JMIR Serious Games

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

Volume 7 (2019), Issue 3 ISSN: 2291-9279

Contents

Original Papers

Corrigenda and Addendas

Original Paper

Development and Evaluation of a Virtual Research Environment to Improve Quality of Care in Overcrowded Emergency Departments: Observational Study

Charles-Henri Houze-Cerfon[1,2,3], MD; Christine Vaissié[4], MD; Laurent Gout[1], MD; Bruno Bastiani[3], DEd; Sandrine Charpentier[1], MD, PhD; Dominique Lauque[1], MD, PhD

[1]Emergency Department, Toulouse University Hospital, Toulouse, France
[2]Institut Toulousain de Simulation en Santé, Toulouse, France
[3]UMR Education, Formation, Travail, Savoir, University Toulouse 2 Jean Jaures, Maison de la Recherche, Toulouse, France
[4]Emergency Department, Albi University Hospital, Albi, France

Corresponding Author:
Charles-Henri Houze-Cerfon, MD
Institut Toulousain de Simulation en Santé
ITSIMS - pavillon U2000
Place du Dr Baylac
Toulouse,
France
Phone: 33 561772147
Email: houze-cerfon.ch@chu-toulouse.fr

Abstract

Background: Despite a wide range of literature on emergency department (ED) overcrowding, scientific knowledge on emergency physicians' cognitive processes coping with overcrowding is limited.

Objective: This study aimed to develop and evaluate a virtual research environment that will allow us to study the effect of physicians' strategies and behaviors on quality of care in the context of ED overcrowding.

Methods: A simulation-based observational study was conducted over two stages: the development of a simulation model and its evaluation. A research environment in emergency medicine combining virtual reality and simulated patients was designed and developed. Afterwards, 12 emergency physicians took part in simulation scenarios and had to manage 13 patients during a 2-hour period. The study outcome was the authenticity of the environment through realism, consistency, and mastering. The realism was the resemblance perceived by the participants between virtual and real ED. The consistency of the scenario and the participants' mastering of the environment was expected for 90% (12/13) of the participants.

Results: The virtual ED was considered realistic with no significant difference from the real world with respect to facilities and resources, except for the length of time of procedures that was perceived to be shorter. A total of 100% (13/13) of participants deemed that patient information, decision making, and managing patient flow were similar to real clinical practice. The virtual environment was well-mastered by all participants over the course of the scenarios.

Conclusions: The new simulation tool, Virtual Research Environment in Emergency Medicine, has been successfully designed and developed. It has been assessed as perfectly authentic by emergency physicians compared with real EDs and thus offers another way to study human factors, quality of care, and patient safety in the context of ED overcrowding.

(JMIR Serious Games 2019;7(3):e13993) doi:10.2196/13993

KEYWORDS

virtual reality; interprofessional relations; emergency medicine

Introduction

Background

Analysis of medical errors shows that among human factors, psychological, cognitive, and organizational features are directly related to the quality of care. Emergency department (ED) overcrowding and patient boarding result in delays in care, increased short-term mortality, and worsened patient experiences and safety [1,2]. Emergency physicians have to deal with several patients in a short period of time and anticipate risk of errors linked to interruptions and disruptions [3]. To alleviate the problem of ED overcrowding, various conceptual models of care have been invented by researchers, which mainly focus on organizational strategies and input-throughput-output processes [4-7]. However, few have analyzed the cognitive processes used by emergency physicians coping with overcrowding, or the impact of overcrowding on team processes [8]. Team processes are defined as the cognitive, verbal, and behavioral activities of a team working together. They include the nontechnical skills for crisis resource management (CRM), such as leadership, teamwork, situation awareness, task management, and decision making. In a systematic literature review, Schmutz has shown that valid measures and adequate tools are required for investigating the effects of team processes on clinical performance and safety [9].

Team process behavior (eg, how the team functions and whether the team and its members grow, develop, and improve over time) is influenced by patient volume, the number and attitude of other caregivers, as well as changes in the department's organization [7]. Variations of these confounding factors make it difficult to analyze the direct relationship between overcrowding, team processes, and quality of care. Simulation in a virtual environment could expose emergency physicians to flow variation only and show its impact on CRM skills and quality of care (test relevance, working diagnosis, decision-making time, and patient orientation) [10]. Online virtual worlds such as serious games have shown their effectiveness concerning medical education and training of teamwork skills, including leadership, coordination, and communication [11,12]. They offer a safe and controlled setting without real patients' risks and issues.

To our knowledge, no study has yet been conducted on the development and evaluation of a virtual ED related to quality care.

Objectives

The goal of this study was to develop a virtual ED intended to replicate real patient flow management by emergency physicians and evaluate the authenticity of this new simulated research model.

Methods

Study Design

This simulation-based observational study was conducted in two stages. The first stage was the design and development of a virtual environment in emergency medicine. The second stage was an observational study to evaluate this environment.

Design and Development of the Virtual Environment

Virtual Research Environment in Emergency Medicine (VIREM) is a hybrid simulation that combines a virtual environment (virtual multiplayer world), technical skills simulation, and simulated patients (Figure 1) . Second life® was used as virtual world online because it is an interactive, 3-dimensional, low-cost environment easily accessed by many participants who interact and communicate together (see Multimedia Appendix 1). Patients and caregivers were represented as characters called avatars that interact with their environment (see Multimedia Appendix 2).

A research team designed and developed the virtual ED. The team consisted of 3 emergency physicians and 1 nurse who had more than 10 years of experience in emergency medicine, a game developer, and an expert in virtual simulation didactics. They designed the simulated ED premises and equipment in accordance with the French Emergency Medicine Society's guidelines on ED architecture and operation [13]. In a scenario, the virtual ED team included avatars for 1 emergency physician, 1 resident in postgraduate years 2, and 3 nurses for triage, ambulatory care, and nonambulatory care, respectively (see Multimedia Appendix 3). The emergency physician worked both as team leader and caregiver. The medical conditions of patient avatars were selected from a French regional registry of 37 ED populations based on the most prevalent conditions; the final choice of the 13 cases was based on team agreement as representative of the usual patients in an urban busy ED. The history, clinical signs, test results, and course were developed for each patient. Operational procedures were designed for the nurse avatars who could triage, draw blood samples, perform electrocardiograms, and deliver medications [12]. The median length of time of each procedure and intervals from test ordering to data-ready time were measured in patients admitted in 3 different EDs and reported in the scenario (Table 1).

The patient flow was set at a level higher than the recommended maximum with 4 patients per physician per hour to simulate overcrowding conditions. The virtual ED was developed and iteratively tested by the research team to correct bugs and improve patient flow, processes, and interactions between the avatars and their environment.

Figure 1. Design, development, and assessment of Virtual Research Environment in Emergency Medicine.

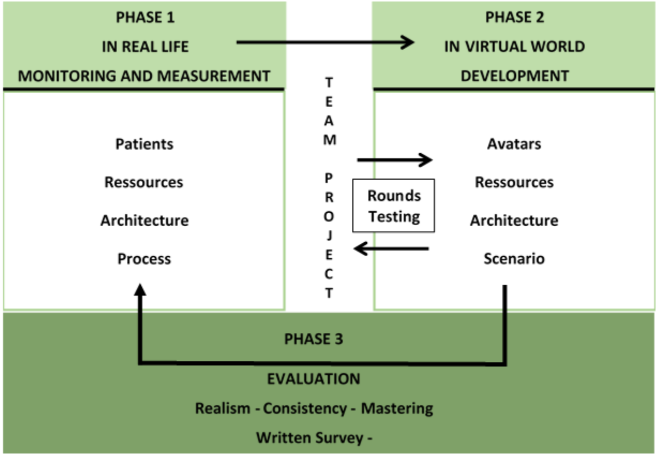


Table 1. Length of time of real emergency procedures used in Virtual Research Environment in Emergency Medicine.

Real emergency procedures (n)	Duration (min), median (interquartile range)
Nurse 2	
Administrative reception time (20)	4 (2-4)
Nurse 1 tasks	
Triage time (20)	4 (3-5)
Time to prepare and give the medications (17)	5 (4-6)
Electrocardiogram recording (16)	5 (4-5)
Blood drawn to data-ready time (18)	88 (67-115)
Specialty physician consultation, data-ready time (16)	45 (36-50)
Resident tasks	
History and examination of patient (20)	14 (10-15)
Perform cast (10)	15 (12-18)
Sutures (10)	21 (13 -39)
Medical records (20)	10 (8-12)
Prescription (18)	7 (6-8)
Patient information (16)	5 (5-5)
Radiological test ordering to data-ready time	
X-ray (19)	40 (37-48)
Computed tomography scan (11)	60 (60-120)
Ultrasounds (15)	105 (102-108)
Specialty physician consultation	
Data ready time (16)	45 (36-50)

Selection of Participants for Virtual Research Environment in Emergency Medicine Evaluation

We recruited emergency physicians in 2 EDs of a multifacility academic hospital, which together treated more than 164,059 patients aged 15 years or older in 2017 [14]. They had at least 3 years' experience in emergency medicine and were not involved with the research team. Upon enrollment, confidentiality was guaranteed, and all physicians gave their written informed consent. This study was carried out with the agreement of the ethics committee of the University and the French National Council for Information Technology and Liberties (Reference 19880021v0).

Study Protocol

Data collection took place in the University Institute of Health Simulation. Data were collected by short structured interviews and questionnaires by a nonparticipant observer. The participants were alone in a room with a computer and a technical skill simulator to perform the sutures. They each had a 30-min briefing focused on the resources available, how to interact with the environment, and communicate with the patients and caregivers as a physician avatar. The participants interacted with the virtual ED by controlling their avatars and with a headset (Voice over Internet Protocol). They used a phone to call other caregivers or specialists who were out of reach. The scenario lasted 2 hours and was followed by a short, structured interview lasting 20 min. A total of 4 instructors from the research team were required to pilot, respectively, the patients, nurse 1 avatar, resident avatar, and nurse 2 avatars or other caregiver avatars (dispatcher, specialist). The instructors, according to their role, triggered the different procedures whose duration had already been defined (Table 1). Participants and instructors used Dell Precision T1700 computers, Nvidia GTX 970 type graphics card with 22.5“ full high-definition screens.

Evaluation of the Environment

The study outcome was the authenticity perceived by participants working in this virtual ED. The perception of authenticity by users results from the fidelity of the environment (setting and available resources), the situations experienced, and the temporal fidelity of procedures [15]. VIREM was adjudicated as authentic if the 3 components, realism, consistency, and mastering, were validated.

The realism was defined as the resemblance of the virtual ED with a real ED. Physicians have been asked to evaluate their own ED and the virtual environment after the scenario on a 7-item questionnaire about facilities, resources available, and delays. Each question was scored according to a 5-point Likert scale from very dissatisfied to very satisfied. The virtual ED was deemed realistic if the satisfaction level of the 7 items was not different between the virtual and real ED.

The consistency of the proposed rules and situations was assessed during individual directive interviews performed after the scenario. Participants gave their opinion on the quality of information provided by the scenario, their ability to make decisions, and whether the scenario was proper to manage the patient flow and preserve the quality of care. The consistency of VIREM was deemed to be good if these 3 criteria were positive for at least 90% (12/13) of participants.

The mastering of the environment was assessed by an independent observer during the scenario. A total of 5 cognitive tasks were rated: (1) acceptance and participation in the scenario; (2) environment testing; (3) ability to make decisions; (4) anticipation; and (5) use of the environment, procedures, and resources. We considered that mastering the VIREM was good when more than 90% (12/13) of participants fulfilled these 5 criteria.

Statistical Analysis

All data were entered into an Excel spreadsheet, (Microsoft Corp) and the statistical analysis was performed with Stata (Version 12.0, Stata Corp). Categorical data were presented as frequencies and percentages (%) and continuous data as mean with SDs or median with interquartile ranges if not normally distributed. The realism satisfaction score between the virtual ED and the real ED environment was compared using a Wilcoxon signed-rank test. Results with a $P<.05$ were considered significant.

Results

Stage 1: Virtual Emergency Department

The virtual ED was designed as a level 1 trauma center in a major adult referral hospital. It consisted of a reception and triage area, 5 examination rooms for monitoring with a computer, a medication preparation room, a waiting room, a medical office, and a radiology department with ultrasound and computed tomography scan equipment. The participants took up their shift when the scenario began. A total of 3 patients were already undergoing care in the examination rooms. Thereafter, 4 and 6 patients attended the ED during the first and second hours, respectively (Table 2).

During the scenario, the participant physician led his avatar and fully reviewed the patient's history and complaints in the available examination room. He selected parts of the patient's body for physical examination, prescribed biological tests, x-rays, and treatment and checked test results. He communicated by Voice over Internet Protocol with the instructor who led the patient avatars to explain procedures and care to patients, assess their pain, and discuss the case with other specialists. The participant could coordinate activities with team members, exchanging information, communicating, making decisions, keeping others aware of the situation, planning, and prioritizing the tasks. The participant could perform the task himself or ask the resident or nurse avatar to do it (Table 1).

A total of 12 enrolled emergency physicians completed the study. The sex ratio was 1:4, median age 30 years (range 29-35), and median experience in emergency medicine as a resident and attending physician was 4 years (range 3-10; Table 3). All participants used software and networks at work and had basic computer skills. Only 1 participant was familiar with the virtual environment.

Table 2. Screenplay of Virtual Research Environment in Emergency Medicine.

Min	Instructor 1: Resident[a]	Instructor 2: Nurse 1/dispatcher/other caregivers[a]	Instructor 3: Nurse 2/patients (P)[a]
T-1	__[b]	—	P1 —> room 1/P2—> room 2
T+0	Physician 1	—	—
T+5	—	Physician family calls	—
T+10	—	—	P3 —> ED[c] hall
T+14	—	—	P3 —> triage room
T+15	—	Radiologist calls for P2	—
T+19	—	—	P3 —> room 3a and call EP[d]
T+21	—	—	P4/P13 —> ED hall
T+24	—	—	P4 —> waiting room
T+25	—	—	P13 —> triage room
T+30	MD (gastro) calls for P1	—	—
T+35	—	Gastroenterologist calls	P13 —> room 2b
T+36	—	—	Call nurse 1 for IV line P13
T+37	—	—	P4 —> triage room
T+42	—	IV[e] line P13	Call EP: “where do I install P4”
T+43	—	—	P5 —> ED hall
T+44	—	Dispatcher calls EP	—
T+46	—	—	P5 —> triage room
T+50	—	Nurse calls EP: “blood test sent”	—
T+51	—	—	P5 —> room 3b
T+52	—	Nurse calls EP: “what is the blood test for p5?”	—
T+62	—	—	P6 —> ED hall
T+66	—	—	P6 —> triage room
T+70	—	—	P7 —> ED hall
T+71	—	—	P6 —> room available or call EP
T+72	—	Dispatcher calls EP	—
T+74	—	—	P7 —> triage room
T+78	Call EP: “neurologist is ready”	—	—
T+79	—	—	P7 —> room available or call EP
T+81	—	—	P8 —> ED hall
T+84	—	—	P8 —> triage room
T+89	—	—	P8 —> room available or call EP
T+90	Call EP: “P13 have a Troponin at 45”	—	—
T+91	—	—	P9 —> ED hall
T+95	—	—	P9 —> triage room
T+96	—	—	P10 —> ED hall
T+99	—	—	P9 —> room available or call EP
T+101	—	—	P10 —> triage room
T+105	—	—	P11 —> ED hall
T+106	—	—	P10 —> room available or call EP
T+110	—	—	P11 —> triage room

Min	Instructor 1: Resident[a]	Instructor 2: Nurse 1/dispatcher/other caregivers[a]	Instructor 3: Nurse 2/patients (P)[a]
T+111	—	—	P12 —> ED hall
T+114	—	—	P11 —> room available or call EP
T+115	—	—	P12 —> triage room

[a]Each column describes the actions carried out by the 3 instructors according to their function in the scenario. The actions are displayed chronologically after the beginning of the scenario. For example, at the 42nd min the second instructor who controls nurse 1 must place an IV line to patient 13, and the third instructor who controls nurse 2 calls the emergency physician to ask him in which emergency room patient 4 would be installed.

[b]No action.

[c]ED: emergency department.

[d]EP: emergency physician.

[e]IV: intravenous.

Table 3. Characteristics of the participants (n=12).

Characteristics	Value
Age (years), median (IQR[a])	30 (29-35)
Males, n (%)	7 (58)
Experience in emergency medicine (years), median (IQR)	4 (3-10)
Basic level of computer experience[b], n (%)	12 (100)
Rate of playing computer games[c], n (%)	
Never	10 (83)
Occasionally	2 (16)
First time with experimental virtual reality system, n (%)	11 (91)
Basic level of knowledge on 3D images production[b], n (%)	12 (100)
Level of knowledge on virtual reality[b], n (%)	
None	11 (91)
Basic	1 (8)

[a]IQR: interquartile range.

[b]Items were graded on Likert-type scale (1=none, 2=basic, 3=intermediate, 4=expert).

[c]Items were graded on Likert-type scale (1=never, 2=occasionally, 3=less than 50% days, 4=more than 50% days, 5=every day).

Stage 2: Evaluation of Authenticity

Realism

Participant satisfaction with the facilities (premise and examination room) and resources (team members) was similar in the virtual ED and in their real ED (Table 4).

The participants considered the timeliness of procedures in the virtual ED satisfactory, and they perceived the lengths of time shorter. Therefore, our virtual ED was realistic for facilities and resources but not for procedures, although lengths of time used in the virtual ED were based on real-life data.

Table 4. Evaluation of the realism of the virtual emergency department versus the real environment (n=12).

Categories	Satisfaction, median (interquartile range)		*P* value[a]
	Real environment, range 1-5[b]	Virtual environment, range 1-5[b]	
Facilities			
Premises	3 (3-4)	4 (3-4)	.41
Examination rooms	3 (3-4)	3 (3-4)	.96
Procedures			
Duration of blood drawing and electrocardiogram	4 (3.6-4)	4 (4-5)	.04[c]
Blood drawn or radiological test ordering to data-ready time	3 (2.7-3)	3.5 (3-4)	.008[c]
Specialty physician consultation ordering to data-ready time	2 (2-3)	3 (3-3.2)	.003[c]
Resources			
Number of nurses	3 (3-3)	3 (2.7-3)	.08
Number of residents	3 (3-4)	3 (3-3.2)	.32

[a]Wilcoxon signed rank test.

[b]Items were graded on Likert-type scale (1: very dissatisfied, 2: dissatisfied, 3: neither, 4: satisfied, 5: very satisfied).

[c]Significant difference.

Consistency

The virtual ED was considered consistent because all participants felt that patient information was consistent with reality and allowed for medical decisions similar to their clinical practice (100%). Strategies for managing patient flow and preserving quality of care in the context of overcrowding was deemed by all participants to be consistent with those used in their real ED (100%).

Mastering

All participants (N=13) performed the cognitive tasks assigned by the scenario, thus reflecting their mastery of this virtual environment. They actively participated in the simulation (100%), used the virtual environment (100%), made decisions (100%), anticipated actions (100%), and used the environment, procedures, and resources (100%).

Discussion

Principal Findings

This was the first study to propose a method for developing and validating a simulated research environment combining different simulation techniques as procedural simulation and virtual simulation [16]. Virtual environments are usually targeted toward the acquisition of skills and knowledge [11,12,17]. Our goal was to build an authentic research environment to study the cognitive strategies and CRM skills of physicians in a situation of high cognitive load because of patient flow in an ED. We relied on the definition of authenticity given by Petraglia: “Authenticity is not an intrinsic property of an object, but a judgment, a decision on the part of the user from the point of view of his or her past experiences and the socio-cultural context. Authentic virtual simulation is a game perceived as authentic by the learners” [15].

Our approach designed and assessed the virtual environment in the 3 dimensions of authenticity: realism, consistency, and mastering.

All participants found that the virtual environment was realistic and easy to master. A previous research study showed that knowledge of virtual reality was related to engagement. More knowledge about virtual reality leads to a greater sense of engagement [18]. Although the population in our study had a moderate level of computer experience and lacked knowledge about virtual reality, the physicians were effectively able to perform the cognitive tasks and get engaged in the simulation. This finding could be explained by the easy use of this virtual environment that does not require users to have specific skills or characteristics.

The realism of the premises and resources was validated by all participants. Perception of time for physician and nurse tasks, as well as evaluation of delays for test or consultation results were different in the scenario than in real life. All participants agreed that virtual care seemed shorter than real life one. This difference of perception between virtual and real ED has been explained by Block et al who found that duration judgments are affected by cognitive load. If participants are aware that duration judgments must be made (prospective paradigm), a greater cognitive load decreases the perception of duration [19]. Therefore, subjective assessment of procedure duration is not an accurate criterion of realism but instead may be an indicator of cognitive overload caused by patient flow in this scenario.

All participants agreed that VIREM was a valuable research environment for studying the cognitive processes of emergency physicians and the effect of patient flow on quality of care. Observation of participants’ attitudes during the test showed that they rapidly mastered this virtual environment. As a preliminary observation, we noticed that the 12 emergency physicians used different attitudes and strategies for managing the patient flow.

Flowerdew et al have shown that workload management is a complex skill that covers a wide range of behaviors [8]. This systemic review identified a link among workload management, safety, and error. For example, diagnostic errors increase with work overload.

