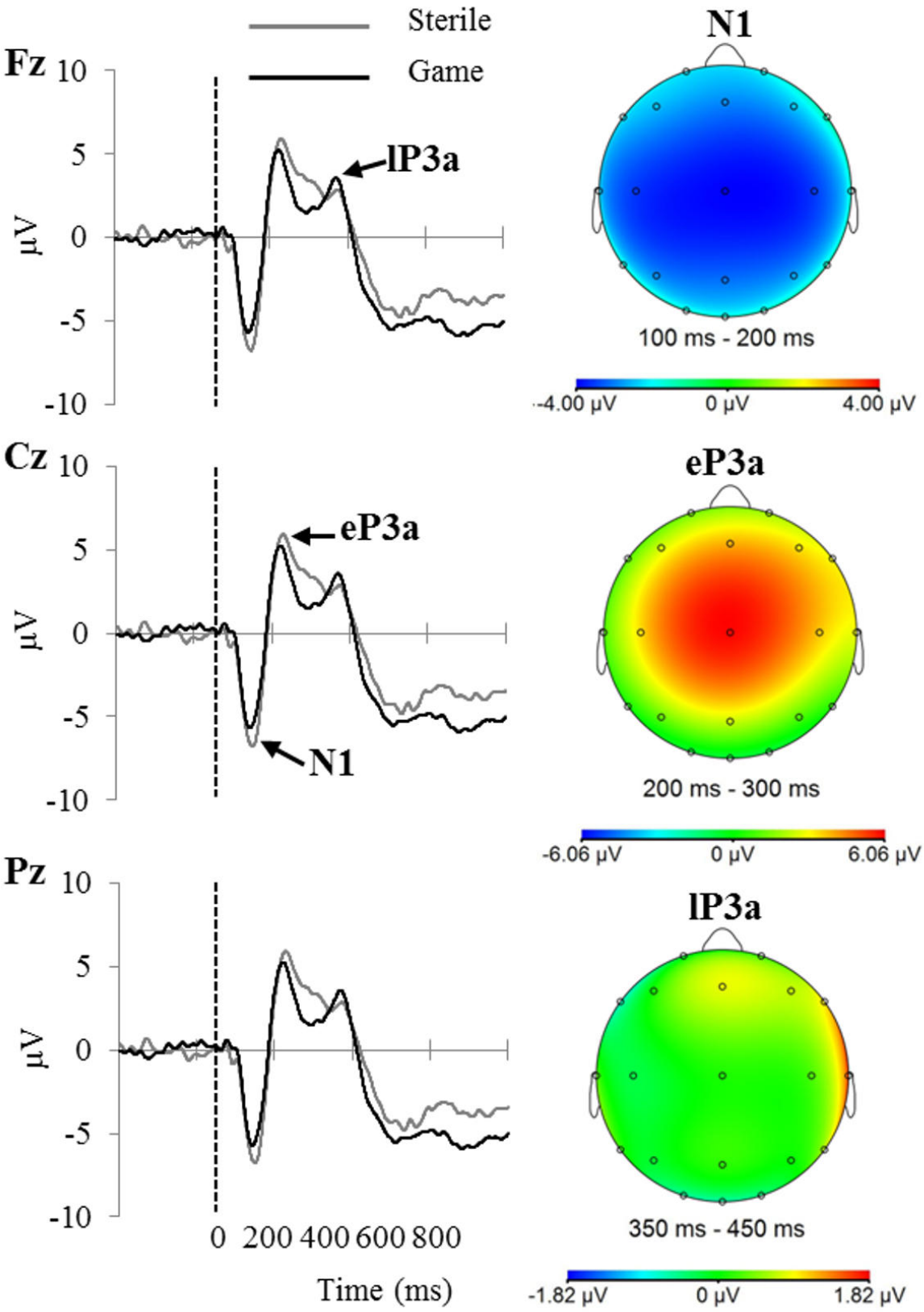
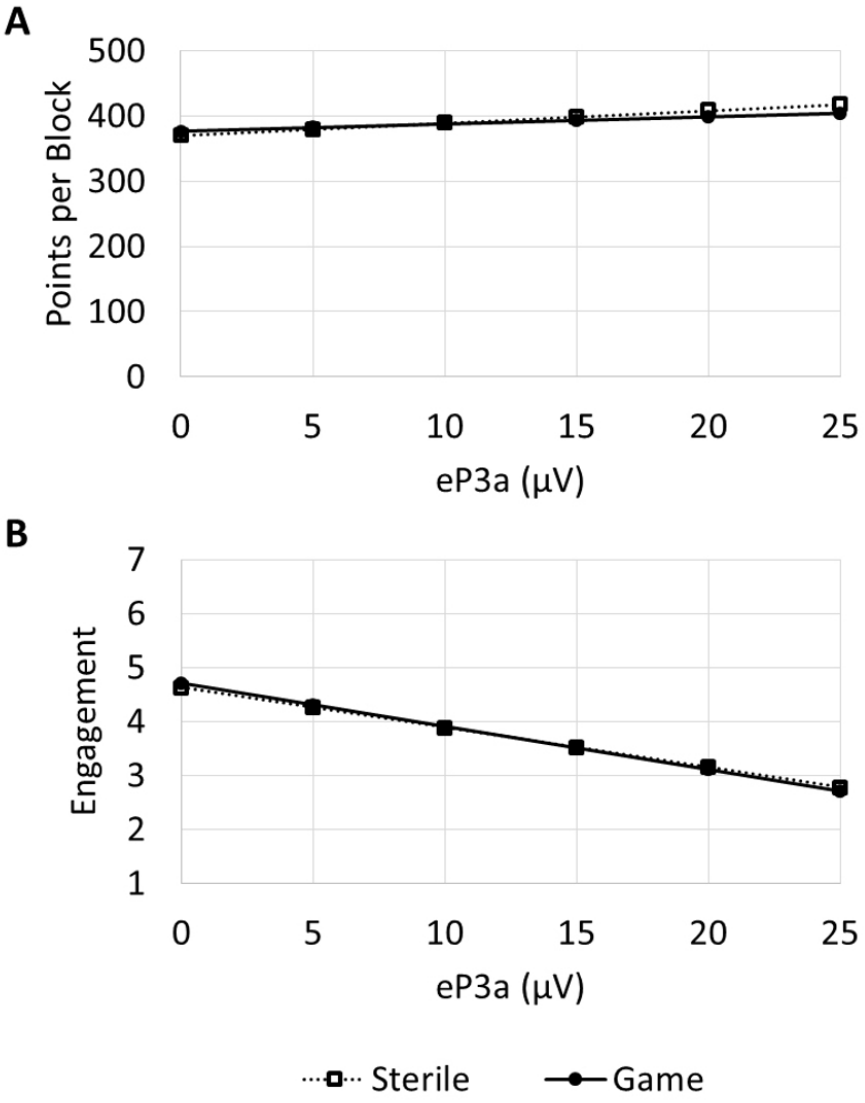


Figure 4. Predicted points per block on the posttest as a function of training condition and eP3a (top). Predicted engagement scores as a function of training condition and eP3a (bottom).



Discussion

Principal Findings

No significant learning or engagement effects were found between the game and sterile groups. Playing in an aesthetically pleasing game environment was not advantageous to learning; both groups showed similar levels of performance on the delayed retention and transfer tests. However, our data provide evidence that attentional reserve decreased proportional to self-reported engagement levels in both game environments. The significant negative relationship of the eP3a to self-reported levels of engagement adds to the theoretical understanding of engagement by providing physiological data showing that engagement elicits increased information processing. These new physiological data suggest increasing engagement is not simply a change in affective state but a change in cognitive processing as well.

Although the learning and engagement effects of Lohse et al [10] were not replicated, there are potential explanations for this lack of replication. First, the use of probe tones in the current experiment meant sounds could not be played in the game condition, which included both background music and action-specific sounds in the previous experiment. Second, in the current experiment participants were seated rather than standing to accommodate the EEG equipment. A limited movement space may have affected engagement during gameplay; participants in the previous experiment had greater freedom to move. A third possibility is that practicing 400 trials in one sitting, as opposed to 200 trials per day for two days in the previous experiment, may have led to boredom, especially if the participant wasn't particularly challenged by the task.

Ultimately, this negative result complicates our previous conclusions about beneficial effects of engagement on learning. However, we cannot consider this study a failure to replicate the work of Lohse et al [10] because of the various experimental differences. The fact that neither the learning effect nor the engagement effect were found in this study may be reassuring, as there may still be a relationship between the two. Specifically, in the previous study gamification enhanced engagement and learning, whereas in the present study gamification failed to

enhance engagement, possibly explaining why learning was not improved. However, the current null results do cast some doubt on the robustness of the previous effect, if not its validity. In order to validate the initial results, we are currently conducting a direct replication to see if the original learning and engagement effects can be obtained in a new sample of participants.

The negative relationship between engagement and the eP3a suggests that engagement causes a fundamental change in information processing and is not just an affective experience. The decreased amplitude of the eP3a in relation to higher self-reported levels of engagement indicates that more attentional resources are being used when players are more engaged in the game. In addition, this relationship was similar in both the game and sterile conditions. These physiological data give us a new insight to engagement. Although engagement is generally discussed with respect to affective consequences [4,9], we have empirically demonstrated a neural correlate of cognitive resources being consumed with increased engagement. The eP3a in response to task-irrelevant tones is an objective, physiological correlate of engagement that could be used to measure engagement in many different populations across many different tasks. Future research may further examine the direct effect of engagement on learning through the eP3a.

Within the context of rehabilitation or other applications of serious games, the eP3a may provide a useful and relatively objective index of engagement. For instance, the eP3a could be a useful source of biofeedback allowing participants to "see" how much attention they have been allocating to their therapeutic tasks or a biomarker for adjusting difficulty, allowing the therapist to dynamically adjust the difficulty of practice to promote long-term learning (for a conceptually related study see Shirzad & Van der Loos [20]). Logistically, the constraints of collecting EEG data to measure engagement might make it a tool better suited for game designers; designers could play-test various game mechanics and measure corresponding changes in eP3a magnitude. As the benefits of using EEG systems increase (less expensive, less time-consuming, and more user friendly), there may be a place for EEG biofeedback in a routine clinical setting. At the moment, however, it is probably more feasible to measure motivation and engagement using validated survey measures. Although there are rehabilitation-specific measures of motivation [21], to our knowledge the only validated measures of engagement come from the human-computer interaction literature [6,7]. Future research should adapt existing or develop new engagement scales specific to rehabilitation.

Limitations

A limitation of the present study is that participants were not as engaged in either condition as they were in the previous experiment [10]. Overall engagement scores were lower, which may have been affected by the changes in the experimental paradigm to accommodate additional EEG measurement (even the simple act of wearing the EEG cap for a prolonged period could have negatively affected engagement). Although we cannot make strong conclusions comparing across experiments, it is likely that structural differences between the two practice conditions (eg, no sounds vs sounds, sitting vs standing, and limited range of motion vs freedom of movement) might explain why engagement scores were generally lower in the current experiment than in our previous research.

Conclusion

Although there is some evidence that performing a complex motor skill in a stimulating game environment increases engagement and learning [10], the present study found no differences in engagement or learning between the game and sterile groups. While further research is needed to better understand the potential effects of engagement on learning, the current findings suggest we can predict individual differences in engagement with the event-related potential component, eP3a. Not only do these results hold theoretical importance because they give more information about the nature of engagement, the results can contribute to real-world solutions for health and rehabilitation research. The eP3a has the potential to become an objective measure of engagement in studies of games for rehabilitation patients adjusting to numerous disabilities.

Acknowledgments

The authors would like to thank Kris De Asis and Bryan Luu for their assistance in programming the game and Mariane Faria Braga Bacelar for her assistance with the collection and processing of the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Details of regression models exploring the relationship between the eP3a, posttest performance, and self-reported engagement.

[PDF File (Adobe PDF File), 25KB - games_v4i1e4_app1.pdf]

References

1. Levac D, Espy D, Fox E, Pradhan S, Deutsch JE. Kinect-ing with clinicians: a knowledge translation resource to support decision making about video game use in rehabilitation. Phys Ther 2015 Mar;95(3):426-440 [FREE Full text] [doi: 10.2522/ptj.20130618] [Medline: 25256741]

2. Donoso Brown EV, Dudgeon B, Gutman K, Moritz CT, McCoy SW. Understanding upper extremity home programs and the use of gaming technology for persons after stroke. Disabil Health J 2015 Oct;8(4):507-513. [doi: 10.1016/j.dhjo.2015.03.007] [Medline: 25953349]
3. Lohse KR, Hilderman CGE, Cheung KL, Tatla S, Van der Loos HFM. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
4. Hunicke R, Leblanc M, Zubek R. MDA: A formal approach to game design and game research. 2004 Presented at: Challenges in Games AI Workshop, Nineteenth National Conference of Artificial Intelligence; July 2004; San Jose, California p. 1-5.
5. Zimmerli L, Jacky M, Lünenburger L, Riener R, Bolliger M. Increasing patient engagement during virtual reality-based motor rehabilitation. Arch Phys Med Rehabil 2013 Sep;94(9):1737-1746. [doi: 10.1016/j.apmr.2013.01.029] [Medline: 23500181]
6. O'Brien H, Toms E. The development and evaluation of a survey to measure user engagement in e-commerce environments. J Am Soc Inf Sci Technol 2010;61:50-69. [doi: 10.1002/asi.v61:1]
7. O'Brien H, Toms E. What is user engagement? A conceptual framework for defining user engagement with technology. J Am Soc Inf Sci Technol 2008;59(6):938-955. [doi: 10.1002/asi.20801]
8. Shirzad N, Van der Loos HFM. Evaluating the user experience of exercising reaching motions with a robot that predicts desired movement difficulty. J Mot Behav 2016;48(1):31-46. [doi: 10.1080/00222895.2015.1035430] [Medline: 25945816]
9. Lohse K, Shirzad N, Verster A, Hodges N, Van der Loos HFM. Video games and rehabilitation: using design principles to enhance engagement in physical therapy. J Neurol Phys Ther 2013 Dec;37(4):166-175. [doi: 10.1097/NPT.0000000000000017] [Medline: 24232363]
10. Lohse K, Boyd L, Hodges N. Engaging environments enhance motor skill learning in a computer gaming task. J Mot Behav 2015 Aug 21:1-11. [doi: 10.1080/00222895.2015.1068158] [Medline: 26296097]
11. Dyke FB, Leiker AM, Grand KF, Godwin MM, Thompson AG, Rietschel JC, et al. The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity. Int J Psychophysiol 2015 Jan;95(1):56-62. [doi: 10.1016/j.ijpsycho.2014.12.008] [Medline: 25528402]
12. Suzuki J, Nittono H, Hori T. Level of interest in video clips modulates event-related potentials to auditory probes. Int J Psychophysiol 2005 Jan;55(1):35-43. [doi: 10.1016/j.ijpsycho.2004.06.001] [Medline: 15598514]
13. Takeda Y, Okuma T, Kimura M, Kurata T, Takenaka T, Iwaki S. Electrophysiological measurement of interest during walking in a simulated environment. Int J Psychophysiol 2014 Sep;93(3):363-370. [doi: 10.1016/j.ijpsycho.2014.05.012] [Medline: 24892726]
14. Miller MW, Rietschel JC, McDonald CG, Hatfield BD. A novel approach to the physiological measurement of mental workload. Int J Psychophysiol 2011 Apr;80(1):75-78. [doi: 10.1016/j.ijpsycho.2011.02.003] [Medline: 21320552]
15. Miller MW, Presacco A, Groman L, Bur S, Rietschel J, Gentili R, et al. The effects of team environment on cerebral cortical processes and attentional reserve. Sport Exercise Perform Psychol 2014;3(1):61-74. [doi: 10.1037/spy0000001]
16. Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol 2001 Apr;112(4):713-719. [Medline: 11275545]
17. Clayson PE, Baldwin SA, Larson MJ. How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. Psychophysiology 2013 Feb;50(2):174-186. [doi: 10.1111/psyp.12001] [Medline: 23216521]
18. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exercise Sport 1989 Mar;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
19. 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]
20. Shirzad N, Van der Loos HFM. Evaluating the user experience of exercising reaching motions with a robot that predicts desired movement difficulty. J Mot Behav 2016;48(1):31-46. [doi: 10.1080/00222895.2015.1035430] [Medline: 25945816]
21. Tatla S, Jarus T, Virji-Babul N, Holsti L. The development of the Pediatric Motivation Scale for rehabilitation. Can J Occup Ther 2015 Apr;82(2):93-105. [Medline: 26281433]

Abbreviations

EEG: electroencephalography
eP3a: early P3a
ERP: event-related potential
ICA: independent component analysis
IMI: Intrinsic Motivation Inventory
lP3a: late P3a

Edited by G Eysenbach; submitted 27.12.15; peer-reviewed by D Levac, M D. Patterson; comments to author 20.01.16; revised version received 25.01.16; accepted 26.01.16; published 21.04.16.

Please cite as:
Leiker AM, Miller M, Brewer L, Nelson M, Siow M, Lohse K
The Relationship Between Engagement and Neurophysiological Measures of Attention in Motion-Controlled Video Games: A Randomized Controlled Trial
JMIR Serious Games 2016;4(1):e4
URL: http://games.jmir.org/2016/1/e4/
doi:10.2196/games.5460
PMID:27103052

Original Paper

A Serious Game for Massive Training and Assessment of French Soldiers Involved in Forward Combat Casualty Care (3D-SC1): Development and Deployment

Pierre Pasquier[1], MD, MSc; Stéphane Mérat[2], MD; Brice Malgras[3], MD, MSc; Ludovic Petit[4], MD; Xavier Queran[5], MD; Christian Bay[6], MD; Mathieu Boutonnet[1], MD, MSc; Patrick Jault[1], MD; Sylvain Ausset[1,7], MD; Yves Auroy[7,8], MD, PhD; Jean Paul Perez[7,8], MD, MSc; Antoine Tesnière[9], MD, PhD; François Pons[7], MD; Alexandre Mignon[9], MD, PhD

[1]Percy Military Teaching Hospital, Department of anesthesiology and intensive care, French Military Health Service, Clamart, France
[2]Marie Lannelongue Surgical Center, Department of anesthesiology and intensive care, Le Plessis Robinson, France
[3]Bégin Military Teaching Hospital, Department of visceral and vascular surgery, French Military Health Service, Paris, France
[4]Military Medical Center, 8ème Régiment Parachutiste d'Infanterie de Marine, French Military Health Service, Castres, France
[5]Military Medical Center, 3ème Régiment Parachutiste d'Infanterie de Marine, French Military Health Service, Carcassonne, France
[6]French Military Health Service Academy, Tactical Care Training Department, French Military Health Service, Paris, France
[7]École du Val-de-Grâce, French Military Health Service Academy, French Military Health Service, Paris, France
[8]Hospital and Research Division, Head Office of French Military Health Service, French Military Health Service, Vincennes, France
[9]iLUMENS, Department of Simulation, Université Sorbonne Paris Cité, Paris, France

Corresponding Author:
Pierre Pasquier, MD, MSc
Percy Military Teaching Hospital
Department of anesthesiology and intensive care
French Military Health Service
101 avenue Henri Barbusse
Clamart, 92141
France
Phone: 33 141466000
Fax: 33 141465319
Email: pasquier9606@me.com

Abstract

Background: The French Military Health Service has standardized its military prehospital care policy in a ''Sauvetage au Combat'' (SC) program (Forward Combat Casualty Care). A major part of the SC training program relies on simulations, which are challenging and costly when dealing with more than 80,000 soldiers. In 2014, the French Military Health Service decided to develop and deploy 3D-SC1, a serious game (SG) intended to train and assess soldiers managing the early steps of SC.

Objectives: The purpose of this paper is to describe the creation and production of 3D-SC1 and to present its deployment.

Methods: A group of 10 experts and the Paris Descartes University Medical Simulation Department spin-off, Medusims, coproduced 3D-SC1. Medusims are virtual medical experiences using 3D real-time videogame technology (creation of an environment and avatars in different scenarios) designed for educational purposes (training and assessment) to simulate medical situations. These virtual situations have been created based on real cases and tested on mannequins by experts. Trainees are asked to manage specific situations according to best practices recommended by SC, and receive a score and a personalized feedback regarding their performance.

Results: The scenario simulated in the SG is an attack on a patrol of 3 soldiers with an improvised explosive device explosion as a result of which one soldier dies, one soldier is slightly stunned, and the third soldier experiences a leg amputation and other injuries. This scenario was first tested with mannequins in military simulation centers, before being transformed into a virtual 3D real-time scenario using a multi-support, multi–operating system platform, Unity. Processes of gamification and scoring were applied, with 2 levels of difficulty. A personalized debriefing was integrated at the end of the simulations. The design and

production of the SG took 9 months. The deployment, performed in 3 months, has reached 84 of 96 (88%) French Army units, with a total of 818 hours of connection in the first 3 months.

Conclusions: The development of 3D-SC1 involved a collaborative platform with interdisciplinary actors from the French Health Service, a university, and videogame industry. Training each French soldier with simulation exercises and mannequins is challenging and costly. Implementation of SGs into the training program could offer a unique opportunity at a lower cost to improve training and subsequently the real-time performance of soldiers when managing combat casualties; ideally, these should be combined with physical simulations.

(JMIR Serious Games 2016;4(1):e5) doi:10.2196/games.5340

KEYWORDS

serious games; forward combat casualty care; care under fire; interdisciplinary collaboration; simulation; medical simulation; virtual simulation; training; education; assessment

Introduction

The survivability of battlefield casualties has recently risen to an unequalled historical level of 90%, compared to 84% in Vietnam and 80% in World War II [1]. Factors most likely related to this improved survivability include technical improvements in body armor and deployment of a comprehensive trauma management system (involving first aid and basic life support on the field and subsequent prompt surgery and critical care). Tactical Combat Casualty Care (TCCC) is now considered as a reference for management of combat casualties at the point of injury. It combines a set of trauma management guidelines designed for use on the battlefield [2,3]. Improved training of soldiers and military caregivers based on the concepts of TCCC plays an important role in the improved survival of combat casualties [4].

In 2007, the French Military Health Service standardized this TCCC concept in its military prehospital care training policy through a specific program entitled ‘‘Sauvetage au Combat’’ (SC, ‘‘forward combat casualty care’’). After delivery of first aid to soldiers in the under fire stage, forward medicalization on the battlefield is one characteristic of the SC. The medical team is sent as close as possible to the casualty at the time of injury. In the SC training program, emphasis is placed on simulations, which are considered a gold standard in team training for improvement of both technical and nontechnical skills, in both civilian and military trauma settings [5-8]. However, the logistics involved in training and testing each French soldier with simulation exercises and mannequins make them challenging and costly. Moreover, there might be important delays between the training period and the actual operations. However, regardless of these delays, knowledge of both adequate procedures and skills has to be maintained.

Computer-based technologies, such as e-learning, massive online open courses, or serious games (SGs), have become increasingly prevalent in education, training, and simulation. SGs are digital simulations similar to video games that are engaging, rewarding, and fun as they simultaneously educate and train [9-10]. SGs use the gamification concept for training applications. Gamification is the application of game-based elements to nongame mechanisms, including education. SGs have drawn much attention over the last decade because they have become more realistic and engaging, owing to technological improvements such as better graphics and new gaming interfaces and gameplays. SGs are also more affordable, owing to their reduced costs of production. Thus, SGs could address a larger population of trainees to increase the frequency of cognitive training, allowing training to occur anytime and anywhere, and can be used to assess retention of procedural skills in a more practical manner and at a lesser cost [11].

Therefore, the French Military Health Service considered in 2014 the development of 3D-SC1, an innovative SG devoted to the training of soldiers for casualty care under fire; 3D-SC1 constitutes the first part of a virtual simulation platform. Combined with the next 3D-SC versions, it will allow training of combat lifesavers (SC2) and nurses and physicians (SC3) for forward combat casualty care applications on the battlefield.

The purpose of this paper is to describe the design and production processes of 3D-SC1 and to discuss its deployment.

Methods

Medusims

Medusims are virtual digital medical simulations using videogame technology to simulate realistic medical situations for training and assessment purposes. They were produced by a French startup, Medusims, created in 2011 [12]. Medusims has already produced 6 SGs in the cardiology area (Staying Alive after cardiac arrest for the general public [www.stayingalive.fr], acute coronary syndrome, atrial fibrillation, and pulmonary embolism), and 2 SGs in the perinatology area (Born to be Alive for the general public [www.borntobealive.fr] and postpartum hemorrhage [12]).

Medusims were used to assess how effective practices are, emphasizing on the importance of having reference material and procedures reflecting real-life scenarios. The trainees used Medusims as an active learning method to familiarize themselves with procedures, without any risk to the patients or casualties. Medusims produced virtual simulations in a 3D studio, using the Unity engine, a multi–operating system and a multiplatform tool allowing trainees to access the digital experiences on personal computers or tablets. Medusims has developed an internal process of in-house production, integrating medicine and technology, to deliver high-quality products, with a production studio panel of game design, human engineering, pedagogic engineering, graphics computing, and 3D animation, working closely with medical and military experts on the topic.

Preparation of 3D-SC1 project

The French Military Health Service designed an advisory board of SC experts, most of them being existing members of the Comité Opérationel d'Enseignement du Sauvetage au Combat (COESC, the Operational Committee for SC teaching). Several steps before the beginning of 3D-SC production were also validated, which were as follows.

- Definition of the trainees involved as first player shooter in the experience (soldiers for SC1, health-qualified soldiers, or combat lifesavers for SC2, physicians and nurses for SC3, to come later). In the first version of the experience, there was only 1 person playing (mono user).
- Definition of training objectives for SC1 (survival positioning, compression bandage, tactical tourniquet, and morphine auto-injector use), SC2 (bleeding shock management including intraosseous access, tactical tourniquet assessment, airway, and respiratory management including pneumothorax decompression and cricothyroidotomy), and SC3 (medical trauma management in a hostile remote environment, with tracheal intubation, sedation, early use of tranexamic acid, and vasopressor agents in case of hemorrhagic shock, digital thoracotomy with chest tube insertion, handheld focused assessment with sonography for trauma, and strategies for crisis resource management) [13-15].
- Design of a scenario involving an improvised explosive device (IED) attack on a patrol of 3 soldiers, followed by a second patrol among which was the trainee, allowing the trainee to test the correct procedural skills execution, in the right order and as quickly as possible.
- Test of the realistic features of this scenario in real simulation with mannequins, in order to observe right and wrong actions delivered.
- Design of 2 levels of difficulty—beginner and advanced.
- Design of a correcting grid for scoring and personalized debriefing.
- Design of a rewarding process (bronze, silver, and gold medal graduation) according to the scoring system, integrating time, and actions delivered.
- Design of a gameplay, interface, and the look and feel of the SG.
- Finally, definition of the different animations to be produced, including using 3D motion capture.

Different iterations were developed between the experts and the studio, which allowed production of a beta version in 8 months. After debugging, a final gold version was delivered after 9 months, and it was successfully submitted to the French Army in January 2015.

Results

Creation of 3D-SC1

A 10-member expert advisory group was formed to identify training priorities for which 3D-SC1 could be appropriately used. All these experts held local or national roles in organization and training of SC programs. Experienced physicians and soldiers in operational units, with significant experience of deployment in combat zones, also participated. A scenario was created to illustrate a real-life–based experience, the explosion of an IED during a reconnaissance mission. The same panel of experts promoted the scientific validation of the scenario according to the SC French guidelines.

In the 3D-SC1 scenario, the explosion of an IED creates 3 casualties: one is dying, one is slightly stunned; and the last one presents with a traumatic amputation of the limb, difficult breathing, and another hemorrhagic injury under the arm. At the beginning of the experience, the care under fire stage is illustrated by hostile fire after the IED explosion, in a stressful, hostile, and noisy ambience (explosion, firearms shooting, shouts of pain), and an austere setting (desert; Figure 1). The trainee has to choose between several tactical options: return fire, determine which casualty is dead or alive, determine which casualty can return fire, or pick up the casualties and run to cover (Figure 2). When no longer under direct enemy fire, the trainee has to deal with different forward combat life-saving procedures from SC programs (Figure 3): tourniquet application, casualty survival positioning, hemostatic dressing, and morphine auto-injector use. In addition, the trainee has to deal with weapons security, management of personal protective equipment, call for a 9-line medical evacuation (MEDEVAC) and a MIST request (which includes mechanism of injury, type of injury, signs, and treatment given).

Figure 1. Tactical options in 3D-SC1.

Figure 2. Pick and run in 3D-SC1.

Figure 3. Combat casualty management with tactical tourniquet application and survival positioning in 3D-SC1.

Development and Production of 3D-SC1

The development of 3D-SC1 was performed in a 3D studio (Figure 4). Particular attention was paid to essential movements and postures, such as tactical tourniquet application, recorded in a motion capture mode to provide to the trainee a high-quality reproduction of the movement in the 3D-SC1 experience. Development included a gamification process where players are challenged to keep on playing to reach the game's objective. Two levels were defined: beginner and advanced modes.

In case of failure in applying the right procedure at the right time and in the right order, an automatic virtual instructor takes control of the experience and simulates the right procedure. In the advanced difficulty level, the trainee has to face a more restrictive time limit, combined with more challenging procedures: an inefficient first tactical tourniquet and delayed MEDEVAC arrival or frequent changes in tactical context.

Most importantly, at the end of the 3D-SC1 simulation, a personalized debriefing is proposed (Figure 5), highlighting good performance achieved in the experience, the procedures for which the trainee has to improve, and the missed procedures for which the automatic virtual instructor had to take the control of the experience to perform the procedure [16-17].

Finally, on the basis of a scoring process, the trainee graduates with either a gold, silver, or bronze medal. In case of fatal outcome in the virtual experience due to an inappropriate response of the trainee, an automatic virtual instructor saves the combat casualties, and the trainee obtains nothing but a training certificate. They are invited to participate in the 3D-SC1 experience once again to improve their knowledge in the application of SC procedures.

Figure 4. Three-dimensional studio-development of 3D-SC1.

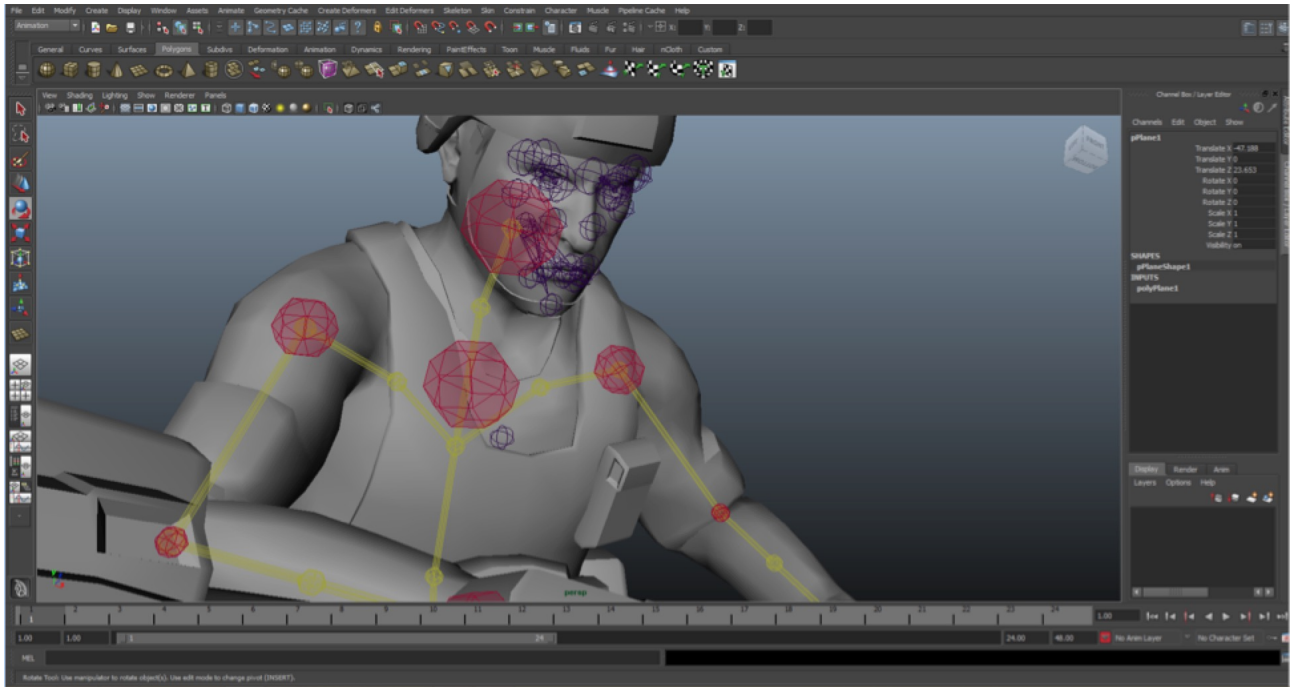
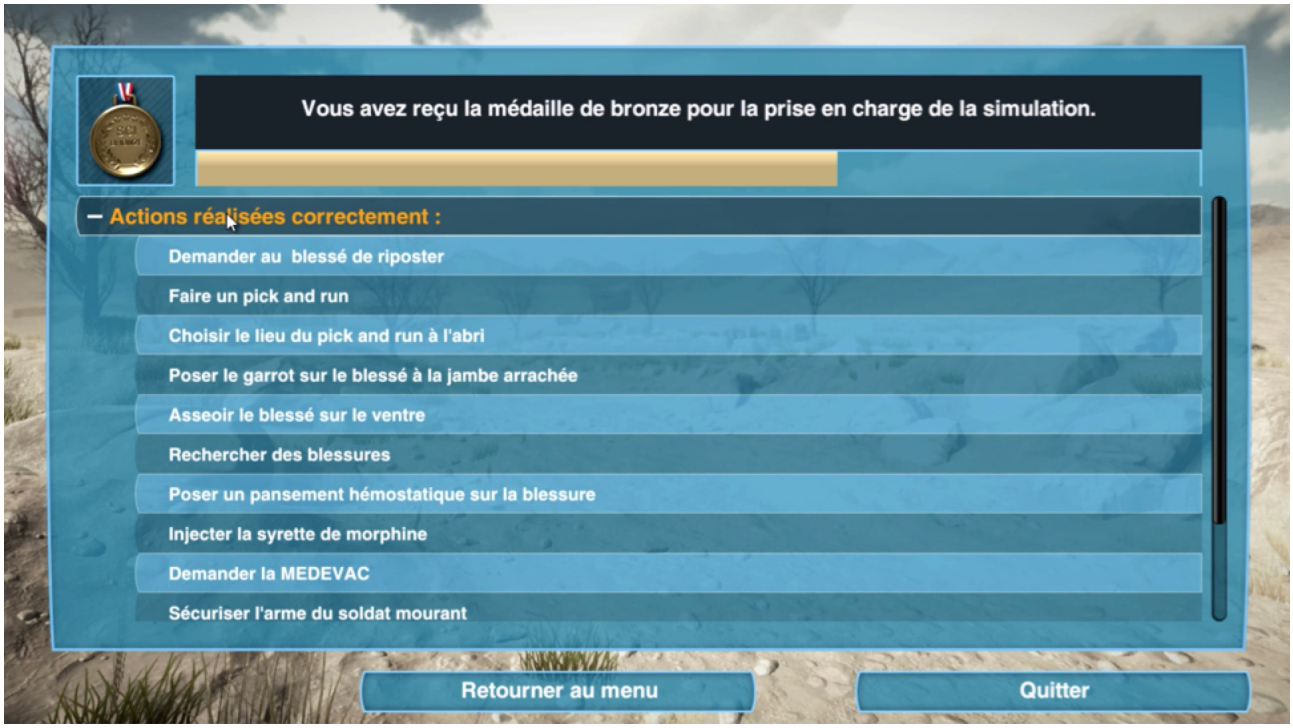


Figure 5. Personalized debriefing at the end of 3D-SC1.