As it has been suggested by Leblanc et al, a simulation model such as VIREM could be used to identify the safest workload management [20]. In addition, its accessibility with distributed teams and its low cost with only an up-to-date computer and internet connection would permit widespread studies to be conducted.

One of the main aspects of VIREM is that it also allows ED management training. The learnings from the implementation of the emergency response highlight the need for adapting a holistic approach by using core competencies in the domain of emergency management, such as preparedness, response system, patient care, and resource management (human and material). Most competencies can be obtained through traditional accredited education and training programs; on the other hand, education for personnel operating in emergency situations should be based on the acquisition of task-related, profession-specific, and cross-disciplinary competencies that cannot be easily acquired through those programs. VIREM will allow us to standardize competency-based education in emergency and resource management.

Similar to VIREM, other virtual environments could easily be designed for any health organization that would be interested in analyzing the links among human factors, diseases, resources available to health professionals in their workplaces, and quality of care. Next, it could also be a good way to deliver an effective and up-to-date interprofessional training in many countries and more particularly in low and medium income countries where access to training is a challenge for financial, geographical, political, and/or institutional reasons. Afterwards, it could facilitate the implementation of interprofessional training for health workers in even the remotest regions.

Limitations

The major limitation of this study was the sample size. Although only 12 physicians took part in the observational study, the duration of the simulation session was too long to involve more participants from the same hospital.

Second, we did not measure the participant stress level during the scenario. Simulations can induce stress that may influence decision making and performance [21,22]. Also, the perception of nonreal scenario is likely to alter the alertness and stress level of the participant, which is unlikely to be the same as the real-life situation. Future studies using VIREM will have to integrate the stress level, which affects emotion, perception, memory, judgment, and reasoning [23]. Third, the physical activity–induced fatigue for emergency physicians working in a real ED was not reproducible in this virtual environment. Another potential limitation of the virtual environment would be the requirement for concentration and energy. The simulation would be less stressful as well. However, working in a real ED during a period of time as short as 2 hours presumably would not cause enough fatigue to influence decision making and behavior.

Conclusions

We developed a new virtual environment, VIREM, to study strategies set in place by emergency physicians to manage patient flow in EDs. This virtual environment was assessed as authentic by emergency physicians and thus, offers a new research tool to study CRM skills and improve quality of care and patient safety in the context of ED crowding. Its use could be extended to a different geographical, sociological, and economic context, but further studies will be needed to validate it.

Acknowledgments

The authors would like to thank Vicki Leblanc, PhD, Chair and Professor of the Department of Innovation in Medical Education at the University of Ottawa and Director of the University of Ottawa Skills and Simulation Centre for her copy editing. They also thank Vanessa Houzé-Cerfon, Clinical Research Coordinator at the emergency department of Toulouse Hospital for her involvement in the proofreading of the manuscript.

The work is attributed to Institut Toulousain de simulation en santé, CHU Purpan, Toulouse.

Authors' Contributions

CHHC, CV, LG, and BB were responsible for the study concept and design. CHHC and CV were responsible for conducting the analysis and interpretation of the data. CHHC and DL drafted the manuscript. DL and SC contributed to the critical revisions of the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Multiview of Virtual Research Environment in Emergency Medicine.

[PDF File (Adobe PDF File), 190KB - games_v7i3e13993_app1.pdf]

Multimedia Appendix 2

Staff in Virtual Research Environment in Emergency Medicine.

[PDF File (Adobe PDF File), 571KB - games_v7i3e13993_app2.pdf]

Multimedia Appendix 3

Emergency room in Virtual Research Environment in Emergency Medicine.

[PDF File (Adobe PDF File), 349KB - games_v7i3e13993_app3.pdf]

References

1. Sun BC, Hsia RY, Weiss RE, Zingmond D, Liang LJ, Han W, et al. Effect of emergency department crowding on outcomes of admitted patients. Ann Emerg Med 2013 Jun;61(6):605-11.e6 [FREE Full text] [doi: 10.1016/j.annemergmed.2012.10.026] [Medline: 23218508]
2. Chiu IM, Lin YR, Syue YJ, Kung CT, Wu KH, Li CJ. The influence of crowding on clinical practice in the emergency department. Am J Emerg Med 2018 Jan;36(1):56-60. [doi: 10.1016/j.ajem.2017.07.011] [Medline: 28705743]
3. Ratwani RM, Fong A, Puthumana JS, Hettinger AZ. Emergency physician use of cognitive strategies to manage interruptions. Ann Emerg Med 2017 Nov;70(5):683-687. [doi: 10.1016/j.annemergmed.2017.04.036] [Medline: 28601266]
4. Yarmohammadian MH, Rezaei F, Haghshenas A, Tavakoli N. Overcrowding in emergency departments: a review of strategies to decrease future challenges. J Res Med Sci 2017;22:23 [FREE Full text] [doi: 10.4103/1735-1995.200277] [Medline: 28413420]
5. Wang H, Ojha RP, Robinson RD, Jackson BE, Shaikh SA, Cowden CD, et al. Optimal measurement interval for emergency department crowding estimation tools. Ann Emerg Med 2017 Nov;70(5):632-9.e4. [doi: 10.1016/j.annemergmed.2017.04.012] [Medline: 28688771]
6. Chen TL, Wang CC. Multi-objective simulation optimization for medical capacity allocation in emergency department. J Simul 2017 Dec 19;10(1):50-68. [doi: 10.1057/jos.2014.39]
7. Paul JA, Lin L. Models for improving patient throughput and waiting at hospital emergency departments. J Emerg Med 2012 Dec;43(6):1119-1126. [doi: 10.1016/j.jemermed.2012.01.063] [Medline: 22902245]
8. Flowerdew L, Brown R, Vincent C, Woloshynowych M. Identifying nontechnical skills associated with safety in the emergency department: a scoping review of the literature. Ann Emerg Med 2012 May;59(5):386-394. [doi: 10.1016/j.annemergmed.2011.11.021] [Medline: 22424651]
9. Schmutz J, Manser T. Do team processes really have an effect on clinical performance? A systematic literature review. Br J Anaesth 2013 Apr;110(4):529-544 [FREE Full text] [doi: 10.1093/bja/aes513] [Medline: 23454826]
10. Lorello GR, Cook DA, Johnson RL, Brydges R. Simulation-based training in anaesthesiology: a systematic review and meta-analysis. Br J Anaesth 2014 Feb;112(2):231-245 [FREE Full text] [doi: 10.1093/bja/aet414] [Medline: 24368556]
11. Olszewski AE, Wolbrink TA. Serious gaming in medical education: a proposed structured framework for game development. Simul Healthc 2017 Aug;12(4):240-253. [doi: 10.1097/SIH.0000000000000212] [Medline: 28027076]
12. Wang R, DeMaria S, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Feb;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
13. Bleichner G, Bleichner G, Braun F, Burnouf J, Cerfontaine C, Delaire L, et al. [Architecture des services d'urgence]. Journal Européen des Urgences 2005 Jun;18(2):102-126. [doi: 10.1016/s0993-9857(05)82478-1]
14. Azema O, Onde O, Mouron H. ORUMIP: Accueil. 2018. Panorama Occitanie: Activité des Structures D'urgences 2017 URL: https://www.orumip.fr/wp-content/uploads/2018/07/Panorama2017.pdf [accessed 2020-07-17]
15. Petraglia J. The real world on a short leash: the (mis) application of constructivism to the design of educational technology. Educ Technol Res Dev 1998 Sep;46(3):53-65. [doi: 10.1007/BF02299761]
16. Chiniara G, Cole G, Brisbin K, Huffman D, Cragg B, Lamacchia M, Canadian Network For Simulation In Healthcare, Guidelines Working Group. Simulation in healthcare: a taxonomy and a conceptual framework for instructional design and media selection. Med Teach 2013 Aug;35(8):e1380-e1395. [doi: 10.3109/0142159X.2012.733451] [Medline: 23121247]
17. 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]
18. de Leo LG, Diggs LA, Radici E, Mastaglio TW. Measuring sense of presence and user characteristics to predict effective training in an online simulated virtual environment. Simul Healthc 2014 Feb;9(1):1-6. [doi: 10.1097/SIH.0b013e3182a99dd9] [Medline: 24310164]
19. Block RA, Hancock PA, Zakay D. Physical load affects duration judgments: a meta-analytic review. Acta Psychol (Amst) 2016 Mar;165:43-47. [doi: 10.1016/j.actpsy.2016.01.002] [Medline: 26922615]
20. LeBlanc VR, Manser T, Weinger MB, Musson D, Kutzin J, Howard SK. The study of factors affecting human and systems performance in healthcare using simulation. Simul Healthc 2011 Aug(6 Suppl):S24-S29. [doi: 10.1097/SIH.0b013e318229f5c8] [Medline: 21817860]

21. LeBlanc VR. The effects of acute stress on performance: implications for health professions education. Acad Med 2009 Oct;84(10 Suppl):S25-S33. [doi: 10.1097/ACM.0b013e3181b37b8f] [Medline: 19907380]
22. Geeraerts T, Roulleau P, Cheisson G, Marhar F, Aidan K, Lallali K, et al. Physiological and self-assessed psychological stress induced by a high fidelity simulation course among third year anesthesia and critical care residents: an observational study. Anaesth Crit Care Pain Med 2017 Dec;36(6):403-406. [doi: 10.1016/j.accpm.2017.06.002] [Medline: 28648752]
23. LeBlanc VR, McConnell MM, Monteiro SD. Predictable chaos: a review of the effects of emotions on attention, memory and decision making. Adv Health Sci Educ Theory Pract 2015 Mar;20(1):265-282. [doi: 10.1007/s10459-014-9516-6] [Medline: 24903583]

Abbreviations

CRM: crisis resource management
ED: emergency department
VIREM: Virtual Research Environment in Emergency Medicine

Edited by T Rashid Soron; submitted 15.03.19; peer-reviewed by M Das, S Zheng; comments to author 02.04.19; revised version received 16.04.19; accepted 31.05.19; published 08.08.19.

Please cite as:
Houze-Cerfon CH, Vaissié C, Gout L, Bastiani B, Charpentier S, Lauque D
Development and Evaluation of a Virtual Research Environment to Improve Quality of Care in Overcrowded Emergency Departments: Observational Study
JMIR Serious Games 2019;7(3):e13993
URL: http://games.jmir.org/2019/3/e13993/
doi:10.2196/13993
PMID:31397292

Original Paper

An Epiduroscopy Simulator Based on a Serious Game for Spatial Cognitive Training (EpiduroSIM): User-Centered Design Approach

Junho Ko[1*], PhD; Jong Joo Lee[2*], MD; Seong-Wook Jang[1], PhD; Yeomin Yun[2], MS; Sungchul Kang[3], PhD; Dong Ah Shin[2*], MD; Yoon Sang Kim[1*], PhD

[1]BioComputing Lab, Institute for Bio-Engineering Application Technology, Korea University of Technology and Education, Cheonan, Republic of Korea

[2]Department of Neurosurgery, Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Robot Center, Samsung Research, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Yoon Sang Kim, PhD
BioComputing Lab
Institute for Bio-Engineering Application Technology
Korea University of Technology and Education
1600, Chungjeol-ro, Byeongcheon-myeon, Dongnam-gu
Cheonan, 31253
Republic of Korea
Phone: 82 5601496
Email: yoonsang@koreatech.ac.kr

Abstract

Background: Performing high-level surgeries with endoscopy is challenging, and hence, an efficient surgical training method or system is required. Serious game–based simulators can provide a trainee-centered educational environment unlike traditional teacher-centered education environments since serious games provide a high level of interaction (feedback that induces learning).

Objective: This study aimed to propose an epiduroscopy simulator, EpiduroSIM, based on a serious game for spatial cognitive training.

Methods: EpiduroSIM was designed based on a serious game. For spatial cognitive training, the virtual environment of EpiduroSIM was modeled based on a cognitive map.

Results: EpiduroSIM was developed considering user accessibility to provide various functions. The experiment for the validation of EpiduroSIM focused on psychological fidelity and repetitive training effects. The experiments were conducted by dividing 16 specialists into 2 groups of 8 surgeons. The group was divided into beginner and expert based on their epiduroscopy experience. The psychological fidelity of EpiduroSIM was confirmed through the training results of the expert group rather than the beginner group. In addition, the repetitive training effect of EpiduroSIM was confirmed by improving the training results in the beginner group.

Conclusions: EpiduroSIM may be useful for training beginner surgeons in epiduroscopy.

(JMIR Serious Games 2019;7(3):e12678) doi:10.2196/12678

KEYWORDS
medical education; endoscopy; surgery; catheters

Introduction

Background

An epiduroscopy is a highly effective minimally invasive surgery (MIS) used for chronic lumbago and lumbar disc herniation [1]. In an MIS, a subminiature endoscope (1-mm diameter) and a laser are inserted into a cavity; the endoscope is maneuvered by viewing a 2-dimensional (2D) computed tomography image. Performing high-level surgeries with endoscopy is challenging, and hence, an efficient surgical training method or system is required [2].

Conventional surgical training is performed using surgery observations, animals, cadavers, and plastic models. Surgery observation is achieved by assisting expert surgeons in the operation room. Such surgical training has limited training opportunities, and the interaction between surgical instruments and human organs is difficult to train [3]. Surgical training using animals and cadavers is costly [4], and surgical training using plastic models cannot provide realistic visual and tactile feedback [5].

Recently, surgical training has been performed using a simulator. The simulator provides an opportunity for the trainee to repeatedly attempt various actions, including mistakes, in a virtual environment [6]. In addition, the simulator allows the objective evaluation of the trainee's training level and can teach rare and dangerous complications at a low cost [4]. As the popularity of video games has increased, serious game technology has been applied to surgical training using a simulator [7]. As serious game provides a high level of interaction (feedback that induces learning) [8], serious game–based simulators can provide a trainee-centered educational environment unlike the traditional teacher-centered education environments [9].

Objectives

In this study, we proposed an epiduroscopy simulator (EpiduroSIM) based on a serious game for spatial cognitive training. EpiduroSIM aims to provide a trainee-centered educational environment for teaching the insertion paths of the surgical instrument (catheter).

Methods

Study Design

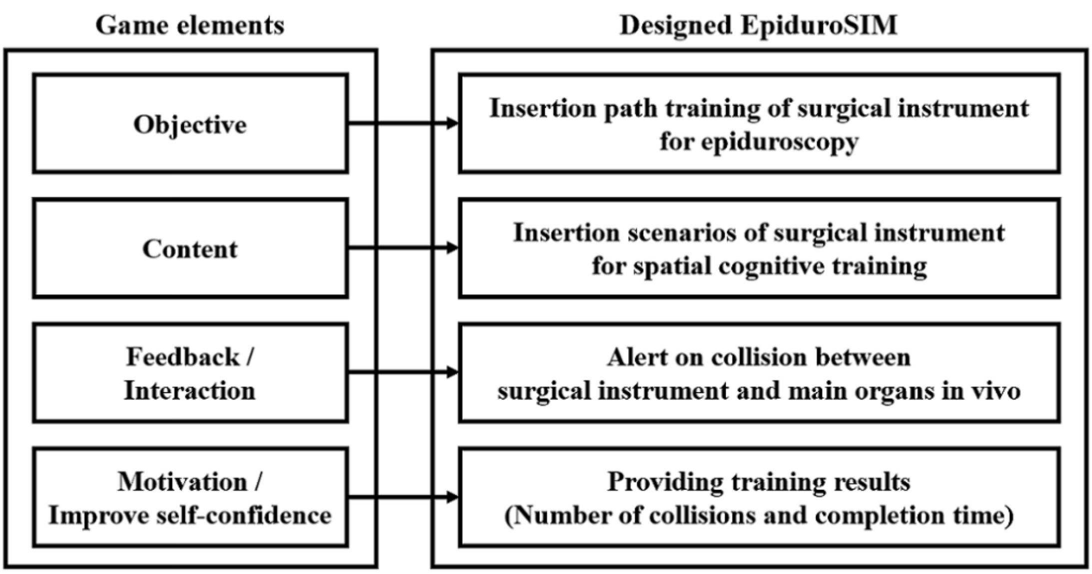
In this study, we focused on EpiduroSIM based on a serious game and the cognitive map for self-directed learning of the beginner. Herein, the beginner means a surgeon who has no experience or has low understanding of epiduroscopy. Design of EpiduroSIM and modeling of the virtual environment were achieved through consultation with neurosurgeons.

Analysis and Design

EpiduroSIM was designed based on the game elements of the serious game, as shown in Figure 1. The objective of the designed EpiduroSIM was to teach the insertion paths of the surgical instrument for epiduroscopy, and the content is the insertion scenario of the surgical instrument for spatial cognition training. The feedback and interaction of the designed EpiduroSIM is an alert on the collision between the surgical instrument and the primary organs in vivo. As surgery is generally evaluated based on the operation time (completion time) and mistakes (number of collisions) in surgery [10], the designed EpiduroSIM provides the number of collisions and the completion time as the training results.

Humans form a cognitive map through a series of processes that recognize the space or environment that they experience. In general, the cognitive map consists of an important point called a landmark and a surrounding area connecting these points [11]. This is because it is more effective in human information processing to hierarchically memorize the regions around the important points than to memorize the entire region [12].

Figure 1. The design of EpiduroSIM based on the game elements of the serious game.



The cognitive map is divided into the route map and survey map based on the form and characteristics [13]. The route map means a linear path connecting the primary landmarks between the starting point and destination. The route map is obtained through limited experience with the surrounding environment (if only 1 path is used). People with only the linear path information (route map) are more likely to get lost if they leave the proficient path [14]. The survey map means a map in which the geographical information between the starting point and the destination is configured in a network. The survey map is obtained through various experiences in the surrounding areas and data (maps and aerial photographs) that can illustrate the entire area. People with the survey map are more likely to search for an alternate path if they leave the proficient path. As the experience of the region increases, the route map generally changes to the survey map [15].

The virtual environment of EpiduroSIM is modeled based on the cognitive map as shown in Figure 2. The modeled virtual environment is divided into a 3-dimensional (3D) virtual environment for the route map and a 2D virtual environment for the survey map. The 3D virtual environment for the route map consists of bones, a dura mater, and discs that serve as landmarks. Through these objects serving as landmarks, the trainee can learn the catheter insertion paths. The 2D virtual environment for the survey map consists of a projected image (fluoroscopy). This allows the trainee to learn the perspective knowledge of all areas of the catheter insertion paths.

Hierarchical spatial cognition [16] does not determine a detailed path (A→S4→S3→B→S2→S1→C in scenario 1 of Figure 3) from the beginning when a person selects a path but selects a primary landmark (A→B→C in scenario 1 of Figure 3) first and subsequently selects a detailed path (A→S4→S3→B→S2→S1→C in scenario 1 of Figure 3). In this path selection process, landmarks with different levels of importance act as partial regions that constitute the whole space. Therefore, hierarchical spatial cognition can be thought as the divide-and-conquer method that resolves a significant problem.

EpiduroSIM provides various catheter insertion scenarios for spatial cognitive training. Figure 3 shows the hierarchical spatial cognitive process in various catheter insertion scenarios. In catheter insertion scenarios in Figure 3, the primary landmark is important for the catheter insertion procedures. In catheter insertion scenarios in Figure 3, the primary landmark A is the starting point where the catheter insertion paths are split into the dorsal side and ventral side in the sacral hiatus. In scenario 1 and scenario 2, the primary landmark B is the point where the catheter insertion paths detour from the dorsal side to the ventral side. In scenario 3 and scenario 4, the primary landmark B is absent because the catheter insertion paths in scenarios 3 and 4 enter the ventral side from the sacral hiatus to reach the lesion disc located on the ventral side without a detour. In catheter insertion scenarios, the primary landmark C is the destination, which is the lesion disc. In catheter insertion scenarios, the secondary landmark is the sacral nerves (S1-4), which is used as a reference point for the trainee to select the detailed paths.

Figure 2. The modeled virtual environments (2D and 3D) based on the cognitive map.

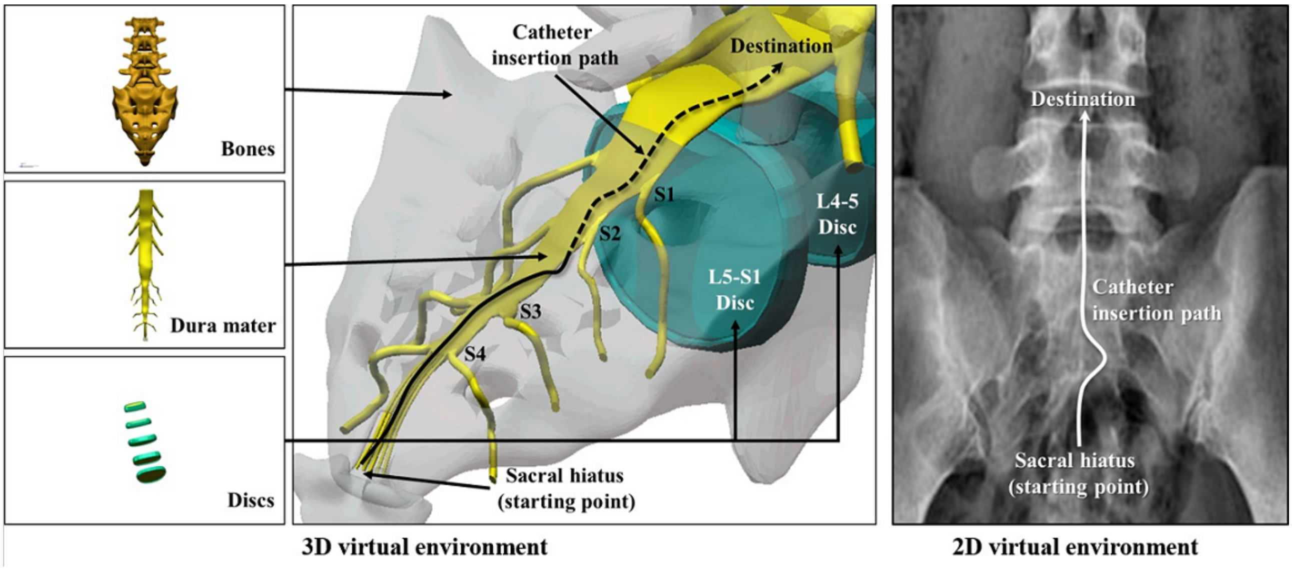
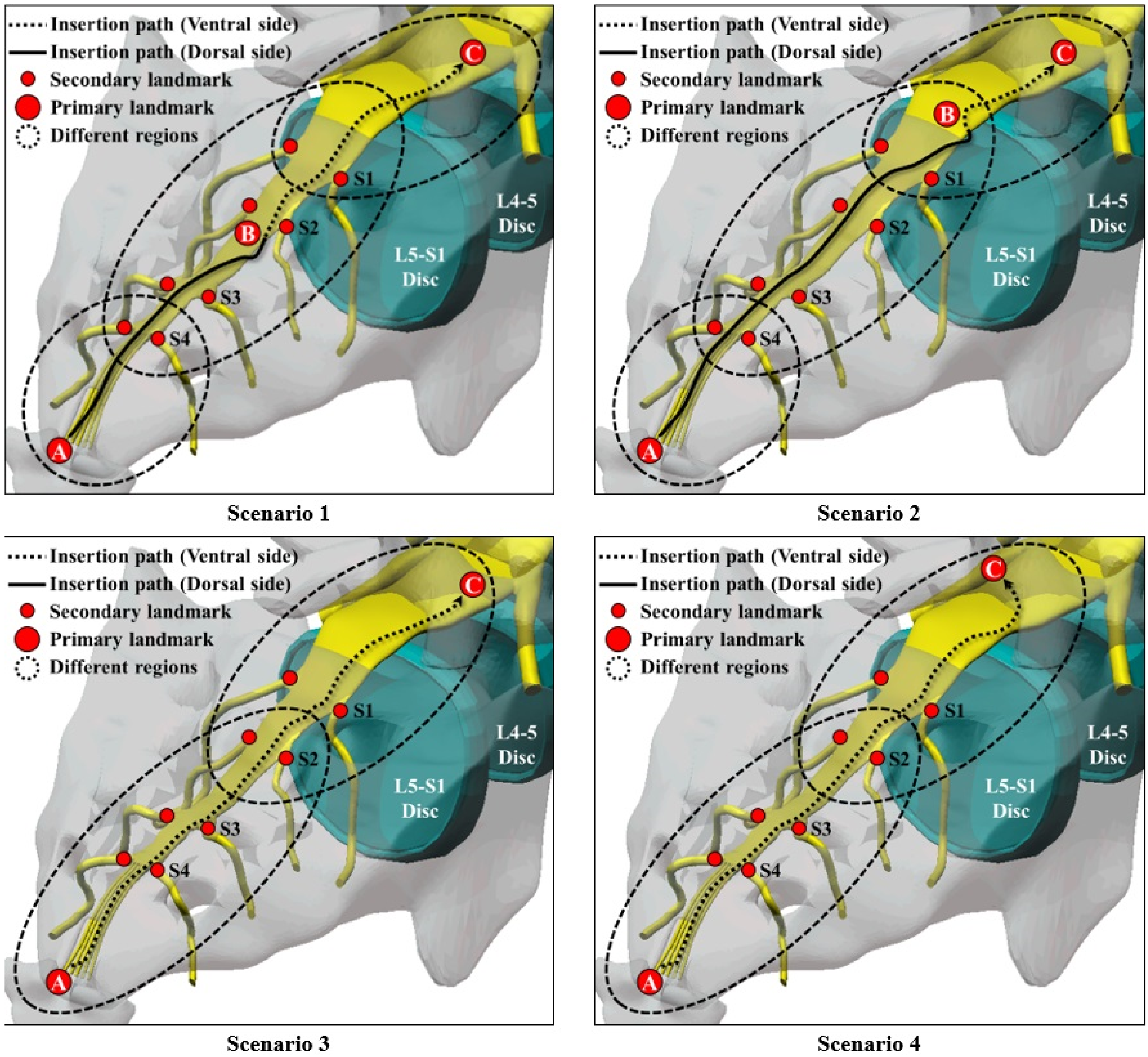


Figure 3. The hierarchical spatial cognitive process in the various catheter insertion scenarios.



Results

Development

EpiduroSIM consists of the input part, processing part, and output part, as shown in Figure 4. The input part transmits the position and rotation data received from the keyboard or mouse, gamepad [17], and master device [18] to the processing part. The processing part controls the surgical instrument in the virtual environment according to the received position and rotation data. When the surgical instrument collides, it stores the crash record and provides the trainee with visual feedback. Furthermore, the force data calculated at the time of the collision are transmitted to the master device, and the feedback is provided to the trainee. The output part provides visually the virtual environments (Figure 2) and training results in the Database.

EpiduroSIM was developed using Unity3D (Unity technologies Corp), a commercial game engine. The developed EpiduroSIM provides an easy operation method and various functions considering user accessibility, as shown in Figure 5. The functions provided by the developed EpiduroSIM are the fluoroscopic view, the catheter insertion path, the recording, and the miscellaneous menu. The fluoroscopic view function (① in Figure 5) provides a screen similar to the C-arm view used in actual surgery, and the catheter insertion path function (② in Figure 5) provides 4 training scenarios (Figure 3). The recording function (③ in Figure 5) measures and records the training results (number of collisions and completion time) in real time, and the miscellaneous menu function (④ in Figure 5) includes virtual fixture visualization, catheter movement speed control (1x/2x), and position correction at the catheter deviation.

The catheter insertion training in the developed EpiduroSIM proceeds in the following order: (1) select a training scenario, (2) begin training, (3) move the catheter tooltip to the destination (lesion), and (4) validate the training results. If the trainee selects a training scenario other than the free scenario, a virtual fixture in the form of a tube is created, as shown in Figure 6. A virtual fixture is a technique that restricts the motion of a control object to a specified path [19]. The trainee learns the catheter insertion paths along the virtual fixture to the destination. During the training, the catheter tooltip will collide with the virtual fixture if the catheter is out of the insertion path owing to the intentions or mistakes of the trainee. When the catheter tooltip collides, the developed EpiduroSIM records the number of collisions in the training results and provides visual feedback by changing the color (from white to red) of the catheter tooltip, as shown in Figure 7. The trainee can learn the catheter insertion paths in the desired field of view by selecting the external view, as shown in Figure 8, or the endoscopic view, as shown in Figure 9, while moving the catheter tooltip. At the end of the training, the developed EpiduroSIM provides the user with the cumulative number of collisions and completion time.

Evaluation

The experiments were designed to focus on the fidelity and repetitive training effects. Fidelity represents the degree to which the simulator matches the real environment and skills [20,21]. Fidelity can be divided into physical fidelity and psychological fidelity [22]. Physical fidelity refers to the degree of similarity between a real environment and virtual environment [21,22]. Psychological fidelity refers to the degree to which the skills of real tasks are reflected in the simulator [22]. EpiduroSIM is designed based on a serious game, focusing on psychological fidelity rather than physical fidelity. Therefore, we examined the psychological fidelity of EpiduroSIM in the experiments.

The training result in the simulator based on iterations generally appears as learning curves. This means that the training result is improved as the training using the simulator is repeated. This improvement in the training result does not necessarily imply that the trainee is acquiring actual skills. However, improved training results in simulators that train basic skills such as hand and tool movement can lead to the acquisition of actual skills [23]. This type of verification has been used in many studies as a simulator evaluation tool [24,25]. Therefore, we examined the repetitive training effects of EpiduroSIM in the experiments.

The experiments were conducted by dividing 16 specialists into 2 groups of 8 surgeons. The group was divided into beginner and expert based on their epiduroscopy experience. The sample size (n=16) of the subject was calculated with a .05 significance level, 0.9 power, and 0.3 effect size using G*Power 3 (Heinrich Heine University Düsseldorf) [26].

Figure 4. The structure of the proposed EpiduroSIM.

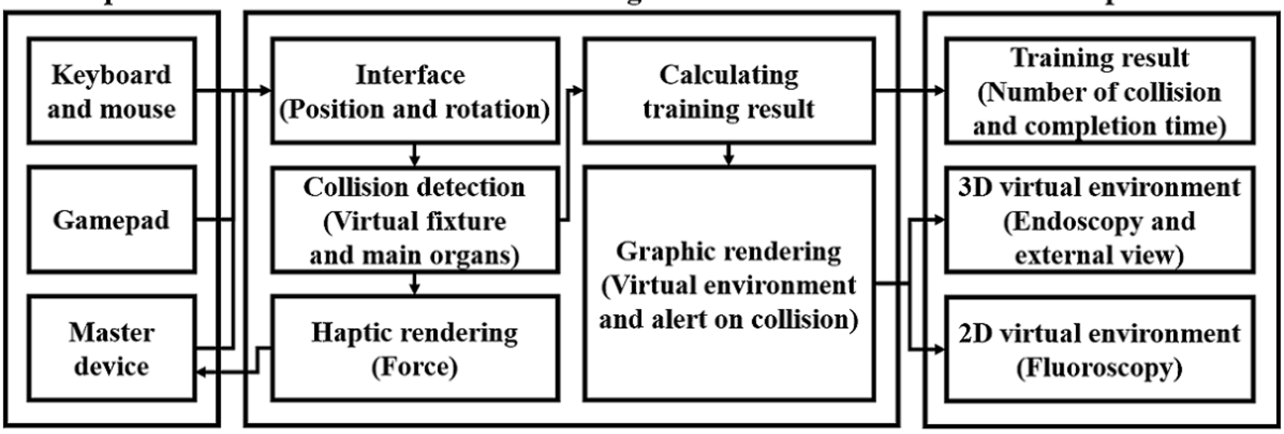


Figure 5. The developed EpiduroSIM considering user accessibility.

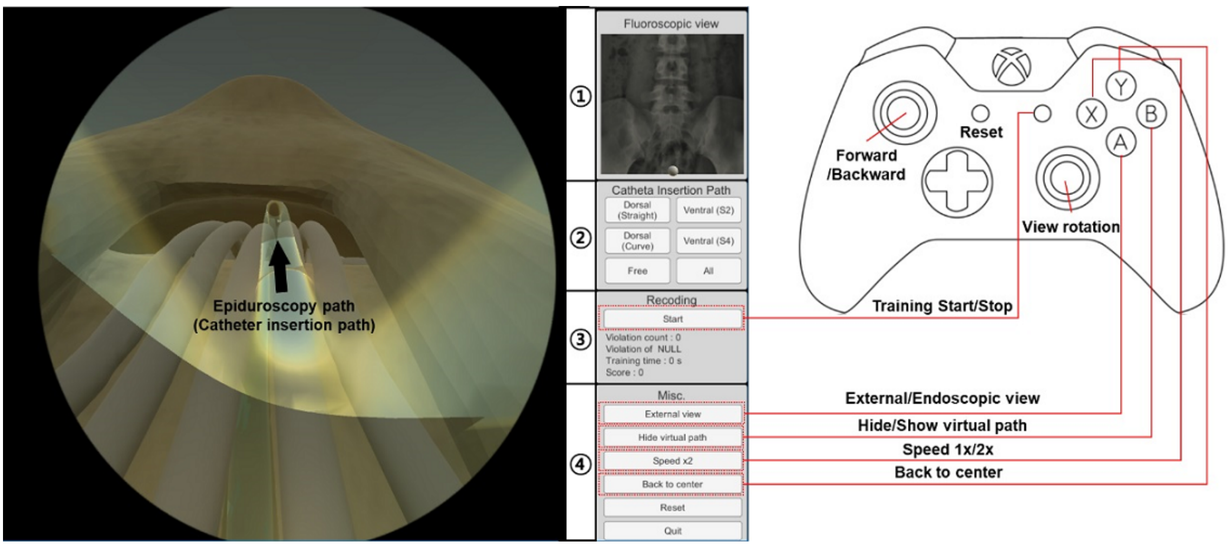


Figure 6. The catheter insertion training without collision.

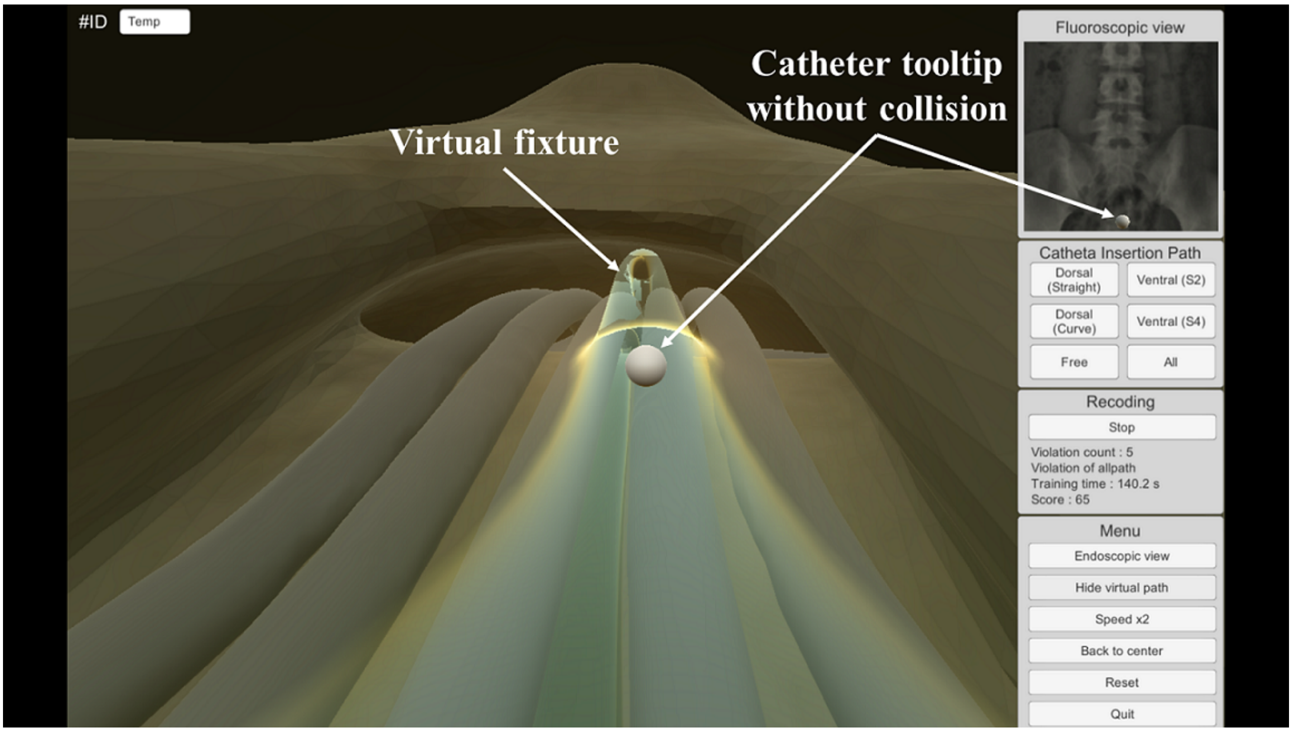


Figure 7. The catheter insertion training with collision.

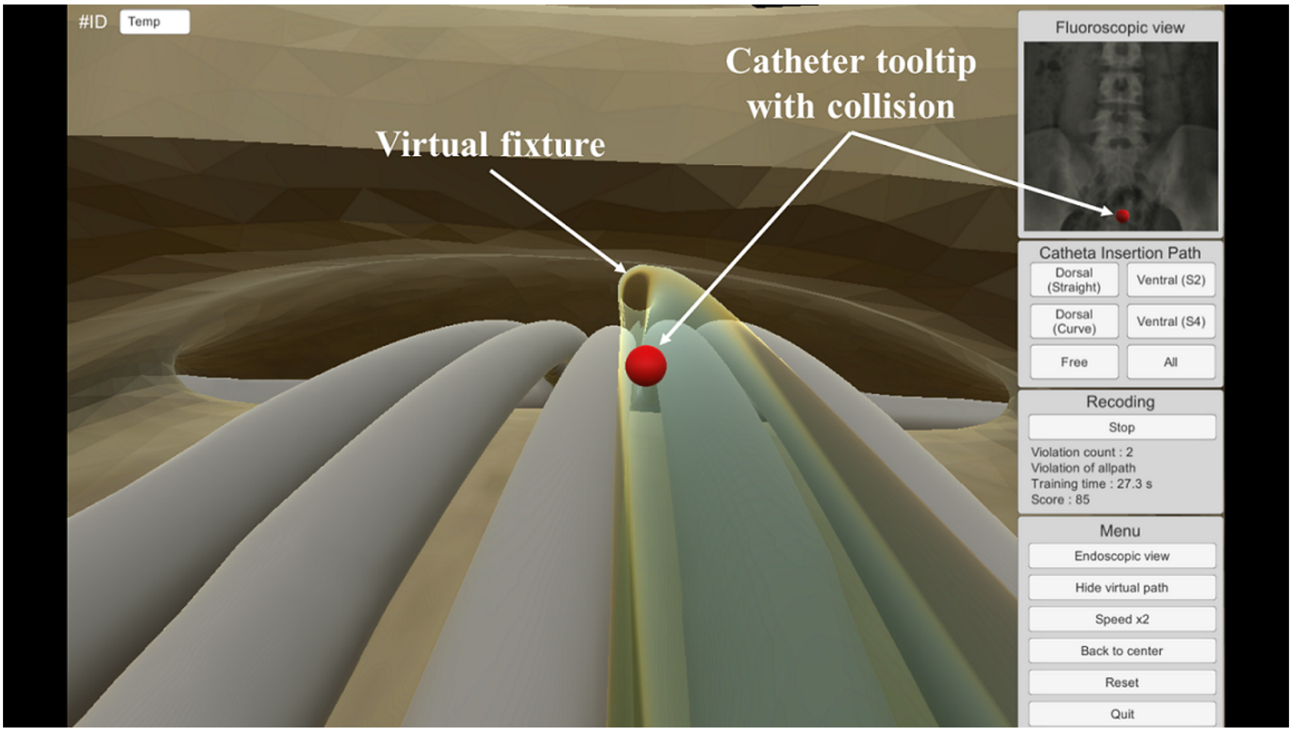


Figure 8. The external view.