Deployment of 3D-SC1

Because SC programs are now completely incorporated into the predeployment training course of all combatants, the diffusion of 3D-SC1 is planned to reach at least all combatants before their deployment in a theater of operations. Successful deployment of 3D-SC1 was conducted in only 3 months and has already reached 84 of 96 (88%) French Army regiments. Solutions for massive implementation of 3D-SC1 include, for now, personal computers and laptops in a basic classroom, with a total of 818 hours of connection in the first 3 months (in hours, for each training center, mean 19.9, [standard deviation 35.2]). However, 3D-SC1 is also scheduled to be deployed on tablets or smartphones. Such an individual deployment of the SG could let the trainees have their own 3D-SC1 experiences anywhere and anytime. The personalized debriefings are planned to be collected in an extensive databank, allowing global statistical analysis, in order to improve the SC training programs according to the results obtained in different groups of trainees.

Discussion

The development of 3D-SC1, a SG for SC, the French Forward Combat Casualty Care program, has been completed.

Here, we have described an innovative tool for learning and assessing care under fire and its collaborative preparation for design and production by interdisciplinary experts and actors. This led to a fast deployment of the SG, and the integration of a comprehensive collection of data for further assessment of the French SC program.

Limitations

A major limitation of this study is that no objective data regarding the pedagogic value of the SG were assessed. In fact, we have not measured most benefits of simulation, such as that with 3D-SC1, described in this section. However, a randomized observational comparative study is ongoing. How these SGs could be assessed, as new tools for learning and training, is not clear. When applying the concept of 4 levels in evaluating training programs, as described by Kirkpatrick, we believe that studying preventable deaths as an outcome (the fourth level in the Kirkpatrick model, ie, did the change in behavior positively affect the patient) could be really challenging, especially in a combat setting [18]. We could probably only reach level 3 in the Kirkpatrick model by placing the soldier in a physical simulated environment to assess whether learning is transferred from cognition to skills and attitudes. Other points of discussion include that 3D-SC1 concerns a single-center production. However, as highlighted in the description of the process of creation and production, the development of 3D-SC1 involved a collaborative platform with interdisciplinary actors.

Actually, realization of an SG for SC training is based on the concept that an SG has the potential to increase observance and engagement with training programs in military institutions. There is an increasing interest for medical virtual simulation in numerous settings, including trauma, surgery, anesthesia, emergency medicine, women's health care, and even patients' education [19-27].

The benefits of complete virtual training include the ability to create unlimited training scenarios and to repeatedly try and fail in a consequence-free environment. In addition, when training is conducted in a virtual environment, each performed step is considered as essential data. The computer reproduces each movement a trainee makes, and thus, can track and record each and every one of these. Using these data, performance can be automatically quantified to a level that was not achievable ever before [28]. Furthermore, through the processes of scoring and gamification applied in 3D-SC1, the trainee is motivated to improve his personal experience. He also shares his scores with his peers in a competitive and engaging challenge [29-31].

Different virtual simulations have been developed in TCCC programs and several military medical settings, with a very strict adherence to TCCC guidelines depending on the nations' specific policies [8,28,32-38]. Indeed, 3D-SC1 is following the SC guidelines created by the French Military Health Service Academy (École du Val-de-Grâce) in an official 2007 publication, which has been updated every year by the COESC (the Operational Committee for SC teaching) [13].

Besides SGs, live simulation exercises are the accepted "gold standard" for military preparedness before deployment in a combat zone. However, they are costly and time consuming to organize and may be disrupt local services [39,40]. In contrast, 3D-SC1 , used on a standalone computer with a suitable connection for data collection, enables multiple trainees to learn in a real-time, immersive environment, regardless of physical location. Finally, SGs are gradually taking an important place beside physical simulation (high-fidelity manikin simulation combined with casualty simulation moulaged actors), although the incomplete application of strong artificial intelligence, which would allow full mixed-initiative dialogue, can limit their applicability. However, growth in the use of SGs is likely to continue because of their ability to scale inexpensively to large numbers of physically dispersed learners, adapt quickly to prior knowledge and other individual characteristics of learners, and be available anytime and anywhere via a global information infrastructure [41].

A common criticism about SGs is that the dynamic colorful world of a computer game will distract the trainee's attention from the learning process. However, current soldiers grew up with digital media and have developed a much better aptitude to relate a virtual world to reality. However, this criticism must be addressed for 3D-SC1 to become well accepted. It is also necessary to demonstrate its educational value and its clinical effectiveness in real combat casualty situations.

Virtual environments are not a substitute for hands-on training; they can neither simulate the physical elements of tactical response nor provide training in the dexterity of performing procedures on a casualty. To potentiate existing training, 3D-SC1 can be used as an. For example, it can improve knowledge of adequate procedures before participating in complete military exercises or local SC stages, conducted in lifelike conditions, such as MedicHos Médicalisation en milieu Hostile (Medicalization in a Hostile environment) or ExOSAN Exercice Opération Sanitaire (Exercise for Operation Sanitary) exercises [13,23].

Multiplication of 3D-SC1 scenarios and the next production of SC2 and SC3 could lead to the building of a virtual simulation platform for SC training, incorporated as part of a large military medical simulation-training program combining both SGs and physical simulation. Multiplayer large-scale virtual exercises could be included in predeployment training [42,43]. Finally, upcoming technologies such as augmented reality or haptic simulation could add to the definite 3D-SC program development [34,43,44].

Conclusions

3D-SC1 is a new SG module, dedicated to the cognitive training of soldiers for forward combat casualty care, during the care under fire stage. The development of 3D-SC1 involved a collaborative platform with interdisciplinary actors, including forward combat casualty care experts (the French SC), programmers, and game designers. The creation and production of 3D-SC1 outlines the applicability and acceptability of using

virtual environments for training in SC procedures in dedicated scenarios. This type of simulation could easily be adapted to address different training needs of SC1 (such as urban combat or convoy attacks), SC2, and SC3 (such as combat casualty management during the tactical field care phase and preparation for casualty evacuation). 3D-SC1, actually designed for the French Army, could also find many applications to train people involved in the management of trauma casualties in other settings such as tactical emergency care, disaster and crisis management, or mass casualty events. Further possibilities related to the application of 3D-SC1 are unlimited. The observational comparative trial addressing its educational value is ongoing, and its integration in a mandatory official military degree validation is the next step [23,45-51].

Conflicts of Interest

None declared.

References

1. Holcomb JB, Stansbury LG, Champion HR, Wade C, Bellamy RF. Understanding combat casualty care statistics. J Trauma 2006 Feb;60. Medline 2006:397-401.
2. Butler F. Tactical combat casualty care: combining good medicine with good tactics. J Trauma 2003 May;54(5 Suppl):S2-S3. [doi: 10.1097/01.TA.0000064504.55800.73] [Medline: 12768093]
3. Butler FK, Blackbourne LH. Battlefield trauma care then and now: a decade of Tactical Combat Casualty Care. J Trauma Acute Care Surg 2012 Dec;73(6 Suppl 5):S395-S402. [doi: 10.1097/TA.0b013e3182754850] [Medline: 23192061]
4. Kotwal RS, Montgomery HR, Kotwal BM, Champion HR, Butler FK, Mabry RL, et al. Eliminating preventable death on the battlefield. Arch Surg 2011 Dec;146(12):1350-1358. [doi: 10.1001/archsurg.2011.213] [Medline: 21844425]
5. Thorson CM, Dubose JJ, Rhee P, Knuth TE, Dorlac WC, Bailey JA, Van Haren Robert M, et al. Military trauma training at civilian centers: a decade of advancements. J Trauma Acute Care Surg 2012 Dec;73(6 Suppl 5):S483-S489. [doi: 10.1097/TA.0b013e31827546fb] [Medline: 23192074]
6. Sohn VY, Runser LA, Puntel RA, Sebesta JA, Beekley AC, Theis JL, et al. Training physicians for combat casualty care on the modern battlefield. J Surg Educ 2007;64(4):199-203. [doi: 10.1016/j.jsurg.2007.05.007] [Medline: 17706571]
7. Butler FK, Smith DJ, Carmona RH. Implementing and preserving the advances in combat casualty care from Iraq and Afghanistan throughout the US Military. J Trauma Acute Care Surg 2015 Aug;79(2):321-326. [doi: 10.1097/TA.0000000000000745] [Medline: 26218704]
8. Mercer SJ, Howell M, Simpson R. Simulation training for the frontline--realistic preparation for role 1 doctors. J R Army Med Corps 2010 Jun;156(2):87-89. [Medline: 20648943]
9. Stokes B. Video games have changed: time to consider ''serious games''. The Development Education Journal 2005;11:108.
10. Zyda M. From simulation to virtual reality to games. IEEE Computer 2005;38:25-32.
11. Graafland M, Dankbaar M, Mert A, Lagro J, De WL, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014 Nov;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
12. Mignon A. Medusims. Producing Serious Games for Medical Education and Assessment URL: http://medusims.com/ [accessed 2015-11-12] [WebCite Cache ID 6czn42zcQ]
13. Pasquier P, Dubost C, Boutonnet M, Chrisment A, Villevieille T, Batjom E, et al. Predeployment training for forward medicalisation in a combat zone: the specific policy of the French Military Health Service. Injury 2014 Sep;45(9):1307-1311. [doi: 10.1016/j.injury.2014.05.037] [Medline: 24952973]
14. Brooks AJ, Price V, Simms M. FAST on operational military deployment. Emerg Med J 2005 Apr;22(4):263-265 [FREE Full text] [doi: 10.1136/emj.2004.014308] [Medline: 15788825]
15. Midwinter MJ, Mercer S, Lambert AW, de RM. Making difficult decisions in major military trauma: a crew resource management perspective. J R Army Med Corps 2011 Sep;157(3 Suppl 1):S299-S304. [Medline: 22049811]
16. Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. Simul Healthc 2007;2(2):115-125. [doi: 10.1097/SIH.0b013e3180315539] [Medline: 19088616]
17. Paige JT, Arora S, Fernandez G, Seymour N. Debriefing 101: training faculty to promote learning in simulation-based training. Am J Surg 2015 Jan;209(1):126-131. [doi: 10.1016/j.amjsurg.2014.05.034] [Medline: 25497438]
18. Kirkpatrick D, Kirkpatrick J. Evaluating training programs: the four levels. San Francisco, CA: Berrett-Koehler; 2006.
19. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002 Oct;236(4):458-63; discussion 463. [doi: 10.1097/01.SLA.0000028969.51489.B4] [Medline: 12368674]
20. Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004 Feb;91(2):146-150. [doi: 10.1002/bjs.4407] [Medline: 14760660]
21. McLaughlin S, Fitch MT, Goyal DG, Hayden E, Kauh CY, Laack TA, SAEM Technology in Medical Education Committeethe Simulation Interest Group. Simulation in graduate medical education 2008: a review for emergency medicine.

Acad Emerg Med 2008 Nov;15(11):1117-1129 [FREE Full text] [doi: 10.1111/j.1553-2712.2008.00188.x] [Medline: 18638028]
22. Cumin D, Weller JM, Henderson K, Merry AF. Standards for simulation in anaesthesia: creating confidence in the tools. Br J Anaesth 2010 Jul;105(1):45-51 [FREE Full text] [doi: 10.1093/bja/aeq095] [Medline: 20507857]
23. Cohen D, Sevdalis N, Taylor D, Kerr K, Heys M, Willett K, et al. Emergency preparedness in the 21st century: training and preparation modules in virtual environments. Resuscitation 2013 Jan;84(1):78-84. [doi: 10.1016/j.resuscitation.2012.05.014] [Medline: 22659598]
24. de Wit-Zuurendonk L D, Oei SG. Serious gaming in women's health care. BJOG 2011 Nov;118 Suppl 3:17-21 [FREE Full text] [doi: 10.1111/j.1471-0528.2011.03176.x] [Medline: 22039888]
25. Thompson D, Baranowski T, Buday R. Conceptual model for the design of a serious video game promoting self-management among youth with type 1 diabetes. J Diabetes Sci Technol 2010 May;4(3):744-749 [FREE Full text] [Medline: 20513343]
26. Inglés-Camats G, Presno-Rivas M, Antonijoan M, Garcia-Panella O, Forrest T. Yummy tricks: a serious game for learning healthy eating habits. Stud Health Technol Inform . Medline 2012;172:185-190.
27. 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]
28. Matt RE, Bowyer MW. Simulation for trauma and combat casualty care. Minim Invasive Ther Allied Technol 2005;14(4):224-234. [doi: 10.1080/13645700500243703] [Medline: 16754168]
29. Nevin CR, Westfall AO, Rodriguez JM, Dempsey DM, Cherrington A, Roy B, et al. Gamification as a tool for enhancing graduate medical education. Postgrad Med J 2014 Dec;90(1070):685-693 [FREE Full text] [doi: 10.1136/postgradmedj-2013-132486] [Medline: 25352673]
30. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
31. Lin DT, Park J, Liebert CA, Lau JN. Validity evidence for Surgical Improvement of Clinical Knowledge Ops: a novel gaming platform to assess surgical decision making. Am J Surg 2015 Jan;209(1):79-85. [doi: 10.1016/j.amjsurg.2014.08.033] [Medline: 25454955]
32. Hemman EA. Improving combat medic learning using a personal computer-based virtual training simulator. Mil Med 2005 Sep;170(9):723-727. [Medline: 16261973]
33. Walker RB, Underwood PK, Bernhagen M, Markin N, Boedeker BH. Virtual intubation training at a remote military site. Stud Health Technol Inform 2012;173:540-542. [Medline: 22357054]
34. Gao K, Wiederhold BK, Kong L, Wiederhold MD. Methodology case study of the application of haptics to combat medic training programs. Stud Health Technol Inform 2013;191:53-57. [Medline: 23792842]
35. Friedl KE, O'Neil HF. Designing and using computer simulations in medical education and training: an introduction. Mil Med 2013 Oct;178(10 Suppl):1-6. [doi: 10.7205/MILMED-D-13-00209] [Medline: 24084299]
36. Laporta AJ, Hoang T, Moloff A, Tieman M, Schwartz BD, Slack S, et al. From trauma in austere environments to combat or medical school: how blended hyper-realism in the real and virtual worlds can better prepare surgeons. Stud Health Technol Inform 2014;196:233-237. [Medline: 24732513]
37. Magee H. Use of Advanced Technologies and New Procedures in Medical Field Operations: RTO-MP-HFM-182. A New Era in Medical Training through Simulation-Based Training Systems URL: https://www.cso.nato.int/pubs/rdp.asp?RDP=RTO-MP-HFM-182 [accessed 2015-11-12] [WebCite Cache ID 6cznNqQxe]
38. Sotomayor TM. Teaching tactical combat casualty care using the TC3 sim game-based simulation: a study to measure training effectiveness. Stud Health Technol Inform 2010;154:176-179. [Medline: 20543293]
39. Patel V, Lee H, Taylor D, Aggarwal R, Kinross J, Darzi A. Virtual worlds are an innovative tool for medical device training in a simulated environment. Stud Health Technol Inform 2012;173:338-343. [Medline: 22357014]
40. 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;12(1):e1 [FREE Full text] [doi: 10.2196/jmir.1337] [Medline: 20097652]
41. Fletcher JD, Wind AP. Cost considerations in using simulations for medical training. Mil Med 2013 Oct;178(10 Suppl):37-46. [doi: 10.7205/MILMED-D-13-00258] [Medline: 24084304]
42. Arora S, Cox C, Davies S, Kassab E, Mahoney P, Sharma E, et al. Towards the next frontier for simulation-based training: full-hospital simulation across the entire patient pathway. Ann Surg 2014 Aug;260(2):252-258. [doi: 10.1097/SLA.0000000000000305] [Medline: 24263325]
43. Goolsby C, Vest R, Goodwin T. New Wide Area Virtual Environment (WAVE) medical education. Mil Med 2014 Jan;179(1):38-41. [doi: 10.7205/MILMED-D-13-00310] [Medline: 24402983]
44. Zhu E, Hadadgar A, Masiello I, Zary N. Augmented reality in healthcare education: an integrative review. PeerJ 2014;2:e469 [FREE Full text] [doi: 10.7717/peerj.469] [Medline: 25071992]
45. Gates JD, Arabian S, Biddinger P, Blansfield J, Burke P, Chung S, et al. The initial response to the Boston marathon bombing: lessons learned to prepare for the next disaster. Ann Surg 2014 Dec;260(6):960-966. [doi: 10.1097/SLA.0000000000000914] [Medline: 25386862]
46. Biddinger PD, Baggish A, Harrington L, d'Hemecourt P, Hooley J, Jones J, et al. Be prepared--the Boston Marathon and mass-casualty events. N Engl J Med 2013 May 23;368(21):1958-1960. [doi: 10.1056/NEJMp1305480] [Medline: 23635020]

47. Wall PL, Welander JD, Smith HL, Buising CM, Sahr SM. What do the people who transport trauma patients know about tourniquets? J Trauma Acute Care Surg 2014 Nov;77(5):734-742. [doi: 10.1097/TA.0000000000000433] [Medline: 25494426]
48. Passos E, Dingley B, Smith A, Engels PT, Ball CG, Faidi S, Canadian Trauma Trials Collaborative. Tourniquet use for peripheral vascular injuries in the civilian setting. Injury 2014 Mar;45(3):573-577. [doi: 10.1016/j.injury.2013.11.031] [Medline: 24360743]
49. Jenkins D, Stubbs J, Williams S, Berns K, Zielinski M, Strandenes G, et al. Implementation and execution of civilian remote damage control resuscitation programs. Shock 2014 May;41 Suppl 1:84-89. [doi: 10.1097/SHK.0000000000000133] [Medline: 24662783]
50. Galante JM, Smith CA, Sena MJ, Scherer LA, Tharratt RS. Identification of barriers to adaptation of battlefield technologies into civilian trauma in California. Mil Med 2013 Nov;178(11):1227-1230. [doi: 10.7205/MILMED-D-13-00127] [Medline: 24183771]
51. Jacobs L, Burns KJ. The Hartford Consensus to improve survivability in mass casualty events: Process to policy. Am J Disaster Med 2014 Jan;9(1):67-71. [doi: 10.5055/ajdm.2014.0143] [Medline: 24715646]

Abbreviations

COESC: Comité Opérationel d'Enseignement du Sauvetage au Combat
MEDEVAC: medical evacuation
SC: Sauvetage au Combat
TCCC: Tactical Combat Casualty Care

Edited by G Eysenbach; submitted 12.11.15; peer-reviewed by M Schijven, I Masiello; comments to author 07.02.16; revised version received 20.02.16; accepted 20.02.16; published 18.05.16.

Please cite as:
Pasquier P, Mérat S, Malgras B, Petit L, Queran X, Bay C, Boutonnet M, Jault P, Ausset S, Auroy Y, Perez JP, Tesnière A, Pons F, Mignon A
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;4(1):e5
URL: http://games.jmir.org/2016/1/e5/
doi:10.2196/games.5340
PMID:27194369

Original Paper

Crave-Out: A Distraction/Motivation Mobile Game to Assist in Smoking Cessation

Kathryn L DeLaughter[1,2], MA; Rajani S Sadasivam[2,3], PhD; Ariana Kamberi[3], MBA; Thomas M English[3], PhD; Greg L Seward[4], MSHCA; S Wayne Chan[3], MS; Julie E Volkman[3,5], PhD; Daniel J Amante[3], MPH; Thomas K Houston[2,3], MD MPH

[1]University of Massachusetts Medical School, Worcester, MA, United States

[2]CHOIR ENRM VAMC, Bedford, MA, United States

[3]University of Massachusetts Medical School, Quantitative Health Sciences, Worcester, MA, United States

[4]University of Massachusetts Medical School, Psychiatry, Worcester, MA, United States

[5]Bryant University, Department of Health Communication, Smithfield, RI, United States

Corresponding Author:
Kathryn L DeLaughter, MA
CHOIR ENRM VAMC
200 Springs Rd.
Bedford, MA,
United States
Phone: 1 781 687 2559
Fax: 1 781 687 7831
Email: kathryn.delaughter@va.gov

Abstract

Background: Smoking is still the number one preventable cause of death. Cravings—an intense desire or longing for a cigarette—are a major contributor to quit attempt failure. New tools to help smokers' manage their cravings are needed.

Objective: To present a case study of the development process and testing of a distraction/motivation game (Crave-Out) to help manage cravings.

Methods: We used a phased approach: in Phase 1 (alpha testing), we tested and refined the game concept, using a Web-based prototype. In Phase 2 (beta testing), we evaluated the distraction/motivation potential of the mobile game prototype, using a prepost design. After varying duration of abstinence, smokers completed the Questionnaire of Smoking Urge-Brief (QSU-Brief) measurement before and after playing Crave-Out. Paired *t* tests were used to compare pregame and postgame QSU-Brief levels. To test dissemination potential, we released the game on the Apple iTunes App Store and tracked downloads between December 22, 2011, and May 5, 2014.

Results: Our concept refinement resulted in a multilevel, pattern memory challenge game, with each level increasing in difficulty. Smokers could play the game as long as they wanted. At the end of each level, smokers were provided clear goals for the next level and rewards (positive reinforcement using motivational tokens that represented a benefit of quitting smoking). Negative reinforcement was removed in alpha testing as smokers felt it reminded them of smoking. Measurement of QSU-Brief (N=30) resulted in a pregame mean of 3.24 (SD 1.65) and postgame mean of 2.99 (SD 1.40) with an overall decrease of 0.25 in cravings (not statistically significant). In a subset analysis, the QSU-Brief decrease was significant for smokers abstinent for more than 48 hours (N=5) with a pregame mean of 2.84 (SD 1.16) and a postgame mean of 2.0 (SD 0.94; change=0.84; P =.03). Between December 22, 2011, and May 29, 2014, the game was downloaded 3372 times from the App-Store, with 1526 smokers visiting the online resource www.decide2quit.org linked to the game.

Conclusions: Overall, playing the game resulted in small, but nonsignificant decreases in cravings, with changes greater for those had already quit for more than 48 hours. Lessons learned can inform further development. Future research could incorporate mHealth games in multicomponent cessation interventions.

Trial Registration: Clinicaltrials.gov NCT00797628; https://clinicaltrials.gov/ct2/show/NCT00797628 (Archived by WebCite at http://www.webcitation.org/6hbJr6LWG)

(JMIR Serious Games 2016;4(1):e3) doi:10.2196/games.4566

KEYWORDS

smoking cessation; Internet; secondary prevention; health behavior

Introduction

Smoking is the number one preventable cause of death [1]. In the United States alone, each year over 480,000 deaths are due to smoking, including from second hand smoke [2]. While the majority of smokers want and have tried to quit, quitting for good is challenging. After quit attempts, relapse rates are as high as 90% [3,4]. While combination FDA-approved pharmacotherapy and behavioral interventions can assist during quit attempts, cravings, an intense desire or longing for a cigarette [5], are a major contributor to quit attempt failure [6]. Most smokers report experiencing cravings or urges to smoke while quitting [7]. Because cravings play a major role in quit attempt failure [8], additional tools to help smokers manage cravings are needed.

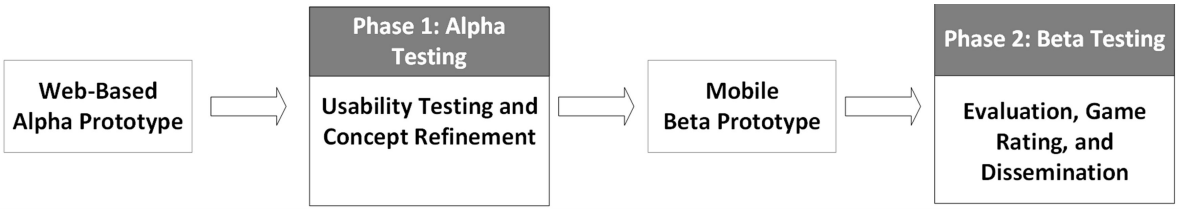
Behavioral techniques such as substitution, relaxation, and distraction have been used to help smokers manage cravings [9]. Cravings can occur at any time and are often triggered by a cue, which could be anything from persons, places, things, or mood states. Thus, craving management tools should be readily available at the point of need [10]. The proliferation of mobile phones, and the increasing number of users who play games on their mobile phones presents a new opportunity to design point-of-need relapse prevention tools [11]. Games on a mobile phone can be engaging and distracting, and smokers could play them until their cravings subside. While any mobile game may help distract smokers, a game specifically designed for cravings could also remind smokers of reasons for quitting and for remaining quit. Developing games for health is challenging as they should be educational, incorporate evidenced-based behavioral strategies, and be engaging and fun [12]. Can a fun, challenging distraction/motivation game be developed to help smokers manage cravings?

We developed and pilot tested a mobile phone game—Crave-Out—for distracting smokers during a craving and motivating them by reinforcing the benefits of quitting. Consistent with the goals of a pilot study [13-17], this small-scale study was designed to assess the feasibility of the game development process, evaluate users' experience in game play, and measure changes in urges to smoke, an important targeted behavioral process. As recommended by game developers [18-21], we developed Crave-Out by using a mixed-methods approach. This study collected qualitative and quantitative data and included two phases (alpha and beta testing; Figure 1):

- Phase 1: Alpha testing (usability testing and concept refinement) encompassed the predevelopment phase in which the game concept was refined using a prototype. Game concepts are crucial to the success of the game; they determines what type of a game it is, how much fun it is, and how much it supports the goals of the game [22]. Prototypes are a recommended approach for concept testing because they collect the thoughts of the users by providing a window into how they interact with the game [22].
- Phase 2: During beta testing (evaluation and dissemination), we developed a mobile version of the game based on the refined concept and evaluated it using a prepost change in cravings assessment. We also tested whether smokers were satisfied with the game and thought it was fun. We also tested the potential of disseminating the game on the Apple iTunes App Store [23], Apple's official repository for downloading apps for the iPhone and iPad.

Since the methods of Phase 2 depended on the results of Phase 1, to improve readability, we first describe the methods and results of Phase 1, followed by the methods and results of Phase 2. Finally, we summarized the main findings of these two studies in the discussion. Our paper demonstrates how an iterative, user-driven pilot testing approach can be used to develop a game. Our protocol was approved by the University of Massachusetts Medical School Institutional Review Board.

Figure 1. Game development phases.



Phase 1: Alpha Testing (Usability Testing and Concept Refinement)

To refine the distraction/motivation concept, we tested the alpha version of our game. The results of concept refinement informed the development of our beta version. We describe the methods and results of our alpha testing below.

Phase 1: Methods

To develop our prototype for concept refinement, we adapted an existing Web-based game that required clicking on pictures to win points. Our intent was to create a distraction tool. On the basis of classical conditioning, including the counterpurpose and relapse prevention model, we redesigned the existing Web-based game to reinforce the benefits of quitting by having smokers click on good things associated with quitting or negative things associated with smoking. We enlisted the aid of students to create artwork, and over about 4 1-hour meetings created images to use for the game. We then enlisted an

additional high school programmer to develop the game on the Web-based platform. The development of the game took approximately 30 hours to complete. During game play, participants started the game by selecting whether during play they wanted to click on "bad" things associated with smoking (eg, cigarettes and yellow teeth), or "good" things associated with quitting (eg, money, health, and time with family; see Multimedia Appendix A). In each version, there was reinforcement for clicking: If a good thing was clicked, the smoker received positive feedback (a green tick mark and a pleasant sound). Similarly, if a bad thing was clicked, the smoker received negative feedback (display of a red cross mark and a harsh sound).

We tested our concept with current or prior smokers aged 19 years or older recruited using flyers, from the University of Massachusetts Medical School campus and UMass Medical Memorial Center. Flyers were distributed across the campus and a coordinated effort with Department of Preventive Medicine was made to have the flyers available in waiting areas and specifically on posted boards at a connected treatment center where counseling for tobacco and other drugs of use was held. In addition, word of mouth was utilized by contact with tobacco treatment specialists on the staff at the University of Massachusetts Medical School. Participants were asked to report to the usability laboratory to complete participation. We used a think aloud protocol [24-26]: while the participants were playing the game, we asked them to vocalize their thoughts, feelings, and opinions about the game. The think aloud protocol provides information about how a user approaches the interface and what they have in their minds when utilizing the interface. We conducted usability testing with 5 users to test our game per existing recommendations regarding think aloud protocol testing [27], which indicate that the majority of issues can be identified with 4±1 users, each with progressively diminishing returns. After playing the game, participants answered a survey and provided open-ended comments. There are several approaches for evaluating qualitative data, including thematic analysis, narrative summary, and grounded theory [28]. Because one of primary goals is to describe the processes of playing the game, we used a combination of thematic summary and narrative summary. This involved an open coding process (without a predefined set of codes) to develop themes.

Phase 1: Results

Prototype Testing

Out of the 5 Phase 1 participants, 3 were between the ages of 19 and 24, with a median number of cigarettes per day of 5. All 5 participants had tried to quit and relapsed. Smokers made comments on the following themes:

- Engagement: Participants indicated that the game was fun and distracting. One participant commented that he "got into" the game, while another participant commented that he was "curious" about the game. Participants also felt that the game should be expanded with additional scenarios to make it more interesting.
- Positive reinforcements: All the participants indicated that the positive reinforcement was beneficial. Example comments included "the game gave reasons but did not necessarily push me" and the game "gave me motivation to quit." A third participant said that he was considering quitting after playing the game.
- Negative reinforcement: Participant disliked the images of cigarettes and other negative stimuli with the accompanying warning sounds. The participants observed that looking at the bad things associated with smoking, while informational, made them think about smoking."

In the follow-up survey, all the participants strongly agreed or agreed that the game would help them think about reasons to quit smoking, 4 out of 5 participants agreed that they would recommend the game, and 3 out of 5 strongly agreed or agreed that they would play the game again. Participants disagreed on whether the game would distract them from cravings, but all 5 agreed that it would help them think about reasons to quit smoking.

Concept Refinement

We first removed negative reinforcements and designed the game to present smokers a motivational token highlighting a benefit of staying abstinent after each level. We then conceptualized the following functions:

1. Multilevel game with a Crave-Out function: The game was designed with multiple levels of game play. The initial levels were simpler, followed by levels of increasing difficulty. Instead of the traditional end game function, we designed the "Crave-Out" function that allows smokers to choose to end the game after their craving has reduced. The game will not complete until the Crave-Out button is pressed.
2. Pattern memory challenge game: We redesigned the game as a pattern memory challenge game fashioned like the old "Simon" game, in which players must rely on memory of the order of objects, which in this instance was specific fruit to choose or "catch" with a virtual bucket at the bottom of the screen. We chose fruit and the virtual bucket for their simplicity and because they were not related in any way to common triggers to smoking. We chose an unrelated concept for our primary game construct, based on findings from Phase 1. To prevent triggering thoughts of smoking, the fruits-in-the-bucket was chosen as a neutral construct for the pattern recognition game.
3. Only positive motivational reinforcements: In the new version, the participant during game play uses a virtual bucket to move back and forth and catch the "good" objects (fruit) as they fell downward from the top of the screen. (See Multimedia Appendix B) The game started simple and then progressively became more difficult. In the first level, the requirement was to catch one "good" object (one of the five fruits) that is named before the level starts. If the participant avoided the other falling fruit and caught the correct fruit, the level was completed. As the level increased, the number of good objects to catch in sequence also increased, and the speed at which the good object fell also increased. After each level, a motivational token that represents a benefit of quitting was awarded to the participant and was continuously displayed in the form of a colorful sticker or stamp on the upper section of the screen

until the participant clicked on the link to go to the next level. Each sticker or stamp included a visual representation of the benefits of quitting and reinforced the many reasons to not smoke. This token appeared in the top center of the screen (eg, a colorful $ sign representing the money saved by not smoking) once the fruit for each level was "caught." Additional tokens included more time with family and friends, whiter teeth, better health, more leisure time, and a cleaner home (Appendix A). Following each completed level, a feedback screen was displayed to the smokers. This included a summary of the level completed and their total score, as well as all the motivational tokens that they received during game play. On each feedback screen, there were two buttons at the bottom: "Crave-Out" (used to indicate that one has finished playing) and "Decide2quit" (links to an evidence-based Web-assisted tobacco intervention: www.Decide2quit.org). The Crave-Out button took the player to a summary screen that contained total time played and both current and running total, which enabled them to see the "reduced" amount of time for which they had been distracted from cigarettes (Appendix B).

4. Referral to other cessation tools: We also provided a link to the Decide2Quit.org Web-assisted tobacco intervention from the summary page [29,30]. The Decide2Quit link allowed them to register with the Decide2Quit.org Web site. Decide2Quit.org is an evidence-based Web-assisted tobacco intervention that includes information about quitting smoking; secure asynchronous messaging with a certified tobacco treatment specialist; an online support group; and a motivational, pushed-email, tailored messaging system. Decide2Quit.org is designed for both current and former smokers.

Phase 2: Beta Testing Pil ot Study (Evaluation of Cravings Reduction and Dissemination)

As noted, on the basis of the results of Phase 1, we developed a mobile version of the game by using the Unity3D game development framework [18]. To develop the game, we first developed the 3D artifacts of the various objects (ie, fruits, bucket, and background scenery items). Once these were developed, we had to program the artifacts to respond to various events or interactions of the game player (eg, the randomness of the different fruits, their speeds and frequencies of recurrences and their tumbling effects, the movement of the bucket from side to side and catching of the fruits in response to the game players' actions, and the appearance of the reward stickers). Another programming action was to end the level if all the fruits were caught in the right sequence. The developed game was then exported to multiple platforms, including the iPhone and iPad mobile platform, for evaluation using a prepost design. We describe the methods and results of our beta testing below.

Phase 2: Methods

Study Design

Our Phase 2 pilot study employed a mixed-methods (qualitative and quantitative) convergent prepost design including measurement of cravings before and after playing the game [31]. Consistent with the goals of a pilot study, our study was designed to provide (1) data on intervention development feasibility and (2) estimates of effect sizes for key measures (cravings) related to smoking cessation. Importantly, our study was designed to provide estimates of effect on cravings to guide future studies, and was not designed to have definitive results on this measure [13-17]. Although we do conduct statistical tests, our results are designed to inform more definitive research and should be carefully considered within the context of a pilot.