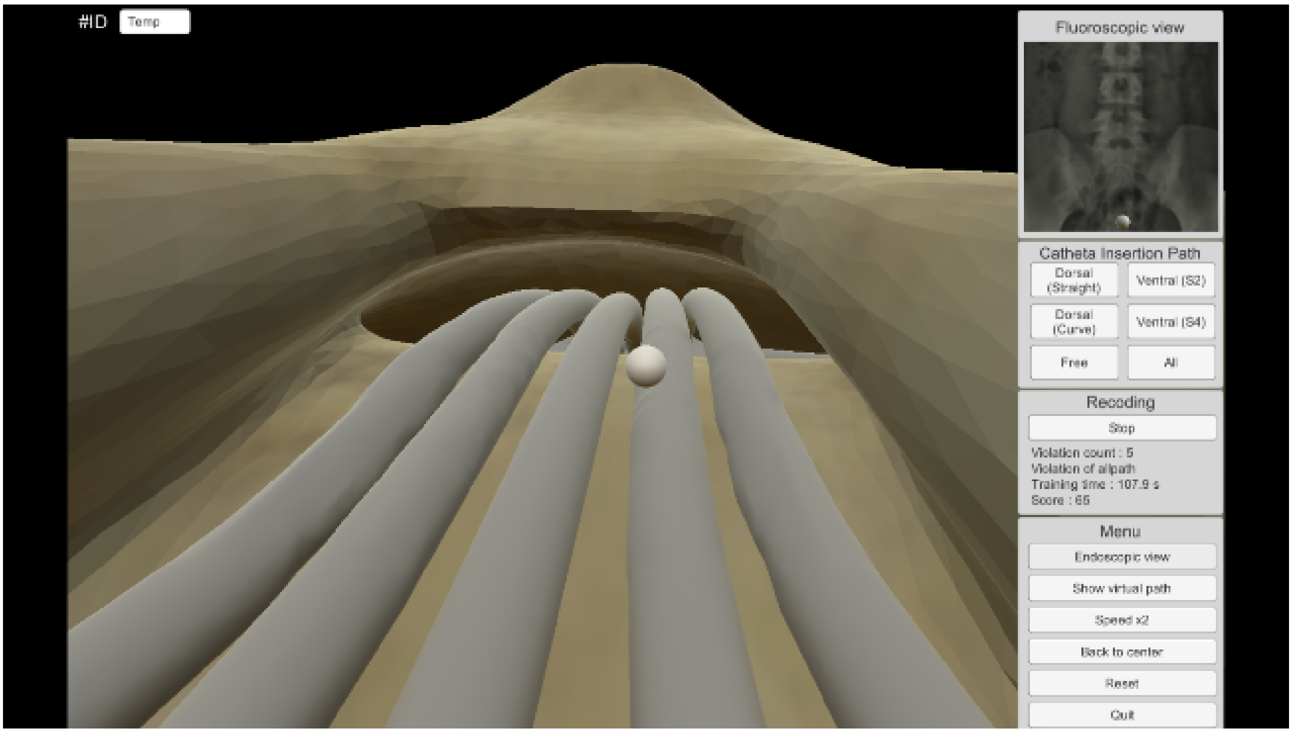
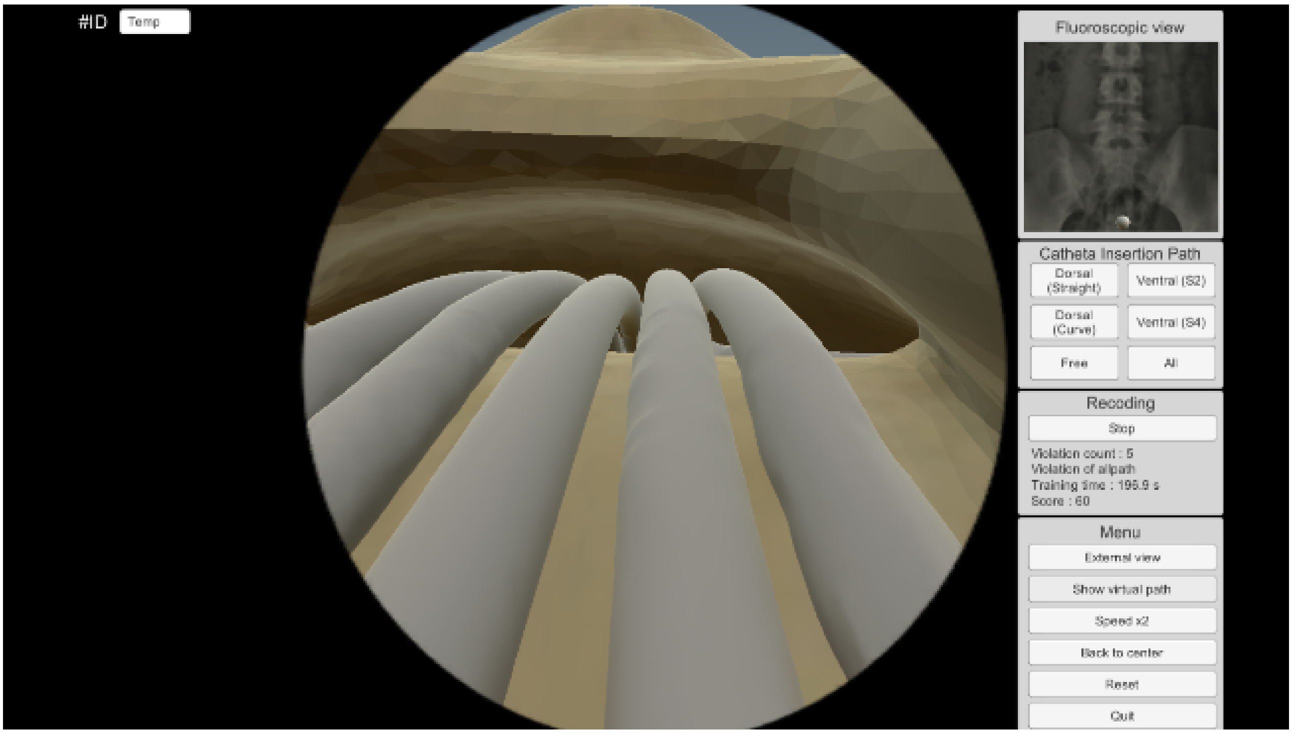


Figure 9. The endoscopic view.



The experimental environment consists of EpiduroSIM and gamepad as shown in Figure 10. A gamepad is an input device with higher precision and user preference than a mechanical master device and joystick [27]. In addition, considering a serious game and user accessibility, the gamepad is suitable as the input device of EpiduroSIM. The experiment was conducted through repetitive training (20 times) in the free scenario (an environment without a virtual fixture) of EpiduroSIM. To ensure the reliability of the experimental results, we set the number of repetitions to be higher than the conventional studies [28,29] on the surgical training simulator. The experimental results of repetitive training were analyzed using the repeated-measures analysis of variance. These analyses were performed using SPSS 21 (IBM Corp). A 2-sided significance level of $P<.05$ was set.

The results of comparison between the 2 groups (beginner and expert) are shown in Figure 11. The mean number of collisions for the beginner group and expert group was 4.91 and 3.56, respectively. The mean completion time of the beginner group and expert group was 84.14 seconds and 83.40 seconds, respectively. The expert group had fewer collisions than the beginner group, and the completion time was faster. This implies that the understanding of actual surgery is well reflected (high psychological fidelity) in EpiduroSIM.

Figure 12 shows the experimental results of repetitive training. As the training was repeated, the beginner group showed a decrease in the number of collisions as a whole and the expert group did not show any significant variation. As the training was repeated, the beginner group had a shorter completion time and the expert group did not show any significant change. From the experimental results, we confirmed that the beginner group exhibited the repetitive training effects, in contrast to the expert group.

Figure 10. The experimental environment.

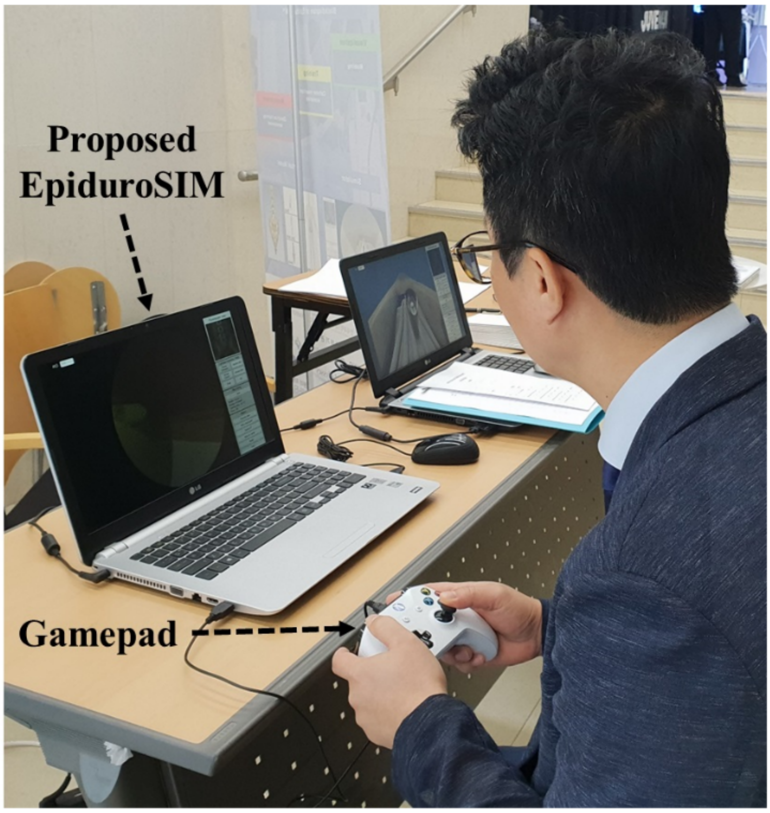


Figure 11. The comparison results (number of collisions and completion time) between the two groups (beginner and expert).

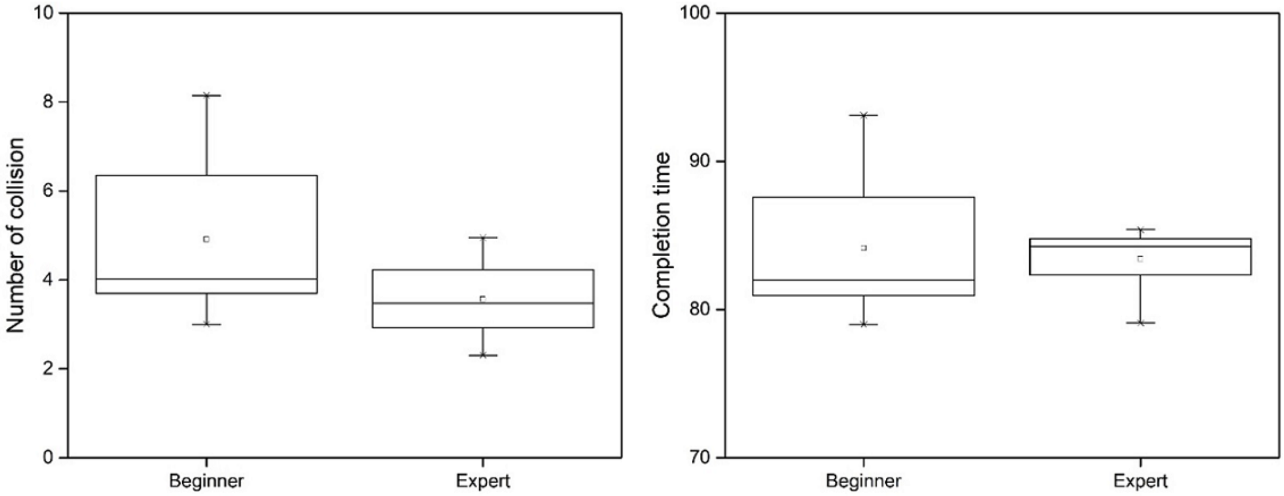
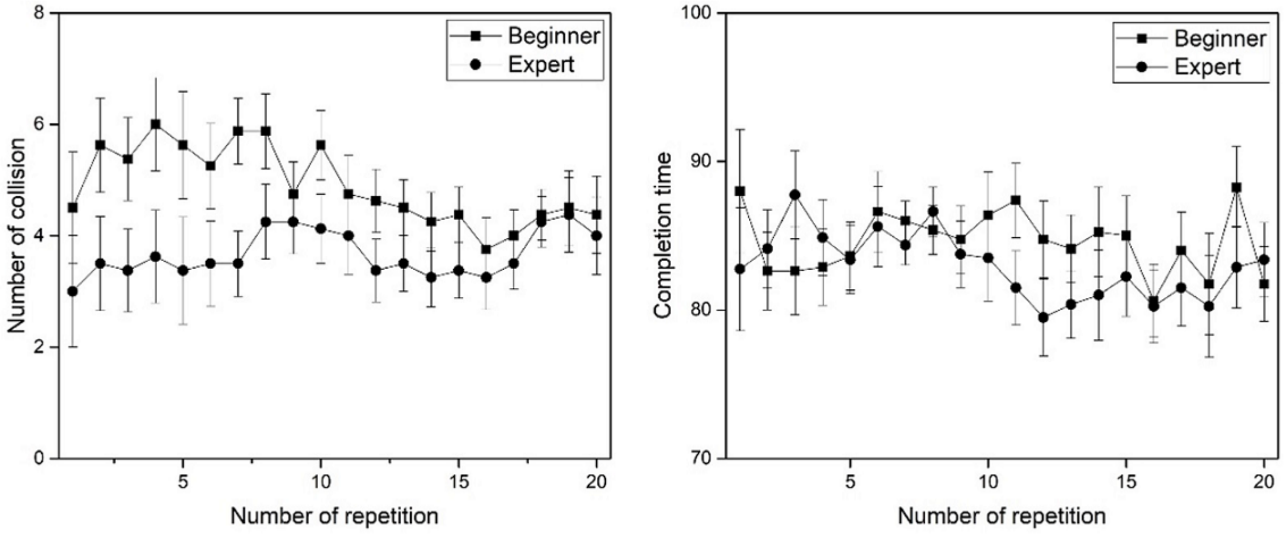


Figure 12. The experimental results (number variation of collision and completion time variation) of repetitive training.



Discussion

Principal Findings

In this study, we proposed an epiduroscopy simulator (EpiduroSIM) based on a serious game for spatial cognitive training. EpiduroSIM was designed based on a game element of a serious game. The objective of EpiduroSIM was to help beginners learn the insertion path of the catheter inside the human body. Therefore, EpiduroSIM provides various scenarios of catheter insertion to the trainee as content in the virtual environment modeled by the cognitive map. The interaction between EpiduroSIM and the trainee was achieved through visual feedback based on the virtual fixture. Through this interaction, the trainee learns a correct insertion path. Finally, the EpiduroSIM provides the training results of the insertion completion time of the catheter and the mistakes (number of collisions) that occurred during the insertion process to improve the motivation and self-confidence of the trainee.

The experiment for the validation of EpiduroSIM focused on the psychological fidelity and repetitive training effects. On the basis of a comparative result between the expert and beginner groups, it can be concluded that the expert group completed catheter insertion faster with fewer collisions than the beginner group. The result indicates that the understanding of actual surgery is well applied (high psychological fidelity) in EpiduroSIM. From the repeated training results for each group, the beginner group showed a decrease in the number of collisions and completion time. Generally, surgery is evaluated based on the surgical time (completion time) and mistakes (number of collisions) that occur during the surgery. Therefore, the decrease in the completion time and the number of collisions in the beginner group means that the understanding of the surgery has increased. However, there was no repeated training effect in the expert group. The reason could be that experts have a higher understanding of the surgery before training by actual surgical experience compared with the beginners.

Comparison With Past Work

Cognition is a general term used to describe mental activities related to thinking, learning, and memory [30]. Cognitive training is applied in the form of mental imagery without actual physical movements [31]. Cognitive training has been used in sports and has become an important tool used by athletes at the elite level [32]. In addition, cognitive training has been applied in the form of simulators in the aeronautical field to improve flight performance through pilot training [33].

Owing to the proven efficacy in the sport and aviation sectors, studies have been conducted on using cognitive training for surgical training. In a study on the relationship between surgical proficiency and cognitive processes, surgical competence reportedly combined the intellectual movements of decision making with the ability to perform mechanical tasks [34]. In a study involving 58 expert surgeons, 70% agreed that cognitive skill was one of the best traits required by the trainee surgeons [35].

Owing to MIS, the ability to perform mechanical tasks has become important but the importance of cognitive skills has not diminished. As laparoscopic surgeries and robotic surgeries require significant levels of cognitive and decision-making abilities [24], a few simulators have been studied to provide cognitive training to surgeons [7,23,36]. In these studies, a variety of virtual reality tasks were designed to improve specific cognitive skills such as movement planning, working memory, and preparatory attention. In other studies, error recognition and the cognitive simulator feedback were evaluated [37,38]. Although these cognitive simulators have been developed focusing on surgical procedures, spatial cognitive training for learning the insertion paths of surgical instruments in the human body was not considered.

We herein proposed EpiduroSIM based on a serious game for spatial cognitive training. To learn the catheter insertion paths in the human body, EpiduroSIM was designed based on the game elements of the serious educational game. For spatial cognitive training, the virtual environment of EpiduroSIM was modeled based on a cognitive map. EpiduroSIM was developed to provide various functions considering user accessibility. The experiment for the validation of EpiduroSIM focused on the psychological fidelity and repetitive training effects. The psychological fidelity of EpiduroSIM was confirmed through the training results of the expert group rather than the beginner group. In addition, the repetitive training effect of EpiduroSIM was confirmed by improving the training results in the beginner group.

Limitations

Serious games based on the game elements for entertainment are differentiated from simulators for reflecting reality [39]. A serious game with game elements such as competition, self-confidence, environment, objectives, and rules has the advantage of improving the self-direction, personality, and persistence of learning compared with the simulator [40]. Despite the advantages of a serious game, the following 7 areas of study are required for the effective development of a serious game for medical education [41]: (1) disposition to engage in learning, (2) impact of realism and fidelity on learning, (3) threshold for learning, (4) process of cognitive development during knowledge gain, (5) stability of knowledge gain (retention), (6) capacity for knowledge transfer to related problems, and (7) disposition toward sensible action within clinical settings. Among these areas, our study focused on the process of cognitive development during knowledge gain.

Therefore, EpiduroSIM provides high-level accessibility to beginners for self-directed learning without past knowledge of the simulator, but it has a limitation that it does not provide the same experience as actual surgery such as the kinesthetic sense.

Conclusions

Serious games have begun to be applied to various applications, including medical education. However, further studies are still required to popularize serious games. We herein applied the cognitive map to the insertion paths training of a surgical instrument to study the cognitive development process during knowledge gain. In the future, we expect the importance of cognitive training in surgical education to increase, and we hope that the results of this study will be used in cognitive training studies.

Acknowledgments

This research was financially supported by the Global Collaborative R&D Program through the Ministry of Trade, Industry & Energy and Korea Institute for Advancement of Technology.

Conflicts of Interest

None declared.

References

1. Roberts KE, Bell RL, Duffy AJ. Evolution of surgical skills training. World J Gastroenterol 2006 May;12(20):3219-3224 [FREE Full text] [doi: 10.3748/wjg.v12.i20.3219] [Medline: 16718842]
2. Pan JJ, Chang J, Yang X, Liang H, Zhang JJ, Qureshi T, et al. Virtual reality training and assessment in laparoscopic rectum surgery. Int J Med Robot 2015 Jun;11(2):194-209. [doi: 10.1002/rcs.1582] [Medline: 24700686]
3. Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R. Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. Ann Surg 2006 Jul;244(3):400-409. [doi: 10.1097/01.sla.0000234808.85789.6a] [Medline: 16926566]
4. Muchinsky PM. Psychology Applied to Work. Summerfield, NC, USA: Hypergraphic Press; 2012.
5. Heng P, Cheng C, Wong T, Xu Y, Chui Y, Chan K, et al. A virtual-reality training system for knee arthroscopic surgery. IEEE Trans Inf Technol Biomed 2004 Jun;8(2):217-227. [Medline: 15217267]
6. Ziv A, Ben-David S, Ziv M. Simulation based medical education: an opportunity to learn from errors. Med Teach 2005 May;27(3):193-199 [FREE Full text] [doi: 10.1080/01421590500126718] [Medline: 16011941]
7. Cowan B, Sabri H, Kapralos B, Cristancho S, Moussa F, Dubrowski A. SCETF: Serious game surgical cognitive education and training framework. : IEEE; 2011 Nov Presented at: 2011 IEEE International Games Innovation Conference (IGIC); November 2-3, 2011; Orange, CA, USA p. 5117 URL: https://ieeexplore.ieee.org/abstract/document/6115117 [doi: 10.1109/IGIC.2011.6115117]
8. Susi T, Johannesson M, Backlund P. DiVA. Sweden: University of Skövde School of Humanities and Informatics; 2007 Feb. Serious games-an overview URL: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A2416&dswid=6821 [accessed 2019-08-08]
9. Stapleton AJ. Serious Games: Serious Opportunities. 2004 Jan Presented at: Australian Game Developers' Conference, Academic Summit; 2-4 December; Melbourne, Victoria, Australia URL: http://www.agdc.com.au/04presentations/acad_andrew_stapleton2.pdf
10. Aggarwal R, Grantcharov TP, Eriksen JR, Blirup D, Kristiansen VB, Funch-Jensen P, et al. An evidence-based virtual reality training program for novice laparoscopic surgeons. Ann Surg 2006 Aug;244(2):310-314 [FREE Full text] [doi: 10.1097/01.sla.0000218094.92650.44] [Medline: 16858196]
11. Couclelis H, Golledge RG, Gale N, Tobler W. Exploring the anchor-point hypothesis of spatial cognition. J Environ Psychol 1987 Jun;7(2):99-122 [FREE Full text] [doi: 10.1016/S0272-4944(87)80020-8]
12. Wiener JH, Mallot HA. 'Fine-to-Coarse' route planning and navigation in regionalized environments. Spat Cogn Comput 2003 Dec;3(4):331-358 [FREE Full text] [doi: 10.1207/s15427633scc0304_5]

13. Shemaykin F. US Office of technical Reports. Arlington, VA, USA: Psychological Science in the USSR; 1962. General problems of orientation in space and space representations URL: https://tinyurl.com/y4k323qv [accessed 2019-08-08]
14. Prestopnik J, Roskos–Ewoldsen B. The relations among wayfinding strategy use, sense of direction, sex, familiarity, and wayfinding ability. J Environ Psychol 2000 Jun;20(2):177-191 [FREE Full text] [doi: 10.1006/jevp.1999.0160]
15. Appleyard D. Styles and methods of structuring a city. Environ Behav 2016 Jul;2(1):100-117 [FREE Full text] [doi: 10.1177/001391657000200106]
16. Chown E, Kaplan S, Kortenkamp D. Prototypes, location, and associative networks (PLAN): towards a unified theory of cognitive mapping. Cogn Sci 1995 Jan;19(1):1-51 [FREE Full text] [doi: 10.1016/0364-0213(95)90003-9]
17. Jang S, Ko J, Choi Y, Kim Y. A development of an epiduroscopy training simulator based on spatial cognition learning. J Vibroeng 2017 Oct 21;14:250-255 [FREE Full text] [doi: 10.21595/vp.2017.19166]
18. Ko J, Jang S, Kim Y. Development of epiduroscopy training simulator using haptic master device. 2017 Jun Presented at: International Conference on Ubiquitous Robots and Ambient Intelligence (URAI); June 28-July 1; Jeju, South Korea p. 542-543 URL: https://ieeexplore.ieee.org/document/7992664 [doi: 10.1109/URAI.2017.7992664]
19. Kim J, Jang J, Jung G, Kim J, Jang S. Epiduroscopic laser disc and neural decompression. J Neurosurg Rev 2011;1(S1):14-19 [FREE Full text]
20. Farmer E, Rooij J, Riemersma J, Moraal J, Jorna P. Handbook of Simulator-Based Training. London: Routledge; 1999.
21. Hays RT, Singer MJ. Simulation Fidelity In Training System Design: Bridging The Gap Between Reality And Training (recent Research In Psychology). New York, USA: Springer-Verlag; 1988.
22. Ker J, Bradley P. Simulation in medical education. In: Understanding Medical Education: Evidence, Theory And Practice. London, UK: Wiley-Blackwell; 2010:164-180.
23. Kahol K, Vankipuram M, Smith M. Cognitive simulators for medical education and training. J Biomed Inform 2009 Aug;42(4):593-604 [FREE Full text] [doi: 10.1016/j.jbi.2009.02.008] [Medline: 19269350]
24. Kohls-Gatzoulis J, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. Can J Surg 2004 Aug;47(4):277-283 [FREE Full text] [Medline: 15362330]
25. Gallagher A, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001 Nov;25(11):1478-1483. [Medline: 11760752]
26. 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 [FREE Full text] [doi: 10.3758/BF03193146] [Medline: 17695343]
27. Marcia K, Gupta A. Passive and active assistance for human performance of a simulated underactuated dynamic task. : IEEE; 2003 Apr Presented at: Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems; March 22-23; Los Angeles, CA, USA p. 348 URL: https://ieeexplore.ieee.org/abstract/document/1191308 [doi: 10.1109/HAPTIC.2003.1191308]
28. Andersen SA, Konge L, Cayé-Thomasen P, Sørensen MS. Learning curves of virtual mastoidectomy in distributed and massed practice. JAMA Otolaryngol Head Neck Surg 2015 Oct;141(10):913-918. [doi: 10.1001/jamaoto.2015.1563] [Medline: 26334610]
29. Connor E, Raker C, Wohlrab K. Effects of repetition and inactivity on laparoscopic skills training. J Minim Invasive Gynecol 2016 Feb;23(2):194-197 [FREE Full text] [doi: 10.1016/j.jmig.2015.09.008] [Medline: 26391057]
30. Coren S, Ward L, Enns J. Sensation and perception. In: Handbook of psychology. New York, USA: John Wiley & Sons; Apr 2003.
31. Munzert J, Lorey B, Zentgraf K. Cognitive motor processes: the role of motor imagery in the study of motor representations. Brain Res Rev 2009 May;60(2):306-326 [FREE Full text] [doi: 10.1016/j.brainresrev.2008.12.024] [Medline: 19167426]
32. Sevdalis N, Moran A, Arora S. Mental imagery and mental practice applications in surgery: state of the art and future directions. In: Multisensory Imagery. New York, USA: Springer; May 2014:343-363.
33. Gopher D, Weil M, Siegel D. Practice under changing priorities: an approach to the training of complex skills. Acta Psychologica 1989 Aug;71(1-3):147-177 [FREE Full text] [doi: 10.1016/0001-6918(89)90007-3]
34. Hall J, Ellis C, Hamdorf J. Surgeons and cognitive processes. Br J Surg 2003 Jan;90(1):10-16. [doi: 10.1002/bjs.4020] [Medline: 12520568]
35. Cuschieri A, Francis N, Crosby J, Hanna G. What do master surgeons think of surgical competence and revalidation? Am J Surg 2001 Aug;182(2):110-116. [doi: 10.1016/s0002-9610(01)00667-5] [Medline: 11574079]
36. Kahol K, Deka M, Mayes S, Smith M, Ferrara JJ, Panchanathan S. The effect of fatigue on cognitive and psychomotor skills of surgical residents. 2007 Presented at: 3rd International Conference on Foundations of Augmented Cognition; Jul 22-27; Beijing, China p. 304. [doi: 10.1016/j.amjsurg.2007.10.004]
37. Loveday B, Oosthuizen G, Diener B, Windsor J. A randomized trial evaluating a cognitive simulator for laparoscopic appendectomy. ANZ J Surg 2010 Sep;80(9):588-594 [FREE Full text] [doi: 10.1111/j.1445-2197.2010.05349.x] [Medline: 20840399]
38. Palter V, Grantcharov TP. Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. Ann Surg 2012 Jul;256(1):25-32 [FREE Full text] [doi: 10.1097/SLA.0b013e318258f5aa] [Medline: 22664557]

39. Greitzer F, Kuchar O, Huston K. Cognitive science implications for enhancing training effectiveness in a serious gaming context. J Educ Resour Comput 2007 Nov;7(3):- [FREE Full text] [doi: 10.1145/1281320.1281322]
40. Chan W, Qin J, Chui Y, Heng P. A serious game for learning ultrasound-guided needle placement skills. IEEE Trans Inf Technol Biomed 2012 Nov;16(6):1032-1042 [FREE Full text] [doi: 10.1109/TITB.2012.2204406] [Medline: 22711779]
41. Tashiro J, Dunlap D. The impact of realism on learning engagement in educational games. In: Proceedings of the 2007 conference on Future Play. 2007 Nov Presented at: Future Play '07; November 14-17; Toronto, Canada p. 113-120 URL: https://dl.acm.org/citation.cfm?doid=1328202.1328223 [doi: 10.1145/1328202.1328223]

Abbreviations

2D: 2-dimensional
3D: 3-dimensional
MIS: minimally invasive surgery

Edited by G Eysenbach; submitted 05.11.18; peer-reviewed by R Ciptaningtyas, P Stourac; comments to author 10.01.19; revised version received 31.03.19; accepted 16.04.19; published 27.08.19.

Please cite as:
Ko J, Lee JJ, Jang SW, Yun Y, Kang S, Shin DA, Kim YS
An Epiduroscopy Simulator Based on a Serious Game for Spatial Cognitive Training (EpiduroSIM): User-Centered Design Approach
JMIR Serious Games 2019;7(3):e12678
URL: http://games.jmir.org/2019/3/e12678/
doi:10.2196/12678
PMID:31456580

Original Paper

Effects of Social Interaction Mechanics in Pervasive Games on the Physical Activity Levels of Older Adults: Quasi-Experimental Study

Luciano Henrique De Oliveira Santos[1], MSc; Kazuya Okamoto[1,2], PhD; Silvana Schwerz Funghetto[3], PhD; Adriana Schüler Cavalli[4], PhD; Shusuke Hiragi[1,2], PhD; Goshiro Yamamoto[2], PhD; Osamu Sugiyama[5], PhD; Carla Denise Castanho[6], PhD; Tomoki Aoyama[7], PhD; Tomohiro Kuroda[1,2], PhD

[1]Graduate School of Informatics, Kyoto University, Kyoto, Japan
[2]Division of Medical Information Technology and Administration Planning, Kyoto University Hospital, Kyoto, Japan
[3]Faculty UnB Ceilândia, University of Brasília, Brasília, Brazil
[4]Escola Superior de Educação Física, Federal University of Pelotas, Pelotas, Brazil
[5]Preemptive Medicine & Lifestyle-Related Disease Research Center, Kyoto University Hospital, Kyoto, Japan
[6]Graduate Program in Informatics, University of Brasília, Brasília, Brazil
[7]Graduate School of Medicine, Kyoto University, Kyoto, Japan

Corresponding Author:
Luciano Henrique De Oliveira Santos, MSc
Graduate School of Informatics
Kyoto University
54 Shogoin Kawahara-cho, Sakyo-ku
Dept of Medical Informatics, Kyoto University Hospital
Kyoto, 606-8507
Japan
Phone: 81 75 366 7701
Fax: 81 75 366 7704
Email: lhsantos@kuhp.kyoto-u.ac.jp

Abstract

Background: The novel genre of pervasive games, which aim to create more fun and engaging experiences by promoting deeper immersion, could be a powerful strategy to stimulate physical activity among older adults. To use these games more effectively, it is necessary to understand how different design elements affect player behavior.

Objective: The aim was to vary a specific design element of pervasive games for older adults, namely social interaction, to test the effect on levels of physical activity.

Methods: Over 4 weeks, two variations of the same pervasive game were compared: social interaction for the test group and no social interaction for the control group. In both versions, players had to walk to physical locations and collect virtual cards, but the social interaction version allowed people to collaborate to obtain more cards. Weekly step counts were used to evaluate the effect on each group, and the number of places visited was used as an indicator of play activity.

Results: A total of 32 participants were recruited (no social interaction=15, social interaction=17); 18 remained until the end of the study (no social interaction=7, social interaction=11). Step counts during the first week were used as the baseline (no social interaction: mean 17,099.4, SE 3906.5; social interaction: mean 17,981.9, SE 2171.1). For the following weeks, changes to individual baseline were as follows for no social interaction (absolute/proportional): 383.8 (SE 563.8)/1.1% (SE 4.3%), 435.9 (SE 574.5)/2.2% (SE 4.6%), and −106.1 (SE 979.9)/−2.6% (SE 8.1%) for weeks 2, 3, and 4, respectively. For social interaction they were 3841.9 (SE 1425.4)/21.7% (SE 5.1%), 2270.6 (SE 947.1)/16.5% (SE 4.4%), and 2443.4 (SE 982.6)/17.9% (SE 4.7%) for weeks 2, 3, and 4, respectively. Analysis of group effect was significant (absolute change: η^2=.19, *P*=.01; proportional change: η^2=.27, *P*=.009). Correlation between the proportional change and the play activity was significant (*r*=.34, 95% CI 0.08 to 0.56), whereas for absolute change it was not.

Conclusions: Social interaction design elements of the pervasive game may have some positive effects on the promotion of physical activity, although other factors might also have influenced this effect.

Trial Registration: Japan Medical Association Clinical Trial Registration Number JMA-IIA00314; https://dbcentre3.jmacct.med.or.jp/JMACTR/App/JMACTRS06/JMACTRS06.aspx?seqno=7274 (Archived by WebCite at http://www.webcitation.org/761a6MVAy)

(JMIR Serious Games 2019;7(3):e13962) doi:10.2196/13962

KEYWORDS

aged; physical activity; pervasive games; social interaction

Introduction

As the proportion of elderly adults increases in populations worldwide, supporting their quality of life has become a pressing social challenge [1]. Many studies have pursued this goal by using electronic games as a novel strategy to address specific issues, such as the rehabilitation of psychomotor functions [2,3], prevention of age-related diseases [4,5], or promotion of active lifestyles [6], all with varied levels of success [7].

Recently, researchers began to explore the new genre of pervasive games [8-12] in this context. A pervasive game is an electronic game that incorporates elements from the real world in its mechanics, blurring the edges of the so-called "magic circle" [13] (ie, the perceived boundaries of the playing space). Among the real-world elements used in such games, two of the most common are physical location and social connections, and both can be beneficial to senior players. When people are invited to walk around and visit places in the real world, they are stimulated to have regular physical activity; when they interact with other people via the social mechanics of the game, it may be possible to reduce or prevent social isolation. Both effects are strongly correlated with higher quality of life among elderly adults and a lower incidence of age-related diseases [14-18].

Few studies have used pervasive games or gamified apps targeting older adults; they usually focus on specific goals, such as cognitive training [19] or the promotion of physical activity using social incentives [20,21]. A successful commercial example that does not target elderly adults specifically that became extremely popular among people of all ages is Pokémon GO [22]. Different studies have analyzed its effects on levels of physical activity and found overall positive results, especially in the first weeks of use [23-25].

However, to effectively use pervasive games to help older players, it is necessary to better understand how different elements of game design can affect their experience. In this work, we investigate whether using social interaction elements in a pervasive game can increase players' levels of physical activity. Because the main element of mechanics of the game is walking, we assume that if higher levels of physical activity are observed in association with more frequent play activity that also implies that the game experience was more fun and engaging.

Methods

Design

The main focus was on evaluating the effects of change in the game design; therefore, we compared two versions of the same pervasive game. The only design element that was changed was social interaction. It was not possible to blindly assign individual participants to different groups because players would be aware of the different interaction options in the game if they interacted with other players. Thus, we compared the versions of the game in two isolated groups in a quasi-experimental study design.

Participants

Participants were recruited in collaboration with the University of Brasília in Brazil among students who attended classes in a university-run community project that targeted older adults. Students incurred no financial cost to join the project, and no educational background was required, except for being able to read and write. Classes were offered at sites in different regions of the city (more than 30 km apart). Students at different sites did not have contact with one another within the project, but students at the same site attended classes together. Two different sites were chosen to recruit students and form the intervention groups: one group played the version of the game without social interaction and the other group played the version with social interaction. There were no identifiable differences between sites regarding participants' social, educational, or economic backgrounds.

The inclusion criteria adopted a broader age range of 50 years or older, aiming for middle-aged and older adults, because this research was contextualized as a preventive health intervention and it is expected that experience with games will become increasingly common among older adults in the future. Additional criteria included healthy people with independent ambulation and no cognitive or physical impairment preventing them from understanding the instructions of the game or taking short walks. All participants signed informed consent forms, and the research protocol was approved by the Ethical Committee of the University of Brasília and by the Kyoto University Hospital's Ethical Committee; both boards report compliance with the Declaration of Helsinki.

Game

Participants played a pervasive location-based mobile game called *Trilhas* [26,27]. This game has been previously evaluated for its feasibility and adaptability to allow for the testing of different design elements [28].

The basic (no social interaction) version of the game (Figure 1) invited players to visit different real-world locations and collect virtual cards. The home screen of the game showed a map that indicated the player's current location and nearby hotspots (ie, places they can visit). Hotspots were defined using information from Google Maps and included publicly accessible places, such as drugstores, bakeries, coffee shops, churches, and government buildings. Hotspots were spread in a fixed area of the city that included both experiment sites and a range of 20 km, with the distance between hotspots a minimum of 100 m and a maximum of 500 m. If a certain region did not have enough known locations to ensure this distribution, abstract locations were assigned to random physical positions in publicly accessible areas.

After visualizing nearby hotspots, players had to walk toward them. When they were within 50 m of their locations, an “Enter” option appeared on the screen, and players could register their visit. For safety reasons, players were not expected to keep the game screen open while walking, and they were instructed to access the game only when they arrived at their destination. If the game was open and the player’s speed exceeded a certain threshold, the game warned the player not to walk while looking at their mobile phone.

Each day, players received “cards” for the first level proportional to how much they walked (measured by the number of steps) and how many places they visited in previous days. Later, they could trade a certain number of cards from one level for one card of the next level. The goal was to obtain a card of the maximum level for every animal in the game.

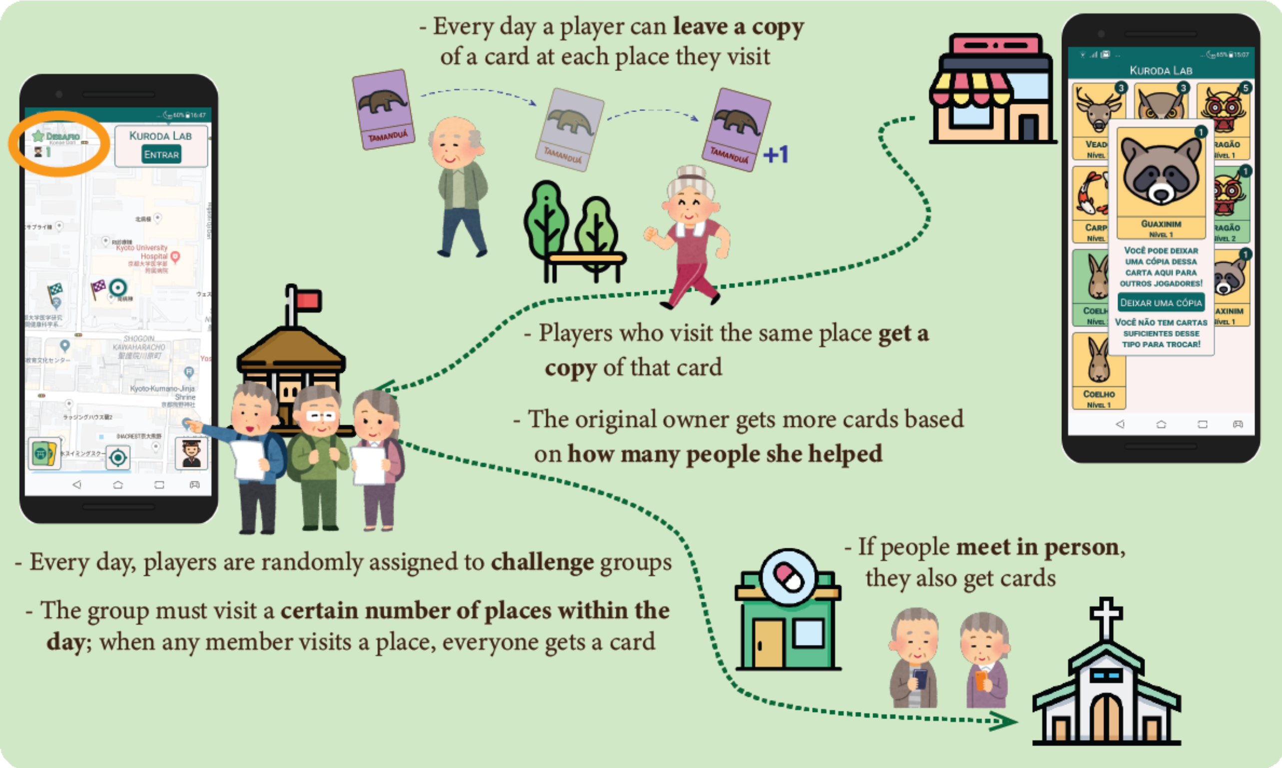
In the social interaction version of the game (Figure 2), the following social elements were added to stimulate players to interact and collaborate to obtain more cards:

1. Players could leave copies of their cards on the places they visited, and when other players passed by, they received a copy and the original owner received more cards. When a certain hotspot had a card left on it, its icon exhibited an exclamation mark.
2. Every day, players were randomly assigned to a challenge group, and when a person in the group collected a card, all other members also received a card. Members who contributed to the challenge on that day were shown, whereas players who could not contribute were not shown to avoid “social shaming” (ie, negative reinforcement from other players).
3. When players met in person and scanned each other’s phones, they also received cards.
4. All players could choose a public avatar and nickname and make a short self-introduction. When players received cards from other players’ actions, they had a chance to give them a “like.”

The feasibility study and follow-up evaluations [27,28] suggested that these mechanics allowed players to feel more engaged in playing the game by working together with other people. We hypothesized that this setup would result in a higher positive effect on levels of physical activity.

Figure 1. Version of the game without social interaction.

Figure 2. Social interaction version of the game.



Outcome Measures

The main observed outcome was the level of physical activity, measured by the mean number of steps per week over a 4-week period. The step count was measured by the game using a background service that operated whenever the mobile phone was turned on. The software used was Google Android's Sensor API, which is the same as Google Fit, an app previously shown to have accuracy equivalent to or better than that of wearable devices [29]. During the first week, participants did not play the game, but their step count was still monitored. This monitoring was performed to assess their baseline level of physical activity. After that, they played the game for an additional 3 weeks. To evaluate how much the participants played the game, the weekly mean number of visits to hotspots was also observed. Within a single day, this observation represented the number of unique hotspots visited by the player, whereas within a week, it was the sum of visits each day of the week (ie, the same hotspot was not counted twice for the same day, but it could be counted twice for a week). This measurement was used because players were directed to not keep the game open while walking, so play time was not a good measurement of how much a person played.

As a secondary evaluation, participants were also asked to answer two questionnaires: one assessing their previous experience with games and technology and another evaluating their experience using the game. This second questionnaire was based on the Game Experience Questionnaire [30] and the System Usability Scale [31]. Although these questionnaires are widely used in previous work, they have not been statistically validated yet. For that reason, they served only as complementary information, and the results are reported here for completion. Because the Game Experience Questionnaire was designed to evaluate a broad range of games, including nonelectronic ones, we included only questions related to the gameplay elements present in *Trilhas*. Items used a 5-level Likert scale, indicating the mean agreement level (0=not at all, 4=extremely). They were grouped into categories, with the mean value calculated for each category, as follows:

- Usability:
  - Controls, with items such as "I found the game too complicated to use"
  - Learning curve, with items such as "I could learn how to use the game quickly"
  - Game rules, with items such as "I could understand the game rules"
- Game experience:
  - Theme and visual style, with items such as "I found the game esthetically pleasing"
  - Feeling of immersion, with items such as "I forgot everything about me"
  - Feeling of enjoyment, with items such as "I found the game fun"
  - Feeling of engagement, with items such as "I felt stimulated"
  - Feeling of freedom or ability to explore, with items such as "I felt that I could explore things"
  - Feeling of positive challenge, with items such as "I felt challenged to reach the game's goals"

The last question of the questionnaire asked participants to freely write comments, criticism, or suggestions. All questionnaires were anonymous.

Procedures

The sites for social interaction and no social interaction were chosen at random; participants were blinded to group assignment. At the beginning of the study, participants at each site signed the informed consent form and answered the first questionnaire (previous experience with games and technology). Their mobile phone was checked for compatibility, and the game was then installed. Compatible systems included Android-based mobile phones with OS version 5.0 or above with a GPS (Global Positioning System) sensor and an internet connection. Participants who did not have an Android mobile phone or who could not or did not want to use their personal devices were lent a previously prepared one by the researchers.

Participants were told to keep the mobile phone turned on and carry it with them whenever they left their homes throughout the study. There was a follow-up meeting on the same weekday every week, in which researchers were available to clarify any questions or solve technical problems. On the last meeting, after 4 weeks, participants answered the final questionnaire to evaluate their experience while playing.

All questionnaires were administered by researchers, who were available to clarify possible questions about the items.