After providing consent, participants completed a pregame questionnaire to assess their cravings and then were provided the game to play. As in Phase 1, we again used think aloud protocol testing, and recorded each session of game play. Once they finished playing the game for 10 minutes, participants then completed the same questionnaire to assess their cravings. Participants also completed a brief survey about the game after they played it. This survey consisted of questions related to how they liked the game and how challenging they felt it was; it included several open-ended questions. Again, as in Phase 1, we compiled the message content and created themes on the basis of the narrative analysis technique.

Further, to test dissemination, once the Crave-Out version was ready for Phase 3 pilot testing, we released the game on the Apple iTunes App Store [21] and tracked downloads. We also tracked visits and registrations to the Decide2Quit.org Web site from the game.

Setting and Sample

Current and former smokers from the University of Massachusetts Medical School campus and UMass Medical Memorial Center were recruited using flyers. Smokers admitted to the hospital were referred by a tobacco treatment specialist through the University of Massachusetts Medical School's psychiatry department.

As with Phase 1, flyers were distributed across the campus and a coordinated effort with Department of Preventive Medicine was made to have the flyers available in waiting areas and specifically on posted boards at a connected treatment center where counseling for tobacco and use of other drugs was held. In addition, word of mouth was utilized by contact with tobacco treatment specialists on the staff at the University of Massachusetts Medical School. Participants were asked to report to the usability laboratory to complete participation. Some participants were recruited while they were inpatients; for those recruited inpatients, the tobacco treatment specialist who was assigned to meet with the patient would ask if he or she was interested in participating. The tobacco treatment specialist would then contact the project coordinators with information regarding patients interested and provided room number and appropriate time for visiting the patient. The project coordinator would take a hard copy of the survey along with a laptop to play the game to the patient's room and obtain his or her consent before participation.

We chose to recruit a sample of 30 individuals to test the feasibility of our methods, test game usability, and to generate effects on the outcome of cravings.

Measures

We used the Questionnaire of Smoking Urge-Brief (QSU-Brief), a 10-item questionnaire on smoking urges utilizing a 7-point Likert scale for responses to assess cravings [32]. The QSU-Brief is a valid and reliable assessment that has become widely used in the measurement of cravings [32-35]. Factor analyses conducted by Cox et al indicated that this brief assessment showed very good reliability as per Cronbach alpha [37}. After providing consent, participants completed the pregame QSU-Brief questionnaire and then were given instructions and access to play the game. In addition to the QSU-Brief, we used a brief demographic questionnaire, including questions on smoking status, and a brief survey about game playing experience for data collection. Smoking status was assessed before game play. Recommendations from the Society for Research on Nicotine and Tobacco on the need for biochemical verification state that the degree of misclassification is moderated by characteristics of the smoking cessation intervention. The more intensive the intervention is, the higher the potential for misclassification due to social desirability. In line with these recommendations, as our intervention (a cravings game) was of low intensity and also because the goal of the intervention was not to induce cessation but to pilot test, we did not use biochemical verification of smoking status [36,37].

Twenty-four qualitative questions on experience playing the game were developed by our team; 6 of these were open-ended questions (see Appendix C). In addition, we collected verbal feedback provided while participants played the game. Using the Apple iTunes App Store dashboard, we recorded download data for each week of the study.

Data Analysis

Following a descriptive analysis, we used paired *t* tests to determine differences between pre- and postgame craving levels. We calculated the change in the craving level by subtracting the pregame craving score of each individual from their postgame craving score, to calculate a mean prepost difference. We also conducted stratified analysis by demographic characteristics, smoking status, and game experiences. Note that for our primary analysis and sensitivity analyses, we have presented *P* values “as-is,” recognizing that the stratified sensitivity analyses represent an instance of multiple comparison testing. The statistical literature review of methods for accounting for multiple comparisons has noted that adjustment is controversial and may be over-conservative in some instances [38-40]. Adjusting for multiple comparisons is highly appropriate in exploratory analyses that are not following a specific research question. For studies where the measures follow a specified research goal, recommendations vary. Although we believe that presenting the *P* values "as-is" provides the reader useful information, we recognize the challenge of interpretation in the setting of multiple comparisons and encourage readers to consider these results in the context of this pilot experiment, not as definitive results. Data analysis was performed using STATA 12, Copyright 1996–2016 StataCorp LP.

Phase 2: Results

Demographics

Participants were 30 smokers. The majority were male (20/30, 67%), between ages 25 and 44 (20/30, 67%), college-educated (21/30, 70%), and abstinent for more than 48 hours (5/30 17%). The mean number of cigarettes smoked per day was 13.8 (SD 10.0, range 0-40). Most had attempted to stop smoking at least once in the past 12 months (17/30, 59%) and were willing to stop smoking (20/30, 67%). See Table 1 below regarding these demographics.

Table 1.

Variable		N=30	%
Age			
	19-24	2	7%
	25-34	8	27%
	35-44	10	33%
	45-54	6	20%
	55-64	2	7%
	65+	2	7%
Gender			
	Male	20	67%
	Female	10	33%
Race and ethnicity			
	White/Non-Hispanic or Latino	30	100%
Education			
	Grades 9-11 (some high school)	1	3%
	Grade 12 or GED[a]	8	27%
	College 1-3 years	12	40%
	College 4 years or more	9	30%
Recruitment source			
	Outpatient	22	73%
	Inpatient	8	27%
Ever visited a smoking cessation Web site			
	Yes	7	23%
	No	23	77%
Successful quitter (smokers who had 0 cigarettes per day & had stopped smoking in the past 12 months)			
	Yes	4	23.5%
	No	13	76.5%
Number of cigarettes smoked per day			
	None	5	17%
	1-6	4	13%
	10-18	9	30%
	20-25	10	33%
	30-40	2	7%
	Mean number of cigarettes per day	13.8	
	StdSD	10.0	
Stopped smoking in the past 12 months			
	Yes	17	59%
	No	13	41%
Desire to stop smoking			
	[b]I do not smoke now	5	17%
	Yes	20	67%

[a]GED: General Educational Development

[b]Participants qualified as "do not smoke now" if they had been abstinent for 48 hours or more at the time of the visit.

Game Rating

Smokers reported that Crave-Out was fun (22/30, 73%), challenging (20/30, 67%), and would help distract them from cravings (17/30, 57%; Figure 2). Most also said that they would play the game outside of the study environment (22/30, 73%). The majority of smokers also responded that they found the game would motivate them to quit smoking (16/30, 53%). While playing the game, smokers commented that the game was fun and distracting. Some commented on how they enjoyed playing, for example, "This is amazing," "I'm just having fun," and "It takes your mind off smoking." Most of the smokers played longer than 10 minutes. The following quote indicated one smoker's desire to continue playing the game, "No, don't want to quit the game, I just won." Smokers also noted that the game was challenging. Example comments included "It's complicated," "I have to pay attention," "It's fast," "It's hard," and "It's difficult actually." In addition, one user-interface issue with the beta version was discovered; the 3D effect made it difficult for participants to move the cursor left-right completely accurately in catching the fruit. This was commented on during think aloud, and noted for future consideration.

Pregame and Postgame Cravings

Among these 30 smokers, measurement of QSU-Brief established a pregame mean of 3.24, (SD 1.65 on a 7-point Likert scale) and postgame mean of 2.99 (SD 1.40), a decrease of 0.25 points in cravings, which is a measurable but not significant difference (*P* =.11; Table 2). Across the ten individual items, two items had a significant improvement, "Smoking would make me less depressed" and "A cigarette would taste good now" (*P* =.03 and .002, respectively). As noted, estimating *P* values for each item represents an instance of multiple comparisons testing for a single research question. *P* values are presented "as-is" but should be considered measures of the relative strength of items, not definitive results.

Table 2. Pregame and postgame cravings of Phase 2 smokers.

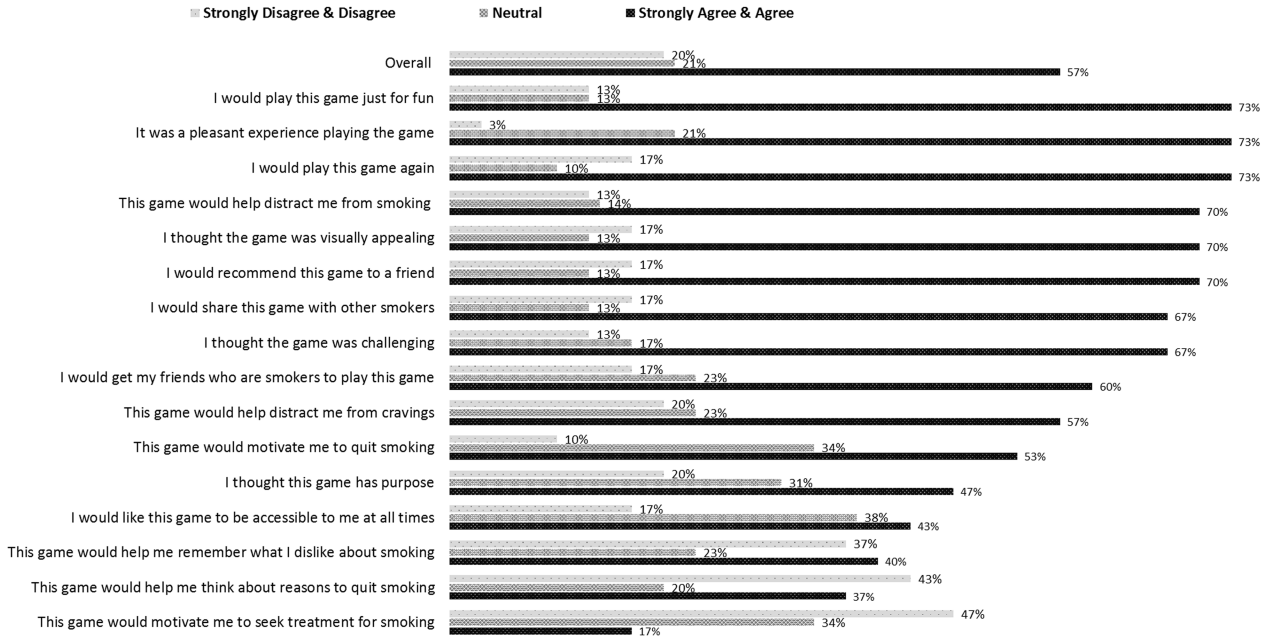
Craving scale	Pregame		Postgame		Change (Δ)	*P* value
	Mean	SD	Mean	SD		
Smoking would make me less depressed	3.36	1.93	2.7	1.51	-0.66	.03
A cigarette would taste good now	3.86	2.03	3.24	1.84	-0.62	.002
All I want now is a cigarette	2.86	1.81	2.56	1.56	-0.30	.28
I would do almost anything for a cigarette now	2.4	1.65	2.13	1.33	-0.27	.24
I could control things better right now if I could smoke	2.96	1.9	2.7	1.55	-0.26	.32
I have an urge for a cigarette	3.48	2.15	3.31	1.91	-0.17	.57
If it were possible, I would probably smoke now	3.8	1.98	3.73	1.85	-0.07	.77
I have a desire for a cigarette now	3.4	2.16	3.5	1.85	0.10	.75
Nothing will be better than smoking a cigarette now	2.86	1.97	3	1.44	0.14	.66
I am going to smoke as soon as possible	3.23	2.02	3.06	1.92	0.16	.32
Overall	3.24	1.65	2.99	1.40	-0.25	.11

In statistical analysis, the QSU-Brief decrease in prepost measurement was significant for those who quit for more than 48 hours (pregame mean 2.84, SD 1.16; postgame mean 2.0, SD 0.94; change 0.84; *P* = . 03). Cravings reduction was not significant for those smoking in last two days (pregame mean 3.14, SD 1.55; postgame mean 2.91, SD 1.06; change 0.23, *P* = . 24). Comparing those who reported that the game was challenging with those who did not, we observed a greater decrease in craving levels (pregame mean 3.26, SD 2.17; postgame mean 2.69, SD 1.66; change 0.57, *P* =.09; Table 3)

Table 3. Stratified analysis of prepost cravings of Phase 2 smokers.

QSU-Brief Question/Response	N	Pregame craving level Mean (SD)	Postgame craving level Mean (SD)	Change (Δ)	*P* value
I thought this game was fun					
Agree	22	3.3 (1.70)	2.96 (1.31)	- 0.32	.10
Disagree	8	3.11 (1.61)	3.07 (1.71)	- 0.04	.10
I thought the game was challenging					
Strongly agree	10	3.26 (2.17)	2.69 (1.66)	- 0.57	.09
Others	19	3.08 (1.24)	2.97 (1.03)	- 0.11	.52
It was a pleasant experience playing the game					
Agree	22	3.26 (1.70)	2.95 (1.30)	- 0.31	.11
Disagree	7	3.48 (1.46)	3.35 (1.71)	- 0.13	.57
This game would help me distract from cravings					
Agree	17	3.55 (1.98)	3.28 (1.63)	-0.27	.27
Disagree	13	2.83 (1.01)	2.61 (0.96)	-0.22	.02
Do you want to stop smoking?					
Abstinent >48 hours	5	2.84 (1.16)	2.0 (0.94)	-0.84	.03
Yes	20	3.14 (1.55)	2.91 (1.06)	-0.23	.24
No	5	4.04 (2.46)	4.32 (2.12)	+0.28	.29

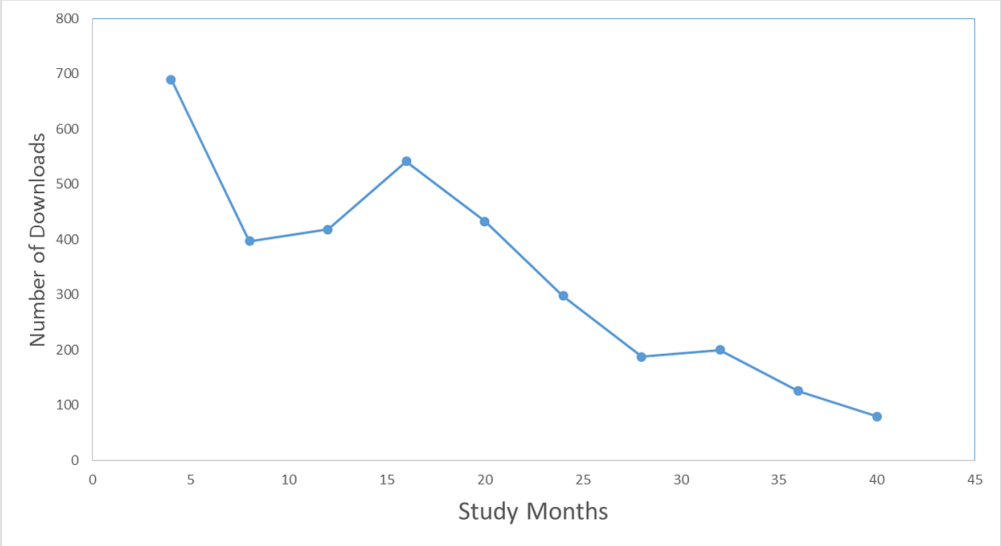
Figure 2. Game experience of Phase 2 smokers.



Phase 2 Dissemination Results

Between December 22, 2011, and May 29, 2014, the game was downloaded 3372 times from the Apple iTunes App Store. As shown in Figure 3 , there was an initial spike in the number of downloads (690) and then again in quarter 4 (542). After quarter 4, the number of downloads tailed off. From January 25, 2012, to May 29, 2014, there were 1526 visits to the Decide2Quite.org Web site from the game, out of which 22 smokers registered on D2Q.

Figure 3. Number of downloads every quarter study period (December 2011 to June 2014).



Discussion

Our pilot study provides preliminary data on the potential of using a mobile game for cravings management, and demonstrates the challenges of developing games for health. Our two-phased approach was successful in pilot testing a game that was challenging and fun and reduced cravings. Smokers found the game enjoyable and that it would motivate them to quit smoking. Cravings reduction was significant in participants abstinent for more than 48 hours. Cravings reduction was also accentuated among those smokers who found the game challenging.

Research on games for smoking cessation is new [41]. Placing our evaluation in context of other studies, games are being developed to change smokers' attitudes toward tobacco addiction to help smokers, and enhance their coping skills [42,43]. A qualitative study assessed the use of existing games (eg, Tetris) as part of a suite of tools to distract smokers from cravings [10]. To our knowledge, Crave-Out is the first mobile game developed and pilot tested specifically as a distraction/motivating game for managing smoking cravings.

Developing a health game is challenging, as game developers have to intermix evidence-based strategies with fun and game play elements [23,41,44]. While behavioral scientists are trained in incorporating evidence-based strategies in interventions, finding ways to intermix these strategies with something fun and game play requires new approaches. Adapting from game development practices, our prototype-based concept refinement approach allowed us to identify important issues early, before time and effort was spent in fully developing the game. For example, while we initially presented bad things associated with smoking, this had the unintended consequence of triggering the participants' thoughts toward smoking. We were also able to make the game fun and challenging for our players by developing it as a multilevel game, with each level increasing in difficulty. Using multiple levels with increasing difficulty is a standard practice in computer games. The experience of facing and overcoming increasing difficulty has been shown to increase engagement and enjoyment of playing the game [23,41,44].

The reduction in cravings was accentuated in those participants abstinent for more than 48 hours. The reduction in cravings was also pronounced among smokers who found the game to be challenging. As noted, creating a game that is challenging may be key for distraction of cravings. Careful consideration should be used when attempting to balance "challenge" with "frustration" [23,41,44]. If the game is too challenging, it can lead to frustration, and trigger the smoker to smoke. If the game is not sufficiently challenging, it can lead to boredom. As noted above, we were able to identify strategies to increase the challenge element of the game, including designing the game as a pattern memory challenge game and the use of levels of increasing difficulty.

The majority of participants (53%) also responded that the game would motivate them to quit smoking. While any game might be distracting, having the ability to add motivational messages might represent the biggest advantage of creating a game specifically for smoking cessation. Adding tailored messages might have increased the motivational potential of our game. Tailored health messages can be highly effective in assisting individuals in understanding and responding to health concerns [45]. Future research is needed to further understand how to add these messages to a game without reducing the fun aspect of the game.

Engaging new smokers is an identified national priority in the State-of-the-Science Conference Statement on Tobacco Use [46], and our pilot study suggests that games might be a useful tool for this purpose. Without any advertisements, smokers found and downloaded our game from the Apple iTunes App Store. The game also engaged smokers to visit the Decide2Quit.org Web-assisted tobacco intervention. About half of the smokers who downloaded the game to the intervention also visited the Web site. Future studies are needed to test this engagement potential.

Limitations

Limitations of this pilot study include a small sample size and a lack of physiological or biological assessments of participants' pregame and postgame cravings levels. Selection bias is another

potential limitation as participants were recruited through tobacco treatment specialists in a hospital setting and potential participants needed to be able to visit the usability laboratory to participate. The game may need to be made more accessible for smokers with different characteristics.Further research is needed to develop games that are easily accessed as distraction/motivation tools.

Conclusions

Developing a distraction/motivation game to help smokers manage cravings is an iterative process. Using a 2-phased, user-testing approach helped us identify some important issues in game development. In a laboratory setting, our game resulted in reduced cravings, accentuated among those abstinent for more than 48 hours. Our next steps include a large study to assess craving management in a real-world setting.

Acknowledgments

The authors would like to acknowledge the students who contributed in the programming, Matthew Schmidt, and artwork, Zachary Bokhour and Madeline Hettich. We would like to thank Dr Thomas Brandon for his advice on the development in this project. Funding for these studies was received from the National Cancer Institute grants R01 CA129091, R01 CA190866, and R21 (R21CA158968) and from the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR000161. Dr Sadasivam is also funded by a National Cancer Institute Career Development Award (K07CA172677). Dr Houston is also supported by the VA eHealth Quality Enhancement Research Initiative (eHealth QUERI) that he directs. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the US government.

Conflicts of Interest

None declared.

Authors' Contributions

KD, RS, and AK contributed to the design of the game, data collection, analysis, and writing of the manuscript; TE contributed to the analysis and review of manuscript; GS contributed to the interpretation of data and review of manuscript; WC contributed to the design of the game and review of manuscript; JV contributed to the writing of the manuscript; DA contributed to review of the manuscript; and TH contributed to the design of the game, analysis, and writing of the manuscript.

Multimedia Appendix 1

Phase 1 Game Play.

[WMV File (Windows Media Video), 951KB - games_v4i1e3_app1.wmv]

Multimedia Appendix 2

Phase 2 Game Play.

[WMV File (Windows Media Video), 3MB - games_v4i1e3_app2.wmv]

References

1. McClure EA, Gray KM. The remote monitoring of smoking in adolescents. Adolesc Psychiatry (Hilversum) 2013 Apr 1;3(2):156-162 [FREE Full text] [doi: 10.2174/2210676611303020006] [Medline: 24834379]
2. CDC. Adult Cigarette Smoking in the United States: Current Estimates. URL: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm [WebCite Cache ID 6VSObiw5j]
3. Partos TR, Borland R, Yong H, Hyland A, Cummings KM. The quitting rollercoaster: How recent quitting history affects future cessation outcomes (data from the International Tobacco Control 4-country cohort study). Nicotine Tob Res 2013 Sep;15(9):1578-1587 [FREE Full text] [doi: 10.1093/ntr/ntt025] [Medline: 23493370]
4. Tonstad S, Tønnesen P, Hajek P, Williams K, Billing C, Reeves K, Varenicline Phase 3 Study Group. Effect of maintenance therapy with varenicline on smoking cessation: A randomized controlled trial. JAMA 2006 Jul 5;296(1):64-71. [doi: 10.1001/jama.296.1.64] [Medline: 16820548]
5. Carter BL, Bordnick P, Traylor A, Day SX, Paris M. Location and longing: The nicotine craving experience in virtual reality. Drug Alcohol Depend 2008 May 1;95(1-2):73-80 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.12.010] [Medline: 18243586]
6. Hayashi T. [The neural mechanisms of cigarette craving and self-control]. Brain Nerve 2014 Jan;66(1):33-39. [Medline: 24371129]
7. Ussher M, Beard E, Abikoye G, Hajek P, West R. Urge to smoke over 52 weeks of abstinence. Psychopharmacology (Berl) 2013 Mar;226(1):83-89. [doi: 10.1007/s00213-012-2886-7] [Medline: 23052572]

8. Killen JD, Fortmann SP. Craving is associated with smoking relapse: Findings from three prospective studies. Exp Clin Psychopharmacol 1997 May;5(2):137-142. [Medline: 9234050]
9. Koegelenberg CFN, Noor F, Bateman ED, van Zyl-Smit RN, Bruning A, O'Brien JA, et al. Efficacy of varenicline combined with nicotine replacement therapy vs varenicline alone for smoking cessation: A randomized clinical trial. JAMA 2014 Jul;312(2):155-161. [doi: 10.1001/jama.2014.7195] [Medline: 25005652]
10. Ploderer BS, Smith W, Pearce J, Borland R. A mobile app offering distractions and tips to cope with cigarette craving: A qualitative study. JMIR mHealth uHealth 2014;2(2):e23 [FREE Full text] [doi: 10.2196/mhealth.3209] [Medline: 25099632]
11. Abroms LP, Padmanabhan N, Thaweethai L, Phillips T. iPhone apps for smoking cessation: A content analysis. Am J Prev Med 2011 Mar;40(3):279-285 [FREE Full text] [doi: 10.1016/j.amepre.2010.10.032] [Medline: 21335258]
12. Payne HE, Moxley VB, MacDonald E. Health behavior theory in physical activity game apps: A content analysis. JMIR Serious Games 2015;3(2):e4 [FREE Full text] [doi: 10.2196/games.4187] [Medline: 26168926]
13. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
14. Conn VS, Algase DL, Rawl SM, Zerwic JJ, Wyman JF. Publishing pilot intervention work. West J Nurs Res 2010 Dec;32(8):994-1010. [doi: 10.1177/0193945910367229] [Medline: 20702685]
15. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
16. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, et al. A tutorial on pilot studies: The what, why and how. BMC Med Res Methodol 2010;10:1 [FREE Full text] [doi: 10.1186/1471-2288-10-1] [Medline: 20053272]
17. Feeley N, Cossette S, Côté J, Héon M, Stremler R, Martorella G, et al. The importance of piloting an RCT intervention. Can J Nurs Res 2009 Jun;41(2):85-99. [Medline: 19650515]
18. Unity 3D Gaming Platform. 2014. URL: http://unity3d.com/ [accessed 2016-03-01] [WebCite Cache ID 6fhEWIrMI]
19. Kanode C, Haddad H. Software Engineering Challenges in Game Development. 2009 Presented at: Sixth International Conference on Information Technology: New Generations; 27-29 April 2009; Las Vegas, Nevada.
20. Barbara C, Jesús T, Rachel G. The Learning Games Design Model: Immersion, collaboration, and outcomes-driven development. International Journal of Game-Based Learning (IJGBL) 2012;2(3):87-110.
21. Game Creation Process. 2014. URL: http://webdocs.cs.ualberta.ca/~games/299/Lecture2.pdf [accessed 2016-03-01] [WebCite Cache ID 6fhF4rY41]
22. Chandler H, Chandler R. Fundamentals of Game Development. Sudbury, MA: Jones & Bartlett Publishers; 2011.
23. Thompson D, Baranowski T, Buday R, Baranowski J, Thompson V, Jago R, et al. Serious video games for health how behavioral science guided the development of a serious video game. Simul Gaming 2010 Aug 1;41(4):587-606 [FREE Full text] [doi: 10.1177/1046878108328087] [Medline: 20711522]
24. Crowley RS, Naus GJ, Stewart J, Friedman CP. Development of visual diagnostic expertise in pathology -- an information-processing study. J Am Med Inform Assoc 2003;10(1):39-51 [FREE Full text] [Medline: 12509356]
25. Kushniruk A, Patel V. Cognitive computer-based video analysis: its application in assessing the usability of medical systems. Medinfo 1995;8 Pt 2:1566-1569. [Medline: 8591502]
26. Kushniruk AW. Analysis of complex decision-making processes in health care: Cognitive approaches to health informatics. J Biomed Inform 2001 Oct;34(5):365-376 [FREE Full text] [doi: 10.1006/jbin.2001.1021] [Medline: 12123153]
27. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. International Journal of Human-Computer Studies 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]
28. Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesising qualitative and quantitative evidence: A review of possible methods. J Health Serv Res Policy 2005 Jan;10(1):45-53. [Medline: 15667704]
29. Sadasivam RS, Allison JJ, Ray MN, Ford DE, Houston TK. Using a resource effect study pre-pilot to inform a large randomized trial: the Decide2Quit.Org Web-assisted tobacco intervention. 2012 Presented at: Annual Symposium proceedings/AMIA Symposium; 3-7 November, 2012; Chicago, USA p. 789-798.
30. Sadasivam RS, Kinney RL, Delaughter K, Rao SR, Williams JH, Coley HL, National Dental PBRN Group, et al. Who participates in Web-assisted tobacco interventions? The QUIT-PRIMO and National Dental Practice-Based Research Network Hi-Quit studies. J Med Internet Res 2013;15(5):e77 [FREE Full text] [doi: 10.2196/jmir.2385] [Medline: 23635417]
31. Creswell J. Research Design: Qualitative, quantitative, and mixed methods approaches. 4th edition. Thousand Oaks, CA: Sage Publications; 2014.
32. Tiffany ST, Drobes DJ. The development and initial validation of a questionnaire on smoking urges. Br J Addict 1991 Nov;86(11):1467-1476. [Medline: 1777741]
33. Cox LS, Tiffany ST, Christen AG. Evaluation of the brief questionnaire of smoking urges (QSU-brief) in laboratory and clinical settings. Nicotine Tob Res 2001 Feb;3(1):7-16. [doi: 10.1080/14622200020032051] [Medline: 11260806]
34. Toll BA, Katulak NA, McKee SA. Investigating the factor structure of the Questionnaire on Smoking Urges-Brief (QSU-Brief). Addict Behav 2006 Jul;31(7):1231-1239 [FREE Full text] [doi: 10.1016/j.addbeh.2005.09.008] [Medline: 16226843]

35. West R, Ussher M. Is the ten-item Questionnaire of Smoking Urges (QSU-brief) more sensitive to abstinence than shorter craving measures? Psychopharmacology (Berl) 2010 Feb;208(3):427-432. [doi: 10.1007/s00213-009-1742-x] [Medline: 20033133]
36. Rabe-Hesketh S, Skrondal A, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. Stata Journal 2002;2(1):1-21.
37. Velicer WF, Prochaska JO, Rossi JS, Snow MG. Assessing outcome in smoking cessation studies. Psychol Bull 1992 Jan;111(1):23-41. [Medline: 1539088]
38. Nakagawa S. A farewell to Bonferroni: The problems of low statistical power and publication bias. Behavioral Ecology 2004 Nov 01;15(6):2004-1045.
39. Perneger T. What's wrong with Bonferroni adjustments. BMJ 1998 Apr 18;316(7139):1236-1238 [FREE Full text] [Medline: 9553006]
40. O'Keefe D. Searching for a defensible application of alpha-adjustment Tools. Human Communication Research 2003 Jul 01;29(3):464-468. [doi: 10.1093/hcr/29.3.464]
41. Baranowski T, Buday R, Thompson D, Lyons EJ, Lu AS, Baranowski J. Developing games for health behavior Ccange: Getting started. Games Health J 2013 Aug;2(4):183-190 [FREE Full text] [doi: 10.1089/g4h.2013.0048] [Medline: 24443708]
42. Krebs PBJ, 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(2):e35 [FREE Full text] [doi: 10.2196/resprot.2416] [Medline: 24025236]
43. Khazaal YC, Chatton A, Prezzemolo R, Hoch A, Cornuz J, Zullino D. A game for smokers: A preliminary naturalistic trial in a psychiatric hospital. Patient Educ Couns 2008 Feb;70(2):205-208. [doi: 10.1016/j.pec.2007.10.006] [Medline: 18093790]
44. Mellecker R, Lyons EJ, Baranowski T. Disentangling fun and enjoyment in exergames using an expanded design, play, experience framework: A narrative review. Games Health J 2013 Jun;2(3):142-149 [FREE Full text] [doi: 10.1089/g4h.2013.0022] [Medline: 24761322]
45. Kreuter M. Tailoring health messages: customizing communication with computer technology. Mahwah, N.J: Lawrence Erlbaum Associates; 2000.
46. National Institutes of Health. NIH consensus and state-of-the-science statements. 2006 Presented at: NIH State-of-the-Science Conference on Tobacco Use: Prevention, Cessation and Control; June 12-14, 2006; Bethesda, MD p. 1-26 URL: https://consensus.nih.gov/2006/tobacco.htm [WebCite Cache ID 6hbI1ZLRT]

Abbreviations

FDA: Food and Drug Administration
GED: General Educational Development
QSU-Brief: Questionnaire of Smoking Urge-Brief

Edited by G Eysenbach; submitted 28.04.15; peer-reviewed by P Haring, L Laranjo, S Baker, T Irizarry, B Baskerville; comments to author 19.08.15; revised version received 23.10.15; accepted 08.01.16; published 26.05.16.

Please cite as:
DeLaughter KL, Sadasivam RS, Kamberi A, English TM, Seward GL, Chan SW, Volkman JE, Amante DJ, Houston TK
Crave-Out: A Distraction/Motivation Mobile Game to Assist in Smoking Cessation
JMIR Serious Games 2016;4(1):e3
URL: http://games.jmir.org/2016/1/e3/
doi:10.2196/games.4566
PMID:27229772

Original Paper

A Serious Game for Clinical Assessment of Cognitive Status: Validation Study

Tiffany Tong[1,2]; Mark Chignell[1,2]; Mary C. Tierney[3,4]; Jacques Lee[5]

[1]Interactive Media Lab, Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON, Canada

[2]Knowledge Media Design Institute, Faculty of Information, University of Toronto, Toronto, ON, Canada

[3]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada

[4]Primary Care Research Unit, Sunnybrook Health Sciences Centre, Toronto, ON, Canada

[5]Clinical Epidemiology Unit and Department of Emergency Services, Sunnybrook Health Sciences Center, Toronto, ON, Canada

Corresponding Author:
Tiffany Tong
Interactive Media Lab
Department of Mechanical and Industrial Engineering
University of Toronto
5 King's College Road
Toronto, ON,
Canada
Phone: 1 416 978 7581
Fax: 1 416 978 7581
Email: tiffany.tong@mail.utoronto.ca

Abstract

Background: We propose the use of serious games to screen for abnormal cognitive status in situations where it may be too costly or impractical to use standard cognitive assessments (eg, emergency departments). If validated, serious games in health care could enable broader availability of efficient and engaging cognitive screening.

Objective: The objective of this work is to demonstrate the feasibility of a game-based cognitive assessment delivered on tablet technology to a clinical sample and to conduct preliminary validation against standard mental status tools commonly used in elderly populations.

Methods: We carried out a feasibility study in a hospital emergency department to evaluate the use of a serious game by elderly adults (N=146; age: mean 80.59, SD 6.00, range 70-94 years). We correlated game performance against a number of standard assessments, including the Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA), and the Confusion Assessment Method (CAM).

Results: After a series of modifications, the game could be used by a wide range of elderly patients in the emergency department demonstrating its feasibility for use with these users. Of 146 patients, 141 (96.6%) consented to participate and played our serious game. Refusals to play the game were typically due to concerns of family members rather than unwillingness of the patient to play the game. Performance on the serious game correlated significantly with the MoCA (r=–.339, P <.001) and MMSE (r=–.558, P <.001), and correlated (point-biserial correlation) with the CAM (r=.565, P <.001) and with other cognitive assessments.

Conclusions: This research demonstrates the feasibility of using serious games in a clinical setting. Further research is required to demonstrate the validity and reliability of game-based assessments for clinical decision making.

(JMIR Serious Games 2016;4(1):e7) doi:10.2196/games.5006

KEYWORDS

cognitive assessments; cognitive screening tools; computerized assessments; games; human computer interaction; human factors; neuropsychological tests; screening; serious games; tablet computers; technology assessment; usability; validation studies; video games

Introduction

The rapidly aging population and high prevalence of age-related conditions, such as delirium and dementia, are placing increasing burdens on health care systems (eg, [1]). More frequent and accessible methods for cognitive screening are needed to detect early signs of impairment and to prevent or better manage further decline. We envision future development of independent patient-administered methods of cognitive screening that can be completed within a hospital or home. Demonstrating that serious games are highly correlated with other methods of cognitive assessment is necessary but not sufficient to justify their use. In order to ensure adequate motivation and realistic assessment of ability, game-based cognitive assessments should be interactive and engaging. They should also be enjoyable so that patients are willing to complete the assessment task at regular intervals.

Background

In geriatric health care, there are standard mental status tools that screen for cognitive impairment, such as the Mini-Mental State Examination (MMSE) [2], Montreal Cognitive Assessment (MoCA) [3], and Confusion Assessment Method (CAM) [4].