Data Analysis

Questionnaire data were consolidated to report percentages in each item, whereas means and standard errors were calculated for demographic data using Google Sheets.

Dropout and step count data were preprocessed using Python (mainly the *pandas* and *matplotlib* packages) to generate graphs and format the data into a suitable format for R. To analyze the effect, we used the change on the number of steps for each week, when compared to the baseline week. This measurement was made for each participant in relation to their own individual baseline, and the proportional change was also calculated (ie, the absolute change divided by the baseline value).

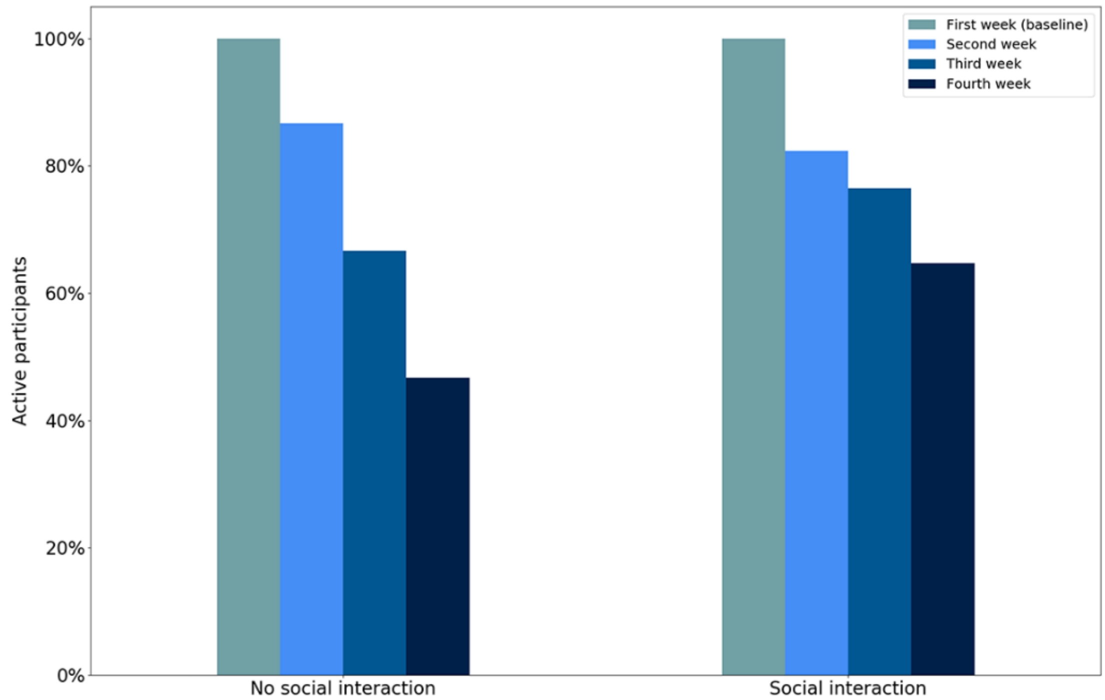
In the statistical model, the change for each week after baseline was considered to be a repeated measure, and an analysis of variance (ANOVA) was performed, with group and week as factors, for participants who remained until the end of the study. Because the experimental setting did not allow for the balancing of data, a type III strategy was used to account for unbalanced data. The relationship between the change in step count and the number of hotspot visits was evaluated using a Pearson correlation coefficient (*r*). This analysis was performed using R.

Results

Participants

The graph in Figure 3 shows the proportion of remaining participants over time. People could drop out at the weekly follow-up meetings, so the data reported here considers participants who stayed for at least the baseline week; otherwise, it was as if they did not join the experiment. The initial number of participants in the no social interaction group was 15 (females: n=11) and 17 (females: n=14) in the social interaction group. At the end of the fourth week, there were 7 (females: n=5) people in the no social interaction group and 11 (females: n=8) people in the social interaction group, indicating dropout rates of 53% and 35%, respectively. Most participants used their own devices. For the no social interaction group, three devices were lent, whereas two were lent for the social interaction group.

Figure 3. Active participants over time.



One person borrowed a device because they did not own a mobile phone, whereas all others borrowed devices because their own device was incompatible (different OS or unsuitable OS version). Only one person who borrowed a device dropped out (from the no social interaction group), and all other participants who borrowed devices stayed until the end. There was no follow-up to verify the reasons for dropping out, but three participants gave spontaneous reports. One person from the no social interaction group said they did not find the game interesting. One person from the social interaction group said they had back pains that prevented them from walking and another one said they did not have time to play.

Information about participants' previous experience with technology and games is reported in Table 1. These data include answers collected at the beginning of the study from all participants.

Main Outcome

Step count data are shown in Table 2 for participants who remained until the end of the experiment. With the absolute change, the analysis of the effect of group as a factor resulted in P=.01 (η^2=.19). No relevant relationship was found with week as a factor (P=.65). For proportional change, taking group as a factor resulted in P=.009 (η^2=.27), whereas taking week as a factor resulted in P=.54.

For hotspot visits, the group without social interaction had mean 8.4 (SE 2.1) visits in week 2, mean 8.9 (SE 1.7) in week 3, and mean 5.1 (SE 3.0) in week 4. In comparison, the social interaction group had mean 14.2 (SE 1.9) visits in week 2, mean 9.5 (SE 2.0) in week 3, and mean 12.8 (SE 3.4) in week 4.

The correlation analysis between the absolute change in the number of steps and the number of visits resulted in a correlation factor of r=.21 (95% CI −.06 to .45). When proportional change was considered, the correlation factor was r=.34 (95% CI .08 to .56).

Game Experience

The scores for the usability and game experience questionnaires are summarized in Table 3. The score of component items could go from 0 to 4; therefore, a value of 2 or greater indicates a positive evaluation. The data included are of only those participants who stayed until the final week.

The questionnaire also included an open-response item in which participants could freely make suggestions and comments.

For the no social interaction group, one participant reported that they often played competitive online games and that *Trilhas* could benefit from a competitive factor. Two participants said they could not play often but wished to have helped more in the research. One participant said the game was boring.

For the social interaction group, six participants said they enjoyed the chance to get more exercise. Three participants said they liked the look of the game, using adjectives such as "cute" and "pleasing." During the follow-up meetings, five participants commented on the fact they received cards from other participants at specific locations. One participant reported concern about other people knowing their whereabouts.

Table 1. Basic data for participants (N=32).

Participant data	Baseline		End of week 4	
	No social interaction	Social interaction	No social interaction	Social interaction
Demographics				
Participants	15	17	7	11
Sex (female), n (%)	11 (73)	14 (82)	5 (71)	9 (82)
Age (years), mean (SD)	64.3 (6.0)	61.1 (7.4)	63.9 (5.1)	60.1 (6.0)
Dropouts, n (%)	8 (53)	6 (35)	—[a]	—
PC usage frequency, n (%)				
Every day	6 (40)	5 (29)	3 (43)	2 (18)
≥2 times/week	4 (27)	4 (24)	1 (14)	3 (27)
≤1 time/week	4 (27)	6 (35)	2 (29)	5 (46)
Never	1 (6)	3 (12)	1 (14)	1 (9)
PC[b] experience, n (%)				
Able to check email	14 (93)	12 (71)	6 (86)	8 (73)
Able to do Web searches	14 (93)	12 (71)	6 (86)	8 (73)
Able to read news online	12 (80)	13 (76)	6 (86)	7 (64)
Able to use social networks	14 (93)	13 (76)	6 (86)	10 (91)
Able to install apps	7 (47)	4 (24)	4 (57)	3 (27)
Mobile phone experience, n (%)				
Never used before	1 (7)	1 (6)	1 (14)	1 (9)
Able to make calls	14 (93)	13 (76)	6 (86)	8 (73)
Able to check email	13 (87)	13 (76)	6 (86)	9 (82)
Able to browse the Web	14 (93)	12 (71)	6 (86)	9 (82)
Able to use social networks	13 (87)	16 (94)	6 (86)	10 (91)
Able to install apps	6 (40)	8 (47)	2 (29)	6 (55)
Electronic games play frequency, n (%)				
Every day	1 (7)	3 (18)	1 (14)	3 (27)
≥2 times/week	3 (20)	0 (0)	2 (29)	0 (0)
≤1 time/week	1 (7)	0 (0)	0 (0)	0 (0)
Very rarely	1 (7)	2 (12)	1 (14)	1 (9)
Never plays	9 (60)	12 (71)	3 (43)	7 (64)
Devices used to play,[c,d,e] n (%)				
Computer	3 (50)	2 (40)	1 (25)	2 (50)
Mobile phone	3 (50)	4 (80)	2 (50)	4 (100)
Portable console	0 (0)	2 (40)	0 (0)	2 (50)
Conventional console	0 (0)	3 (60)	0 (0)	3 (75)
Play partners,[c,d,e] n (%)				
Plays alone	4 (67)	4 (80)	2 (50)	4 (100)
Plays with friends	0 (0)	1 (20)	0 (0)	1 (25)
Plays with adult family members	1 (17)	2 (40)	1 (25)	1 (25)
Plays with young family members	1 (17)	3 (60)	0 (0)	1 (25)
Plays with strangers online	1 (17)	0 (0)	0 (0)	0 (0)

[a]Not applicable.
[b]PC: personal computer.
[c]Respondents could indicate more than one item.
[d]Some people reported playing but did not indicate any option on this item.
[e]Percentages are relative to the number of people who reported any play activity.

Table 2. Mean number of steps at baseline for each group and mean of the individual variations in subsequent weeks.

Group and value	Baseline, mean (SE)	Mean change at week 2 (SE)	Mean change at week 3 (SE)	Mean change at week 4 (SE)
No social interaction				
Absolute[a]	17,099.4 (3906.5)	383.8 (563.8)	435.9 (574.5)	−106.1 (979.9)
Proportional (%)[b]	—[c]	1.1 (4.3)	2.2 (4.6)	−2.6 (8.1)
Social interaction				
Absolute[a]	17,981.9 (2171.1)	3841.9 (1425.4)	2270.6 (947.1)	2443.4 (982.6)
Proportional (%)[b]	—	21.7 (5.1)	16.5 (4.4)	17.9 (4.7)

[a]Absolute values indicate the change in the weekly number of steps compared with the user's own baseline.
[b]Proportional values indicate the absolute value divided by the user's own baseline.
[c]Not applicable.

Table 3. Results from the usability and game experience questionnaires (N=18).

Category	No social interaction (n=7), mean score	Social interaction (n=11), mean score
Usability		
Controls	2.2	2.7
Learn curve	2.2	2.6
Game rules	1.9	2.2
Game experience		
Theme and visual style	2.3	2.5
Feeling of immersion	2.4	2.2
Feeling of enjoyment	2.4	2.9
Feeling of engagement	1.6	2.3
Feeling of freedom/ability to explore	1.5	2.9
Feeling of (positive) challenge	1.4	2.1

Discussion

Principal Results

At the beginning of the experiment, the majority of participants in both groups had experience with both personal computers and mobile phones. This was also true of those participants who remained until the end of the experiment. When previous experience with games was considered, the majority of participants in both groups reported never playing or playing very rarely. For participants who remained until the end, the ratio of people who played at least once a week increased in both groups. Additionally, participants in both the social interaction and no social interaction groups reported using personal computers and mobile phones to play, but only participants in the no social interaction group used all the devices listed as options. Most people reported playing alone, with the one remaining participant in the social interaction group reporting also playing with friends; the remaining participants in the no social interaction group reported playing with family members and friends.

For the main outcome, a larger positive effect was observed in the social interaction group compared with the no social interaction group. The statistical analysis regarding the absolute change indicated a medium-to-large effect size (η^2=.19), and the *P* value of .01 indicates a statistically significant difference. There was more variation in the main outcome for the control group, probably due to the higher dropout rate by the end of the third week in that group. A higher number of visits in this group suggests that participants played more, although correlation data were inconclusive: a medium correlation was found for proportional change, but only a small correlation was found for absolute change and that measure was not statistically significant. Therefore, social interaction mechanics may affect player engagement, but other factors may also have influence.

Because hotspots are not uniformly spaced, one other possible explanation could be that players tended to visit the same nearby places more often, or visit faraway places only a few times, thus increasing their physical activity to some extent, but not in a linear relationship with the number of visits.

The evaluation of usability and game experience was used as complementary information only. Statistical analysis was not performed because the questionnaires used are not validated. The results indicated an overall positive evaluation for system usability and game experience, as most items had values of 2 or greater. The average evaluation for the social interaction group was higher for all items except immersion, but the difference between groups was small for most items, preventing any solid conclusions. Players in both groups gave positive feedback for the "visual style" of the game and feelings of "immersion" and "enjoyment." However, there was a larger difference between groups for the categories of "engagement," "exploration," and "challenge." One possible explanation is that the no social interaction group might have had an inferior experience in these categories because the version of the game they played had a subset of the rules of the social interaction group, which could be perceived as less challenging. The usability evaluation for "game rules" was lower in the no social interaction group; however, because the rules in the no social interaction version were a subset of those in the social interaction version, this finding might also be explained by an inferior game experience. Another possible explanation is that some players might not have understood these rules, although players' comments did not indicate such a case.

In the subjective evaluation considering the open-response comments, players in the social interaction group seemed to have had more fun and felt more engaged in the game, specifically enjoying the card exchange mechanics, although they also felt motivated by the chance of being stimulated to do more exercise. One participant from the no social interaction group complained about the lack of competition, which is a modality of social interaction, leading to the belief that the social factor is relevant for some players. Because these reports were voluntary and many players did not make any comments, it is not possible to generalize these impressions.

Limitations

There were limitations to this study. The sample size was small, and although the power analysis indicated a medium-to-large effect for relative change and a statistically significant difference for absolute and relative changes, more data could potentially increase the accuracy of these results. Additionally, the dropout rate was high, which could introduce bias toward a positive effect, because the remaining participants might be those who enjoyed the game and were stimulated to continue playing and, potentially, have more physical activity.

The nature of the game and recruitment context made it impossible to use a double-blind design and individually assign participants to groups, which might introduce two biases. First, researchers were aware of group assignments; therefore, they could involuntarily influence participant's behavior or attitude toward the game during the follow-up meetings. Secondly, even though participants were blind to group assignment, participants in each group had classes together, which might introduce a cohort effect (ie, participants who knew each other and might have a higher tendency to interact using the game and stimulate each other to play). It was also not possible to control for previous experience with technology and games or other possible socioeconomic differences that might have affected the results, although questionnaire data suggest that remaining participants in the no social interaction group had a higher ratio of proficiency to technology, which could have made participants in that group more prone to using the game, in opposition to the observed effect.

The main outcome was measured using mobile phone software. The methodology has been evaluated in previous studies, and the authors of those studies concluded that it is adequate; however, future interventions might test similar settings with a different device, such as external pedometers, and compare the results. In both cases, because the data are not collected in a controlled environment but rather in a user-dependent context, and participant's adherence to carrying the mobile phone with them was not measured, thus measurements for noncompliant participants are not accurate.

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

Although this study is inserted in the more general field of interventions to improve the quality of life of older adults, it focuses specifically on increasing physical activity based on previous results that showed a strong correlation between these variables. Future interventions could directly focus on these two variables and evaluate their relationship in the context of pervasive games. Also, the questionnaires used to assess usability and game experience were not statistically validated. They could not be used to draw conclusions about the effect and were only complementary information. Using validated metrics would allow for increased data comprehension and the ability to test more hypotheses.

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

Comparison With Prior Work

Although games have been used many times previously to promote the well-being of elderly people, there is usually an excessive focus on health benefits, with little attention given to aspects such as motivations to play and overall game experience. Recently, new research has emerged [32,33] that analyzes in greater depth the experience of elderly people in play based on the principle that games, even serious games, should first be fun because the health benefits come later as a natural consequence [34]. This notion aligns with the idea that a deep and meaningful connection with play and fun is an inherent part of human nature [35], and elderly people are no exception. In that respect, few studies have attempted to clarify elderly players' needs and motivations and investigated possible challenges in designing for older audiences, listing common physical and cognitive limitations that should be taken into consideration [36-40]. Other studies have attempted to identify the preferences of elderly people regarding the content or genre of the games [41-44]. In this study, we evaluated social interaction as a design element in the context of pervasive games, which is a new kind of game that is only now being explored. This study was limited and focused on a specific metric, namely physical activity, which was used as a proxy, but the results suggest that this topic should be further investigated, with the consideration of additional variables related to game experience.

Regarding interventions that promoted walking in general [45], the most effective studies that were analyzed achieved a net increase of 30 to 60 minutes of walking per week. Considering the conversion criteria used in that systematic review, this is equivalent to 3000 to 6000 steps. This study achieved those numbers for the social interaction group in the first week, but those results were not sustained over time. More investigation is necessary to explore how player engagement could be maintained for longer periods.

Conclusions

In this work, we investigated whether the new genre of pervasive games could be used to increase physical activity of older adults. Our results indicated that a pervasive game using social interaction had a greater positive effect on levels of physical activity than the same game without social interaction. This study was limited; these results are promising but not conclusive. In future interventions, other types of social interaction or design elements should be evaluated, and additional variables considered, such as indicators of physical and psychological health among others.

Acknowledgments

This study was partially supported by the Hayao Nakayama Foundation for Science & Technology and Culture and by Kyoto University's Collaborative Graduate Program in Design. The authors thank the participants for volunteering and the students and staff for helping conduct the experiment.

Conflicts of Interest

None declared.

References

1. United Nations, Department of Economic and Social Affairs, Population Division. World Population Ageing. New York: United Nations; 2017.
2. Breton Z, Zapirain B, Zorrilla A. KiMentia: kinect based tool to help cognitive stimulation for individuals with dementia. In: 14th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom). 2012 Presented at: 14th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom); Oct 15-18, 2014; Natal, Brazil p. 325-328. [doi: 10.1109/HealthCom.2012.6379430]
3. Smith S, Talaei-Khoei A. Electronic games for aged care and rehabilitation. In: 11th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom). 2009 Presented at: 11th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom); Dec 16, 2009; Sydney, Australia p. 42-47. [doi: 10.1109/HEALTH.2009.5406197]
4. Marin J, Navarro K, Lawrence E. Serious games to improve the physical health of the elderly: a categorization scheme. In: 4th International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services. 2011 Presented at: 4th International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services; Oct 23-29, 2011; Barcelona p. 64-71.
5. Kayama H, Okamoto K, Nishiguchi S, Yamada M, Kuroda T, Aoyama T. Effect of a Kinect-based exercise game on improving executive cognitive performance in community-dwelling elderly: case control study. J Med Internet Res 2014 Feb 24;16(2):e61 [FREE Full text] [doi: 10.2196/jmir.3108] [Medline: 24565806]
6. Bird M, Clark B, Millar J, Whetton S, Smith S. Exposure to "exergames" increases older adults' perception of the usefulness of technology for improving health and physical activity: a pilot study. JMIR Serious Games 2015 Nov 27;3(2):e8 [FREE Full text] [doi: 10.2196/games.4275] [Medline: 26614263]
7. 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]
8. Kasapakis V, Gavalas D. Pervasive gaming: status, trends and design principles. J Netw Comput Appl 2015 Sep;55:213-236. [doi: 10.1016/j.jnca.2015.05.009]

9. Björk S, Holopainen J, Ljungstrand P, Åkesson K. Designing ubiquitous computing games-a report from a workshop exploring ubiquitous computing entertainment. In: Personal and Ubiquitous Computing. 2002 Presented at: Personal and Ubiquitous Computing; Sep 29, 2002; Göteborg, Sweden p. 443-458. [doi: 10.1007/s007790200048]
10. Bonsignore E, Hansen D, Toups Z, Nacke L, Salter A, Lutters W. Mixed reality games. In: ACM 2012 conference on Computer Supported Cooperative Work Companion-CSCW '12. 2012 Presented at: ACM 2012 conference on Computer Supported Cooperative Work Companion - CSCW '12; Feb 11-15, 2012; Seattle, WA p. 7-8. [doi: 10.1145/2141512.2141517]
11. Gutierrez L, Stroulia E, Nikolaidis I. fAARS: A platform for location-aware trans-reality games. In: 11th International Conference on Entertainment Computing. 2012 Presented at: 11th International Conference on Entertainment Computing; Sep 26-29, 2012; Bremen, Germany p. 185-192. [doi: 10.1007/978-3-642-33542-6_16]
12. Buzeto F, Castillo A, Castanho C, Jacobi R. What is going on with ubicomp games. In: XI Brazilian Symposium on Games and Digital Entertainment. 2012 Presented at: XI Brazilian Symposium on Games and Digital Entertainment; Nov 2-4, 2012; Brasília, Brazil p. 1-7.
13. Montola M. Exploring the edge of the magic circle: defining pervasive games. In: Digital Arts and Culture Conference. 2005 Presented at: Digital Arts and Culture Conference; Dec 1, 2005; Copenhagen p. 16-19.
14. Colcombe S, Kramer AF. Fitness effects on the cognitive function of older adults: a meta-analytic study. Psychol Sci 2003 Mar;14(2):125-130. [doi: 10.1111/1467-9280.t01-1-01430] [Medline: 12661673]
15. Larsen LH, Schou L, Lund HH, Langberg H. The physical effect of exergames in healthy elderly-a systematic review. Games Health J 2013 Aug;2(4):205-212. [doi: 10.1089/g4h.2013.0036] [Medline: 26192224]
16. Vagetti GC, Barbosa FV, Moreira NB, Oliveira VD, Mazzardo O, Campos WD. Association between physical activity and quality of life in the elderly: a systematic review, 2000-2012. Rev Bras Psiquiatr 2014 Mar;36(1):76-88 [FREE Full text] [doi: 10.1590/1516-4446-2012-0895] [Medline: 24554274]
17. Tomaka J, Thompson S, Palacios R. The relation of social isolation, loneliness, and social support to disease outcomes among the elderly. J Aging Health 2006 Jun;18(3):359-384. [doi: 10.1177/0898264305280993] [Medline: 16648391]
18. Lee G, Ishii-Kuntz M. Social interaction, loneliness, and emotional well-being among the elderly. Res Aging 1987 Dec;9(4):459-482. [doi: 10.1177/0164027587094001] [Medline: 3438563]
19. Gamberini L, Martino F, Seraglia B, Spagnolli A, Fabregat M, Ibanez F, et al. Eldergames project: an innovative mixed reality table-top solution to preserve cognitive functions in elderly people. In: 2nd Conference on Human System Interactions. 2009 Presented at: 2nd Conference on Human System Interactions; May 21-23, 2009; Catania, Italy p. 164-169. [doi: 10.1109/HSI.2009.5090973]
20. Brox E, Fernandez-Luque L, Evertsen G, González-Hernández J. Exergames for elderly: social exergames to persuade seniors to increase physical activity. In: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare. 2011 Presented at: 5th International ICST Conference on Pervasive Computing Technologies for Healthcare; May 23-26, 2011; Dublin, Ireland p. 546-549. [doi: 10.4108/icst.pervasivehealth.2011.246049]
21. Paul L, Brewster S, Wyke S, McFadyen AK, Sattar N, Gill JM, et al. Increasing physical activity in older adults using STARFISH, an interactive smartphone application (app); a pilot study. J Rehabil Assist Technol Eng 2017 Apr 05;4:205566831769623. [doi: 10.1177/2055668317696236]
22. Niantic L. Google Play. 2016. Pokemon GO URL: https://play.google.com/store/apps/details?id=com.nianticlabs.pokemongo [accessed 2017-11-15] [WebCite Cache ID 6uywbsV3m]
23. Hino K, Asami Y, Lee J. Step counts of middle-aged and elderly adults for 10 months before and after the release of Pokémon GO in Yokohama, Japan. J Med Internet Res 2019 Feb 05;21(2):e10724 [FREE Full text] [doi: 10.2196/10724] [Medline: 30720438]
24. Marquet O, Alberico C, Adlakha D, Hipp JA. Examining motivations to play Pokémon GO and their influence on perceived outcomes and physical activity. JMIR Serious Games 2017 Oct 24;5(4):e21 [FREE Full text] [doi: 10.2196/games.8048] [Medline: 29066423]
25. Althoff T, White RW, Horvitz E. Influence of Pokémon Go on physical activity: study and implications. J Med Internet Res 2016 Dec 06;18(12):e315 [FREE Full text] [doi: 10.2196/jmir.6759] [Medline: 27923778]
26. Santos L, Okamoto K, Yamamoto G, Sugiyama O, Aoyama T, Kuroda T. A pervasive game to evaluate the effectiveness of social interaction to increase daily activity of older adults. In: Entertainment Computing Symposium 2017. 2017 Presented at: Entertainment Computing Symposium 2017; Sep 16-18, 2017; Sendai, Japan p. 232-236.
27. Santos L, Okamoto K, Cavalli A, Hiragi S, Yamamoto G, Sugiyama O, et al. Designing pervasive social interaction mechanics for elderly players: a multicultural study case. In: Innovation in Medicine and Healthcare Systems, and Multimedia. 2019 Presented at: KES InMed 2019; June 17-19, 2019; St Julians, Malta p. 293-303. [doi: 10.1007/978-981-13-8566-7_29]
28. Santos L, Okamoto K, Hiragi S, Yamamoto G, Sugiyama O, Aoyama T, et al. Pervasive game design to evaluate social interaction effects on levels of physical activity among older adults. J Rehabil Assist Technol Eng 2019;6:2055668319844443 [FREE Full text] [doi: 10.1177/2055668319844443] [Medline: 31285836]
29. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. JAMA 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
30. Poels K, de Kort Y, Ijsselsteijn W. D3.3. Game Experience Questionnaire: Development of a Self-Report Measure to Assess the Psychological Impact of Digital Games. Eindhoven, Netherlands: Technische Universiteit Eindhoven; 2007.

31. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation in Industry. Boca Raton, Florida: CRC Press; 1996:4-7.
32. De Shutter B, Malliet S. The older player of digital games: a classification based on perceived need satisfaction. Communications 2014;39(1):67-88. [doi: 10.1515/commun-2014-0005]
33. Gerling K, De Schutter B, Brown J, Allaire J. Ageing playfully: advancing research on games for older adults beyond accessibility and health benefits. In: Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15. 2015 Presented at: Annual Symposium on Computer-Human Interaction in Play - CHI PLAY '15; Oct 5-7, 2015; London, UK p. 817-820. [doi: 10.1145/2793107.2810261]
34. De Schutter B, Vanden Abeele V. Towards a gerontoludic manifesto. Anthropol Aging 2015 Nov 19;36(2):112-120. [doi: 10.5195/AA.2015.104]
35. Huizinga J. Homo Ludens: A Study of the Play-element in Culture. Boston, MA: Beacon Press; 1971.
36. Ijsselsteijn W, Nap H, de Kort Y, Poels K. Digital game design for elderly users. In: Future Play '07. 2007 Presented at: Future Play '07; Nov 14-17, 2007; Toronto, ON p. 17-22. [doi: 10.1145/1328202.1328206]
37. Ogomori K, Nagamachi M, Ishihara K, Ishihara S, Kohchi M. Requirements for a cognitive training game for elderly or disabled people. In: International Conference on Biometrics and Kansei Engineering. 2011 Presented at: International Conference on Biometrics and Kansei Engineering; Sep 19-22, 2011; Takamatsu, Japan p. 150-154. [doi: 10.1109/ICBAKE.2011.30]
38. Barnard Y, Bradley MD, Hodgson F, Lloyd AD. Learning to use new technologies by older adults: perceived difficulties, experimentation behaviour and usability. Comput Hum Behav 2013 Jul;29(4):1715-1724. [doi: 10.1016/j.chb.2013.02.006]
39. Brox E, Konstantinidis S, Evertsen G. User-centered design of serious games for older adults following 3 years of experience with exergames for seniors: a study design. JMIR Serious Games 2017 Jan 11;5(1):e2 [FREE Full text] [doi: 10.2196/games.6254] [Medline: 28077348]
40. Meekes W, Stanmore EK. Motivational determinants of exergame participation for older people in assisted living facilities: mixed-methods study. J Med Internet Res 2017 Jul 06;19(7):e238 [FREE Full text] [doi: 10.2196/jmir.6841] [Medline: 28684385]
41. Carvalho RN, Ishitani L. Motivational factors for mobile serious games for elderly users. In: XI Brazilian Symposium on Games and Digital Entertainment. 2012 Presented at: XI Brazilian Symposium on Games and Digital Entertainment; Nov 2-4, 2012; Brasília, Brazil p. 19.
42. Cota TT, Ishitani L, Vieira N. Mobile game design for the elderly: a study with focus on the motivation to play. Comput Hum Behav 2015 Oct;51:96-105. [doi: 10.1016/j.chb.2015.04.026]
43. Kaufman D, Sauvé L, Renaud L, Sixsmith A, Mortenson B. Older adults' digital gameplay. Simulat Gaming 2016 Apr 21;47(4):465-489. [doi: 10.1177/1046878116645736]
44. Chesham A, Wyss P, Müri RM, Mosimann U, Nef T. What older people like to play: genre preferences and acceptance of casual games. JMIR Serious Games 2017 Apr 18;5(2):e8 [FREE Full text] [doi: 10.2196/games.7025] [Medline: 28420601]
45. Ogilvie D, Foster CE, Rothnie H, Cavill N, Hamilton V, Fitzsimons CF, Scottish Physical Activity Research Collaboration. Interventions to promote walking: systematic review. BMJ 2007 Jun 09;334(7605):1204 [FREE Full text] [doi: 10.1136/bmj.39198.722720.BE] [Medline: 17540909]

Abbreviations

ANOVA: analysis of variance
GPS: Global Positioning System

Edited by G Eysenbach; submitted 08.03.19; peer-reviewed by E Giesbrecht, M Rosly, E Dove, YY Chao; comments to author 03.04.19; revised version received 29.05.19; accepted 07.07.19; published 22.07.19.

Please cite as:
Santos LHDO, Okamoto K, Funghetto SS, Cavalli AS, Hiragi S, Yamamoto G, Sugiyama O, Castanho CD, Aoyama T, Kuroda T
Effects of Social Interaction Mechanics in Pervasive Games on the Physical Activity Levels of Older Adults: Quasi-Experimental Study
JMIR Serious Games 2019;7(3):e13962
URL: http://games.jmir.org/2019/3/e13962/
doi:10.2196/13962
PMID:31333202

Original Paper

Factors Associated with Sustained Exergaming: Longitudinal Investigation

Erin Kathleen O'Loughlin[1,2], MA; Tracie A Barnett[1,3], PhD; Jennifer J McGrath[1], PhD; Mia Consalvo[1], PhD; Lisa Kakinami[1], PhD

[1]Concordia University, Montreal, QC, Canada

[2]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada

[3]Le Centre L'Institut National de la Recherche Scientifique–Institut Armand-Frappier, Laval, QC, Canada

Corresponding Author:
Erin Kathleen O'Loughlin, MA
Concordia University
1455 De Maisonneuve Blvd
Montreal, QC, H3G 1M8
Canada
Phone: 1 514 890 8000
Email: erin_oloughlin@hotmail.com

Abstract

Background: Exergaming is technology-driven physical activity (PA) which, unlike traditional video game play, requires that participants be physically active to play the game. Exergaming may have potential to increase PA and decrease sedentary behavior in youth, but little is known about sustained exergaming.

Objective: The objectives of this study were to describe the frequency, correlates, and predictors of sustained exergaming.

Methods: Data were available in AdoQuest (2005-11), a longitudinal investigation of 1843 grade 5 students in Montréal, Canada. This analysis used data from grade 9 (2008-09) and 11 (2010-11). Participants at Time 1 (T1; mean age 14 years, SD 0.8) who reported past-week exergaming (n=186, 19.1% of AdoQuest sample) completed mailed self-report questionnaires at Time 2 (T2; mean age 16 years, SD 0.8). Independent sociodemographic, psychological, and behavioral correlates (from T2)/predictors (from T1 or earlier) were identified using multivariable logistic regression.

Results: Of 186 exergamers at T1, 81 (44%) reported exergaming at T2. Being female and having higher introjected regulation (ie, a type of PA motivation indicative of internalizing PA as a behavior) were independent correlates. None of the predictors investigated were associated with sustained exergaming.

Conclusions: Almost half of grade 9 exergamers sustained exergaming for 2 years. Exergaming may be a viable approach to help adolescents engage in and sustain PA during adolescence. Sex and PA motivation may be important in the sustainability of exergaming.

(JMIR Serious Games 2019;7(3):e13335) doi:10.2196/13335

KEYWORDS

video games; physical activity; adolescents

Introduction

Background

Only 8% of Canadian youth meet current physical activity (PA) recommendations [1] and they spend up to 9 hours sitting daily, often in screen-time pursuits [1,2]. Youth who do not meet PA recommendations are unlikely to benefit from the positive effects of PA on health [3,4], and sedentary youth may experience a wide range of negative physical and mental health issues [5]. Effective programs and public health policy are needed to address these issues because co-occurrence of low PA and high sedentary time contributes to childhood obesity and its numerous, deleterious health sequelae [6-9]. Despite possible negative impacts on health, screens are an indisputable reality in today's increasingly high-tech world. Underscoring its popularity, 83% of US youth aged 8 to 18 years have a traditional video game console at home, and many of these consoles having add-on capacity for exergaming (ie, video games that are also a form of exercise) equipment. In addition, approximately 50% of adolescents own smartphones, which can download apps for mobile exergaming [10]. Many smartphone apps are free or low-cost, making exergaming and augmented reality (ie, a real-world view of the environment which is “augmented” by computer-generated input) highly accessible to youth.

In 2010, nearly 40% of US students in grades 9 to 12 exergamed at least 1 day per week [10], 25% of Canadian youth aged 15 to 18 years reported exergaming in the past week [11], and up to 20% of young adults exergamed 1 to 3 times per month or more [12]. In addition, in 2014, a representative sample of 20,122 Canadian youth in grades 6 to 10 reported exergaming for an average of 30 min per day [13]. The Xbox 360’s Kinect unit (ie, a motion-sensing device used to capture bodily movement and promote PA) sold approximately 25 million units in 2013 [14], and Pokémon Go, released on July 6, 2016 by Nintendo, had 25 million active players daily after only 1 month of the official release [15]. Leveraging benefit from screen fervor by incorporating PA into video gaming and designing exergames [16-19] that are easily available and appealing holds promise in terms of increasing PA and reducing sedentary time [16-19]. Exergaming may be a healthier pastime than traditional video games because they elicit more energy expenditure (EE) [20-22], and they may improve physical fitness, body composition, and cognitive health [20,23-25]. The effect of exergaming on reducing sedentary behavior has also been studied, but to a much lesser extent [26]. Overall, the data suggest that exergaming may be a viable method to increase PA, change body composition, and improve mental health, and it represents an improvement in EE over sedentary behavior [16,18,19,27].

There are important methodological challenges in the current literature evaluating exergaming including that many intervention studies report high attrition. Short intervention periods (ie, usually 6 or 12 weeks with no follow-up post intervention) may hinder detection of longer-term effects [24,28,29], and most studies are performed under controlled conditions that may not apply to “real-life” settings. For example, a small pilot study that provided African-American and Hispanic children with access to exergaming at home and during an after-school program indicated that fitness improved after 12 weeks [30]. However, 42% of participants lost interest in exergaming within 3 months, indicative that use related partly to the novelty of the consoles. This suggests that improving understanding of PA motivation pertaining to exergaming and sustained exergaming is needed.

Objectives

To inform the development of exergaming interventions that have the potential to be sustained, evidence is needed on the determinants of sustained exergaming in naturalistic settings. Although the study by O’Loughlin et al [12]. reported that exergamers were more likely to be girls, to play nonactive video games, to watch TV 2 hours per day or more, to report weight-related stress, and to be nonsmokers, no reports to date describe factors associated with sustained exergaming in “real-world” contexts. The objectives of this study were (1) to describe the sustainability of exergaming over 2 years in a population-based sample of adolescents and (2) to identify factors associated with sustained exergaming.

Methods

Population Sample

Data were drawn from AdoQuest I (2005-11), a 6-year longitudinal study of grade 5 students (n=1843 aged 10-11 years at baseline), which investigated the natural course of the co-occurrence of health-compromising behaviors such as smoking, PA, sedentary behavior, and substance use in children [11]. A random sample of French-language elementary schools with more than 90 students in grade 5 was recruited in the greater Montreal area. To balance representation of students in high, middle, and low socioeconomic status, all schools were first stratified into groupings defined by tertile cut-offs of a school deprivation indicator based on maternal education, parental employment, and a measure of low family income that accounts for family size and area of residence [30]. An equal number of schools was then selected into each grouping, and 10 schools in the first, 10 in the second, and 9 in the third grouping (72.5% (29/40) of schools invited) agreed to participate. All students in all grade 5 classes in the 29 participating schools were eligible for recruitment. Of 2946 grade 5 students, 61.13% (1801/2946) participated at baseline (42 students joined after baseline data were collected). No data were collected from students who did not participate. Characteristics of the AdoQuest sample were comparable to those of 2 provincially representative samples of similarly aged Québec youth [31]. Parents completed mailed self-report questionnaires in 2006-07 and again in 2008-09. Participants provided assent, and their parents or guardians provided informed consent. The study received approval from the ethics and protection review boards of Concordia University and the Centre de Recherche du Centre Hospitalier de l’Université de Montréal.

Sustainability of exergaming over 2 years was investigated using data collected in the fall and winter of 2008-09 (Time 1 [T1]) when most participants were in grades 8 and 9 (aged 12-14 years) and 2 years later in the fall and winter of 2010-11 (Time 2 [T2]) when most participants were in grades 10 and 11 (aged 14-19 years). Data on exergaming and potential correlates and predictors were collected in mailed self-report questionnaires at both T1 and T2. Most questions used in AdoQuest were drawn from ongoing surveys and studies of youth including the Canadian Youth Smoking Survey [32] and the Nicotine Dependence in Teens study [33].

Study Variables

Sustained Exergaming

At T1, participants were asked: “How many hours a day do you play active video games?” Response choices included 0<1, ≥1-2, ≥2-<3, ≥3-4, or ≥5 hours per day. Participants were categorized as exergamers if they responded ≥1 hour per day. At T2, participants were asked: “Do you play active video games (eg, Wii Fit, Dance Dance, and Revolution)?” (yes, no). Participants were categorized into 1 of 2 groups based on their responses to these 2 questions. “Sustained exergamers” included participants who reported exergaming at T1 and T2; those who "stopped exergaming" reported exergaming at T1 but not T2. "Never exergamers" and those who began exergaming between T1 and T2 were not included in the analysis.

Frequency, Timing, and Intensity of Exergaming

At T2, questions on exergaming were included based on the International Physical Activity Questionnaire (IPAQ), a short, self-administered questionnaire used in cross-national monitoring of usual weekly PA in youth and adults. The IPAQ demonstrates reliability as well as validity against accelerometer [34]. Specifically, exergamers were asked the following questions:

1. How many days a week do you play active video games? *(1-7 days)*
2. How many minutes (on average) do you play each time? *(open-ended)*
3. Physical effort during play? *(light, moderate, vigorous)*

Potential correlates and predictors of sustained exergaming were selected based on factors known to be associated with PA or exergaming in adolescence [12,35] and on the availability of data in AdoQuest.

Potential predictors measured before or at T1 included lifestyle behaviors (ie, level of PA, ever smoked cigarettes, binge drank, marijuana use, hours of TV per day, hours of computer per day, hours of nonactive video games per day, hours of sedentary behavior per day), weight-related indicators (ie, body mass index [BMI], stress about weight, perceived weight too heavy, trying to lose weight, body-related guilt, body-related shame), depressive symptoms, and the Pediatric Daytime Sleepiness Scale (PDSS) [36].

Potential correlates measured at T2 included sociodemographic characteristics (ie, age, sex, mother university-educated, annual household income, participant currently employed), lifestyle behaviors (ie, moderate-to-vigorous PA [MVPA] per week, meeting MVPA guidelines, past-year binge drinking, past-year marijuana use, hours of TV per day, hours of computer per day, hours of nonactive video games per day), weight-related indicators (BMI, stress about weight, perceived weight too heavy, trying to lose weight, body-related guilt, body-related shame), PA motivation, and depressive symptoms. The only variable reported by parents was mother's education; all other variables were reported by participants. Multimedia Appendix 1 describes each variable in detail including response options, coding for analysis, and Cronbach alpha for scales.

Data Analysis

Descriptive statistics were used to compare participants who did and did not sustain exergaming. The analysis considered investigation of each potential correlate or predictor as an independent study that addressed a specific hypothesis so that only 3 statistical tests (ie, univariate, partially adjusted model, and fully adjusted model) were performed for each potential correlate or predictor [34]. Potential confounders were retained in the model if they were not on the causal pathway and if they were correlated with the potential correlate or predictor and outcome at $r>0.20$ [35,37]. A total of 3 logistic regression models including (1) a univariate model examining the unadjusted association between the potential correlate/predictor and outcome with no covariates; (2) a partially adjusted model accounting for age and sex; and (3) a fully adjusted model including age, sex, and (other) potential confounders. Data were analyzed using SPSS version 20.0 (released 2011, SPSS Statistics for Windows; IBM Corp). All statistical tests were 2-sided, with the significance level set at 0.05.

Results

Data on exergaming were collected in AdoQuest for the first time in grade 9 in 2008-09. Of 1801 grade 5 participants at inception in 2005-06, 68.46% (1233/1801) completed questionnaires at T1 in grade 9, and 16.46% (203/1233) of participants reported exergaming. At T2 in grade 11, 69.01% (1243/1801) completed questionnaires, and 23.97% (298/1243) of participants reported exergaming. A total of 54.08% (974/1801) participants completed questionnaires at both T1 and T2. Of the 974, 62.92% (613/974) never exergamed, 17.96% (175/974) began exergaming between T1 and T2, and 8.00% (81/974) and 10.98% (105/974) sustained and stopped exergaming, respectively. The total number of exergamers at T1 who provided data on exergaming at T2 was 19.09% (186/974), and these participants constituted the analytic sample. Table 1 compares the characteristics of the 186 participants retained in the analytic sample with those of the 59 participants who exergamed at T1 but were missing data on exergaming at T2. There were no statistically significant differences between the 2 groups, although 49.46% (92/186) of those retained reported taking action to change their weight compared with 36.02% (21/59) of those not retained (P=.08).

Table 1. Comparison of baseline characteristics of exergamers retained (n=186) and not retained (n=59) in the analytic sample, AdoQuest 2005-12.