Current cognitive screening methods are only minimally interactive, creating little in the way of engagement or entertainment. They are typically initiated by a health care professional rather than sought out by individuals and they are generally not designed for self-administration or for use by nonclinicians. Some tools such as the CAM require subjective assessments, which may result in administrator bias [4]. Additionally, it may not be feasible for the test administrator to repeatedly assess individuals for changes in their cognitive status over time. The resulting lack of frequent assessment may result in underdiagnosis of a condition such as delirium, where cognitive status can fluctuate widely over the course of a day, making it difficult to detect early stages of delirium and initiate preventive interventions [4].

Software suites, such as CogTest [5] and the Cambridge Neuropsychological Test Automated Battery [6], offer computerized versions of traditional cognitive tests. In addition to validation issues when moving a test to the computer medium, there is also the problem of potential lack of motivation when performing somewhat uninteresting tasks on a computer. To deal with the lack of motivation and engagement, games have been promoted as a way to stimulate cognitive activity in elderly users [7] and to improve brain fitness or to preserve cognitive status. For example, the Games to Train and Assess Impaired Persons game suite is composed of eight different games to evaluate motor and cognitive abilities in individuals with impairments [8]. However, such games do not yet provide validated cognitive assessment, have not been used in the health care setting, and evidence about whether they improve broader measures of intelligence is mixed (eg, [9]).

Manera et al [10] performed a pilot study with a serious game involving patients with mild cognitive impairment (MCI) and Alzheimer disease. They were able to demonstrate that their game correlates with the MMSE and other assessments such as the Trail Making Test Part 2 and Victoria Stroop Test. Because this research [10] was carried out on patients with MCI and dementia, and involved a relatively small pilot sample of 21 people using a kitchen and cooking game, there remains a need for a validated game-like screening tool that can be completed rapidly and independently (or with minimal assistance) by a broad range of older adults with varying cognitive ability.

Serious games are games designed with a primary purpose other than entertainment, such as education and training [11]. Specially adapted games can be leveraged to create an interactive and engaging tool that promotes patient-centered cognitive assessment. Mobile phones and tablets are commonly used devices and can be used as platforms for serious gaming. Previous work has demonstrated that elderly users can use mobile phones [12, 13] and touch-based tablets [14]. Many of these technologies also provide the ability to modify contrast/brightness and text size/font to increase readability. Gaming on mobile platforms is already a growing trend that is enjoyed across a wide range of age groups. Thus, the design of a game-based assessment on a mobile platform would likely increase the accessibility of cognitive assessment.

Although there are many potential benefits of designing games for the elderly, there are possible shortcomings to consider. For instance, some elderly users may not be interested in playing games or may be uncomfortable using technology [8]. A brief comparison between paper-and-pencil–based methods and serious games for cognitive assessment is provided in Table 1.

Table 1. Comparison between traditional paper-and-pencil cognitive assessments and the use of serious games for cognitive screening.

Feature	Paper-based assessments	Serious games
Administration method	Trained administrator	Self
Administration bias potential	Yes	No
Equipment	Paper, pencil	Tablet
Repeatability	Limited repeatability—not necessarily if alternate forms are available	Yes
Multiple variations	Few or none	Yes, can be randomized
Motivation/Entertainment	Low	High, if target users enjoy playing the game
Validation	Available	Yet to be completed

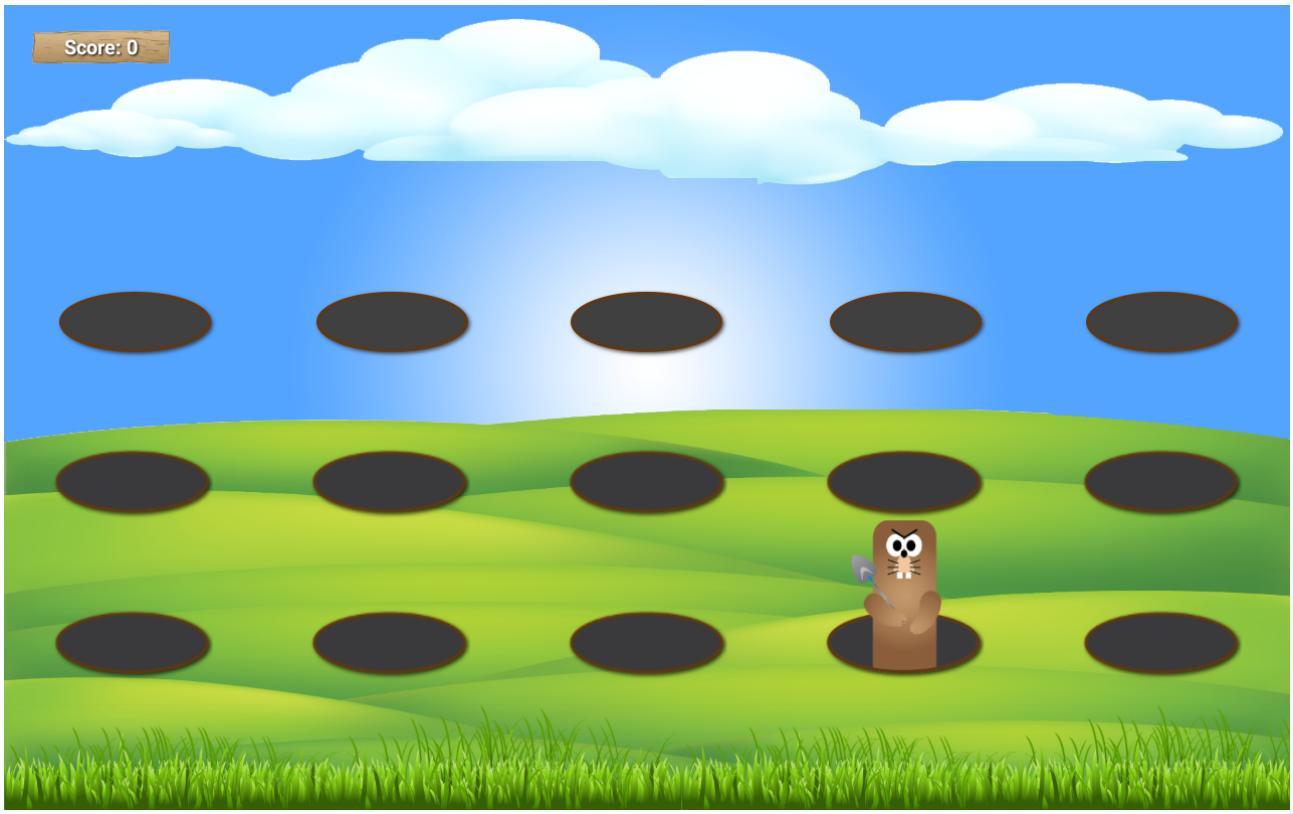
Serious games have been used in health care for the purpose of brain training in projects such as the ElderGames [7], Smart Aging [15], and the work reported by Anguera et al [16]. The ElderGames project uses a large touchscreen tabletop surface as a gaming platform. The goal of this work is to promote social interactions through gameplay with other elderly adults. A limitation associated with this work is that it requires a large apparatus and is not mobile. Moreover, the Smart Aging platform uses a computer and touchscreen monitor to simulate a virtual loft apartment. It is designed to identify MCI through the completion of a series of tasks that simulate daily activities [15]. This project was reported to be in the pilot phase and was evaluated with a relatively small sample of healthy individuals (N=50). A computer-based serious game has been created [16] that simulates driving a vehicle. However, that research compared serious game performance in elderly users with their performance on psychological tasks rather than with standard cognitive assessments. In contrast, we are explicitly developing a game for cognitive assessment.

Development of a Serious Game

We developed a serious game to assess cognitive status in elderly adults with a focus on detecting small changes in cognition for conditions such as delirium. Our serious game mimics features of the classic psychological Go/No-Go Discrimination Task [17], a measure of inhibition ability. As implemented, our game is similar to the carnival game whack-a-mole (see Figure 1). In a previous study with healthy younger adults, we found that our serious game had a significant relationship (r =.60, P <.001) with the Stroop task [14]. The Stroop task is a test of the inhibitory executive function, which declines with age, and the task has been shown to correlate with white matter loss in the brain [18, 19].

After demonstrating that the game-based screening tool was usable by young and older healthy adult samples, and was predictive of inhibition ability, our next step was to evaluate its usability in a clinical sample. In this paper, we present our findings concerning the process of integrating a game-based cognitive assessment into a clinical environment. We demonstrate that our serious game is usable by an elderly population from an emergency department (ED) and is predictive of scores on standard cognitive assessments. The ED is a promising target for serious game-based cognitive assessment because there is a high prevalence of cognitive impairment in that setting compounded by a high rate of underdetection of delirium [20]. Based on the findings from this research, a set of design guidelines is provided in a later section of this paper to assist future researchers in implementing other serious games for assessing cognitive ability.

Figure 1. Screenshot of the whack-a-mole game.



Methods

We conducted a prospective observational clinical study with participants recruited from the Sunnybrook Health Sciences Centre ED (see Figure 2) located in Toronto, Ontario, Canada under a research protocol approved by both the Research Ethics Boards of the Sunnybrook Health Sciences Centre and the University of Toronto. Participants who were 70 years or older and who were present in the ED for a minimum of 4 hours were recruited for the study. Exclusion criteria included patients who were (1) critically ill (defined by the Canadian Triage Acuity Scale score of 1), (2) in acute pain (measured using the Numeric Rating Scale with a score greater than or equal to two out of

10), (3) receiving psychoactive medications, (4) judged to have a psychiatric primary presenting complaint, (5) previously enrolled, (6) blind, or (7) unable to speak English, follow commands, or communicate verbally.

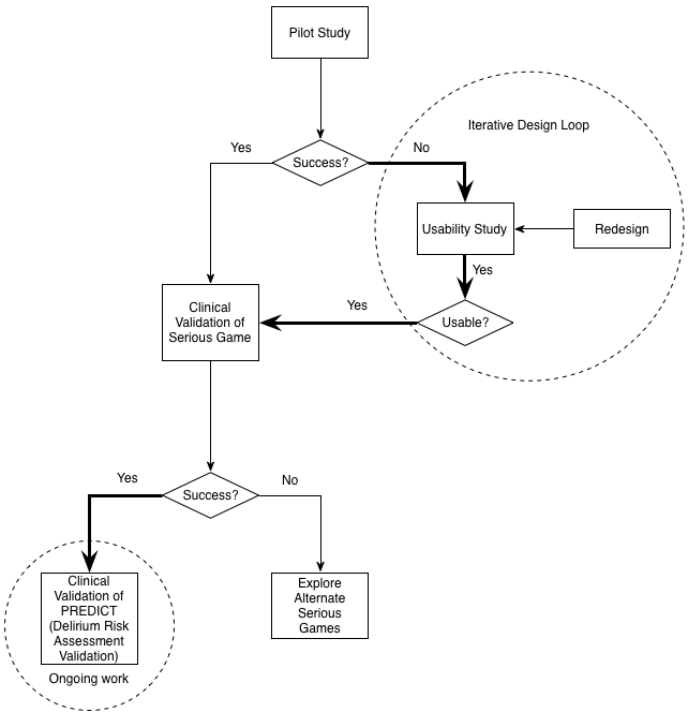
Clinical research assistants (RAs) administered standard cognitive assessments including the MMSE, CAM, Delirium Index (DI) [21], Richmond Agitation-Sedation Scale (RASS) [22], Digit Vigilance Test (DVT) [23], and a choice reaction time (CRT) task. Each participant was then asked to play the serious game and provide feedback. The serious game was played on a 10-inch Samsung Galaxy Tab 4 10.1 tablet. Participants received instructions on how to play the game and interact with the tablet. There was no limit on the number of attempts to play the game. Participants were invited to provide open feedback at the end of the study. At the end of each session, the RA informally interviewed the participant on his/her experience with the game. In addition, RAs provided their own feedback and comments on their experience with the game and their observations of the interaction between each participant and the game.

The RAs recorded the date of the ED visit, whether the cognitive assessments were refused, and the cognitive assessment scores. Usage notes were also recorded and later used to infer usability problems as well as evidence for enjoyment and engagement.

Statistical Analysis

The cognitive data and serious game results were nonnormally distributed based on visual inspection of the data. Tests for normality, including the Kolmogorov-Smirnov and Shapiro-Wilk tests [24], were not used due to the large sample size in this study because they are known to result in oversensitivity to relatively small departures from normality [24]. Transformations of the data were not performed because some of the measures, such as the CAM and DI, are binary/categorical and cannot follow a normal distribution. Our interest was in correlations as a measure of the effect size of the underlying relationship between game performance and the cognitive assessments, but we used nonparametric correlation measures for some of the comparisons [25] that involved categorical or narrow ordinal scales. Correlations between the dichotomous CAM and the other measures were assessed using point-biserial correlations [24]. Correlations involving the DI and RASS (and not involving the CAM) were assessed using Spearman rho because the DI and RASS use a small number of ordered categories. The remaining comparisons were done using Pearson correlations. In order for readers to judge strengths of relationships involving game performance, scatterplots of the relationship between game performance and the MMSE, MoCA, and CAM, respectively, are also presented.

Figure 2. Diagram of studies in this research. The thick line highlights the path taken in this study.



Results

Description of Sample

We recruited 147 participants (80 males and 67 females) between the ages of 70 and 94 years (mean 80.61, SD 6.08). One participant was excluded for not completing any of the cognitive assessments and five people did not play the serious game (of whom two were CAM-positive), leaving 141 participants who completed the study (age range 70-94, mean age 80.64, SD 6.09; 79 males and 67 females).

Some participants declined to complete some of the cognitive assessments entirely or declined to answer certain questions. The completion rate of each test is shown in Table 2. All participants completed the CAM, DI, and RASS. The serious game had a combined completion rate of 96.6% (141/146), whereas the completion rates for the other assessments were lower with DVT being the worst at 36.3% (37/102) overall.

Because the DVT and CRT assessments were initiated partway through the study, the denominators in calculating completion rates for those measures (102 and 99, respectively) were lower than for the other tests (which were initiated at the start of the study).

Table 2. Summary of completion rates for standard cognitive assessment scores.

Cognitive assessment	Completion rate, n (%)
Mini-Mental State Examination (MMSE)	145/146 (99.3)
Montreal Cognitive Assessment (MoCA)	108/146 (73.9)
Confusion Assessment Method (CAM)	146/146 (100.0)
Delirium Index (DI)	146/146 (100.0)
Richmond Agitation-Sedation Scale (RASS)	146/146 (100.0)
Digit Vigilance Test (DVT)[a]	37/102 (36.3)
Choice Reaction Task (CRT)[a]	82/99 (83)
Serious game	141/146 (96.6)

[a]This assessment was introduced later in the study.

There were a number of people in the sample with low MMSE and MoCA scores (down to 9 and 8, respectively). There were 129 participants who were negative for the CAM and 12 participants who were positive (a positive result on the CAM suggests that the participant has delirium). Moreover, the DI scores ranged from 0 to 10 (the score indicates the severity of delirium), RASS scores ranged from –2 to 1 (a score >0 suggests that the patient is agitated and a score <0 suggests that the patient is sedated), DVT scores ranged from 81 to 103, and CRT choice accuracy ranged from 34% to 95%. The combined median response time (RT) on the CRT was 1.2 sec (IQR 0.4). The overall median RT on the serious game was 0.9 sec (IQR 0.2), and the mean accuracy was a deviation of 328.5 pixels (SD 59.7) from the center of the target. A summary of the scores on the cognitive assessments can be found in Table 3.

Table 3. Summary of study sample demographics and cognitive assessment scores.

Variable	Males (n=80)		Females (n=66)		Total (N=146)	
	Mean (SD) / median (IQR)[a]	Range	Mean (SD) / median (IQR)[a]	Range	Mean (SD) / median (IQR)[a]	Range
Age (years)	80.6 (6.3)	70-94	80.6 (5.7)	70-94	80.6 (6.0)	70-94
MMSE	28.2 (1.5)	25-30	27.7 (2.2)	9-30	26.7 (3.9)	9-30
MoCA	24.5 (2.6)	8-30	23.2 (3.8)	15-30	23.2 (4.6)	8-30
CAM	0.1 (0.3)	0-1	0.1 (0.3)	0-1	0.1 (0.3)	0-1
DI	0.5 (0.7)	0-10	0.5 (0.8)	0-8	1.3 (2.3)	0-10
RASS	–0.1 (0.4)	–2 to 1	–0.1 (0.4)	–2 to 1	–0.1 (0.3)	–2 to 1
DVT	97.5 (5.7)	81-103	98.7 (4.0)	92-103	97.8 (5.3)	81-103
CRT RT (sec)	1.2 (0.3)	0.87-1.98	1.2 (0.5)	0.78-3.23	1.2 (0.4)	0.78-3.40
CRT accuracy (%)	87 (1)	50-95	87 (13)	34-95	87 (1)	34-95
Game RT (sec)	0.8 (0.1)	0.65-2.46	0.9 (0.3)	0.65-2.65	0.9 (0.2)	0.65-2.65
Game accuracy (pixels)	331.9 (49.0)	140-449	327.8 (69.9)	81-424	328.5 (59.7)	81-449

[a]For CRT RT and game RT, the median (IQR) is reported. All others are mean (SD).

Comparison Between Serious Game Performance and Standard Cognitive Assessments

Game performance was measured based on a participant's RT and accuracy. In our serious game, RT was measured from the time the target appeared to the time of the user's response and accuracy was measured as the pixel distance between the center of the target and the center of the user's touch.

Correlation analysis revealed significant relationships between game median RT and scores on the six cognitive assessments: MMSE, MoCA, CAM, DI, RASS, DVT, and CRT RT (see Table 4). In contrast to the RT results, the corresponding relationships between game accuracy and the standard cognitive assessments were not statistically significant, except for the relationship with DVT. Note that information about which types of correlation were used for each comparison are shown in the footnotes to Table 4.

Table 4. Correlations comparing game performance to the standard cognitive assessments..

Measure	Correlation[a](*P*-value)									
	Game RT	Game accuracy	MMSE	MoCA	CAM	DI	RASS	DVT	CRT RT	CRT accuracy
Game RT	1	.132 (.12)	–.558 (<.001)	–.339 (<.001)	.565 (<.001)	.280 (<.001)	–.296 (<.001)	–.122 (.48)	.625 (<.001)	–.325 (.003)
Game accuracy		1	–.104 (.22)	–.042 (.67)	.071 (.40)	.048 (.46)	–.108 (.12)	.432 (.008)	–.053 (.64)	.004 (.97)
MMSE			1	.630 (<.001)	–.693 (<.001)	–.689 (<.001)	.339 (<.001)	.200 (.24)	–.503 (<.001)	.307 (.005)
MoCA				1	–.505 (<.001)	–.339 (<.001)	.193 (.01)	.192 (.28)	–.296 (.01)	.148 (.22)
CAM					1	.515 (<.001)	–.644 (<.001)	—[b]	.434 (<.001)	–.237 (.03)
DI						1	–.418 (<.001)	–.037 (.79)	.272 (.002)	–.160 (.06)
RASS							1	—[b]	–.124 (.17)	.129 (.16)
DVT								1	.045 (.80)	–.237 (.18)
CRT RT									1	–.503 (<.001)
CRT accuracy										1

[a]Correlations involving the CAM were calculated using point-biserial correlations. Correlations involving the DI and RASS (and not involving the CAM) were assessed using Spearman rho. All other correlations were calculated using Pearson *r*.

[b]Cannot be computed because at least one of the variables is constant.

As a follow-up to our correlation analyses in Table 4, we carried out the same analysis using Spearman rho correlations instead of Pearson correlations. All significant correlations between the cognitive assessments and game RT and game accuracy, respectively, were also observed to be significant using Spearman rho.

In order to examine the separate contributions of speed of processing and executive functioning on cognitive assessment scores, we looked at the partial correlations of serious game and CRT performance (controlling for each other) with the clinical assessments (see Table 5). The partial correlations with game RT (controlling for CRT) remained significant for the MMSE, CAM, and DI, but not for the MoCA and DVT. There was one significant relationship for the partial correlation of game accuracy (controlling for CRT) with DVT. On the other hand, the partial correlations involving CRT, but controlling for serious game performance RT, were not significant except for the MMSE (see Table 5). In addition, the partial correlations involving CRT but controlling for game accuracy were significant for the DI only (Table 5).

Table 5. Partial correlations that control for CRT RT on game performance and standard cognitive assessments and control for game RT on standard cognitive assessments.

Assessment	Control for CRT RT				Control for game RT			
	Serious game median RT		Serious game median accuracy		CRT RT		CRT Accuracy	
	ρ	*P*	ρ	*P*	ρ	*P*	ρ	*P*
MMSE	–.313	.005	–.024	.84	–.241	.03	.221	.52
MoCA	–.068	.58	.160	.19	–.197	.11	.063	.61
CAM	.516	<.001	–.112	.33	–.040	.73	.014	.01
DI	.412	<.001	.066	.56	.215	.06	–.255	.02
RASS	.173	.13	–.088	.44	–.179	.11	.135	.24
DVT	.39	.440	.01	.105	.57	–.227	.21	–.159

Detection of Abnormal State Using Serious Game Performance

A Mann-Whitney *U* test (see Table 6) was performed to investigate the difference between cognitive ability and serious game performance when the MMSE score was 24 and above (normal cognitive function or possible MCI) versus when that score was below 24 (signs of dementia) [2, 26]. The MMSE was chosen as the grouping criterion because it was a standard in screening for dementia at the time this research was carried out. The test results suggest that there was a significant difference on the CRT in terms of RT between participants with dementia (MMSE <24) and no dementia (MMSE ≥24) [26]. In addition, there was a significant difference between MMSE groups in terms of game RT (U =348.5, z =–4.7; P <.001), but not for game accuracy. For Table 6, the corresponding scatterplot (Figure 3) is also shown. Figure 3 shows the distribution of game RT versus MMSE ("dementia" scores are indicated by triangles) where a tendency for lower MMSE scores to be associated with longer RTs can be seen.

Table 6. Mann-Whitney *U* test results comparing cognitive assessment performance based on the absence (≥24) or presence (≤24) of dementia as assessed by the MMSE.[a]

Assessment[b]	MMSE <24		MMSE ≥24		*U*	*P*	*z*	*r*	IQR
	n	Mean (SE)	n	Mean (SE)					
Game RT	18	327.6 (17.6)	122	317.2 (5.2)	348.5	<.001	–4.7	.4	0.9-1.1
CRT RT	8	2.2 (0.3)	73	1.3 (0.0)	104.0	.003	–2.9	.3	1.0-1.4
CRT accuracy	8	0.7 (0.0)	73	0.8 (0.0)	181.0	.08	–1.7	.1	0.8-0.9
Game accuracy	18	0.7 (0.0)	122	0.8 (0.0)	980.5	.46	–0.7	.0	299.0-328.5

[a]Table has been reordered based on the *U* statistic value according to estimated *P* value.

[b]RT measures are reported in seconds, CRT accuracy reflects proportion of responses that were correct, and game accuracy reflects deviation in pixels from the center of the target.

Similar to the analysis reported in Table 6, a Mann-Whitney *U* test (see Table 7) was performed to investigate the difference between cognitive ability and serious game performance when the MoCA score was 23 and above (normal cognitive function) versus below 23 (MCI) [27]. The MoCA was chosen as the criterion in this comparison because it is a de facto standard in screening for MCI versus normality. There was a significant difference (U =947.5, z =–2.7; P =.001) on the CRT RT between participants with cognitive impairment (MoCA <23) and no impairment (MoCA ≥23). There was also a significant difference between MoCA groups for game RT (U =370.0, z =–3.2; P =.03). For Table 7, the bivariate relationship is illustrated in the scatterplot in Figure 4. This figure illustrates a tendency for lower MoCA scores to be associated with longer RTs, although that relationship appeared to be weaker for the MoCA than it was for the MMSE.

Table 7. Mann-Whitney *U* test results comparing game performance based on the absence (≥23) or presence (≤23) of cognitive impairment as assessed by the MoCA.[a]

Assessment[b]	MoCA, <23		MoCA ≥23		*U*	*P*	*z*	*r*	IQR
	n	Mean (SE)	n	Mean (SE)					
Game RT	38	1.0 (0.07)	67	0.9 (0.02)	307.0	.03	–3.2	.31	0.7-117
CRT RT	26	1.6 (0.1)	44	1.2 (0.08)	947.5	.001	–2.7	.32	1.0-1.1
CRT accuracy	26	0.8 (0.02)	44	0.9 (0.02)	439.5	.11	–1.6	.19	0.8-0.9
Game accuracy	38	317.5 (9.2)	67	3222.4 (5.6)	1240.0	.83	–0.2	.02	299.0-352.5

[a]Table has been reordered based on the *U* statistic value according to significance.

[b]RT measures are reported in seconds, CRT accuracy reflects proportion of responses that were correct, and game accuracy reflects deviation in pixels from the center of the target.

Another Mann-Whitney *U* test (see Table 8) was performed to investigate the difference between cognitive ability and serious game performance when delirium was present (CAM positive) versus absent (CAM negative). The CAM was chosen as the grouping factor as it is the gold standard in screening for delirium. The test indicated a significant difference on the MMSE, MoCA, RASS, and DI between participants with delirium (CAM positive) and no delirium (CAM negative). In addition, there was a significant difference between CAM groups in terms of RT on the serious game (U =–4.5, P <.001). For Table 8, this relationship is shown in Figure 5. These between-group differences in game RT and MMSE are consistent with findings by Lowery [28], where CAM-negative participants demonstrated faster RT and higher MMSE scores compared to CAM-positive participants.

Table 8. Mann-Whitney *U* test results comparing cognitive assessment performance based on the absence (CAM negative) or presence (CAM positive) of delirium as assessed by the CAM.[a]

Assessment[b]	CAM Negative		CAM Positive		*U*	*P*	*z*	*r*	IQR
	n	Mean (SE)	n	Mean (SE)					
RASS	14	–0.03 (0.02)	142	–0.8 (0.2)	288.0	<.001	–7.8	.62	0.0-0.0
Game RT	12	0.9 (0.02)	129	1.7 (0.2)	158.0	<.001	–4.5	.38	0.7-1.1
MoCA	7	23.8 (0.4)	101	14.3 (2.0)	60.5	<.001	–3.7	.36	21.0-26.0
MMSE	14	27.6 (0.2)	131	18.4 (1.3)	38.0	<.001	–5.9	.49	26.0-29.0
DI	14	0.6 (0.1)	131	6.9 (0.5)	24.5	<.001	–6.6	.55	0.0-1.0
CRT RT	4	1.3 (0.06)	78	2.6 (0.5)	45.0	.02	–2.4	.26	1.0-1.4
CRT accuracy	4	0.8 (0.01)	78	0.7 (0.1)	91.5[c]	.17	–1.4	.15	0.8-0.9
Game accuracy	12	317.3 (5.3)	129	332.4 (15.2)	708.0	.63	–0.5	.04	299.0-352.5

[a]Table has been reordered based on the *U* statistic value according to significance. No Mann-Whitney *U* test analysis was carried out for the DVT because there were no CAM-positive participants who completed the DVT. Additional assessments are included in this table for the purpose of comparison.

[b]RT measures are reported in seconds, CRT accuracy reflects proportion of responses that were correct, and game accuracy reflects deviation in pixels from the center of the target. Other measures shown reflect the scores on the instruments (MoCA, MMSE, DI, RASS).

[c]The independent samples *t* test was nonsignificant for this comparison (t_{80}=1.5, *P* =.21).

As a check, we replicated all the Mann-Whitney *U* tests in Tables 6-8 with their parametric equivalent, in this case the independent samples *t*-test. The pattern of significant and nonsignificant effects was identical for both tests, with the exception of the comparison of CRT RT between CAM-positive and CAM-negative participants (Table 8). For that comparison, the independent samples *t*-tests did not show a significant effect, whereas the Mann-Whitney *U* test did.

Figure 3. Scatterplot illustrating the differences on game RT performance based on MMSE score (≥24=normal cognitive function or possible MCI; <24=signs of dementia).

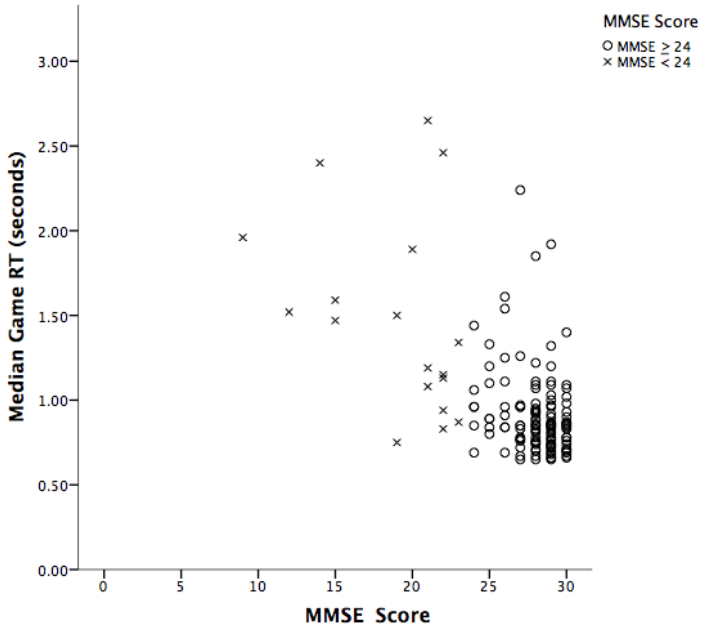


Figure 4. Scatterplot illustrating the differences on game RT based on MoCA score (≥23=normal cognitive function; <23=cognitive impairment).

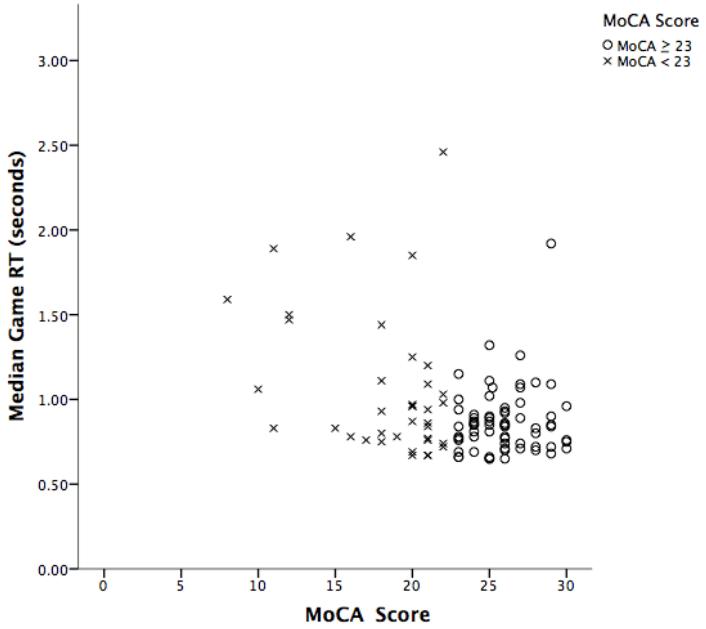
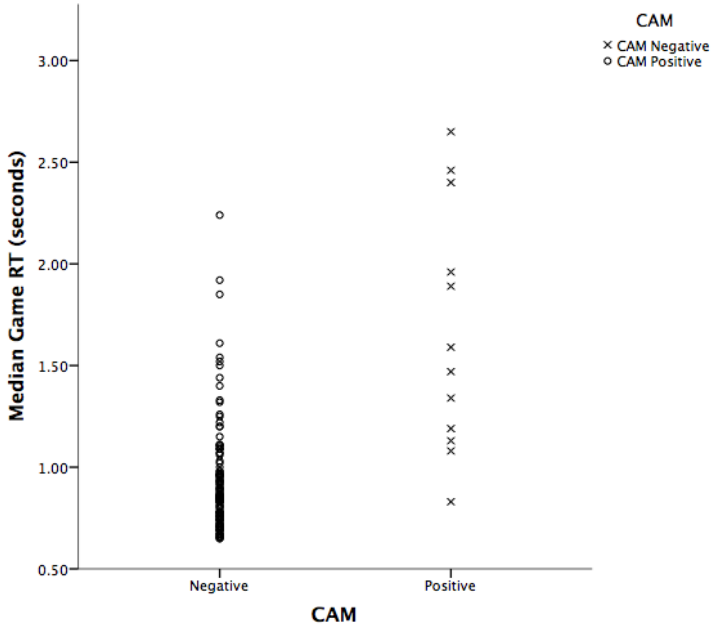


Figure 5. Scatterplot illustrating the differences on game RT based on CAM groups (CAM negative=delirium absent; CAM positive=delirium present).



Predicting Delirium Status Using Serious Game Performance

In the preceding section, we examined the relationship between game performance and current standards for clinical assessment with respect to MCI, delirium, and dementia. In this section, we examine the question of how well the serious game performance predicted CAM status (delirium).

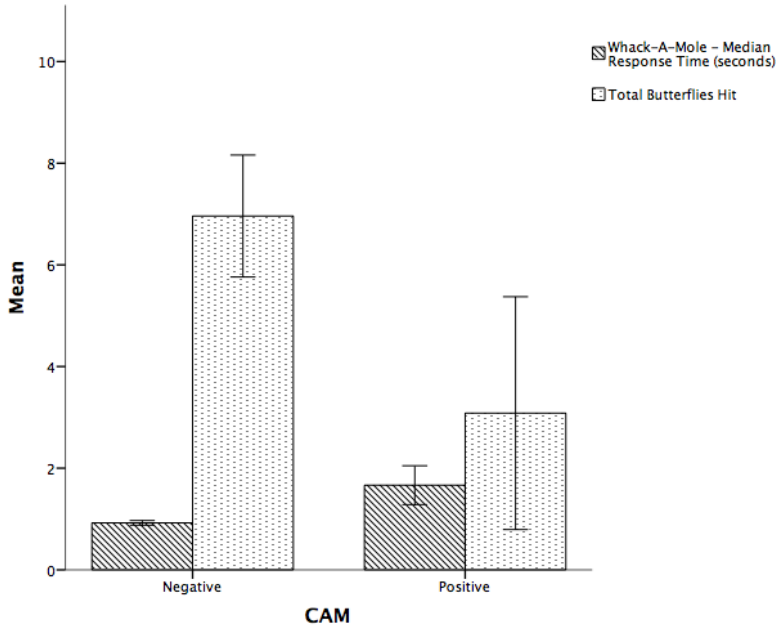
Discriminant analysis was carried out to see how well game performance could predict CAM status. The two predictors were game RT and accuracy. Game accuracy provided no benefit in prediction and received a zero weight in the discriminant function. Thus, we focused on game RT as a potential screener for further evaluation using the CAM. We examined different possible cutoff values for distinguishing between people who should be screened for possible delirium (using the CAM) and those who should not.

Setting a relatively long median RT for the decision threshold (≥1.88 seconds) resulted in good specificity (127/129, 98.4% CAM-negative patients were correctly identified), but relatively poor sensitivity (only 5/12, 41% CAM-positive patients were correctly identified).

On the other hand, using a more stringent median RT cutoff of 1.13 seconds, there was both good sensitivity (10/12, 83% CAM-positive patients were correctly identified) and good specificity (114/129, 88.3% CAM-negative patients were correctly identified).

We also found that CAM-positive patients hit fewer distractors by mistake (as shown in Figure 6). Since CAM-positive participants had fewer hits in general (to both moles and butterflies), it seems likely that their apparently lower error rate was due to a lower response rate rather than to the presence of a speed-accuracy tradeoff.

Figure 6. Mean of median RTs and mean number of butterflies hit for CAM-negative and CAM-positive patients. Error bars indicate 95% CI.



Usability Issues and Evidence of Enjoyment and Engagement

The following brief notes recorded by the RAs during patient use of the serious game are indicative examples of enjoyment and engagement that were observed: “Loved the game, she was playing games on her iPhone before I approached her” “Enjoyed the game, he would play on his own,” “Too easy but don’t make it too challenging, like the game,” and “Really loved the tablet, wanted to keep playing even after testing was over.” However, usability problems were also observed. Some participants placed their palm on the tablet while trying to interact with the serious game. This confused the software because it was unclear which hit points were intentional versus accidental. Some participants claimed that the game was too easy and suggested that we include more difficult levels to make it more interesting. Elderly users also expressed an interest in playing games such as crossword puzzles. Anecdotally, the RAs who supervised the data collection at the hospital reported that this game was easier to administer and more fun to complete compared to standard cognitive assessments such as the MoCA and DVT.

Ergonomic Issues

While interacting with the tablets, the elderly participants assumed numerous positions, such as being seated, lying down, standing, or walking around. Each of these positions had different ergonomic requirements and some brief recommendations based on our experience in this study are provided in the Discussion. Some participants were also frail and required the assistance of the RA to hold the tablet for them.

Discussion

Performance on the serious game in terms of median RT was significantly correlated with MMSE, MoCA, CAM, DI, RASS, DVT, and CRT scores for elderly ED patients and differences were in the expected direction (slower game RT for people with possible MCI and dementia). The correlations suggest a relationship between longer RT on the game and lower cognitive assessment scores. These correlations demonstrate the potential value of serious games in clinical assessment of cognitive status. The correlations between the standard cognitive tests observed in this study are similar to results seen in other research. For example, correlations of r =.43 and r =.60 between MMSE and MoCA scores for healthy controls and patients with MCI, respectively, have been found [29]. In our study, we observed a correlation of r =.63 (P <.001) between the MMSE and MoCA scores. Overall, the correlation of our serious game with existing methods of clinical cognitive assessment appears to be almost as strong as the correlations of the clinical assessment methods with themselves.

In our partial correlation analysis, we observed that our serious game correlates with the MMSE and DI, but that part of that correlation is attributable to speed of processing (CRT speed). Thus, serious game performance in this case involved both speed of processing and executive functioning components. Both components are involved in the correlation of the serious game with the MMSE. However, only the speed of processing component appears to be involved in the correlation with the MoCA. Crucially, the partial correlations of serious game performance (controlling for CRT RT) were higher than the corresponding partial correlations for CRT (controlling for serious game performance) indicating that the serious game is

an overall better predictor of cognitive status than simple processing speed as measured by the CRT task.

We found that there was a lack of association between serious game accuracy and scores on cognitive assessments. This may be due to variations in interaction methods where some users used their fingers instead of a stylus to interact with the tablet device. Another reason may be that some users preferred responding more quickly over being accurate in their responses.

One of the goals of this research was to develop a method for predicting the presence of delirium using this serious game. In this study, we found that a median RT cutoff of 1.13 seconds implied relatively good sensitivity and specificity in the clinical decision. However, 25 of the 129 (19.4%) participants were above the median cutoff and only 10 of these were CAM-positive. Thus, in a clinical setting the question remains of how to deal with people who are identified as CAM-positive using this RT cutoff value. One approach would be to give those people full CAM assessment and then treat the CAM-positive patients accordingly. The value of the serious game in this case is that it would allow (based on screening with the serious game) a high rate of delirium detection using CAM assessment in only around 20% of patients (assuming that the current results generalize to other contexts). Ideally, a suitably adapted serious game would also detect risk of delirium onset so that prevention strategies could be used on targeted patients before they developed delirium, but that prospect was beyond the scope of the research reported in this paper.

During our studies, we observed many ergonomic issues that could arise during the administration of the serious game. For instance, there were a variety of positions and methods used to interact with the tablet-based serious game. For participants who are sitting down, we recommend a tablet case that has a hand holder or kickstand to allow them to interact with the tablet in multiple ways. In contrast, for participants lying down on a bed, it may be difficult for them to hold the tablet to play the serious game; thus, a stand affixed to a table or intravenous pole that holds up the tablet would be appropriate. Furthermore, the ergonomic solutions that are adopted should meet hospital standards on hygiene and sanitization for technology. For patients with hand injuries or visual disabilities, the serious game may not be a usable option.

User-centered design and ergonomic interventions were both key in ensuring that the serious game was usable with a challenging user group (elderly patients) and in the fairly unique and demanding context of a hospital ED. The touch interface was modified so that it was more forgiving of the kinds of gestures made by elderly users when interacting with the game and the gameplay was modified so that users with a wide range of ability could play the game. Ergonomic issues that were dealt with in our research included the form factor of the device and the selection and use of accessories to facilitate interactions with the device in different postures and contexts.

Based on our research experience, we present the following recommendations for enhancing tablet-based user interaction between elderly adults and touch-based technologies:

1. Accept multiple gestures, including taps and swipes, as input to maximize interaction.

2. Provide a stylus for users who have difficulties interacting with the tablet with their fingers.

3. For time-sensitive tasks, the time limit should be increased to allow older or more frail users a chance to interact with the software.

4. Tablet screen protectors should be installed to provide more friction between a user's hand and the screen.

5. A variety of ergonomic stands and mounts should be available to accommodate various interaction positions.

6. Serious games for cognitive assessment should incorporate validated psychological task components (eg, executive functions) and should be easily playable for independent use.

7. Assess the validity of the game across different subgroups of patients. Consider the possibility of using multiple versions of a game, or multiple games, to accommodate the different characteristics and needs of different types of patient.

Limitations

The usability and validation results obtained apply to elderly adults in an emergency setting. Further research would be needed to generalize these results to different types of patient and different clinical settings. The design of this study was cross-sectional, so each participant/patient was only studied during one ED visit and played the game only once. Future research may assess the reliability of the game when played repeatedly by the same patient in the ED. One other limitation is that only one game was examined in this research (the whack-a-mole game that we developed). Other serious games should also be explored to determine which games work best with different types of patients.

This work is an initial validation study of our serious game for cognitive screening, where the game was only administered once. One of the goals of this research is frequent cognitive screening, which can potentially lead to learning effects on the game. Future research that assesses the reliability of the game-based screening tool will need to address how to overcome and differentiate between learning effects on a patient's game performance on our serious game versus their actual cognitive status. Because we are interested in changes in cognitive status, we are not as concerned with a patient's improved performance due to learning effects from repeated gameplay, but would aim to track deviations in their performance over time due to cognitive decline.

Conclusions

We believe that serious games are a promising methodology for cognitive screening in clinical settings, including the high-acuity time-pressured ED environment. This work demonstrates the feasibility of implementing a serious game for cognitive screening in a health care environment. To the best of our knowledge, this is the first time that a serious game for cognitive assessment has been tested in an ED and with a full battery of standard cognitive assessment methods for comparison. Based on these results, ergonomically appropriate

serious games can potentially revolutionize cognitive assessment of the elderly in clinical settings, allowing assessments to be more frequent, more affordable, and more enjoyable.

This research provides a case study in the development of an interactive serious game for cognitive screening that may be used independently and repeatedly, thus promoting patient-centered health and safety. We have demonstrated in this study that elderly adults older than age 70 years can successfully play our serious game in an ED and that RT performance on the game can be used as an initial screen for cognitive status.

These findings do not yet demonstrate that the serious game evaluated here is ready to be used to screen for delirium in the ED. Only 12 CAM-positive patients were observed in the study and of the game performance measures (RT, accuracy, number of targets hit, number of distractors hit), only game RT was predictive of CAM status. However, due to the known underreporting of delirium in the ED, an efficient and usable method of screening for delirium is clearly needed. In this study, a game median RT cutoff of 1.13 seconds produced a sensitivity of 83% and a specificity of 88% when used retrospectively as a screen for CAM-positive status. Although further research is needed, it seems possible that a suitably revised and validated game might be able to identify approximately 80% to 90% of CAM-positive cases while requiring the screening of no more than approximately 20% of cases.

Outside the ED, the use of the serious game for ongoing patient-administered assessment would ideally involve patients who remain actively engaged with their support network (eg, family and care providers) and with health care professionals. For instance, if patients perform poorly on the serious game or notice a decline in their performance, they could discuss these results with their care providers, which might lead to interventions such as changes to medication or lifestyle that could slow observed declines.

Acknowledgments

The authors would like to thank all volunteers who participated in our research studies. We would also like to thank Janahan Sandrakumar, Jacob Woldegabriel, and Joanna Yeung for assisting with data collection.

MCT is supported by a Clinician Scientist Award from the Department of Family & Community Medicine, University of Toronto. TT is supported by CIHR - STIHR Fellowship in Health Care, Technology, and Place (TGF-53911). MC is supported by a grant from the AGE-WELL National Center of Excellence (WP 6.1). This research was funded by a Canadian Institutes of Health Research Catalyst Grant: eHealth Innovations (application number 316802).

Conflicts of Interest

None declared.

References

1. Schneider E, Guralnik JM. The aging of America. JAMA 1990 May 02;263(17):2335. [doi: 10.1001/jama.1990.03440170057036]
2. Folstein MF, Folstein SE, McHugh PR. “Mini-mental state”: a practical method for grading the cognitive state of patients for the clinician.. J Psychiat Res 1975 Nov;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6]
3. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
4. Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. Ann Intern Med 1990 Dec 15;113(12):941-948. [Medline: 2240918]
5. Barua P, Bilder R, Small A, Sharma T. Standardisation study of Cogtest. Schizophrenia Bull 2005;31(2).
6. Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, et al. A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: implications for theories of executive functioning and cognitive aging. Cambridge Neuropsychological Test Automated Battery. J Int Neuropsychol Soc 1998 Sep;4(5):474-490. [Medline: 9745237]
7. Gamberini L, Alcaniz M, Barresi G, Fabregat M, Ibanez F, Prontu L. Cognition, technology and games for the elderly: an introduction to ELDERGAMES Project. PsychNology J 2006;4(3):285-308.
8. Tso L, Papagrigoriou C, Sowoidnich Y. Universität Stuttgart. 2015. Analysis and comparison of software-tools for cognitive assessment URL: http://elib.uni-stuttgart.de/handle/11682/3569 [accessed 2016-05-08] [WebCite Cache ID 6hMl3byCB]
9. Jacova C, Kertesz A, Blair M, Fisk J, Feldman H. Neuropsychological testing and assessment for dementia. Alzheimers Dement 2007 Oct;3(4):299-317. [doi: 10.1016/j.jalz.2007.07.011] [Medline: 19595951]
10. 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]
11. Charsky D. From edutainment to serious games: a change in the use of game characteristics. Games Cult 2010 Feb 11;5(2):177-198. [doi: 10.1177/1555412009354727]

12. Kurniawan S, Mahmud M, Nugroho Y. A study of the use of mobile phones by older persons. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems. 2006 Presented at: CHI '06; April 22-27, 2006; Montreal, QC p. 989-994. [doi: 10.1145/1125451.1125641]
13. Kurniawan S. Older people and mobile phones: a multi-method investigation. Int J Hum-Comput St 2008 Dec;66(12):889-901. [doi: 10.1016/j.ijhcs.2008.03.002]
14. Tong T, Chignell M. Developing serious games for cognitive assessment: aligning game parameters with variations in capability. In: Proceedings of the Second International Symposium of Chinese CHI. 2014 Presented at: Chinese CHI '14; April 26-27, 2014; Toronto, ON p. 70-79. [doi: 10.1145/2592235.2592246]
15. Zucchella C, Sinforiani E, Tassorelli C, Cavallini E, Tost-Pardell D, Grau S, et al. Serious games for screening pre-dementia conditions: from virtuality to reality? A pilot project. Funct Neurol 2014;29(3):153-158 [FREE Full text] [Medline: 25473734]
16. Anguera JA, Boccanfuso J, Rintoul JL, 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]
17. Yechiam E, Goodnight J, Bates JE, Busemeyer JR, Dodge KA, Pettit GS, et al. A formal cognitive model of the go/no-go discrimination task: evaluation and implications. Psychol Assess 2006 Sep;18(3):239-249 [FREE Full text] [doi: 10.1037/1040-3590.18.3.239] [Medline: 16953727]
18. Ludwig C, Borella E, Tettamanti M, de RA. Adult age differences in the Color Stroop Test: a comparison between an item-by-item and a blocked version. Arch Gerontol Geriatr 2010;51(2):135-142. [doi: 10.1016/j.archger.2009.09.040] [Medline: 19846224]
19. Olsen RK, Pangelinan MM, Bogulski C, Chakravarty MM, Luk G, Grady CL, et al. The effect of lifelong bilingualism on regional grey and white matter volume. Brain Res 2015 Jul 1;1612:128-139. [doi: 10.1016/j.brainres.2015.02.034] [Medline: 25725380]
20. Wilber ST, Lofgren SD, Mager TG, Blanda M, Gerson LW. An evaluation of two screening tools for cognitive impairment in older emergency department patients. Acad Emerg Med 2005 Jul;12(7):612-616 [FREE Full text] [doi: 10.1197/j.aem.2005.01.017] [Medline: 15995092]
21. McCusker J, Cole MG, Dendukuri N, Belzile E. The delirium index, a measure of the severity of delirium: new findings on reliability, validity, and responsiveness. J Am Geriatr Soc 2004 Oct;52(10):1744-1749. [doi: 10.1111/j.1532-5415.2004.52471.x] [Medline: 15450055]
22. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O'Neal PV, Keane KA, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med 2002 Nov 15;166(10):1338-1344. [doi: 10.1164/rccm.2107138] [Medline: 12421743]
23. Kelland D, Lewis RF. The digit vigilance test: reliability, validity, and sensitivity to diazepam. Arch Clin Neuropsych 1996;11(4):339-344. [doi: 10.1016/0887-6177(95)00032-1]
24. Field A. Discovering Statistics using IBM SPSS Statistics, 4th ed. Thousand Oaks, CA: Sage Publications; Jan 23, 2013.
25. Bishara A, Hittner JB. Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches. Psychol Methods 2012 Sep;17(3):399-417. [doi: 10.1037/a0028087] [Medline: 22563845]
26. O'Connor DW, Pollitt PA, Hyde JB, Fellows JL, Miller ND, Brook CP, et al. The reliability and validity of the Mini-Mental State in a British community survey. J Psychiatr Res 1989;23(1):87-96. [Medline: 2666647]
27. Luis CA, Keegan AP, Mullan M. Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. Int J Geriatr Psychiatry 2009 Feb;24(2):197-201. [doi: 10.1002/gps.2101] [Medline: 18850670]
28. Lowery DP, Wesnes K, Brewster N, Ballard C. Subtle deficits of attention after surgery: quantifying indicators of sub syndrome delirium. Int J Geriatr Psychiatry 2010 Oct;25(10):945-952. [doi: 10.1002/gps.2430] [Medline: 20054840]
29. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin A. Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. BMC Geriatrics 2015;15(107). [doi: 10.1186/s12877-015-0103-3]

Abbreviations

CAM: Confusion Assessment Method
CRT: choice reaction time
DI: Delirium Index
DVT: Digit Vigilance Test
ED: emergency department
MCI: mild cognitive impairment
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
RA: research assistant

RASS: Richmond Agitation-Sedation Scale
RT: response time

Edited by E Lettieri; submitted 05.08.15; peer-reviewed by K Assmann, J Anguera, PC Masella; comments to author 01.09.15; revised version received 30.11.15; accepted 29.02.16; published 27.05.16.

Please cite as:
Tong T, Chignell M, Tierney MC, Lee J
A Serious Game for Clinical Assessment of Cognitive Status: Validation Study
JMIR Serious Games 2016;4(1):e7
URL: http://games.jmir.org/2016/1/e7/
doi:10.2196/games.5006
PMID:27234145

Original Paper

Effects of Playing a Serious Computer Game on Body Mass Index and Nutrition Knowledge in Women

Mariya Shiyko[1], PhD (Quantitative Psychology); Sean Hallinan[1], MA; Magy Seif El-Nasr[2], PhD; Shree Subramanian[3], PhD; Carmen Castaneda-Sceppa[4], MD, PhD

[1]Northeastern University, Applied Psychology, Boston, MA, United States

[2]Northeastern University, College of Computer and Information Science, Boston, MA, United States

[3]Northeastern University, Center for Advancing Teaching and Learning through Research, Boston, MA, United States

[4]Northeastern University, Department of Health Sciences, Boston, MA, United States

Corresponding Author:
Mariya Shiyko, PhD (Quantitative Psychology)
Northeastern University
Applied Psychology
360 Huntington Ave
404 INV
Boston, MA, 02115
United States
Phone: 1 6173737527
Fax: 1 6173738892
Email: m.shiyko@neu.edu

Abstract

Background: Obesity and weight gain is a critical public health concern. Serious digital games are gaining popularity in the context of health interventions. They use persuasive and fun design features to engage users in health-related behaviors in a non-game context. As a young field, research about effectiveness and acceptability of such games for weight loss is sparse.

Objective: The goal of this study was to evaluate real-world play patterns of SpaPlay and its impact on body mass index (BMI) and nutritional knowledge. SpaPlay is a computer game designed to help women adopt healthier dietary and exercise behaviors, developed based on Self-Determination theory and the Player Experience of Need Satisfaction (PENS) model. Progress in the game is tied to real-life activities (e.g., eating a healthy snack, taking a flight of stairs).

Methods: We recruited 47 women to partake in a within-subject 90-day longitudinal study, with assessments taken at baseline, 1-, 2-, and 3- months. Women were on average, 29.8 years old (±7.3), highly educated (80.9% had BA or higher), 39% non-White, baseline BMI 26.98 (±5.6), who reported at least contemplating making changes in their diet and exercise routine based on the Stages of Change Model. We computed 9 indices from game utilization data to evaluate game play. We used general linear models to examine inter-individual differences between levels of play, and multilevel models to assess temporal changes in BMI and nutritional knowledge.

Results: Patterns of game play were mixed. Participants who reported being in the preparation or action stages of behavior change exhibited more days of play and more play regularity compared to those who were in the contemplation stage. Additionally, women who reported playing video games 1-2 hours per session demonstrated more sparse game play. Brief activities, such as one-time actions related to physical activity or healthy food, were preferred over activities that require a longer commitment (e.g., taking stairs every day for a week). BMI decreased significantly (P<.001) from baseline to 3-month follow-up, yielding a large effect size of 1.28. Nutritional knowledge increased significantly (P<.001) from first to third month follow-ups, with an effect size of .86. The degree of change in both outcomes was related to game play, baseline readiness to change, and the extent of video game play in general.

Conclusions: This work demonstrates initial evidence of success for using a serious game as an intervention for health behavior change in real world settings. Our findings also highlight the need to understand not only game effectiveness but also inter-individual differences. Individualizing content and the intervention medium appears to be necessary for a more personalized and long-lasting impact.

(JMIR Serious Games 2016;4(1):e8) doi:10.2196/games.4977

KEYWORDS

serious games; games for health; weight loss; body mass index; nutritional knowledge; game play; self-determination theory; Player Experience of Need Satisfaction (PENS) model; women

Introduction

Obesity and weight gain are critical public health concerns in the United States. More than one third of all US adults are overweight, a total of 78.6 million [1]. Obesity has been strongly linked to several preventable health conditions, including heart disease, stroke, type 2 diabetes, certain types of cancers, sleep apnea, and hypertension, among others [2]. An estimated 300,000 deaths a year are linked to obesity [3]; and almost 21% of the total healthcare budget (around $190.2 billion) is spent annually towards obesity-related illnesses [4].

Given the scope of the problem, the Internet and ubiquitous technology present a unique opportunity for behavior change intervention and reach. Serious games targeting health behavior change represent a new field of research. Video games are a growing medium in the United States and are becoming more popular than the motion picture industry. Zynga, a large game developer, claims 148 million unique users [5]. Video games are now appearing on computers, phones, toys, and even medical devices and kitchen appliances [6]. As an intervention tool, video games are appealing due to the adaptability and customizability of the user's experience, and tend to be relatively low-cost [6]. Serious games are a genre of video games that employ playful design strategies to encourage users' engagement in a non-game context. They are considered to be uniquely suited for increasing individuals' motivation and, thus, have a potential to reach individuals to whom traditional modalities of behavior change may not be appealing or available [7].They can also be scaled up to reach a large audience. Based on a recent meta-analysis, online behavior change technologies (which include video games) are more successful than public health campaigns at initiating behavior change, reaching 10% of users, compared to public health's 5% [8].

While promising, there is scarce empirical evidence of the efficacy of serious games for weight loss in adults, as only a few studies have evaluated games in the adult population. Moller et al. [9] assessed the acceptability and initial effectiveness of an online fantasy sports game on physical activity in two small pilot studies, spanning 13 (N=9) and 17 (N=10) weeks respectively. An overall positive effect was found on maintenance and increase in walking behavior in a convenience sample of adults. Another study [10] evaluated the effect of a 3-week role-play educational game with a sample of 40 undergraduates (80% women) and found an improvement in knowledge of healthy foods, an enhanced understanding of the perceived benefits of and barriers to healthy eating, and increased self-efficacy and intention to engage in healthy eating behaviors. Finally, in the context of diabetes management, adults (N=41) participating in a virtual island game aimed to promote knowledge of health-related self-management behaviors demonstrated a modest significant weight loss over the period of 6 months [11]. The field is gradually developing, and a description of several other games is available (eg, Dance Dance Revolution [12], an exercise bicycle linked to a computer game [13], and a pedometer linked to a fish avatar [14]). However, there is no empirical data to evaluate effectiveness of these games, although a few initial studies provide some evidence for the effectiveness of serious games as ways to increase weight loss-related health behaviors in adults.

Despite the initial promise of serious games, there is a need to understand their role, applications, limitations, and types of individuals for whom they are most suitable. Technology on its own is unlikely to make games effective unless strongly grounded in the principles and theories of health behavior change. Some games are developed in collaboration between industry and academia, with theories of health behavior change serving as a foundation for game design [15]; however, this trend is still in its infancy. Based on empirical evidence and theory [16-20], the following elements are considered important for building a persuasive product: goal setting, capacity to overcome challenges, providing feedback on performance, reinforcement, progress comparison, social connectivity, and fun and playfulness [21]. In games, the elements translate through gamification principles, defined as the use of game design elements in a non-game context [22]. Some examples of gamification include an engaging story line [23]; provision of clear goals and challenges through game principles of leveling up, earning points, badges, and rewards; a regular performance feedback through visualization; and community support through an in-game social network [21].

Weight loss and maintenance require regular engagement in healthy eating and physical activity over a long period. While some intervention programs have been shown to be effective in the short term [24], a relapse to old habits is common. For example, an estimated 94% relapse has been observed among people engaged in dieting [25]. Motivation to engage in health-related behaviors is essential and has been shown to be more effective long-term (ie, lasting 6-18 months) than a skills-based approach [26]. Self-determination theory is a global theory of human motivation that has been actively applied to video games [10,27-29]. The theory postulates that continuous motivation can be sustained through meeting 3 primary human needs: *competence*, *autonomy*, and *relatedness*. *Competence* is defined as the innate desire to learn new skills and gain mastery over them. It can be cultivated through presenting new challenges that progressively build learned skills, creating opportunities for participants to meet those challenges and progress through game leveling. *Autonomy* is defined as an innate desire to be in control over goals and behaviors. In a game, this can be achieved by allowing users to choose personal goals and individual behaviors that can meet the desired goals. Finally, *relatedness* taps into the human propensity for a social connection and belonging. Relatedness can be targeted through

real or virtual social networks, players' community support through feedback, encouragement, and competition.

Overall, self-determination theory is one of the frameworks that has been studied in the field of video games [30-31] and explicitly adopted as the Player Experience of Need Satisfaction (PENS) model, which outlines game elements that tap into needs for competence, autonomy, and relatedness. Based on research from thousands of players, video games that include elements that meet these needs are predictive of emotional, behavioral, and objective outcomes, including self-reported fun and enjoyment, game immersion, game values and sales, length of play, and recommendations to others [31]. To our knowledge, no existing games have explicitly incorporated elements of the PENS model into their design to reduce the Body Mass Index (BMI) and increase nutrition knowledge. Given that weight loss requires a long-term commitment, motivation towards a healthy weight is an important factor to initiate and sustain new behaviors. It is our hypothesis that explicitly designing a game based on the PENS principles will also translate into real-life behavioral outcomes.

In this paper, we present SpaPlay–a serious game for encouraging and sustaining healthy living in women [27]. The game SpaPlay was developed based on the outlined principles of the PENS model [31] and gamification strategies [21,31]. Our previous qualitative study demonstrated overall acceptability of the game [32-33]. The focus of the current study is to evaluate the extent of game play derived from objective gamification data and examine initial evidence of play effectiveness on BMI and nutritional knowledge. In the following sections, we provide an overview of the game, describe the within-subject longitudinal pilot study, and summarize major findings and conclusions.

SpaPlay–Game Description

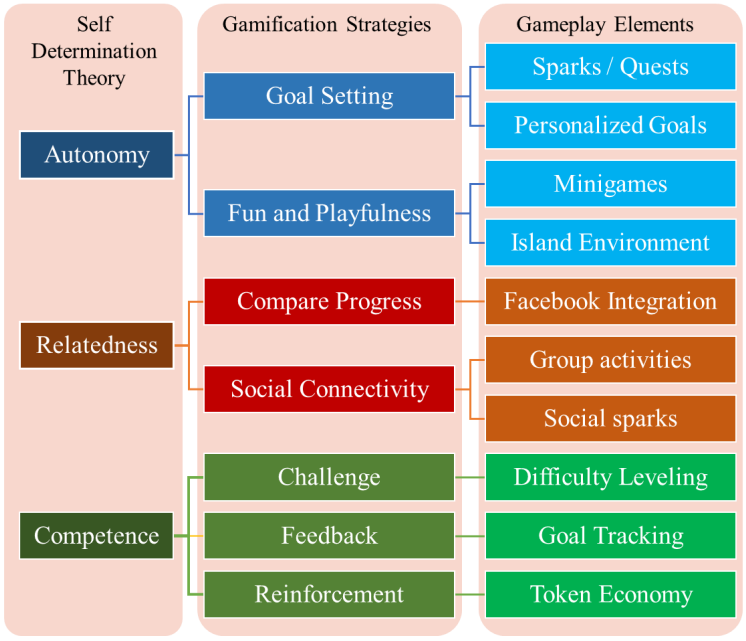
SpaPlay is a digital social online game developed to motivate women to make healthy eating choices and to exercise. It was developed through a close partnership between academia and industry, spanning more than 5 years, and involving all stages of game design and development. The current version of SpaPlay is a browser-based video game accessible to players via personal computers and laptops. It requires an Internet connection and a user-generated username and password. The content of the game centers on a virtual spa resort that needs to be developed and maintained, similar to other popular games like Farmville or We Rule. Figure 1 presents several screen shots of game elements. SpaPlay bridges real and virtual worlds, and game progress is contingent on activities completed in real life around physical activity and healthy eating. For example, eating a salad or taking a 10-minute walk would earn players points towards developing the spa (eg, building facilities, accumulating ratings, playing mini games). Gamification and principles of the Self-Determination theory are used to sustain players' interest and engagement. Further, since weight loss takes time to self-observe, more immediate rewards in the game (eg, power-ups and customization options for the user's avatar) or real life (eg, coupons from associated vendors) are used to keep players engaged.

Figure 2 explicitly demonstrates the relationship between game play elements, gamification principles and the self-determination theory. The two core game mechanics through which experience points are earned are Quests and Sparks. Quests are sets of physical activities or dietary tasks that the players complete within the span of a week and target longer-term commitment to health behaviors. An example includes taking 1-2 flights of stairs twice within a week; substituting a sugary beverage for water 5 times a week; or eating a fruit instead of a snack twice in a week. Sparks are short, single-time tasks that can include stretching for 5 minutes or adding spinach to a sandwich. Social Sparks are dietary or exercise activities in which the user engages with others in the real world. Both Sparks and Quests allow users to set their own short and longer-term goals and chose from a variety of activities that the game offers. They promote autonomy in the way that the players design and enact their own program. They also build social connectivity through in-game and real-life group-based activities (eg, taking a walk with a friend), promoting relatedness.

It is important to note a few additional game features. For instance, healthy behaviors are introduced through the game interface, which both teaches exercise methods, such as yoga through the yoga mini game, and proper diet through recommendations and diet mini games, such as the chef game, which challenges players to create healthy dishes with healthy recipes. Self-tracking of physical activity is enabled through a connected Fitbit sensor, allowing users to log and measure external activities. Finally, the game incorporates social features that allow users to interact with friends and display measures of their progress.

Figure 1. Screenshots of SpaPlay.

Figure 2. Self-Determination theory-based game elements in SpaPlay.



Methods

Participants

To evaluate the game, we recruited women from an urban New England area. Participants self-selected for the study in response to online solicitations sent through listservs and handouts and flyers posted at several public health and non-for-profit organization sites. Inclusion criteria were being ≥ 21 years old, speaking English, and reporting being at least at the contemplation stage of change for engaging in exercise and nutrition on the transtheoretical model of change [34], as described below. Of the 60 women initially recruited, 13 (21.7%) dropped out within the first week of the study. Reasons cited included lack of gaming experience and inability to provide adequate time commitment to game play. The final sample was

comprised of 47 women who partook in the 90-day longitudinal study, with global assessments taken at baseline, 1-, 2-, and 3-month follow-ups. The study was approved by the Institutional Review Board at Northeastern University.

Procedure

After the online consent, participants completed a baseline survey and were given instructions with game access. Following the baseline, participants were contacted via emails at months 1, 2, and 3 with a request to fill out follow-up surveys assessing main study outcomes (BMI and nutrition knowledge). A research assistant emailed participants once a month to check on progress and address any questions or concerns. All participants were awarded a $20 gift certificate at study completion.

Measures

Telemetry game play data were automatically recorded and time-stamped throughout the game play period. Data consisted of time-stamped actions, such as logins, type of Quests or Sparks selected, completion of Sparks or Quests, game activities such as picking up trash, score changes, and social interactions.

Two major study outcomes were BMI and nutritional knowledge. Participant self-reported weight and height was collected at baseline, 1-, 2-, and 3- months. BMI was calculated using the formula: weight (kg)/height $(m)^2$ [35].

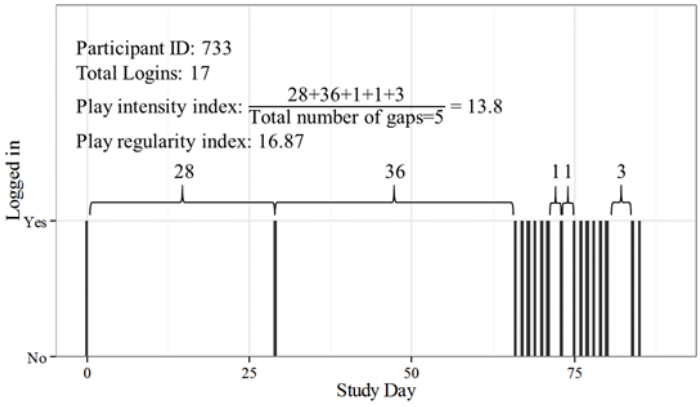
Nutritional knowledge was assessed at 1, 2, and 3 months with the General Nutrition Knowledge questionnaire for adults. The questionnaire is comprised of 53 items evaluating an individual's knowledge of nutritional and dietary needs. It has high test-retest reliability and construct validity established through expert review [36]. To reduce burden, due to the length of the questionnaire, we purposefully omitted assessing participants on their nutrition knowledge at baseline.

Readiness to change behaviors in domains of exercise, nutrition and consumption of sweetened beverages was assessed at baseline by the *Readiness to Change Questionnaire*, a 16-item instrument with high levels of test-retest reliability and predictive validity for behavior change [37]. For this study, we omitted information on the beverage-related items, and readiness to change on exercise and nutrition domains were used as covariates.

Statistical Analyses

First, we examined play patterns based on objective telemetry data. Frequency of play was determined from daily game logins, which were tracked from the automatically recorded telemetry data. Daily play was noted as present (1) or absent (0). The total number of play days was computed for each person. In addition, the *play intensity index* was computed as the length of time between logins. With daily play, an individual mean would be expected to equal 0. Less frequent play would naturally translate into a higher mean value. *Play regularity index* was computed as the standard deviation between logins. For a regular play pattern (eg, every 2 days), it would be expected to be 0, and increase with irregular play. Figure 3 presents an example of login data from one study participant with corresponding play intensity and play regularity indices. Further, to capture the nature of game activities for each individual, we computed the total number of food- and exercise- related Sparks and Quests, and the number of game and social activities. Descriptive statistics were computed for all 9 indicators of game play for the entire sample.

Figure 3. Example log-in data from one study participant. Vertical bars across the 90-day timeline represent occurrences of play, with quantitative summaries of adherence statistics.



Second, we examined whether background baseline variables differentiated between play patterns. To account for data non-normality, log transformations were applied to total days of play and the total number of food and exercise Quests and Sparks. General linear models were run with 7 predictors including education, ethnicity, age, history of weight loss program participation, baseline BMI, history of gaming 1-2 hours per session, and baseline stage of change on exercise and healthy food intake questionnaires. While repeated testing jeopardizes the overall type I error rate, given the preliminary nature of the study, we did not adjust the p-value and kept it at the .05 level for each model.

Third, to examine changes in BMI and nutritional knowledge over the course of gameplay, multilevel modeling (MLM) [38,39] was used to test for effects of time and game adherence on the study outcomes while controlling for major demographics (ie, age, ethnicity, and educational level), history of game play, and baseline readiness to change stage. MLM is designed to account for nested data structure with observations clustered within participants, for missing data with some participants

omitting certain observations, and for inter-individual differences in trajectories with possible variability in intercept and slope values. The following model, as specified in Figure 4 was fit to the data, where the residuals are normally distributed, $\varepsilon_{it} \sim N(0,R_i)$, with variance R_i following the compound symmetry structure, and u_{0i} and u_{1i} are random intercept and slope parameters with corresponding variances, τ_{00} and τ_{10}. In this equation, BMI for a given person i measured at time t is modeled as a function of, where time is centered at the 3rd month follow-up (with values of -1, -2, and -3 corresponding to observations at 60 days, 30 days, and baseline, respectively). γ_{00} is the BMI value at 3 months, controlling for study covariates Cov_i, γ_{10} is the effect of game play on BMI, and is a linear rate of change in BMI in increments of 30 days (1 month). A similar model was fitted to Nutrition Knowledge as the outcome. Both were fitted in the nonlinear mixed effects (nlme) package in R, which is free, open-source statistical software [40]. Final models included predictors significant at the .05 alpha level.

Figure 4. MLM equation.

$$BMI_{it} = \gamma_{00} + \gamma_{01} * GamePlay_i + \gamma_{02} * Covariate_i + \gamma_{10} * time_{it} + u_{0i} + u_{1i} + \varepsilon_{it}$$

Results

Sample Description

The sample was comprised of women with an average age of 30 years (SD 7.3), 81% (38/47) reporting a Bachelors or Masters degree, 61% (30/47) being White, 20% (10/47) Asian, 8% (4/47) Black, and 2% (1/47) Latina. Of these participants, 6% (3/47) were concurrently enrolled in a weight loss program, while 32.7% (16/47) had previously completed weight loss programs. Average BMI at the beginning of the study was 26.98 (SD 5.6), which is considered "overweight" by the CDC's criteria [35]. A number of participants (8/47, 17%) reported daily video game usage, 33% (26/47) reported less than daily but more than weekly usage, 25% (12/47) reported playing weekly or less frequently, and 6% (3/47) did not provide information about their video game behavior. Of those who did report playing video games, 72% (34/47) reported playing 1-2 hours per gaming session.

For behavior change stages on nutrition, 60% (28/47) of participants were in the contemplation stage, 34% (16/47) were in the preparation stage and 4% (247) were in the action stage. For physical activity, 55% (26/47) were in the contemplation stage, 38% (18/47) were in the preparation stage and 4% (2/47) were in the action stage.

Game Play

Table 1 summarizes descriptive statistics for game play indicators. On average, participants played about 7 days (SD 12.5), 25% of participants did not utilize the game at all between the beginning and the end of the study, and 75% logged in fewer than 7 times over the course of the study. Food and exercise Sparks were the most popular activities. Distributions of Sparks were very skewed, with several players engaging in several hundreds of Sparks over the course of 90 days, but the majority engaging in a few (median of 8 for food and 4 for exercise Sparks). The average number of days between logins was 24.65 (SD 14.23), and the play regularity index averaged at 16.48 (SD 7.66).

Table 1. Summary of game play data.

Activity	Mean (SD)	Median
Total Number of Play Days	6.9 (12.5)	2
Play Intensity Index	24.65 (14.23)	22.04
Play Regularity Index	16.48 (7.66)	15.16
Food Sparks	108.2 (349.2)	4
Food Quests	4.5 (10.8)	0
Exercise Sparks	100.6 (294)	8
Exercise Quests	9.3 (27.5)	0
Game Activities	1.8 (3.2)	0
Social Activities	1.2 (2.4)	1

Analyses of game play data revealed several predictors that differentiated between different levels of engagement with the game (Table 2). Two major predictors emerged. First, those who reported being in the preparation or action stages of eating healthy at baseline had a higher total number of logins, had shorted gaps between logins, played the game more consistently, and completed more exercise Sparks (all P values < .05). Second, those who reported at baseline spending 1-2 hours playing video games per session had fewer total logins and completed fewer exercise and food-related Sparks and Quests (all P values < .05).

Table 2. Results of general linear models predicting game play data[a].

Outcomes	Predictors: Parameter Estimate (Standard Error)			
	Intercept	Gaming 1-2 hr/session	TTM Eat Healthy	Below BA
Log (Days Total)	1.589 (.307)	-.943 (.386) *P*=.02	.789 (.321) *P*=.02	
Play Intensity	27.752 (2.497)		-8.953 (3.703) *P*=.02	
Play Regularity	18.403 (1.340)		-5.393 (1.988) *P*=.001	
Log (Food Sparks)	3.531 (1.214)	-3.067 (1.330) *P*=.046		-3.153 (1.535) *P*=.026
Log (Exercise Sparks)	2.573 (.947)	-2.769 (1.192) *P*=.025	2.667 (.991) *P*=.01	
Log (Food Quests)	-.446 (1.119)	-4.310 (1.312) *P*=.002		
Log (Exercise Quests)	-.234 (1.198)	-4.165 (1.404) *P*=.005		

* *P*<.05
** *P*<.01
[a]Results for Log (Game Activities) and Log (Social Activities) were not significant.

Multilevel Modeling

Figure 5 presents observed BMI trajectories, with the large line summarizing the sample average, and thinner lines summarizing data from 10 randomly selected participants. Overall, the trajectory decreased over the course of three months. Average BMI at the beginning of the study was 26.98 (SD 5.6) and 26.09 (SD 5.27) at the end of the study, corresponding to Cohen's *d* effect size of 1.28 (large) based on a paired-sample *t* test.

Table 3 presents MLM results for BMI data. As expected from the graphical summary, the effect of time was significant (*P*<.001), with the average estimated rate of change of .27 per month on the BMI scale. Controlling for age, individuals who completed more exercise Quests tended to have higher final BMI (*P*=.043); and individuals who played with less regularity also tended to have higher final BMI (*P*=.079).

Table 3. Results of multilevel modeling for changes in BMI.

Parameters	Estimate	SE	*P* value
Intercept	17.659	3.189	<.001
Time	-0.271	.036	<.001
Game Activities	.429	.195	.034
Gaming 1-2 hr/session	3.202	1.505	.040
TTM Physical Activity	-2.466	1.084	.029
Age	.223	.089	.017
SD (Intercept)	4.617		<.001
SD (Time)	.207		<.001
Residual	.235		
Rho	.001		

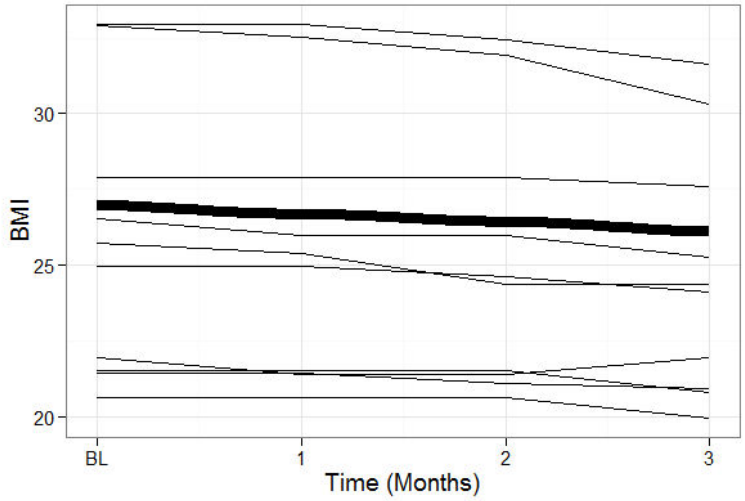
For nutrition knowledge, the average value at 30 days was 73.30 (± 13.59) compared to the mean at 90 days of 78.68 (± 12.66). This increase corresponds to Cohen's d effect size of |.856| (large) based on the paired t-test. Results of the MLM are summarized in Table 4. Overall, knowledge increased with time (*P*<.001), with the estimated increase of 2.003 units per month in the study. Individuals who completed more exercise Quests had a marginally higher final level of nutrition knowledge (*P*=.089).

Table 4. Results of multilevel modeling for changes in nutrition knowledge.

Parameters	Estimate	SE	*P* value
Intercept	71.171	2.367	<.001
Time	1.980	.481	<.001
Food Quests	.926	.348	.011
Exercise Sparks	-.042	.018	.024
TTM[a]Physical Activity	12.656	3.426	<.001
SD (Intercept)	10.322		<.001
SD (Time)	2.324		<.001
Residual	3.022		
Rho	.001		

[a]Transtheoretical Model of Behavior Change for Physical Activity

Figure 5. Graphical summary of BMI trajectories across the 3 study Months: Sample average (thick black line) and individual trajectories of ten randomly selected participants.



Discussion

Despite variable and limited adherence to the video game by the participants, our results demonstrate a relationship between SpaPlay play and changes in BMI and nutrition knowledge. This serves as preliminary evidence of the positive effects of incorporating Self-Determination and PENS theories into video game design to encourage behavior change towards healthy weight in women.

Major Study Outcomes

A very promising result of the study is that the participants lost weight over the course of the three months. A significant and large change of almost one point on the BMI scale was observed, with the sample mean of 27 (SD 5.6) at baseline and 26.1 (SD 5.26) at the 3-month follow-up. The only other study that reported on BMI changes in the context of a game for diabetes management in adults [11] also found a drop in BMI, although of a smaller magnitude, .7 on the BMI scale over the course of 6 months. Sample demographics vary greatly between these 2 studies. At the same time, it is very encouraging to observe such a large effect. Further, individuals who completed more game activities (ie, cleaning the island, playing mini games), reported being in contemplation stage vs. preparation or action, were older, and who played 1-2 hours of video games per session tended to have higher final BMI scores.

According to the Transtheoretical Model [41-42], matching the intervention to cognitions and behaviors of individuals is of utmost important for ensuring success. Individualization of interventions for health behaviors, including diet and exercise, based on the individual's readiness to change helps to move him or her along the behavior change trajectory. This theory found previous empirical support in traditional forms of interventions for dietary behaviors [43-45]. In the context of games, Lin et al. [14] carried out a qualitative study of participants engaged in step count linked to a virtual fish avatar. They found that those in the pre-contemplation or maintenance stages (both ends of the scale) were the least likely to change their daily steps. Findings from our study further support the relationship between behaviors and cognitions assessed through individuals' readiness to change and their BMI. This has implications not only on how individuals are recruited but also clarifies the target audience and raises possibilities for

individualizing game content explicitly depending on individual characteristics.

For nutrition knowledge, we also observed a large increase from the first month of assessment to follow up. While nutritional knowledge was intentionally not assessed at baseline, it is encouraging to see that the increase continued even after the first month of game play. Thus, it is likely that current findings are an underestimate of the game effect on nutritional knowledge. Similar findings of a large increase in knowledge of a food pyramid have been previously observed with a computer game to promote healthy diet in young adults (average age = 20 years) after 1 month of game play [10]. In our study, individuals who completed more food Quests and, interestingly, fewer exercise Sparks demonstrated higher levels of knowledge at post-test. In addition, women further along on the readiness to change their physical activity scale had significantly higher post-scores. These results support the fact that nutritional knowledge can be changed in a course of three months, even sporadic, game play intervention.

SpaPlay Utilization

The overall low utilization of the SpaPlay program by the participants was a disappointing finding. On average, participants played for a week out of the 3 months, with only 50% playing more than 2 days. Gaps between logins were long, averaging 25 days (SD 14.23), and players largely exhibited an irregular play pattern as indicated by sporadic logins, estimated as an average standard deviation between logins of 16.48 (SD 7.66). Our previous usability studies [27,32] provided evidence for more frequent and consistent game play and generally good acceptance. Lower summaries from this study may be an artifact of design differences. The length of the qualitative study was 1 month, with interviews conducted regularly at one-week intervals, and interview questions informed by the play data [33]. This setup likely enforced more play than would occur naturally, without external reinforcement. These results may be a function of the instructions given to participants at baseline, where women were not explicitly told how often they should play the game. Rather, participants were expected to utilize the game as they would in real-life settings.

Further, we have examined game elements that were most frequently used. The participants favored briefer “Spark” activities over longer “Quests” that required a week-long commitment. It appears that short game features with immediate reinforcement were more appealing. A similar trend was found in our qualitative study and was largely expected.

We also found that some players engaged with the game more extensively than others, indicating a greater suitability of the game for some participants. This was supported by high variability scores in almost all game play elements, except for game and social activities. Women who, at baseline, reported being in the preparation and action stages of change on their diet played SpaPlay more frequently and consistently than those in the contemplation stage; they also completed more exercise Sparks. A theoretical application of the thranstheoretical model to game play adherence in the context of narrative-based games for health was proposed by Yin, Bickmore, and Montfort [46]; however, no empirical data were provided. Our findings present evidence that adherence to game play does depend on one’s cognitions and behaviors (as in preparation and action stages of change) and should be considered when designing and potentially individualizing games. Further research is needed to understand the construct of stages of changes more intimately in the context of games for health.

Another finding is that women who report playing video games 1-2 hours at a time had lower play frequency and consistency in the current game and completed fewer Sparks and Quests. Thus, a preference for spending a lot of time playing video games does not naturally translate into playing a health-related game. Taylor [47] discussed motivational factors for women who participate in multiplayer online gaming environments, which include social interaction, mastery, status, team participation, and exploration. The different style and context of SpaPlay seems to not appeal to this subgroup of women. This finding is important in understanding who might adopt or disregard this type of intervention. Interviewing women who engage in heavy bursts of game play could have shed more light into the nuances of the mismatch between their preferences and the current game, and could be a topic of a future investigation.

Limitations

Results of the current study should be interpreted with caution in light of certain limitations. First, this study included a convenience sample of participants interested in playing computer games and who were at least contemplating losing weight. While this is a limitation, the sample is reflective of the population for whom we considered the game to be the most suitable and thus might resemble individuals who would seek a computer game to change their diet and exercise habits.

Second, the lack of a control group precludes complete certainty around the cause of weight loss and increase in nutritional knowledge. Given the pilot nature of the study, participants served as their own controls, which is strong but not absolute evidence of the game effectiveness. This could be an avenue of future research.

Third, more study follow-up data would have been beneficial for learning detailed lessons about players’ experiences. While our qualitative data from the previous study provides insight about satisfaction and acceptability [27,32], more data could provide additional insights into what works and what does not work when playing the game in more naturalistic settings. This should be considered in future studies, while considering participants’ fatigue from questionnaire responses.

Finally, all study measures were based on self-report. Since the current study is a part of an ongoing process of SpaPlay development, a new iteration of the game integrates objective measures of physical activity via Fitbit that is directly linked to game environment. Further investigations will make use of these objective data.

Conclusion

As a concept, “games for health” is relatively new, with very few studies systematically and comprehensively evaluating validity and effectiveness. The current study incrementally contributes to the field, and highlights the complexity of several

issues related to adherence, deployment in real life, and individualization. However, it is encouraging to see the ability of the game to change BMI and nutritional knowledge, both important targets of many health interventions.

Authors' Contributions

MS conceptualized the manuscript, wrote the first draft and incorporated changed from all co-authors, completed data analyses, and served as a co-PI on the original study. SH assisted with data analyses and manuscript preparation and constructed Figures. MSE is the PI on the original study and was involved in conceptualization of the manuscript and contributed to drafts. SD was involved in data collection and processing and contributed to drafts. CCS is a co-PI on the original study, was involved in conceptualization of the manuscript and contributed to drafts.

Conflicts of Interest

None declared.

References

1. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the United States, 2011-2012. JAMA 2014 Feb 26;311(8):806-814. [doi: 10.1001/jama.2014.732] [Medline: 24570244]
2. Rubenstein AH. Obesity: a modern epidemic. Trans Am Clin Climatol Assoc 2005;116:103-111 [FREE Full text] [Medline: 16555609]
3. Allison DB, Fontaine KR, Manson JE, Stevens J, VanItallie TB. Annual deaths attributable to obesity in the United States. JAMA 1999 Oct 27;282(16):1530-1538. [Medline: 10546692]
4. Cawley J, Meyerhoefer C. The medical care costs of obesity: an instrumental variables approach. J Health Econ 2012 Jan;31(1):219-230. [doi: 10.1016/j.jhealeco.2011.10.003] [Medline: 22094013]
5. Eldon E. Social Times.: Adweek; 2011 Jul 01. Zynga Files S-1, Plans to Go Public and Raise $1 Billion, Overcoming Credits Costs URL: http://www.adweek.com/socialtimes/zynga-files-s-1-plans-to-go-public [accessed 2016-01-29] [WebCite Cache ID 6euLqVtxT]
6. Read JL, Shortell SM. Interactive games to promote behavior change in prevention and treatment. JAMA 2011 Apr 27;305(16):1704-1705. [doi: 10.1001/jama.2011.408] [Medline: 21447802]
7. Papastergiou M. Exploring the potential of computer and video games for health and physical education: a literature review. Computers & Education 2009 Nov;53(3):603-622. [doi: 10.1016/j.compedu.2009.04.001]
8. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
9. Moller AC, Majewski S, Standish M, Agarwal P, Podowski A, Carson R, et al. Active fantasy sports: rationale and feasibility of leveraging online fantasy sports to promote physical activity. JMIR Serious Games 2014 Nov;2(2):e13 [FREE Full text] [doi: 10.2196/games.3691] [Medline: 25654304]
10. Peng W. Design and evaluation of a computer game to promote a healthy diet for young adults. Health Commun 2009 Mar;24(2):115-127. [doi: 10.1080/10410230802676490] [Medline: 19280455]
11. Ruggiero L, Moadsiri A, Quinn LT, Riley BB, Danielson KK, Monahan C, et al. Diabetes island: preliminary impact of a virtual world self-care educational intervention for african americans with type 2 diabetes. JMIR Serious Games 2014 Aug;2(2) [FREE Full text] [doi: 10.2196/games.3260] [Medline: 25584346]
12. Lieberman D. Dance Dance Revolution: the most researched serious game ever. Why, and what have we learned? 2005 Presented at: Serious Games Summit; 11/01/2005; Washington, DC.
13. Mokka S, Väätänen A, Heinilä J, Välkkynen P. Fitness computer game with a bodily user interface. 2003 Presented at: The Second International Conference on Entertainment Computing; May 8, 2003; Pittsburgh, PA p. 2003. [doi: 10.1145/958720.958729]
14. Lin J, Mamykina L, Lindtner S, Delajoux G, Strub H. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In: UbiComp 2006: Ubiquitous Computing. New York City: Springer; 2006:261-278.
15. Krebs P, Burkhalter JE, Snow B, Fiske J, Ostroff J. Development and alpha testing of QuitIT: an interactive video game to enhance skills for coping with smoking urges. JMIR Res Protoc 2013;2(2):e35 [FREE Full text] [doi: 10.2196/resprot.2416] [Medline: 24025236]
16. Coonradt C. The Game of Work. Layton, Utah: Gibbs Smith; 2012.
17. Cugelman B. Persuasive Design Toolkit. Toronto, Canada: AlterSpark Corp; 2013.
18. Prensky M. Fun, Play and Games: What Makes Games Engaging. Digital Game-Based Learning. New York City: McGraw-Hill; 2011.
19. Reeves B, Read J. Total Engagement: How Games and Virtual Worlds Are Changing the Way People Work and Businesses Compete. 1st edition. Boston, MA: Harvard Business Review Press; 2009.

20. Zichermann G. Gamification Corp. 2011 Nov 29. The Six Rules of Gamification URL: http://www.gamification.co/2011/11/29/the-six-rules-of-gamification/ [accessed 2016-01-29] [WebCite Cache ID 6euMtXINW]
21. 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]
22. Deterding S, Dixon D, Khaled R, Nacke L. From Game Design Elements to Gamefulness: Defining. New York City: ACM; 2011.
23. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
24. Jeffery RW, Drewnowski A, Epstein LH, Stunkard AJ, Wilson GT, Wing RR, et al. Long-term maintenance of weight loss: current status. Health Psychol 2000 Jan;19(1 Suppl):5-16. [Medline: 10709944]
25. Burke LE, Steenkiste A, Music E, Styn MA. A descriptive study of past experiences with weight-loss treatment. J Am Diet Assoc 2008 Apr;108(4):640-647. [doi: 10.1016/j.jada.2008.01.012] [Medline: 18375220]
26. West DS, Gorin AA, Subak LL, Foster G, Bragg C, Hecht J, Program to Reduce Incontinence by DietExercise (PRIDE) Research Group. A motivation-focused weight loss maintenance program is an effective alternative to a skill-based approach. Int J Obes (Lond) 2011 Feb;35(2):259-269 [FREE Full text] [doi: 10.1038/ijo.2010.138] [Medline: 20680012]
27. El-Nasr M, Andres L, Lavender T, Funk N, Jahangiri N, Sun M. Encouraging and sustaining healthy living through social games. 2011 Presented at: IEEE International Games Innovation Conference (IGIC); 11/3/2011; Orange, CA. [doi: 10.1109/IGIC.2011.6115122]
28. Patrick H, Canevello A. Methodological overview of a self-determination theory-based computerized intervention to promote leisure-time physical activity. Psychol Sport Exerc 2011 Jan;12(1):13-19 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.011] [Medline: 21103069]
29. Rigby S, Ryan R. Glued to Games: How Video Games Draw Us In and Hold Us Spellbound. Santa Barbara, Calif: Praeger; 2011.
30. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
31. Rigby S, Ryan R. Immersyve » Science of Engagement. 2007. The Player Experience of Need Satisfaction (PENS) URL: http://www.immersyve.com/white-paper-the-player-experience-of-need-satisfaction-pens-2007/ [accessed 2016-01-29] [WebCite Cache ID 6euNSPyuT]
32. Durga S, El-Nasr M, Shiyko M, Sceppa C. Investigating game acceptance with data-driven retrospective interviewing: a case study with SpaPlay–a health-based game. 2015 Presented at: Human Computer Interaction; 08/02/2015; Los Angeles, CA.
33. Seif El-Nasr M, Durga S, Shiyko M, Sceppa C. Data-driven retrospective interviewing (DDRI): a proposed methodology for formative evaluation of pervasive games. Entertainment Computing 2015 Nov;11:1-19. [doi: 10.1016/j.entcom.2015.07.002]
34. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. Journal of Consulting and Clinical Psychology 1983;51(3):390-395. [doi: 10.1037//0022-006X.51.3.390]
35. Centers for Disease Control. 2015. About Adult BMI | Assessing Your Weight | Healthy Weight URL: http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/ [accessed 2016-01-29] [WebCite Cache ID 6euNrHfA6]
36. Parmenter K, Wardle J. Development of a general nutrition knowledge questionnaire for adults. Eur J Clin Nutr 1999 Apr;53(4):298-308. [Medline: 10334656]
37. Marcus B, Rakowski W, Rossi J. Assessing motivational readiness and decision making for exercise. Health Psychol 1992;11(4):257-261. [Medline: 1396494]
38. Raudenbush S, Bryk A. Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage; 2002.
39. Hox J, Moerbeek M, van de Schoot R. Multilevel Analysis: Techniques and Applications, Second Edition. New York: Routledge; 2010.
40. Pinheiro J, Bates D, DebRoy S, Sarkar D. CRAN. 2015. nlme: Linear and Nonlinear Mixed Effects Models URL: https://cran.r-project.org/web/packages/nlme/index.html [accessed 2016-01-29] [WebCite Cache ID 6euO44q7Q]
41. Prochaska J. A transtheoretical model of behavior change: Implications for diet interventions. In: Promoting Dietary Change in Communities: Applying Existing Models of Dietary Change to Population-Based Interventions. Seattle, WA: Fred Hutchinson Cancer Research Center; 1992:37-49.
42. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994 Jan;13(1):39-46. [Medline: 8168470]
43. Auld GW, Nitzke SA, McNulty J, Bock MA, Bruhn CM, Gabel K, et al. A stage-of-change classification system based on actions and beliefs regarding dietary fat and fiber. Am J Health Promot 1998;12(3):192-201. [Medline: 10176094]
44. Greene G, Rossi S, Reed G, Willey C, Prochaska J. Stages of change for reducing dietary fat to 30% of energy or less. J Am Diet Assoc 1994 Oct;94(10):1105-1110. [Medline: 7930314]
45. Povey R, Conner M, Sparks P, James R, Shepherd R. A critical examination of the application of the transtheoretical model's stages of change to dietary behaviours. Health Educ Res 1999 Oct;14(5):641-651 [FREE Full text] [Medline: 10510072]

46. Yin L, Bickmore T, Montfort N. An Interactive Narrative System for Narrative-Based Games for Health. 2014 Presented at: Seventh Intelligent Narrative Technologies Workshop; 06/17/2014; Milwaukee, WI.
47. Taylor T. Multiple pleasures: women and online gaming. Convergence: The International Journal of Research into New Media Technologies 2003 Mar 01;9(1):21-46. [doi: 10.1177/135485650300900103]

Abbreviations

BMI: body mass index
MLM: multilevel modeling
nlme: nonlinear mixed effects
PENS: Player Experience of Need Satisfaction
TTM: Transtheoretical Model of Behavior Change

Edited by G Eysenbach; submitted 24.07.15; peer-reviewed by E Seltzer, Z Lv; comments to author 19.08.15; accepted 20.12.15; published 02.06.16.

Please cite as:
Shiyko M, Hallinan S, Seif El-Nasr M, Subramanian S, Castaneda-Sceppa C
Effects of Playing a Serious Computer Game on Body Mass Index and Nutrition Knowledge in Women
JMIR Serious Games 2016;4(1):e8
URL: http://games.jmir.org/2016/1/e8/
doi:10.2196/games.4977
PMID:27255497

Original Paper

Epic Allies: Development of a Gaming App to Improve Antiretroviral Therapy Adherence Among Young HIV-Positive Men Who Have Sex With Men

Sara LeGrand[1], PhD; Kathryn Elizabeth Muessig[2], PhD; Tobias McNulty[3], BA; Karina Soni[4], BA; Kelly Knudtson[4], MPH; Alex Lemann[3], MS; Nkechinyere Nwoko[3], BA; Lisa B Hightow-Weidman[4], MPH, MD

[1]Center for Health Policy and Inequalities Research, Duke Global Health Institute, Duke University, Durham, NC, United States

[2]Department of Health Behavior, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[3]Caktus Consulting Group, LLC, Durham, NC, United States

[4]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Sara LeGrand, PhD
Center for Health Policy and Inequalities Research
Duke Global Health Institute
Duke University
310 Trent Drive, Room 305
Durham, NC, 27710
United States
Phone: 1 9194380448
Fax: 1 9196810714
Email: sara.legrand@duke.edu

Abstract

Background: In the United States, the human immunodeficiency virus (HIV) disproportionately affects young men who have sex with men (YMSM). For HIV-positive individuals, adherence to antiretroviral therapy (ART) is critical for achieving optimal health outcomes and reducing secondary transmission of HIV. However, YMSM often struggle with ART adherence. Novel mobile phone apps that incorporate game-based mechanics and social networking elements represent a promising intervention approach for improving ART adherence among YMSM.

Objective: This study used a multiphase, iterative development process to create an ART adherence app for YMSM.

Methods: The three-phase development process included: (1) theory-based concept development jointly by public health researchers and the technology team, (2) assessment of the target population's ART adherence needs and app preferences and development and testing of a clickable app prototype, and (3) development and usability testing of the final app prototype.

Results: The initial theory-based app concept developed in Phase One included medication reminders, daily ART adherence tracking and visualization, ART educational modules, limited virtual interactions with other app users, and gamification elements. In Phase Two, adherence needs, including those related to information, motivation, and behavioral skills, were identified. Participants expressed preferences for an ART adherence app that was informational, interactive, social, and customizable. Based on the findings from Phase Two, additional gaming features were added in Phase Three, including an interactive battle, superhero app theme, and app storyline. Other features were modified to increase interactivity and customization options and integrate the game theme. During usability testing of the final prototype, participants were able to understand and navigate the app successfully and rated the app favorably.

Conclusions: An iterative development process was critical for the development of an ART adherence game app that was viewed as highly acceptable, relevant, and useful by YMSM.

(JMIR Serious Games 2016;4(1):e6) doi:10.2196/games.5687

KEYWORDS

mobile applications; video games; serious games; HIV; medication adherence; health knowledge, attitudes, practice; youth; men who have sex with men

Introduction

In the United States, youth accounted for approximately 26% of new human immunodeficiency virus (HIV) infections in 2010 [1]. Among youth, young men who have sex with men (YMSM) accounted for 72% of new HIV infections and were the only risk group that experienced a significant increase in HIV incidence [1,2]. Once diagnosed with HIV, youth are less likely to engage in HIV care, receive a prescription for antiretroviral therapy (ART), and have sustained adherence [3]. For those on ART, daily, lifelong adherence is necessary to maximize health benefits and reduce the likelihood of onward HIV transmission [4-8]. Novel, sustainable interventions that improve adherence to ART among HIV-positive YMSM are needed to improve individual health, decrease health care costs, and reduce HIV transmission risk [8].

The near saturation of the smartphone market among youth and young adults in the United States has created opportunities for reaching a large number of young people with health behavior interventions, including those that address ART adherence [9]. In a recent review of the literature on smartphone, Internet, and Web 2.0 interventions for HIV prevention and care, 10 published or ongoing studies were identified that used smartphones to improve ART adherence, including the intervention described in this paper, Epic Allies. Epic Allies was the only intervention identified that was explicitly designed to meet the specific adherence needs of YMSM [10].

Games are increasingly used to address behavioral and psychological factors associated with adherence to medical treatment regimens [11]. Games are goal-oriented, immersive, challenging, and motivating and can be used to improve attitudes and self-efficacy for health behavior change [12-15]. Social engagement and provision of support are also powerful tools for behavior change, particularly for HIV-positive YMSM who often experience social isolation from HIV-related stigma and homophobia [16-18]. Social networking is one tool that can be used to connect individuals around a specific health issue and allow for the provision and receipt of social support [19]. Because youth and young adults are the most avid users of games and social media [20,21], inclusion of gaming and social networking-based elements into mobile phone applications represents a promising health behavior change intervention strategy. The purpose of this study was to develop Epic Allies, a gaming app with behavior tracking features and social networking elements, to improve ART adherence among HIV-positive YMSM. This paper highlights the importance of using an iterative design process, including obtaining feedback from YMSM at each stage of development, to achieve optimal app design and functionality.

Methods

Introduction

The Epic Allies development team was assembled in September 2013. The team consisted of medical and public health researchers from University of North Carolina (UNC) and Duke University (Duke) and a technology team from Caktus Consulting Group (Caktus) with expertise in app design, development, and programming. During the initial meeting, the team agreed to a three-phase approach to app development. The first phase included the development of the initial app concept, which was theoretically grounded in the Information, Motivation, and Behavioral Skills (IMB) model of behavior change (September 2013) [22]. The second phase involved the development and refinement of an early prototype of selected app features and collection of formative data from HIV-positive YMSM, ages 18 to 29, to inform app development (October-December 2013). The third phase consisted of final app prototype development and internal and external usability testing (January-May 2014).

Phase One

Weekly team meetings focused on app concept development based on the IMB model. The IMB, which has frequently been used to guide the development of ART adherence interventions, conceptualizes health behavior change (eg, medication adherence) as a product of mediators including information about the behavior, motivation to change, and the skills needed to achieve change [22]. Using this model as a guide, our multidisciplinary team outlined and iteratively created a paper prototype of potential features, including gaming and social networking elements.

Phase Two

At the beginning of Phase Two, a clickable prototype of the Epic Allies app was developed for the Android operating system. Three focus groups were then conducted in Raleigh, Durham, and Charlotte, North Carolina with 20 HIV-positive YMSM, ages 20 to 28 to (1) assess ART adherence information, motivation, and behavioral skills needs, (2) determine strategies to address these needs via a mobile app, and (3) gather feedback on the evolving features of Epic Allies and future feature concepts. Before the focus groups began, each participant completed a brief survey on sociodemographics, Internet/mobile phone use, and ART adherence and selected a pseudonym to protect participant confidentiality during the focus group. After discussing adherence needs and strategies for addressing them through an app, participants used study smartphones to explore and provide feedback on the current Epic Allies prototype. The duration of the focus groups was approximately 90 minutes. Groups were audio recorded, transcribed, and then analyzed using Dedoose qualitative data analysis software [23] to identify a range of themes across participants. Quotes representing common responses and variations within each theme were identified by study team consensus. After each focus group, the study team met to discuss preliminary findings and themes and to identify key features to add or remove from the Epic Allies prototype. An updated clickable prototype was created prior to each subsequent focus group in an iterative development process.

Phase Three

The team worked collaboratively to modify the initial conceptual design of the app based on the focus group findings and created a final plan for app features and functionality. The final app prototype was developed incrementally with components added and refined weekly. Research staff members from UNC and

Duke conducted ongoing testing of new versions of the prototype to identify bugs and provide feedback on the user interface and navigation experience. After a full prototype was completed, five new internal testers assessed all functionality of the app to identify any remaining technical glitches or usability concerns. Caktus used the detailed feedback from this internal testing to develop a polished version of the final app prototype for external usability testing.

In April and May 2014, external usability testing was conducted with seven HIV-positive YMSM, ages 20 to 28. External usability testing aimed to assess whether users: could successfully navigate features and functions of the app, could comprehend the educational content, and found the app to be engaging and relevant. After providing informed consent and completing a brief survey on sociodemographics, Internet/mobile phone use, and ART adherence, participants met with a member of the development team and a member of the research team to "walk through" the app. Other members of the team (including developers and researchers) watched a live audio-video stream of the testing in a separate room. To protect participant confidentiality during the audio-video streaming, participants used pseudonyms and the video only captured the participant's hand movements and the app screen. Participants were given a checklist of tasks to complete within the app at hypothetical intervention days 1, 21, and 42. These days were selected to give usability participants an opportunity to test features they would see on day 1, representing the first time they had used the app, and on days 21 and 42 after they had been tracking their adherence and gaining power and points within the game. A checklist was provided to guide users through the tasks for each day. Participants were asked to "think aloud" and share thoughts and questions that came up as they were completing tasks. After completing the tasks and providing feedback, participants completed a 9-question survey adapted from the System Usability Scale [24-27]. Usability testing sessions were recorded, transcribed and analyzed using the same methods described for the focus groups. Quantitative survey and usability results were summarized using Excel.

Results

Phase One

The initial Epic Allies app concept included features designed to address potential IMB ART adherence needs of HIV-positive YMSM based on clinical expertise and a review of the literature. The initial proposed features and their relationship to the IMB are highlighted in Table 1.

Phase Two

Descriptive Statistics

The mean age of focus group participants (n=20) was 24 years. All participants identified as black or African American. Most (15/19, 79%) earned less than US$20,999 per year. All participants owned a smartphone (20/20, 100%) and (11/20, 55%) had used a health app in the past 3 months. There was substantial variation in medication adherence patterns among the 19 participants currently on ART: five had missed at least one dose in the past week, five had missed at least one dose in the past month, and four reported never missing any doses. Key focus group findings are organized into three sections: medication adherence information, motivation, and behavioral skills needs; strategies for addressing medication adherence information, motivation, and behavioral skills needs in an app; and strategies to motivate app use.

Medication Adherence Information, Motivation, and Behavioral Skills Needs

Information needs identified by participants included understanding dosing times and information about side effects and side effects management. Participants also expressed confusion about appropriate dosing times and the number of pills required per week to constitute medication adherence. In one focus group, participants discussed differences in provider instructions regarding missed doses.

Goku: *And sometimes it's like my doctor told me that if I miss my time there's like a 4-hour period or something like that, so I just try to catch it in between.*

Grayland: *Goku, it's strange that you say your doctor gave you a 4-hour time window to take your meds. Mine only gave me like an hour. I mean he [doctor] said if you happen to miss, you have an hour window. If you don't take it by then you might as well skip it and take it the next day.* [Goku & Grayland, Raleigh]

Side effects knowledge and management was also a consistent theme across focus groups. Participants felt that those new to medications or switching medications should be aware of potential side effects and how to cope with them.

Acceptance of HIV status was identified as a key adherence motivation need. Participants spoke of the denial and internalized stigma they experienced around their diagnosis and the fear they had around living with a highly stigmatized health condition that would require lifelong medical management.

> *Because it's not so much ignorance of other people, it's the ignorance that you have within yourself that you're battling with and, ya know, learning these new things. You really have to be mentally accepting to what the new reality is. Yeah it's a hard, ongoing process. It really continues [for me], extremely continues. It is happening now actually.* [Ichiban, Durham]

Those who had not accepted their status were less motivated to take their medication, as taking their medication was a daily reminder that they had HIV.

> *Yeah, cause honestly, it was a good few months before I ever took medication. And in that timeframe of diagnosis to taking medication, it was very easy for me to detach. It was very easy for me to say, this is not real, nahhh, whatever. It didn't become real until I had to take a pill. When you take a pill, it's real.* [Brett, Raleigh]

> *A lot of the people who you say do forget to take their medications, those are the people who have the mind set of they don't want the disease, or the health condition, to be a banner for them, they don't want*

that to be the representation of them. [Crowned, Charlotte]

Social support had a strong influence on motivation to adhere to medications. Some of the participants who disclosed their status reported receiving social support that helped them deal with the fear and uncertainty of an HIV diagnosis. The support often motivated individuals to receive HIV care and adhere to medications.

Then I had a moment, I was just like, 'Oh my god, my mom's gonna kill me, she's gonna kick me out of the house.' My mom calls me on the way home from work. She's like, 'So what did the doctor say? Yeah, he said, oh, so you got it [HIV]? Okay, alright, so we'll call the doctor tomorrow and see if we can get you some pills.' I was like, thank you mommy. [Brett, Raleigh]

However, many participants did not feel they had people that they could talk to about their diagnosis due to both the fear of HIV stigma or previously experienced stigma associated with their sexuality. Individuals who had not disclosed their status to their social circle often relied on social workers, case managers, or therapists for adherence motivation.

Participants who experienced medication side effects also had less motivation and lower self-efficacy in their ability to fully adhere to their medication.

Yeah, it makes me like really don't want to take the medications because I really don't want to have to deal with this or if I have to wake up early and I'm still under the medication and I'm still like, I just be drowsy, and like not wanting to. And sometimes like my friends be like you're so much meaner when you are on your medications cause I'll be like, don't want to be bothered sometimes. [Vee, Durham]

Finally, some participants had difficulty integrating medication routines into their daily lives. For example, some men had trouble remembering to take their medications.

Well for me, I know how to take my pill. I do, I think with taking a pill, it's just like anything else. You may forget. You know what I mean? It's something that you do on a daily basis. But for the most part, you may forget to take a pill, it happens. [Jerry, Durham]

Strategies for Addressing Medication Adherence Information, Motivation, and Behavioral Skills Needs in an App

After discussing general adherence needs, participants were asked about the ways an app could help them with adherence. Medication reminders were suggested as an important tool for those who had difficulty remembering to take their medications. Participants emphasized the importance of discreet reminders so that their HIV status would not be revealed if someone saw the message on their phone. Several participants felt that it was important to provide other types of information through the app, including current information about HIV.

It needs something that's going to keep the person's attention…maybe a motivational update section where people could post things that are going on with HIV advances, technologies, empowerments [affirmations]…like a news feed or some thing that will keep people intrigued…staying updated and abreast with all of the new HIV information…That would keep me interested. [Stew, Raleigh]

Participants also suggested that the app could also be used to deliver important information about adherence at critical times, such as starting medications for the first time or switching medications.

[the app] could be like a whole lot of help especially for people who don't know some of the side effects they might have to the medication. Like the previous medication that I was on, I had to eat a certain amount of calories when I took the medicine and that would help with questions of that nature and all. Because a whole lot of people just take it. They don't know its rules to really taking it. Some medicine you got to stay hydrated, some medicine you got to eat, so it's all kinds of different things that I think [the app] should be able to help with. [Orleans, Charlotte]

Several men were in favor of developing an app that could be used to connect with others who may be dealing with similar adherence challenges. Many emphasized the importance of maintaining anonymity in these interactions.

I guess where you can interact with the other users…that way you could find people maybe you can talk to and be in confidence. You wouldn't have to talk to them face to face…but if you're having a bad day you could maybe post it and somebody might be going through the same thing and you guys could kinda talk cause, I mean you never know what somebody else might be going through that's going through the same things that might help. [Batman, Durham]

Strategies to Motivate App Use

Participants were asked about features that would motivate them to use a medication adherence app. Several overarching themes emerged including the importance of creating an app that is interactive, engaging, social, informational, customizable, and personalized. The men noted that these features would help capture their attention, motivate them to use the app regularly, and improve the likelihood of sustained use.

Participants emphasized the importance of interactivity. Games were seen as a good way to increase app use. In addition, many participants thought rewards for activities within the app would be highly motivating and promote ongoing engagement. Customization was important so users could selectively choose the features most relevant for them.

Put it this way: you want people to use the app daily, it needs to be as comfortable for them as possible. They need to be able to do whatever they want to do with it…because it's their phone, it's their medication, it's their health care. [Brett, Raleigh]

Avatars were discussed as a way for users to have a virtual representation within the app and were highly acceptable. There

were different perspectives on the level of customization needed to sustain interest, but all participants preferred a personalized avatar.

Phase Three

Development of Full Prototype for Usability Testing

Focus group findings informed the development of the full prototype for usability testing. Changes to the features initially conceptualized in Phase One are highlighted in Table 2. The app storyline and screenshots of selected features are included in Textbox 1 and Figures 1-3.

Table 1. Initial proposed features of Epic Allies.

Planned feature	Planned function	Relationship to IMB model
Customizable avatars		
	Visual representation of users within the app to facilitate development of an online identity while preserving anonymity to peers.	M[a]: Increases app engagement and facilitates social support.
Dashboard		
	Home screen where users enter daily adherence information. Provides a visualization of historical adherence patterns.	I[b]: Tracking and historical visualization of data provides information on adherence behaviors.
		M: Data display visually reinforces positive behaviors and motivates users to change if adherence is suboptimal.
		B[c]: Adherence achievements increase behavior change self-efficacy.
Reminders and tailored feedback messages		
	Personalized medication reminders and feedback messages based on users' successes, setbacks and progress toward their adherence goals.	I: Reminders specify appropriate times to take medications.
		M: Reminders provide a cue to action. Feedback messages affirm positive adherence behaviors and identify adherence challenges.
		B: Reminders help build skills for integrating medication routines in daily life. Feedback messages increase adherence self-efficacy.
Friends		
	Users select virtual "friends" to interact with by sending preset messages to challenge, praise or encourage others.	M: Increases social motivation to adhere to medications.
		B: Ally interactions create opportunities for peer modeling and reinforcement of adherence.
Information modules		
	Education modules on ART adherence and HIV.	I: Modules provide relevant ART adherence information.
Gamification		
	Users earn points for completing selected tasks within the app.	M: Opportunities to earn points increases motivation for app engagement and ART adherence.
		B: Achieving milestones increase adherence self-efficacy.

[a]motivation.

[b]information.

[c]behavioral skills/self-efficacy.

Table 2. Changes to Epic Allies for Phase Three

Phase One and Two components	Phase Three components	Phase Three component description	Rationale for change
–	Superhero theme, app storyline, virtual guide	An app storyline was developed to explain the superhero theme, provide a backstory for the app and to introduce users to their role in the game (Textbox 1). A virtual guide, Walter, was created to present the storyline to users and provide a tour of the app.	The storyline and superhero theme were added in response to focus group participants' request for an engaging and interactive user experience that would motivate regular app use.
–	The Battle	The objective of the Battle is to collect virtual cards used to defeat the monsters in the city of Medopolis. Each day, players fight a new battle to defeat a monster; success or failure depends on the set of three cards they have in their hand. Players can buy and upgrade their cards using points earned by engaging in other parts of the app. Within the Battle, a new social component called the "Spotlight" was created. Using the spotlight, participants can call on their Allies to help defeat difficult monsters.	The Battle feature was introduced to increase app engagement, motivate participants to complete activities within the app, and encourage interactions with Allies.
Customizable avatars	Profile	The avatar concept was expanded to include more detailed information about users such as interests, hobbies, and current ART use.	Focus group participants emphasized the need for social support for adherence. The profile was changed to create a stronger sense of community among users.
Dashboard	Dashboard	The Dashboard feature expanded to include additional tracking options such as exercise, smoking, drug and alcohol use, and mood.	This change responded to focus group participants' request for customization within the app.
Reminder messages and tailored feedback	Reminder messages and tailored feedback	Reminder messages were made optional. Tailored feedback was expanded to include messages on new tracking options (ie, exercise, smoking).	Focus group participants noted that reminders are an important feature to include in an adherence app. However, those who already have strategies for remembering to take their medications may not need them. The change allows users to customize the app to their needs. Expansion of tailored feedback was designed to provide additional customized feedback on factors related to adherence.
Friends	Allies	The name was changed from "Friends" to "Allies". The feature was expanded to include the "Spotlight" feature described in the Battle section above.	The name was changed to align with the storyline and superhero theme. The "Spotlight" feature enhanced the ability for users to interact with their Allies in an interactive and fun way.
Information modules	Daily Dose	The information modules were replaced with the Daily Dose, an app newspaper that follows a curriculum of daily short articles and tips to address HIV and ART knowledge and promote disease management. Reading an article earns users points that can be used to buy new cards for the Battle.	The Daily Dose replaced the educational modules so that informational needs could be addressed in an engaging and interactive manner (ie, proper dosing guidelines, common medication side effects, coping with side effects, HIV acceptance process, identifying sources of social support). Users are encouraged to log into the app daily to get a new article. Awarding points increases motivation for reading articles and app engagement.
Gamification	Gamification	The gamification principles were expanded so that users earned points for each completed activity in the app. A "leveling up" feature was added so that users could unlock new battles.	The additional gamification features were added to increase motivation for behavior change and app use.

Textbox 1. App Storyline.

The year is 2024, a pharmaceutical plant has just exploded sending poisonous rays throughout the city of Medopolis. Walter, the scientist responsible for the accident, has managed to escape, though not without sustaining some damage. The poison has wreaked havoc in Medopolis and unleashed monsters who are trying to destroy the city. Only you and your superhero allies can help. Walter will be your guide as you work together to gain knowledge and power to fight the monsters and restore Medopolis to its former glory.

Figure 1. App tutorial.

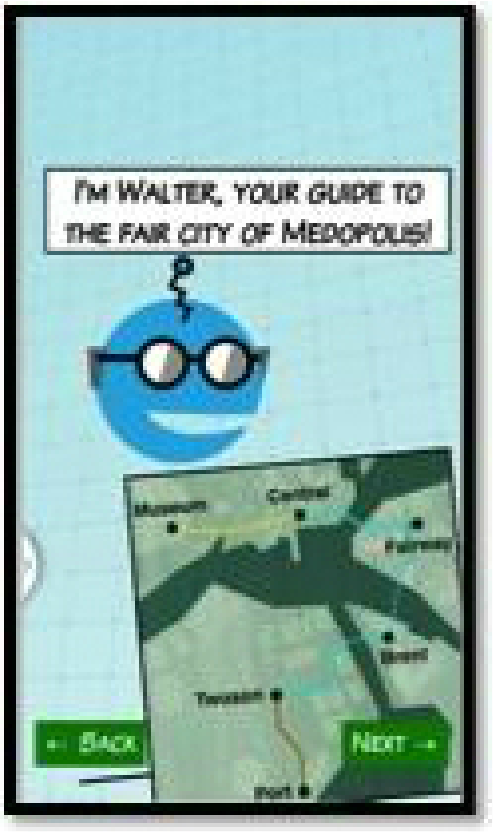


Figure 2. The Daily Dose.

Figure 3. The Battle.

Usability Testing Results

The mean and median age of participants was 23 years (n=7). Six participants identified as black or African American and one identified as white. Over half (4/7) earned less than US$20,999 per year. All participants owned a smartphone and more than half (4/7) had used a health app in the past 3 months. All participants were on ART to treat their HIV infection, but less than a third (2/7) said they never missed any of their medication.

Usability testing assessed participants' ability to successfully navigate the app, comprehend the educational content, and determine if they found the app to be engaging and relevant. Usability scores are presented in Table 3. For all items, the mean responses indicated a favorable evaluation (agree/strongly agree) of Epic Allies. All participants were able to successfully complete the checklist of tasks representing days 1, 21, and 42 of app use. The checklist included tasks such as entering medications, setting up a medication reminder, reading the Daily Dose newspaper, upgrading Battle cards, playing Battles, and calling on an Ally to help them during a battle using the "Spotlight" card (Figure 4).

Table 3. Usability testing score means (n=7).

	Mean	Standard Deviation
Visually appealing	1.9	1.2
Overall impression is favorable	1.6	0.8
Medication tracking features easy to understand	1.7	1.1
Layout and structure easy to understand and navigate	2.0	1.0
Functions were easy to use	1.6	1.1
Interesting	1.3	0.5
Could help with medication taking	1.4	0.8
Can see the benefits of using an app like this	1.6	1.0
Could see myself using an app like this	1.7	1.0

[a]Score key: 1=strongly agree to 5=strongly disagree.

Overall, usability testing participants found the app to be engaging and relevant to their lives. As one man noted:

> *[I]t would be a game that I would play every day and it would make me, you know, it would make me want to join more programs like this to help others with HIV. And it would keep me on my medicine, keep exercising, and keep me motivated, whatever. I would be good. If I ever do get in the mood, I take this game out and play it. And, it's fun. And it actually educates people that don't know anything about HIV.* [Participant 5]

Participants found the informational content delivered through the Dashboard and the Daily Dose features informative and easy

to comprehend. Remarking on the information provided through the Dashboard:

> *To be able to see and realize, ok well last month I had this much bad days and this much good days to be able to go back and think what was I doing last month comparatively this month, maybe thinking about why I'm having better days this month. Just being able to monitor it I guess and seeing the difference.* [Participant 4]

> *The fact that it monitors your mood, your medication, and your exercise is a really good thing. To be able to monitor and to be able to see a pattern of how your days usually go would make me feel a little bit better.*" [Participant 2]

Participants also commented on the Daily Dose content:

> *I thought [the Daily Dose articles] were all informative pieces and they definitely all hit things that [are important]… but looking back early in my diagnosis I can't say I knew all of these things.* [Participant 1]

> *It's something I can relate to, why not read up on it and understand more about something you're living with. This is your life. You wanna be educated on as much as possible on what's going on because this day and age things change. And it gets real really quick.* [Participant 4]

Several participants commented on the importance of the Allies social networking element. The overarching theme was that the feature could help participants feel they are not alone in taking medications for HIV.

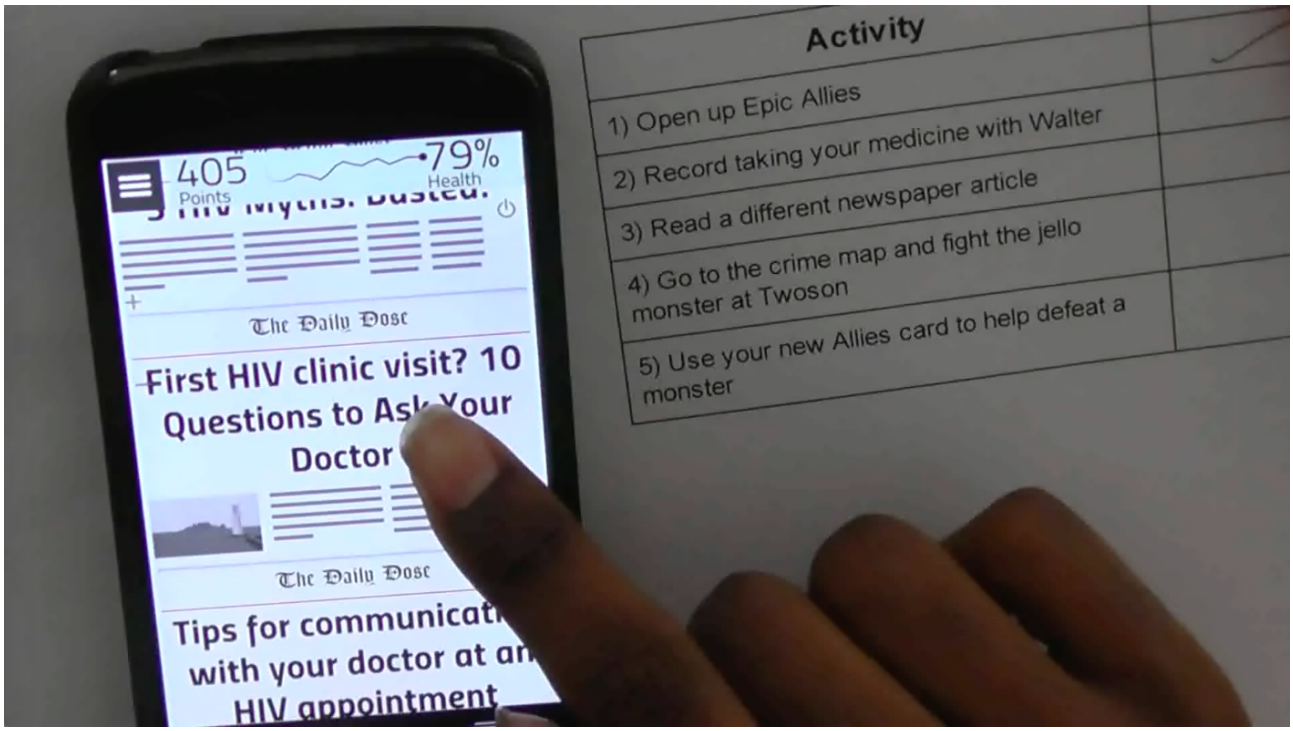
> *The app gives you the sense that you aren't alone. I'm not the only one keeping up with my meds everyday and playing the game everyday.* [Participant 3]

While most feedback about the app was positive, participants also made suggestions for app improvement. Some men felt the Battle feature of the app would be more engaging with increased interactivity. They expressed concerns about the turn-based mechanics in the card game:

> *Well, I don't like the fact that the enemy just automatically 'gets his turn' and I just gotta stand back for the enemy to just demolish me. I would really like, if it's possible, you know how like computer video games are…I guess more of an interactive fight.* [Participant 3]

After completing the tasks assigned for hypothetical Day 1 of app use, over half (4/7) of usability testing participants did not fully understand how the character in the Battle was associated with adherence. By the end of the usability session, all participants were able to describe the relationship between the Battle character and adherence. However, participants noted that it is important to clearly establish this link during the first day of app use.

Figure 4. Usability testing with checklist.



Discussion

Overview

Smartphone app interventions are increasingly used to promote positive health behaviors among individuals living with HIV, including ART adherence [10,28-30]. In this study, a smartphone app was viewed as a highly acceptable tool for the delivery of an ART adherence intervention. Other studies have also found that adult men who have sex with men and YMSM view smartphone apps as an acceptable delivery mechanism for HIV-related interventions [31-34]. However, to our knowledge,

no ART adherence apps have been designed to meet the unique adherence needs of YMSM. This study used an iterative development process to create a theory-based game adherence app for YMSM. Based on input and feedback from study participants, the initial app concept was incrementally modified and enhanced to create the final app prototype. This highlights the importance of seeking target population input and using an iterative process in the development of behavior change apps.

Focus group participants identified key IMB needs and proposed strategies for addressing these needs through a smartphone app. The men identified a need for specific information about adherence-related topics, such as dosing schedules and side effects, and recommended that the app provide general information about HIV and adherence. Medication side effects and side effects management was a specific area of interest because, as in other studies of HIV-positive youth side effects were identified as a barrier to adherence [35,36]. The Daily Dose feature, a virtual newspaper delivered through the app daily, was developed for the final prototype. The daily nature of the feature allowed for an increase in the amount of content delivered and provided an opportunity to address the specific and general informational needs identified in the focus groups.

A key theme associated with motivation for adherence was acceptance of HIV status. Several men noted that those who have not yet accepted their diagnosis might not be motivated to take their medications because doing so reminds them that they have HIV. Pill taking as a reminder of HIV infection has been found to be a key barrier to adherence in other studies of youth and young adults [35]. In contrast, social support was viewed as a motivating factor for adherence and participants recommended that the app include opportunities for users to give and receive social support. Among youth and young adults, social support has been identified as a facilitator for ART adherence while lack of support is a barrier [35,37]. In one study of racial and ethnic minority youth, greater social support predicted greater self-efficacy, which was associated with improved adherence [37]. In other studies, YMSM have indicated a preference for HIV-related apps that provide a platform for giving and receiving support [32,33]. Presenting opportunities for interactions with other HIV-positive YMSM through the Allies feature of Epic Allies may help those struggling with acceptance of HIV status feel like a part of a community and increase perceptions of support for adherence.

Consistent with other studies on ART adherence among youth and young adults [35,38], forgetting to take medications was identified as a challenge. Even though some men had developed strategies for remembering to take their medications and did not need reminders, participants agreed that reminders should be included in an ART adherence app. In response to the desire for customization, the medication reminder feature was made optional in the final prototype developed in Phase Three.

Participants also provided important guidance on app characteristics that would motivate regular and sustained app use. Interactivity through games, points, rewards, and social interactions were discussed as motivating factors for app use. Among adult and young MSM, interactivity has been emphasized as a necessary function of mobile apps [33,39]. The superhero theme and Battle feature introduced during the last focus group was viewed as a promising approach to further enhance app interactivity. While customization was viewed as important for meeting unique adherence needs, it was also important for motivation for app use. For example, participants noted that unnecessary reminders would be annoying and intrusive and increase the likelihood that they would discontinue app use. Another study of YMSM also found a preference for customization in HIV-related app interventions with unnecessary reminders and alerts resulting in app avoidance or deletion [33].

The feedback gathered during usability testing in Phase Three was generally positive and constructive and offered actionable guidance for changes to the app features prior to further testing. In response to suggestions provided during usability testing, the turn-based mechanics in the Battle were eliminated to increase interactivity and engagement. In order to quickly establish the link between real-world medication adherence and the character's strength and performance in the game, the virtual guide's initial app introduction was modified to explicitly explain the relationship. The Dashboard was changed so that it featured more prominently in the app and now serves as a constant reminder of the relationship between the user's adherence behaviors and game play.

Limitations

Our study was limited by its small sample size and geographical region. Though common in qualitative research, this limits the ability to generalize the study findings to all YMSM in the United States. Further, the majority of study participants identified as black or African American, which reflects the high burden of infection among black or African American individuals in study area [40]. While this may limit generalizability of findings to YMSM of other races or ethnicities, it should be noted that none of the app recommendations were unique to black MSM [39,41]. It is highly likely that the findings are relevant for YMSM of all races and ethnicities. However, future studies are needed to test the app with a more racially and ethnically diverse sample.

Epic Allies represents only one type of technology-based intervention that could be developed to address ART adherence among YMSM. Other less costly approaches could be used, such as interventions built on existing social media sites or adaption of currently available medication adherence apps for other medical conditions. However, our formative work with YMSM in the United States found a demand for an app intervention that integrates multiple features including behavior tracking, social networking features, and gaming features [32,33]. Existing commercial HIV apps often lack a theoretical basis, are limited in the topics they address and features they offer, have failed to capture the attention of end users and have not undergone rigorous evaluation [28,29]. If successful, the planned commercialization of Epic Allies would justify the costs invested and address the dearth of theoretically grounded, multicomponent commercial apps designed to address ART adherence.

The IMB is a well-established model of behavior change that has been used by researchers to guide the development of technology-based HIV interventions [42-44]. However,

modifications of existing behavior change models or theories, including the IMB, are needed to more effectively account for the dynamic nature of technology-based interventions [45,46]. Such modifications will help guide future intervention development and evaluation of engagement and effectiveness [45,46]. With an eye toward adapting existing theories, our team has developed alternative measures of engagement in technology-based interventions and will be exploring these and other approaches in the randomized controlled trial of Epic Allies [47]. Furthermore, we will seek to identify the specific elements of Epic Allies that contribute to changes in information, motivation, and behavioral skills and, ultimately, medication adherence.

The results from usability testing indicated a positive response to the Epic Allies app. However, these findings were limited to a walk-through of the app in a controlled environment. This type of testing, while appropriate for app development, may not reveal barriers to implementation in the real world. The app was carefully designed to quickly engage users, sustain motivation for daily, long-term app use, and minimize the usage of phone data; however, the success of these strategies will not be known until the app is tested in a randomized controlled trial. In spite of these limitations, this study provides key insights into a process of app development that engages the target population to develop an app appropriate for end-user needs and preferences.

Conclusion

Focus group and usability findings confirmed the appropriateness of a game-based app with social networking features to address ART adherence for YMSM. As evidenced by the changes made to the app between initial concept and final prototype, an iterative approach is critical for developing an app that is relevant, engaging and useful. This multistage process of development can be used to develop different types of health behavior apps for a wide range of populations.

Acknowledgments

This study was supported by the National Institute of Mental Health (R43MH102096).

Conflicts of Interest

Caktus Consulting Group, LLC developed and owns the Epic Allies app. Tobias McNulty and Alex Lemann are part owners of the company and Nkechinyere Nwoko is an employee.

References

1. Centers for Disease ControlPrevention (CDC). Vital signs: HIV infection, testing, and risk behaviors among youths - United States. MMWR Morb Mortal Wkly Rep 2012;61:971-976 [FREE Full text] [Medline: 23190571]
2. Johnson AS, Hall HI, Hu X, Lansky A, Holtgrave DR, Mermin J. Trends in diagnoses of HIV infection in the United States, 2002-2011. JAMA 2014;312:432-434. [doi: 10.1001/jama.2014.8534] [Medline: 25038362]
3. Zanoni BC, Mayer KH. The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities. AIDS Patient Care STDS 2014;28:128-135 [FREE Full text] [doi: 10.1089/apc.2013.0345] [Medline: 24601734]
4. Bangsberg DR, Perry S, Charlebois ED, Clark RA, Roberston M, Zolopa AR, et al. Non-adherence to highly active antiretroviral therapy predicts progression to AIDS. AIDS 2001;15:1181-1183. [Medline: 11416722]
5. Haubrich RH, Little SJ, Currier JS, Forthal DN, Kemper CA, Beall GN, et al. The value of patient-reported adherence to antiretroviral therapy in predicting virologic and immunologic response. California Collaborative Treatment Group. AIDS 1999;13:1099-1107. [Medline: 10397541]
6. Mannheimer S, Friedland G, Matts J, Child C, Chesney M. The consistency of adherence to antiretroviral therapy predicts biologic outcomes for human immunodeficiency virus-infected persons in clinical trials. Clin Infect Dis 2002;34:1115-1121 [FREE Full text] [doi: 10.1086/339074] [Medline: 11915001]
7. Paterson DL, Swindells S, Mohr J, Brester M, Vergis EN, Squier C, et al. Adherence to protease inhibitor therapy and outcomes in patients with HIV infection. Ann Intern Med 2000;133:21-30. [Medline: 10877736]
8. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, et al. Viral load and heterosexual transmission of human immunodeficiency virus type 1. Rakai Project Study Group. N Engl J Med 2000;342:921-929. [doi: 10.1056/NEJM200003303421303] [Medline: 10738050]
9. Pew RC. U.S. Smartphone use in 2015. 2015. URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6fQ26scWF]
10. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, internet and web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015;12:173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
11. Kato PM. Video games in health care: closing the gap. Review of General Psychology 2010;14:113-121. [doi: 10.1037/a0019441]
12. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22:77-89. [Medline: 9183781]

13. Baranowski T, Baranowski J, Cullen KW, Marsh T, Islam N, Zakeri I, et al. Squire's Quest! Dietary outcome evaluation of a multimedia game. Am J Prev Med 2003;24:52-61. [Medline: 12554024]
14. 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]
15. Lieberman D. Interactive video games for health promotion: effects on knowledge, self-efficacy, social support health. In: Street RL, Gold WR, Manning T, editors. Health Promotion and Interactive Technology: Theoretical Applications and Future Directions. Mahwah, NJ: Lawrence Erlbaum Associates; 1997:103-120.
16. Balaji AB, Oster AM, Viall AH, Heffelfinger JD, Mena LA, Toledo CA. Role flexing: how community, religion, and family shape the experiences of young black men who have sex with men. AIDS Patient Care STDS 2012;26:730-737. [doi: 10.1089/apc.2012.0177] [Medline: 23134543]
17. Lewis LJ, Kertzner RM. Toward improved interpretation and theory building of African American male sexualities. J Sex Res 2003;40:383-395. [doi: 10.1080/00224490209552204] [Medline: 14735412]
18. Stokes JP, Peterson JL. Homophobia, self-esteem, and risk for HIV among African American men who have sex with men. AIDS Educ Prev 1998;10:278-292. [Medline: 9642425]
19. Pew Internet American Life Project. The Social Life of Health Information, 2011. 2011. URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [WebCite Cache ID 6fQ7AfN5O]
20. Pew Research Center . Social media usage, 2005-2015. 2015. 2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2016-02-19] [WebCite Cache ID 6fQ7HX4Tu]
21. Pew Research Center . Adults and video games. 2008. URL: http://www.pewinternet.org/2008/12/07/adults-and-video-games/ [WebCite Cache ID 6fQ7MocxE]
22. Fisher JD, Fisher WA, Amico KR, Harman JJ. An information-motivation-behavioral skills model of adherence to antiretroviral therapy. Health Psychol 2006;25:462-473. [doi: 10.1037/0278-6133.25.4.462] [Medline: 16846321]
23. Dedoose Version 6.1.18, web application for managing, analyzing, and presenting qualitative and mixed method research data. Los Angeles, CA: SocioCultural Research Consultants, LLC; 2015. URL: http://www.dedoose.com/ [accessed 2016-04-19] [WebCite Cache ID 6gsmI2546]
24. Lewis J, Sauro J. The factor structure of the System Usability Scale. In: Human Centered Design: Lecture Notes in Computer Science. Berlin, Germany: Springer; 2009:94-103.
25. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 1995;7:57-78. [doi: 10.1080/10447319509526110]
26. Coursaris C, Kim D. A meta-analytical review of empirical mobile usability studies. Journal of Usability Studies 2011;6:117-171.
27. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: a test of alternative measurement models. Cogn Process 2009;10:193-197. [doi: 10.1007/s10339-009-0268-9] [Medline: 19565283]
28. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, technology, and HIV: recent advances and future directions. Curr HIV/AIDS Rep 2015;12:500-515. [doi: 10.1007/s11904-015-0280-x] [Medline: 26385582]
29. Muessig KE, Pike EC, Legrand S, Hightow-Weidman LB. Mobile phone applications for the care and prevention of HIV and other sexually transmitted diseases: a review. J Med Internet Res 2013;15:e1 [FREE Full text] [doi: 10.2196/jmir.2301] [Medline: 23291245]
30. Sullivan PS, Jones J, Kishore N, Stephenson R. The roles of technology in primary HIV prevention for men who have sex with men. Curr HIV/AIDS Rep 2015;12:481-488. [doi: 10.1007/s11904-015-0293-5] [Medline: 26519083]
31. Levy ME, Watson CC, Wilton L, Criss V, Kuo I, Glick SN, et al. Acceptability of a mobile smartphone application intervention to improve access to HIV prevention and care services for Black men who have sex with men in the District of Columbia. Digit Cult Educ 2015;7:169-191 [FREE Full text] [Medline: 26594251]
32. LeGrand S, Muessig KE, Pike EC, Baltierra N, Hightow-Weidman LB. If you build it will they come? Addressing social isolation within a technology-based HIV intervention for young black men who have sex with men. AIDS Care 2014;26:1194-1200 [FREE Full text] [doi: 10.1080/09540121.2014.894608] [Medline: 24617609]
33. Muessig KE, Pike EC, Fowler B, LeGrand S, Parsons JT, Bull SS, et al. Putting prevention in their pockets: developing mobile phone-based HIV interventions for black men who have sex with men. AIDS Patient Care STDS 2013;27:211-222 [FREE Full text] [doi: 10.1089/apc.2012.0404] [Medline: 23565925]
34. Holloway IW, Rice E, Gibbs J, Winetrobe H, Dunlap S, Rhoades H. Acceptability of smartphone application-based HIV prevention among young men who have sex with men. AIDS Behav 2014;18:285-296 [FREE Full text] [doi: 10.1007/s10461-013-0671-1] [Medline: 24292281]
35. MacDonell K, Naar-King S, Huszti H, Belzer M. Barriers to medication adherence in behaviorally and perinatally infected youth living with HIV. AIDS Behav 2013;17:86-93 [FREE Full text] [doi: 10.1007/s10461-012-0364-1] [Medline: 23142855]
36. Macdonell KE, Naar-King S, Murphy DA, Parsons JT, Huszti H. Situational temptation for HIV medication adherence in high-risk youth. AIDS Patient Care STDS 2011;25:47-52 [FREE Full text] [doi: 10.1089/apc.2010.0172] [Medline: 21162691]

37. Kolmodin MK, Naar S, Isabella FM, ATN 086/106 Protocol Team. Predictors of self-reported adherence to antiretroviral medication in a multisite study of ethnic and racial minority HIV-positive youth. J Pediatr Psychol 2015:24. [doi: 10.1093/jpepsy/jsv097] [Medline: 26498724]
38. Murphy DA, Sarr M, Durako SJ, Moscicki A, Wilson CM, Muenz LR, Adolescent Medicine HIV/AIDS Research Network. Barriers to HAART adherence among human immunodeficiency virus-infected adolescents. Arch Pediatr Adolesc Med 2003;157:249-255. [Medline: 12622674]
39. Goldenberg T, McDougal SJ, Sullivan PS, Stekler JD, Stephenson R. Preferences for a mobile HIV prevention app for men who have sex with men. JMIR Mhealth Uhealth 2014;2:e47 [FREE Full text] [doi: 10.2196/mhealth.3745] [Medline: 25355249]
40. North Carolina HIV/STD Surveillance Unit. Raleigh, North Carolina: North Carolina Department of Health and Human Service. 2015. 2014 North Carolina HIV/STD Surveillance Report URL: http://epi.publichealth.nc.gov/cd/stds/figures/std14rpt.pdf [accessed 2016-02-19] [WebCite Cache ID 6fQADDFy4]
41. Goldenberg T, McDougal S, Sullivan P, Stekler J, Stephenson R. Building a mobile HIV prevention app for men who have sex with men: an iterative and community-driven process. JMIR Public Health Surveill 2015;1:e18. [doi: 10.2196/publichealth.4449]
42. Fisher JD, Amico KR, Fisher WA, Cornman DH, Shuper PA, Trayling C, et al. Computer-based intervention in HIV clinical care setting improves antiretroviral adherence: the LifeWindows Project. AIDS Behav 2011;15:1635-1646. [doi: 10.1007/s10461-011-9926-x] [Medline: 21452051]
43. Mustanski B, Garofalo R, Monahan C, Gratzer B, Andrews R. Feasibility, acceptability, and preliminary efficacy of an online HIV prevention program for diverse young men who have sex with men: the keep it up! intervention. AIDS Behav 2013;17:2999-3012 [FREE Full text] [doi: 10.1007/s10461-013-0507-z] [Medline: 23673793]
44. Aliabadi N, Carballo-Dieguez A, Bakken S, Rojas M, Brown W, Carry M, et al. Using the Information-Motivation-Behavioral Skills Model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015;27:522-537. [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
45. Allison S, Bauermeister JA, Bull S, Lightfoot M, Mustanski B, Shegog R, et al. The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. J Adolesc Health 2012;51:207-212 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.012] [Medline: 22921129]
46. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011;1:53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
47. Baltierra NB, Muessig KE, Pike EC, LeGrand S, Bull SS, Hightow-Weidman LB. More than just tracking time: complex measures of user engagement with an internet-based health promotion intervention. J Biomed Inform 2016;59:299-307. [doi: 10.1016/j.jbi.2015.12.015] [Medline: 26732997]

Abbreviations

ART: antiretroviral therapy
HIV: human immunodeficiency virus
UNC: University of North Carolina
YMSM: young men who have sex with men

Edited by G Eysenbach; submitted 23.02.16; peer-reviewed by S Bull, K Horvath, C Yang; comments to author 16.03.16; revised version received 20.03.16; accepted 21.03.16; published 13.05.16.

Please cite as:
LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, Nwoko N, Hightow-Weidman LB
Epic Allies: Development of a Gaming App to Improve Antiretroviral Therapy Adherence Among Young HIV-Positive Men Who Have Sex With Men
JMIR Serious Games 2016;4(1):e6
URL: http://games.jmir.org/2016/1/e6/
doi:10.2196/games.5687
PMID:27178752

Original Paper

Development and Reliability Evaluation of the Movement Rating Instrument for Virtual Reality Video Game Play

Danielle Levac[1], BScPT, MSc, PhD; Joanna Nawrotek[2], MScPT; Emilie Deschenes[2], MScPT; Tia Giguere[2], MScPT; Julie Serafin[2], MScPT; Martin Bilodeau[2,3,4], BScPT, PhD; Heidi Sveistrup[2,3], PhD

[1]Northeastern University, Department of Physical Therapy, Movement and Rehabilitation Sciences, Boston, MA, United States

[2]University of Ottawa, School of Rehabilitation Sciences, Ottawa, ON, Canada

[3]University of Ottawa, School of Human Kinetics, Ottawa, ON, Canada

[4]Bruyere Research Institute, Ottawa, ON, Canada

Corresponding Author:
Danielle Levac, BScPT, MSc, PhD
Northeastern University
Department of Physical Therapy, Movement and Rehabilitation Sciences
407c Robinson Hall, 360 Huntington Avenue
Boston, MA, 02115
United States
Phone: 1 6173735198
Fax: 1 6173733700
Email: d.levac@neu.edu

Abstract

Background: Virtual reality active video games are increasingly popular physical therapy interventions for children with cerebral palsy. However, physical therapists require educational resources to support decision making about game selection to match individual patient goals. Quantifying the movements elicited during virtual reality active video game play can inform individualized game selection in pediatric rehabilitation.

Objective: The objectives of this study were to develop and evaluate the feasibility and reliability of the Movement Rating Instrument for Virtual Reality Game Play (MRI-VRGP).

Methods: Item generation occurred through an iterative process of literature review and sample videotape viewing. The MRI-VRGP includes 25 items quantifying upper extremity, lower extremity, and total body movements. A total of 176 videotaped 90-second game play sessions involving 7 typically developing children and 4 children with cerebral palsy were rated by 3 raters trained in MRI-VRGP use. Children played 8 games on 2 virtual reality and active video game systems. Intraclass correlation coefficients (ICCs) determined intra-rater and interrater reliability.

Results: Excellent intrarater reliability was evidenced by ICCs of >0.75 for 17 of the 25 items across the 3 raters. Interrater reliability estimates were less precise. Excellent interrater reliability was achieved for far reach upper extremity movements (ICC=0.92 [for right and ICC=0.90 for left) and for squat (ICC=0.80) and jump items (ICC=0.99), with 9 items achieving ICCs of >0.70, 12 items achieving ICCs of between 0.40 and 0.70, and 4 items achieving poor reliability (close-reach upper extremity-ICC=0.14 for right and ICC=0.07 for left) and single-leg stance (ICC=0.55 for right and ICC=0.27 for left).

Conclusions: Poor video quality, differing item interpretations between raters, and difficulty quantifying the high-speed movements involved in game play affected reliability. With item definition clarification and further psychometric property evaluation, the MRI-VRGP could inform the content of educational resources for therapists by ranking games according to frequency and type of elicited body movements.

(JMIR Serious Games 2016;4(1):e9) doi:10.2196/games.5528

KEYWORDS

active video games, virtual reality, physical therapy, movement, reliability

Introduction

There is increasing evidence for the potential of virtual reality active video games to challenge balance, promote active movement, and increase energy expenditure in children with cerebral palsy [1-3]. To support the integration of virtual reality and active video games into clinical practice, physical therapists require educational resources that inform decision making about matching virtual reality active video games with differing client needs [4]. A recent survey of Canadian physical and occupational therapists found that 76% reported the need for knowledge translation resources providing information about specific systems and games [5]. These resources can build on existing information, including a classification framework that categorizes systems based on characteristics such as the ability to track therapeutically relevant variables [6], a Nintendo Wii and Wii Fit game analysis table that describes general movement requirements and feedback provided by each game [7], and a systematic assessment of serious games in health care [8]. However, selecting from among the wide variety of games available in off-the-shelf systems such as the Nintendo Wii and Microsoft Kinect can be overwhelming for busy clinicians. Resources that provide more detailed information about each game are required.

Specifically, information about the type and frequency of movements elicited during game play across differing games and systems can inform decisions about which game may be best suited for an individual client's physical capabilities and rehabilitation needs. For instance, whereas one game might elicit more squats and be ideally suited to strengthening the lower extremities, another game might elicit more upper extremity reaches and be better suited to increasing shoulder range of motion. Distinguishing between games that elicit movements within the base of support (BOS; eg, trunk leans) and those that elicit movements outside of it (eg, steps) is important from a therapeutic perspective because this categorization relates to different levels of functioning. For example, a therapeutic goal may be to enhance energy expenditure, in which case, the number of steps outside of the BOS, squats, and jumps are relevant. A different goal may be to increase weight shifting to one side of the body, and trunk leans or far reaches within the BOS may be important in that case. In addition, understanding the similarities between games across different systems can inform therapist decisions about progressing use of virtual reality active video games from the clinic to the home.

A valid and reliable instrument quantifying the type and frequency of body movements elicited during game play is required. The objectives of this study were to: (1) develop the Movement Rating Instrument for Virtual Reality Game Play (MRI-VRGP); (2) examine the feasibility of using the MRI-VRGP to rate videotaped game play sessions; and (3) evaluate the instrument's inter-rater and intrarater reliability.

Methods

Study Design

A measurement study design was used to develop the MRI-VRGP and evaluate its feasibility and reliability.

Instrument Development

An iterative process of item generation and refinement involving the study authors was undertaken by 3 researchers (2 of whom are also physical therapists) and 4 physical therapy students. The authors began by undertaking a literature search to identify existing instruments to quantify body movements elicited during virtual reality game play. Finding none, we discussed the movement characteristics that might be relevant to physical therapists interested in differentiating between virtual reality active video games. We then watched sample videotapes of typically developing children playing virtual reality active video games to generate an initial list of items. A series of 4 meetings was undertaken. After each meeting, students and investigators went back to practice videos and discussed refinements to the items. The final instrument is shown in Multimedia Appendix 1.

The MRI-VRGP items represent 12 upper extremity, lower extremity, or full-body movements involved in game play. A rater repeatedly views a videotaped game play session and records the frequency with which each movement is observed. Upper extremity movements are identified as unilateral or bilateral (occurring simultaneously) and separated into close reaches and far reaches. Full body weight shifts that occur inside the BOS are identified by direction (anterior, left, or right), and movements that occur outside the BOS such as lower extremity steps are identified by the limb (left or right) and direction (anterior, posterior, lateral, and crossing midline). Full-body movements of squats and jumps are recorded. Rater training materials with operational definitions for each item (summarized in Table 1) were developed in conjunction with instrument creation.

Table 1. Summary of operational movement definitions.

MRI-VRGP items[a]	Definitions
Close reach	Flexion of elbow >90 °
Far reach	Flexion of elbow <90 °
Front weight shift inside base of support	Front inclination >10° without bending knees
Side weight shift inside base of support	Side inclination >10° with or without bending knees
Single leg stance	Single leg stance > 2 seconds
Front weight shift outside base of support	Forward step
Side weight shift outside base of support	Side step
Back weight shift outside base of support	Backward step
Cross midline weight shift outside base of support	Forward or backward step that crosses the midline
Squat	Knee bend with forward inclination of trunk
Jump	Clearing ground with both feet

[a]MRI-VRGP: Movement Rating Instrument for Virtual Reality Game Play.

Reliability and Feasibility Evaluation

Three physical therapist student raters who had participated in instrument creation underwent a training process in which they and the researchers each rated 3 sample videotapes and met to discuss their results, coming to consensus about each item on each videotape. The students subsequently rated 176 videos. Each rater watched each video at least 3 times, focusing on upper extremity, lower extremity inside the BOS, and lower extremity movements outside the BOS movements separately on each viewing. Video pauses and playbacks were encouraged to maximize the quality of scoring. For intrarater reliability evaluations, each student re-rated 58 videos, a minimum of 1 week after the initial rating. The total time period of rating was approximately 2.5 months. A fourth physical therapy student determined the rating schedules. Raters were blinded to the game that the child was playing and whether the child was typically developing or had cerebral palsy. After completion of the MRI-VRGP for each video, the rater completed a feasibility evaluation involving questions determined by study authors. A 5-cm visual analog scale was used to quantify rating difficulty, with the anchor at 0 cm described as “Easy” and the anchor at 5 cm described as “Difficult.” A similar scale was used to quantify rater confidence, with anchors on “low” and “high.” Raters recorded the time taken to watch the video and complete the MRI-VRGP. Raters provided comments if required to identify video-specific rating difficulties.

Videotaping Participants

The videotapes used in this study were recorded in the context of our previous study exploring energy expenditure during virtual reality active video game play in typically developing children and children with cerebral palsy (Levac D, PhD, 2014 unpublished data). Children and parents provided informed assent or consent for videotaping. Overall, 176 prerecorded videotapes of 11 children that were each 90 seconds long, playing the games against a standardized green backdrop were used. The videotapes involved 4 children (3 girls, 1 boy) with cerebral palsy classified at Gross Motor Function Classification System Level 1 (mean age 12.75 years, standard deviation (SD) 2.87 years) and 7 typically developing children (5 girls, 2 boys; mean age 12.86 years SD: 2.97 years). The participants reported minimal exposure (<5 hours) to virtual reality active video games before the study.

Study participants played 8 90-second games on 2 systems: the Interactive Rehabilitation and Exercise System (IREX; GestureTek Health; www.gesturetek.com) and the Microsoft Kinect for Xbox 360. The IREX and the Kinect use similar motion capture technology; in the IREX, the user’s image is embedded in the virtual environment where they can interact with virtual objects [9], whereas Kinect games involve full-body movement represented onscreen by an avatar. The 8 games (IREX: Snowboarding, Shark Bait, Zebra Crossing, Soccer; Kinect: Space Pop, Reflex Ridge, River Rush, and 20,000 Leaks) were chosen to represent the range of movement possibilities across games on each system. Each game was played at its easiest difficulty level.

Statistical Analysis

Analyses were conducted using Statistical Package for the Social Sciences (SPSS; version 21.0). Intrarater and interrater reliability were determined for the total score, category totals, and for each item of the MRI-VRGP. Intraclass correlation coefficient (ICC; type [1,0k] random effects model) and associated 95% CI were calculated. Traditionally used comparators of < 0.40 as low, 0.4 to 0.74 as moderate, and 0.75 and higher as good for ICCs were used [10]. An ICC > 0.75 with a 95% CI lower bound of 0.60 was set a priori as acceptable for each item. Descriptive statistics summarize time, ease, and confidence ratings across raters. Analysis of variances compare differences in time, ease, and confidence ratings between the 3 raters.

Results

Reliability

Tables 2 and 3 summarize intrarater ICCs and 95% confidence intervals as well as the range of observed frequencies for each item per rater. ICCs for rater 1 ranged from 0 to 0.99, rater 2 from 0.54 to 1, and rater 3 from 0.06 to 1. For each rater,

far-reach upper extremity movements and full-body jump movements had the highest ICCs, whereas close-reach upper extremity movements and lower extremity movements outside of BOS had the lowest ICCs.

Table 2. Intrarater intraclass correlation coefficients of upper extremity movements.

Movement Rating Instrument Items	Rater 1			Rater 2			Rater 3		
	Intrarater ICC	CI	Range	Intrarater ICC	CI	Range	Intrarater ICC	CI	Range
Upper extremity movements									
Close-reach right arm	0.967	0.945-0.98	0-45	0.976	0.985-0.96	0-61	0.831	0.732-0.896	0-5
Far-reach right arm	0.993	0.988 to −996)	0-202	0.987	0.978-0.992	0-144	0.977	0.962-0.986	0-214
Total movements right arm	0.992	0.987-0.995	0-202	0.99	0.983-0.994	0-195	0.977	0.962-0.986	0-214
Close-reach left arm	0.989	0.981-0.993	0-46	0.964	0.978-0.94	0-56	0.765	0.635-0.853	0-4
Far-reach left arm	0.995	0.991-0.997	0-189	0.987	0.977-0.992	0-133	0.977	0.962-0.986	0-215
Total movements left arm	0.995	0.991-0.997	0-189	0.986	0.976-0.992	0-182	0.977	0.962-0.986	0-215
Close-reach bilateral	0	−0.255 to 0.256	0-9	0.86	0.775-0.914	0-23	0.064	0.192-0.312	0-7
Far-reach bilateral	0.987	0.962-0.987	0-54	0.983	0.971-0.99	0-42	0.99	0.984-0.994	0-64
Total bilateral movements	0.977	0.962-0.986	0-54	0.985	0.975-0.991	0-51	0.983	0.971-0.99	0-64

Table 3. Intrarater intraclass correlation coefficients (ICC) of lower extremity movements.

Movement Rating Instrument Items	Rater 1			Rater 2			Rater 3		
	Intrarater ICC	CI	Range	Intrarater ICC	CI	Range	Intrarater ICC	CI	Range
Lower extremity movements									
Front lean	0.684	0.52-0.8	0-10	0.703	0.548-0.812	0-12	0.915	0.861-0.948	0-10
Side lean right	0.863	0.78-0.917	0-20	0.827	0.726-0.893	0-16	0.915	0.862-0.949	0-18
Side lean left	0.926	0.878-0.955	0-18	0.786	0.665-0.867	0-16	0.915	0.862-0.949	0-13
Single-leg stance right leg	0.973	0.955-0.984	0-12	0.856	0.77-0.912	0-5	0.643	0.466-0.771	0-12
Single-leg stance left leg	0.957	0.929-0.974	0-8	1	1	0-6	0.622	0.439-0.757	0-5
Front step right leg	0.956	0.927-0.974	0-34	0.94	0.901-0.964)	0-39	0.921	0.871-0.952	0-32
Side step right leg	0.898	0.834-0.938	0-39	0.548	0.343-0.704	0-25	0.932	0.888-0.959	0-46
Back step right leg	0.956	0.927-0.974	0-27	0.936	0.895-0.961	0-31	0.94	0.902-0.964	0-28
Cross midline right leg	0.43	0.196-0.617	0-11	0.983	0.971-0.99	0-8	0.766	0.636-0.853	0-5
Front step left leg	0.957	0.929-0.974	0-35	0.962	0.937-0.977	0-39	0.943	0.906-0.966	0-32
Side step left leg	0.939	0.9-0.963	0-42	0.541	0.334-0.698	0-30	0.947	0.913-0.968	0-49
Back step left leg	0.963	0.939-0.978	0-35	0.899	0.836-0.938	0-34	0.929	0.884-0.957	0-30
Cross midline left leg	0.953	0.923-0.972	0-6	0.962	0.936-0.977	0-13	0.867	0.787-0.919	0-11
Squat	0.98	0.967-0.988	0-24	0.933	0.889-0.959	0-45	0.918	0.867-0.951	0-23
Jump	0.99	0.982-0.994	0-62	0.995	0.992-0.997	0-57	0.997	0.995-0.998	0-56

Table 4 summarizes interrater reliability findings for each item. The ICC was high for far-reach bilateral (ICC=0.94) and low for close reach in both upper extremities (ICC=0.07). For full-body and lower extremity movements, the highest ICC was for the jump item (ICC=0.99) and the lowest for single-leg stance left leg (ICC=0.27).

Table 4. Interrater intraclass correlation coefficients (ICCs).

Movement Rating Instrument Items		Interrater values				
		Interrater ICC	CI	Mean (SD) Rater 1	Mean (SD) Rater 2	Mean (SD) Rater 3
Upper extremity movements						
	Close-reach right arm	0.14	0.05-0.24	1.60 (5.08)	3.89 (8.83)	0.18 (0.69)
	Far-reach right arm	0.92	0.89-0.94	19.39 (32.88)	21.30 (29.77)	20.14 (29.74)
	Total movements right arm	0.93	0.91-0.95	20.99 (33.35)	25.19 (34.25)	20.31 (29.78)
	Close-reach left arm	0.07	−0.02 to 0.16	1.00 (4.02)	3.12 (8.15)	0.16 (0.57)
	Far-reach left arm	0.90	0.88-0.92	17.22 (31.09)	19.63 (28.63)	19.06 (29.43)
	Total movements left arm	0.93	0.91-0.94	18.23 (31.35)	22.74 (32.46)	19.22 (29.42)
	Close-reach bilateral	0.10	0.01-0.20	0.15 (0.89)	0.74 (2.59)	0.09 (0.57)
	Far-reach bilateral	0.94	0.93-0.96	5.35 (9.79)	5.12 (9.38)	5.52 (9.66)
	Total bilateral movements	0.93	0.92-0.95	5.49 (9.98)	5.80 (10.81)	5.61 (9.67)
Lower extremity movements						
	Front lean	0.62	0.54-0.69	0.68 (1.50)	0.52 (1.41)	0.39 (1.35)
	Side lean right	0.72	0.65-0.77	2.39 (3.44)	2.57 (3.80)	1.68 (3.09)
	Side lean left	0.63	0.56-0.70	2.14 (3.18)	2.35 (3.42)	1.43 (2.64)
	Single-leg stance right leg	0.55	0.46-0.62	0.28 (1.24)	0.11 (0.56)	0.40 (1.38)
	Single-leg stance left leg	0.27	0.18-0.37	0.16 (0.88)	0.11 (0.60)	0.20 (0.61)
	Front step right leg	0.60	0.52-0.67	3.04 (4.86)	10.85 (7.79)	2.69 (4.94)
	Side step right leg	0.57	0.49-0.64	13.80 (9.24)	7.66 (5.84)	17.59 (9.70)
	Back step Right Leg	0.67	0.60-0.73	2.61 (4.22)	7.94 (7.02)	2.60 (4.23)
	Cross midline right leg	0.68	0.61-0.74	0.35 (1.19)	0.55 (1.41)	0.38 (0.94)
	Front step left leg	0.55	0.47-0.63	3.26 (4.86)	10.46 (7.94)	2.85 (4.85)
	Side step left leg	0.54	0.46-0.62	14.02 (9.30)	7.74 (6.12)	17.69 (9.80)
	Back step left leg	0.73	0.67-0.79	2.68 (4.41)	7.26 (6.30)	2.35 (4.04)
	Cross midline left leg	0.70	0.64-0.76	0.35 (0.89)	0.68 (1.64)	0.32 (1.04)
	Squat	0.80	0.76-0.85	5.00 (6.28)	7.96 (9.70)	5.01 (6.18)
	Jump	0.99	0.98-0.99	7.48 (11.94)	7.72 (12.14)	7.49 (11.57)

Feasibility

The mean (SD) difficulty of rating score was 1.89 (0.26) of 5. The mean (SD) confidence of rating score was 3.44 (0.24) of 5. Raters took an average of 14.37 (0.77) minutes (range 4-27) per video. There was a significant difference between raters in difficulty ratings (P <.001), with rater 3 finding rating to be more easy as compared with raters 1 and 2 finding ratings. There was a significant difference between raters in confidence ratings (P <.001) with rater 1 being less confident than raters 2 and 3. Finally, there was a significant difference in time to rate the videos (P <.001), with rater 2 taking more time than rater 1 or 3. Comments on the form indicated that raters often were not able to visualize the child's legs or feet because of camera angle and that videos in which movements were occurring faster and at a higher frequency were more challenging to rate.

Discussion

Intrarater reliability estimates for each of the 3 raters indicate that individual raters were consistently able to record frequency of 16 of the 25 items in the MRI-VRGP on repeated viewing of a videotaped game play session at a reliability rate of greater than the predetermined ICC of 0.75 and lower bound CI of 0.60. The 8 items with which 1 or more raters had difficulty were front lean, side step right leg, side step left leg, single-leg stance right leg, single-leg stance left leg, close-reach bilateral, and cross midline right leg. The lower bound of CIs stayed well above the targeted range for acceptable reliability, with the exception of those items. However, interrater reliability estimates were less precise, with ICCs ranging from poor to excellent and wider 95% CIs. Despite this lack of precision, 8 items were above the preidentified ICC and CI range for

acceptance, 12 items were between 0.40 and 0.74, and only 4 items were <0.40. These 4 most problematic items were upper extremity close reach, total body within the BOS, and lower extremity within and outside BOS items.

The upper extremity items that were problematic across both intrarater and interrater reliability estimates included close reaches (both unilateral—left or right—and bilateral). Lack of clarity in item definitions likely contributed to rater inconsistencies. For example, the distinction between close- and far-reach was defined as an elbow flexion angle of greater than 90°, but, the speed of movements made this angle difficult to determine while watching the video, and ICCs were very low for this item (ranging from 0.07 to 0.14). Although each rater identified differing numbers of close- and far-reaches, total combined arm movements (left, right, and bilateral) had good interrater reliability, indicating that the raters reported similar amounts of total arm movements but that problems arose in distinguishing between "close" and "far." We had included this distinction between close- and far-reach based on our discussions of the therapeutic relevance of different reach ranges. Therapists might be interested in knowing how often children are required to make a potentially more challenging (ie, in a greater joint range of motion) upper extremity movement. However, confirming with practicing therapists as to the clinical relevance of categorizing upper extremity movements in this way is an important next step in instrument revisions.

With respect to trunk and lower extremity items, single-leg stances, front leans, and side steps were most problematic. Video quality likely impacted difficulties identifying trunk and lower body movements. Front leans are defined as "an isolated movement that cannot precede a step." Distinguishing leans from steps was problematic because some raters likely included a lean within a step, whereas others may have counted the 2 movements separately. Indeed, in 96 of the 176 videos (54.5%), the camera angle did not allow for the visualization of participants' feet. This was detrimental when rating items such as weight shifts within the BOS, such as single-leg stance or side steps, where seeing whether the foot lifted off the ground was essential for item scoring. The 2 lower extremity items that achieved good interrater reliability—jumps and squats—are clearly distinguishable movements that can be identified appropriately even without visualizing the feet.

Movement speed and differing game play strategies across children are issues that impacted reliability. Two games in particular on the Xbox Kinect system—Space Pop and Rally Ball—required rapid upper extremity movements. Raters needed to slow down the video speed or pause the video repeatedly. In the Space Pop game, arm movements to simulate flying are needed to "pop" the virtual bubbles. These high-speed movements may have led to interrater differences in counts because movements may have been missed or counted twice. In addition to movement speed, differing game play strategies that enhanced the variation across children playing the same game were observed. Although each game was played at the same difficulty level, individual children chose to focus on different components of the game (eg, choosing to go for all the "coins" in Reflex Ridge by moving their arms or choosing to focus only on body movements that avoided the obstacles). In addition, during the 90 seconds, some children advanced further in the game than others; one game in particular (Rally Ball) required quiet standing while it reset to the previous level if a player was unsuccessful, limiting movement options during this resetting time (approximately 3-5 seconds). Despite controlling for difficulty level and duration of play, children's game play abilities and their level of success at each game during those 90 seconds resulted in a wide variation of movements that related both to each child's personal "style" (ie, did they move in a slower, more controlled manner or did they use rapid, flailing movements) and to choice of what to focus on for each game (ie, getting as many points as possible or making as few errors as possible).

From a feasibility perspective, despite these issues, raters found it fairly easy to rate and were fairly confident, although rater 3 found it the most difficult, and rater 2 was the most confident. Interestingly, rater 2, who was the most confident, also had the highest mean rating time. As anticipated, given the protocol requiring a minimum of 3 viewings, rating time was long for such a short video, indicating that raters likely slowed down the video speed and stopped the tapes on a frequent basis while watching and rewatching.

Skjaeret et al [11] were the first to systematically observe movement characteristics of users during videotaped active video game play. Using a 5-point Likert scale, the researchers rated 5 movement characteristics considered relevant to fall prevention exercises in seniors playing 3 virtual reality active video games [11]. Their goal was to inform the design of new virtual reality active video games for this population. Raters also watched each video numerous times to focus solely on a single movement characteristic per viewing. The movement characteristics that they examined included amount of weight shift, temporal variation, step length variation, variation in movement direction, and visual independency [11]. They achieved high interrater reliability across 3 raters (>0.840) for all characteristics. Rating movement characteristics that can be judged in summary after watching a video as opposed to frequency counts of more specific movements may be a method to increase the consistency of observations across raters. For the population of children with cerebral palsy, other global movement characteristics might be more relevant, including cross midline movements and bilateral reaches.

Finally, it is important to consider the amount of error that is acceptable for this type of instrument. The purpose of the MRI-VRGP is to document the frequency of movements elicited during game play. Thus, the magnitude of error that is acceptable for this instrument is greater than would be the case if the purpose was to use it for making decisions about an individual child's treatment or progress. Given that information obtained through the use of this instrument will be used to inform comparisons between virtual reality active video games and systems, subsequent steps in the instrument evaluation process will focus on determining whether items can be made more general (eg, is the magnitude of reach for arm movements important?) and on better defining each movement that is rated through a validity process.

Limitations

MRI-VRGP items were established by a small group of researchers and physical therapy students. The research team arrived at the items and their definitions through a literature search of movement characteristics of children with cerebral, energy expenditure related to different virtual reality active video games, viewing of sample videotapes, and clinical understanding of the movements that physical therapists would be interested in when selecting a particular game for a therapy intervention. However, a more formal face and content validity process with additional experts and clinicians would have determined whether the chosen items are representative of what clinicians would like to capture and may have served to clarify the item operational definitions before reliability testing. In addition, involving raters who were not involved in instrument development would have strengthened the findings.

The MRI-VRGP provides clinicians with a simple count of movements but does not include an analysis of movement quality. This may be an issue if therapists are interested in both how often a game elicits a particular movement and the quality of that movement. Moreover, the scale does not quantify whether the player has used potentially unwanted or therapeutically harmful compensations required to achieve a certain movement. For example, the instrument does not distinguish or document whether a child is using shoulder hiking to reach a target above them or using hip circumduction to take a step. It may be important to include a section where the rater can make note of any perceived maladaptive movement patterns during game play. This is particularly important if therapists are using the instrument to inform development of unsupervised home programs. In supervised situations, maladaptive movements can be monitored by the therapist as the child plays the game. Given that this is not possible in supervised exercise, therapists can use these observations to recommend changes to game parameters that might avoid them (eg, recommending that the child play at a lower difficulty level, which may slow down the game and reduce unwanted movements).

Motion analysis systems were once limited to laboratory use, but, the introduction of the Kinect sensor has made markerless motion analysis feasible on a wider scale. How can an observer-rated measure quantifying movement frequency be a useful adjunct to this low-cost kinematic sensor? Reports exploring the psychometric properties of the Kinect sensor to measure movement across a wide variety of populations and tasks are available; accuracy and reliability are inconsistent and dependent on the type and frequency of movement (eg, [12-14]). As evidence continues to emerge to support use of the Kinect sensor for kinematic analysis, the MRI-VRGP could act as an adjunct to quantify movement frequency as the Kinect provides information to therapists that can be used to assess movement quality.

Finally, videotapes of typically developing children and children with cerebral palsy were included in this study. There was a wide range of frequency of movements observed for each of the items, implying sufficient heterogeneity of the measured construct to enable reliability analyses. The 8 games targeted upper extremity and lower extremity movements to different extents. However, the small sample size of participants reduced the precision of the reliability estimates. This first attempt at developing the instrument and evaluating reliability indicated issues of strengths and weaknesses that can be built on in future work.

Future Recommendations

Given that most items in both intrarater and interrater reliability achieved a minimum of good reliability in this preliminary investigation, further refinements will be undertaken. Subsequent steps include videotaping a greater number of children and youth to use as the basis for adding greater clarity to item definitions. Items will then be put to a Delphi process with pediatric physical therapists to achieve consensus on content and definition. The revised items and definitions will be on the basis of a systematic rater training procedure, involving the new videotapes. Subsequently, psychometric property testing on a larger sample size of typically developing children will be undertaken. If shown to have adequate reliability, therapists could use these numbers as a baseline when making decisions about game use for their clients with cerebral palsy or other diagnoses. The instrument could also be used as a tool to compare movements elicited in different games across different virtual reality active video game systems, adding objective information to include in clinical decision-making tools that help clinicians make decisions about which games to use for different clinical goals. Multiple games from different systems will be included in future work. The result will be a game ranking from most to least elicited movements in each category, allowing clinicians to select the game that elicits the movements most important for an individual child's rehabilitation needs.

Conclusions

The MRI-VRGP demonstrated overall good intrarater reliability and moderate interrater reliability. Poor video quality, rater inconsistencies in terms of interpretation of operational movement definitions, and difficulty quantifying movements occurring at high speed contributed to these findings. With subsequent development and psychometric property evaluation, a valid and reliable instrument could be used to provide objective information about movement quantity across different games and systems, contributing to clinical decision-making tools that will inform game selection by clinicians for a broad range of clients.

Acknowledgments

The authors would like to thank the occupational therapy students involved in the original study during which the videotapes were recorded, as well as the children and their parents for participating and consenting to be a part of this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

[PDF File (Adobe PDF File), 32KB - games_v4i1e9_app1.pdf]

References

1. Dewar R, Love S, Johnston LM. Exercise interventions improve postural control in children with cerebral palsy: a systematic review. Dev Med Child Neurol 2015 Jun;57(6):504-520. [doi: 10.1111/dmcn.12660] [Medline: 25523410]
2. Weiss Patrice L Tamar, Tirosh E, Fehlings D. Role of virtual reality for cerebral palsy management. J Child Neurol 2014 Aug;29(8):1119-1124. [doi: 10.1177/0883073814533007] [Medline: 24799367]
3. Mitchell L, Ziviani J, Oftedal S, Boyd R. The effect of virtual reality interventions on physical activity in children and adolescents with early brain injuries including cerebral palsy. Dev Med Child Neurol 2012 Jul;54(7):667-671 [FREE Full text] [doi: 10.1111/j.1469-8749.2011.04199.x] [Medline: 22283557]
4. Levac DE, Galvin J. When is virtual reality 'therapy'? Arch Phys Med Rehabil 2013 Apr;94(4):795-798. [doi: 10.1016/j.apmr.2012.10.021] [Medline: 23124132]
5. Glegg S, Levac D, Miller P, Colquhoun H, Wright V. A survey of physical and occupational therapists' virtual reality use and learning needs. 2016 Presented at: Canadian Occupational Therapy Association Conference; May 21-24; Victoria, B.C.
6. Galvin J, Levac D. Facilitating clinical decision-making about the use of virtual reality within paediatric motor rehabilitation: describing and classifying virtual reality systems. Dev Neurorehabil 2011;14(2):112-122. [doi: 10.3109/17518423.2010.535805] [Medline: 21410403]
7. Deutsch JE, Brettler A, Smith C, Welsh J, John R, Guarrera-Bowlby P, et al. Nintendo wii sports and wii fit game analysis, validation, and application to stroke rehabilitation. Top Stroke Rehabil 2011;18(6):701-719. [doi: 10.1310/tsr1806-701] [Medline: 22436308]
8. Graafland M, Dankbaar M, Mert A, Lagro J, De WL, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
9. Glegg Stephanie M N, Tatla SK, Holsti L. The GestureTek virtual reality system in rehabilitation: a scoping review. Disabil Rehabil Assist Technol 2014 Mar;9(2):89-111. [doi: 10.3109/17483107.2013.799236] [Medline: 23713408]
10. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
11. Skjæret N, Nawaz A, Ystmark K, Dahl Y, Helbostad JL, Svanæs D, et al. Designing for movement quality in exergames: lessons learned from observing senior citizens playing stepping games. Gerontology 2015;61(2):186-194. [doi: 10.1159/000365755] [Medline: 25402845]
12. Clark RA, Pua Y, Fortin K, Ritchie C, Webster KE, Denehy L, et al. Validity of the Microsoft Kinect for assessment of postural control. Gait Posture 2012 Jul;36(3):372-377. [doi: 10.1016/j.gaitpost.2012.03.033] [Medline: 22633015]
13. Mentiplay BF, Clark RA, Mullins A, Bryant AL, Bartold S, Paterson K. Reliability and validity of the Microsoft Kinect for evaluating static foot posture. J Foot Ankle Res 2013 Jan;6(1):14 [FREE Full text] [doi: 10.1186/1757-1146-6-14] [Medline: 23566934]
14. Galna B, Barry G, Jackson D, Mhiripiri D, Olivier P, Rochester L. Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease. Gait Posture 2014 Apr;39(4):1062-1068. [doi: 10.1016/j.gaitpost.2014.01.008] [Medline: 24560691]

Abbreviations

BOS: base of support
ICC: intraclass correlation coefficient
MRI-VRGP: Movement Rating Instrument for Virtual Reality Game Play

Edited by G Eysenbach; submitted 15.01.16; peer-reviewed by M Graafland, M Franco; comments to author 06.04.16; revised version received 11.04.16; accepted 27.04.16; published 01.06.16.

Please cite as:
Levac D, Nawrotek J, Deschenes E, Giguere T, Serafin J, Bilodeau M, Sveistrup H
Development and Reliability Evaluation of the Movement Rating Instrument for Virtual Reality Video Game Play
JMIR Serious Games 2016;4(1):e9
URL: http://games.jmir.org/2016/1/e9/
doi:10.2196/games.5528
PMID:27251029

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org