Variable	Retained	Not retained[a]	*P* value
Age (years) at baseline, mean (SD)	10.8 (0.5)	10.7 (0.5)	.52
Male, n (%)	90 (48.6)	32 (53.4)	.52
Mother university-educated, n (%)	56 (30.1)	15 (25.5)	.13
French-speaking, n (%)	165 (88.6)	51 (86.4)	.66
Self-esteem, mean (SD)	2.0 (0.7)	2.0 (0.7)	>.99
Depressive symptoms, mean (SD)	2.0 (0.7)	2.0 (0.6)	.87
School connectedness, mean (SD)	2.3 (1.0)	2.3 (0.9)	.97
Taking action to change weight, n (%)	92 (49.4)	21 (35.8)	.08
TV ≥3 hours/day, n (%)	70 (37.6)	23 (39.7)	.78
Perceived academic performance, n (%) above average	72 (38.8)	27 (45.5)	.35

[a]Exergamers at T1 without exergaming data at T2.

The mean age of participants in the analytic sample was 16.7 (SD 0.5) years at T2, 42% (522/1243) were boys, 93% (1156/1243) were white, 76% (945/1243) were in grade 11, 47% (584/1243) were employed full or part-time, and 33% (410/1243) had university-educated mothers. The mean (SD) BMI was 23.6 (5.0) in boys and 22.2 (4.1) in girls. A total of 30% (373/1243) of participants' parents reported an annual household income greater or equal to Can $100,000.

Of the 186 participants retained in the analytic sample, 43.6% (81/186) sustained exergaming (ie, reported exergaming at T1 and T2). At T2, sustained exergamers (n=81) exergamed on a mean (SD) of 1.9 (1.4) days per week, for 49.0 (33.8) min per bout on average. Overall, 35% (65/186) of sustained exergamers reported that they exergamed at light intensity, 42% (78/186) exergamed at moderate intensity, and 23% (32/186) exergamed at vigorous intensity.

In univariate analyses of potential correlates measured at T2 (Table 2), girls were twice as likely to sustain exergaming as boys. With each unit increase in depressive symptoms, the odds of sustained exergaming increased by 40%. Similarly, the odds of sustained exergaming were higher in participants with higher levels of introjected, identified, or intrinsic regulation. Finally, the odds of exergaming increased by 30% with each unit increase in body-related guilt. In partially adjusted models, girls were more likely to sustain exergaming, and for each unit increase in introjected regulation, there was a 60% increase in the odds of sustained exergaming. In fully adjusted models, the only variable retained as statistically significant was introjected regulation (odds ratio [OR] 1.8; 95% CI 1.1-3.2).

Table 2. Odds ratio (OR) and 95% CI for the association between potential correlates and sustained exergaming, AdoQuest 2005-12 (N=186).

Indicators[a]	n (%[b])	Sustained exergaming[c], n (%)	OR_{crude} (95% CI)	Model adjusted for age and sex OR_{adj} (95% CI)	Fully adjusted model OR_{adj} (95% CI)	Covariate(s) included in fully adjusted model
Sociodemographic indicators						
Age (years)[d]						
14.39-16.63	60 (32.3)	25 (41.7)	1.3 (0.8-2.1)	1.2 (0.7-2.0)	1.2 (0.7-2.0)	Sex
16.64-16.98	59 (31.7)	21 (40.7)	—[e]	—	—	—
16.99-19.86	59 (31.7)	28 (47.5)	—	—	—	—
Sex						
Boys	83 (44.6)	30 (34.9)	reference	reference	reference	Age, MVPA[f], nonactive video games
Girls	103 (55.4)	51 (50.5)	2.0 (1.1-3.4)	1.9 (1.0-3.4)	1.6 (0.8-3.3)	—
Mother university-educated						
Yes	49 (26.3)	23 (46.9)	reference	reference	reference	Age, sex, income
No	108 (58.1)	45 (41.7)	0.8 (0.4-1.6)	0.8 (0.4-1.)	0.9 (0.4-2.0)	—
Income, Can $						
<100K	84 (45.2)	33 (38.9)	reference	reference	reference	Age, sex, mother's education
≥100K	60 (32.3)	28 (46.9)	1.4 (0.7-3.0)	1.4 (0.7-3.0)	1.5 (0.7-3.2)	Age, sex, mother's education
Employed (AdoQuest participant)						
Yes	78 (42.0)	34 (43.6)	reference	reference	reference	Age, sex
No	101 (54.3)	43 (42.6)	1.0 (0.5-1.7)	1.0 (0.6-1.9)	1.0 (0.6-1.9)	Age, sex
Lifestyle behaviors						
MVPA/min per week[d]						
0-105	59 (32.8)	21 (35.6)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	1.0 (1.0-1.0)	Age, sex, identified regulation
106-295	57 (30.6)	30 (52.6)	—	—	—	—
≥295	64 (34.4)	28 (43.8)	—	—	—	—
Meets MVPA guidelines						
Yes	42 (22.6)	16 (38.1)	reference	reference	reference	Age, sex, intrinsic, body shame
No	138 (74.2)	63 (45.7)	1.4 (0.7-2.8)	1.1 (0.5-2.3)	1.3 (1.0-1.6)	—
Binge drank past year						
Yes	82 (44.1)	37 (45.1)	reference	reference	reference	Age, sex, marijuana use, intrinsic regulation
No	94 (50.5)	39 (41.5)	0.9 (0.5-1.6)	0.9 (0.5-1.6)	1.2 (0.6-2.4)	—
Used marijuana past year						
Yes	141 (75.8)	71 (50.0)	reference	reference	reference	Age, sex, binge drinking, smoked cigarettes
No	34 (18.3)	14.2 (41.8)	1.0 (0.8-1.3)	0.7 (0.3-1.6)	0.4 (0.1-1.0)	—
Hours of TV/day[d]						
≥0<1	42 (22.6)	19 (45.2)	1.0 (0.8-1.4)	1.1 (0.8-1.4)	1.1 (0.8-1.5)	Age, sex, depressive symptoms
≥1<2	78 (41.9)	31 (39.7)	—	—	—	—
≥2	66 (35.5)	31 (47.0)	—	—	—	—
Hours of computer/day[d]						
≥0<1	38 (20.4)	15 (39.5)	1.1 (0.9-1.3)	1.1 (0.8-1.3)	1.1 (0.8-1.3)[g]	Age, sex
≥1<2	56 (30.1)	22 (39.3)	—	—	—	—
≥2	92 (49.5)	44 (47.8)	—	—	—	—

Indicators[a]			n (%[b])	Sustained exergaming[c], n (%)	OR_{crude} (95% CI)	Model adjusted for age and sex OR_{adj} (95% CI)	Fully adjusted model OR_{adj} (95% CI)	Covariate(s) included in fully adjusted model
	Hours of nonactive video games/day[d]							
		0	82 (44.1)	32 (39.0)	1.0 (0.8-1.3)	1.0 (0.9-1.4)	1.0 (0.9-1.4)[g]	Age, sex
		<1	50 (26.7)	28 (56.0)	—	—	—	—
		≥1	53 (28.5)	21 (39.6)	—	—	—	—
Weight-related indicators								
	BMI[d,h]							
		16.01-20.1	49 (26.3)	23 (46.9)	1.0 (0.9-1.0)	0.9 (0.9-1.1)	0.9 (0.8-1.0)	Age, sex, income, extrinsic motivation, body shame, body guilt
		20.2-23.54	51 (27.4)	20 (39.2)	—	—	—	—
		23.55-40.70	49 (26.3)	18 (36.7)	—	—	—	—
	Stressed about weight							
		Yes	50 (26.9)	25 (50.0)	reference	reference	reference	Age, sex, depressive symptoms, weight perception, introjected regulation, external regulation, body shame, body guilt, trying to lose weight
		No	129 (69.4)	52 (40.3)	0.7 (0.4-1.3)	0.8 (0.3-2.1)	0.9 (0.3-2.0)	—
	Perceived weight status: overweight							
		Yes	49 (26.3)	23 (46.9)	reference	reference	reference	Age, sex, BMI, introjected regulation, external regulation, body shame, body guilt, stress about weight, trying to lose weight
		No	127 (68.3)	53 (41.7)	0.8 (0.2-1.6)	0.8 (0.2-1.6)	0.3 (0.1-1.3)	—
	Trying to lose weight							
		Yes	114 (61.3)	57 (50.0)	reference	reference	reference	Age, sex, BMI, weight perception, introjected regulation, external regulation, identified regulation, stress about weight, weight perception, body shame, body guilt
		No	62 (33.3)	25 (39.5)	0.7 (0.4-1.2)	0.8 (0.4-1.4)	0.8 (0.3-1.2)	—
	Body-related guilt[d]							
		0≤1.0	67 (36.0)	20 (29.9)	1.3 (1.0-1.7)	1.2 (0.9-1.7)	0.9 (0.4-2.0)	Age, sex, BMI, depressive symptoms, weight perception, extrinsic regulation, introjected regulation, identified regulation, body shame, stress about weight, trying to lose weight
		>1.0≤2.0	52 (28.0)	26 (50.0)	—	—	—	—
		>2.0-5.0	58 (31.2)	29 (50.0)	—	—	—	—
	Body-related shame[d]							
		0≤1.0	61 (32.8)	16 (26.2)	1.2 (0.9-1.7)	1.2 (0.9-1.6)	1.0 (0.4-2.3)	Age, sex, stress about weight, BMI, depressive symptoms, weight perception, extrinsic regulation, introjected regulation, body guilt, trying to lose weight
		1.0≤2.0	59 (31.7)	31 (52.5)	—	—	—	—
		>2.0-4.8	57 (30.6)	28 (49.1)	—	—	—	—
Physical activity motivation								
	Amotivation[d]							
		0≤1	130 (69.9)	57 (43.8)	1.3 (0.7-2.4)	1.3 (0.7-2.6)	2.0 (0.9-5.0)	Age, sex, identified regulation, intrinsic regulation, body shame
		>1-4	47 (25.3)	18 (38.3)	—	—	—	—

Indicators[a]	n (%[b])	Sustained exergaming[c], n (%)	OR_{crude} (95% CI)	Model adjusted for age and sex OR_{adj} (95% CI)	Fully adjusted model OR_{adj} (95% CI)	Covariate(s) included in fully adjusted model
External[d]						
0≤1	113 (60.8)	46 (40.7)	1.0 (0.5-2.1)	1.1 (0.5-2.2)	0.9 (0.3-2.8)	Age, sex. BMI, depressive symptoms, stress about weight, weight perception, trying to lose weight, introjected regulation, body shame, body guilt
>1≤1.67	20 (10.8)	11 (55)	—	—	—	—
1.68-3.25	44 (23.7)	18 (40.9)	—	—	—	—
Introjected[d]						
0	67 (36.0)	21 (31.3)	1.7 (1.2-2.4)	1.6 (1.1-2.3)	1.8 (1.1-3.2)	Age, sex, depressive symptoms, weight stress, weight perception, external regulation, identified regulation, intrinsic regulation, trying to lose weight, body shame, body guilt
1-1.67	46 (24.7)	20 (43.5)	—	—	—	—
1.68-5	64 (34.4)	24 (53.1)	—	—	—	—
Identified[d]						
0-2.25	54 (29.0)	29 (37.0)	1.4 (1.0-1.8)	1.3 (1.0-1.8)	1.3 (0.8-2.0)	Age, sex, amotivation, introjected regulation, intrinsic motivation, regulation, body guilt, trying to lose
2.26-3.25	63 (33.9)	22 (34.9)	—	—	—	—
3.26-5	60 (32.3)	33 (55.0)	—	—	—	—
Intrinsic[d]						
0-1.0	56 (30.1)	19 (33.9)	1.3 (1.0-1.6)	1.3 (1.0-1.7)	1.3 (0.9-2.0)	Age, sex, MVPA, amotivation, identified, binge drinking, introjected regulation, body shame
1-1.67	55 (29.6)	23 (41.8)	—	—	—	—
≥1.68	66 (35.5)	33 (50.0)	—	—	—	—
Depressive symptoms[d]						
0-1.67	59 (31.7)	22 (37.3)	1.4 (1.0-2.0)	1.3 (0.9-1.9)	1.4 (0.8-2.2)	Age, sex, introjected regulation, external regulation, body shame, body guilt, stress about weight, TV use
1.68-2.50	66 (35.5)	25 (37.9)	—	—	—	—
2.51-4.5	54 (29.0)	30 (55.6)	—	—	—	—

[a]Potential correlate measured in grade 11.

[b]Percentages may not total 100% due to missing data.

[c]Continuous variables were grouped according to tertile cut-offs for descriptive purposes.

[d]Potential correlate included in the model as a continuous variable. Odds ratio indicates the increase in the probability of the outcome per 1-unit change in the correlate.

[e]Not applicable.

[f]MVPA: moderate-to-vigorous physical activity.

[g]Partially and fully adjusted models were identical.

[h]BMI: body mass index.

In univariate analyses of potential predictors (Table 3), more time spent playing nonactive video games daily was protective against sustained exergaming sustainers. Trying to lose weight was associated with an 80% increase in the odds of sustained exergaming. In partially adjusted models controlling for sex and age and in fully-adjusted models, none of the potential predictors were statistically significantly associated with sustained exergaming.

Table 3. Odds ratio (OR) and 95% CI for the association between potential predictors and sustained exergaming, AdoQuest 2005-12 (N=186).

Predictors[a]	n (%[b])	Sustained exergaming[c], n (%)	OR_{crude} (95% CI)	Model adjusted for age and sex, OR_{adj} (95% CI)	Fully adjusted model, OR_{adj} (95% CI)	Covariates included in fully adjusted model
Lifestyle behaviors						
Physical activity level[d]						
1	37 (20.0)	13 (35.1)	1.0 (0.8-1.3)	1.1 (0.8-1.3)	1.1 (0.8-1.3)[e]	Age, sex
2-3	107 (57.5)	50 (46.7)	—[f]	—	—	—
4-5	40 (21.3)	16 (40.0)	—	—	—	—
Ever smoked cigarettes						
Yes	57 (30.7)	29 (50.9)	reference	reference	reference	Age, sex, cannabis, binge drinking, depressive symptoms, PDSS[g]
No	127 (68.3)	51 (40.2)	0.7 (0.4-1.2)	0.7 (0.4-1.3)	0.5 (0.3-1.2)	—
Binge drank						
Yes	43 (26.3)	16 (37.2)	reference	reference	reference	Age, sex, cigarette, depressive symptoms, cannabis, PDSS
No	143 (76.9)	65 (45.5)	1.4 (0.7-2.8)	1.3 (0.6-2.6)	1.5 (0.6-3.8)	—
Marijuana use						
Yes	21 (11.3)	9 (42.9)	reference	reference	reference	Age, sex, cig try, binge drinking, depressive symptoms
No	164 (88.2)	72 (43.9)	1.0 (0.4-2.6)	0.9 (0.4- 2.4)	1.0 (0.3-3.5)	—
Hours of TV/day						
≤1	20 (10.8)	9 (45.0)	1.1 (0.6-2.2)	0.8 (0.6-1.1)	0.9 (0.6-1.4)	Age, sex, computer, sedentary behavior
≥1≤3	93 (50.0)	41 (44.1)	—	—	—	—
>3	72 (38.8)	30 (41.7)	—	—	—	—
Hours of computer/day[d]						
≤1	31 (16.7)	11 (35.5)	1.3 (1.0-1.7)	1.4 (1.0-2.0)	1.3 (0.9-2.1)	Age, sex, depressive symptoms, video games, TV, sedentary behavior
≥1≤3	75 (40.3)	29 (38.7)	—	—	—	—
>3	79 (42.5)	41 (51.9)	—	—	—	—
Hours of nonactive video games/day[d]						
≤1	107 (57.5)	54 (50.5)	0.6 (0.5-1.0)	0.8 (0.5-1.3)	0.8 (0.4-1.3)	Age, sex, computer, sedentary behavior
>1	73 (39.2)	26 (35.6)	—	—	—	—
Hours of sedentary behavior/day[d]						
0-9.5	60 (32.3)	23 (38.3)	1.1 (1.0-1.2)	1.0 (1.0-1.1)	1.0 (0.9-1.2)	Age, sex, TV, computer, video games
9.6-11.5	52 (27.4)	24 (46.2)	—	—	—	—
>11.5	61 (37.8)	30 (49.2)	—	—	—	—
Weight-related indicators						
BMI[d,h]						
13.7-19.2	35 (18.8)	17 (48.6)	1.0 (0.9-1.1)	1.1 (1.0-1.2)	1.1 (1.0-1.2)	Age, sex, trying to lose weight, video games, perceived weight, stress about weight, sedentary behavior
19.3-22.66	35 (18.8)	14 (40.0)	—	—	—	—
22.67+	36 (19.4)	11 (30.6)	—	—	—	—
Stress about weight						

Predictors[a]	n (%[b])	Sustained exergaming[c], n (%)	OR_{crude} (95% CI)	Model adjusted for age and sex, OR_{adj} (95% CI)	Fully adjusted model, OR_{adj} (95% CI)	Covariates included in fully adjusted model
Yes	80 (43.0)	40 (50.0)	reference	reference	reference	Age, sex, depressive symptoms, trying to lose weight, perceived weight, PDSS, BMI
No	105 (56.5)	41 (39.0)	0.6 (0.4-1.2)	0.8 (0.4-1.5)	0.6 (0.2-1.8)	—
Perceived weight too heavy						
Yes	55 (29.6)	16 (29.5)	reference	reference	reference	Age, sex, BMI, stress about weight, trying to lose weight
No	128 (68.8)	39 (30.8)	1.1 (0.6-2.1)	1.1 (0.6-2.2)	3.0 (0.8-10.8)	—
Trying to lose weight						
No	113 (60.8)	42(37.2)	reference	reference	reference	Age, sex, stress about weight, depressive symptoms, perceived weight
Yes	70 (37.6)	36 (51.4)	1.8 (1.0-3.3)	1.3 (0.7-2.6)	1.6 (0.7-3.5)	—
Depressive symptoms[d]						
0-1.5	57 (30.6)	27 (47.4)	1.0 (0.7-1.4)	0.8 (0.5-1.2)	0.8 (0.5-1.3)	Age, sex, stress weight, binge drinking, trying to lose weight, PDSS, cigarettes, computer, sedentary behavior
1.6-1.17	65 (35.0)	24 (36.9)	—	—	—	—
1.8-5.0	62 (33.3)	30 (48.4)	—	—	—	—
PDSS[d]						
0-7	62 (33.3)	24 (38.7)	1.0 (1.0-1.1)	1.0 (1.0-1.1)	1.0 (1.0-1.1)	Age, sex, stress about weight, binge drinking, depressive symptoms
8-13	60 (32.3)	23 (38.3)	—	—	—	—
≥14	62 (33.3)	34 (54.8)	—	—	—	—

[a]Potential predictor measured in grade 9.

[b]Percentages may not total 100% due to missing data.

[c]Continuous variables were grouped according to tertile cut-offs for descriptive purposes.

[d]Potential predictor included in the models as a continuous variable. Odds ratio (OR) indicates the increase in the probability of the outcome per 1 unit change in the predictor.

[e]Partially and fully adjusted models were identical.

[f]Not applicable.

[g]PDSS: Paediatric daytime sleepiness scale.

[h]BMI: body mass index.

Discussion

Principal Findings

The purpose of this study was to describe sustained exergaming in a population-based sample of adolescents and to identify factors associated with sustained exergaming. To our knowledge, this is the first investigation of sustained exergaming in a population-based sample. Results indicated that 44% of exergamers sustained exergaming for at least 2 years. Female sex and having higher introjected PA behavior regulation were associated with sustained exergaming.

In a cross-sectional study of 200 children who owned consoles, Simons et al [38] found that 11% never exergamed and only 32% exergamed regularly (ie, >1 hour per week). Because of concerns that the novelty of exergaming dissipates over time, there have been numerous calls to investigate the sustainability of exergaming [17,24,28,39-42]. The 44% (81/186) of sustained exergamers in our study compares with the 41% of girls (but not the 69% of boys) who remained involved in team sports for over 5 years, as observed by Belanger et al [41]. Specifically, of 1276 adolescents age 12-13 years who completed a 7-day PA recall every 3 months for 5 years, the authors reported that between 14% and 53% re-engaged in a specific PA after discontinuation. Although not investigated in the Belanger study, exergaming could be an intermittent activity in adolescents linked to the release of new games and consoles. Future studies will need more frequent follow-up to assess the stop-start aspect of exergaming over time.

Three-quarters (77% [143/186]) of participants exergamed at light or moderate intensity. Exergaming at these intensities may be more enjoyable, practical, and achievable among young persons. Thus, the flexibility of exergaming in level of intensity may contribute to sustainability.

Previous studies have reported that females are more likely to exergame than males [12], and in this analysis, girls were more likely to sustain exergaming. The reasons for sex differences in sustained exergaming may mirror the reasons for sex difference in exergaming. Girls, especially those with body image challenges, may be more comfortable being active at home away from the scrutiny of others, while still enjoying the social interaction provided by exergaming. Boys may be attracted by the novelty of exergaming but return to nonactive video games as the novelty wears off. Although sex differences in exergaming should be further investigated, our results suggest that exergaming may be a viable option to help girls in particular remain physically active.

A specific focus of this work was to assess whether PA motivation is associated with sustained exergaming. Self-determination theory (SDT) has been used as a framework to predict intentions to engage in and sustain traditional PA, and investigators have reported that those with higher intrinsic regulation (ie, motivation) report increased intentions to engage in PA than those with external PA regulation (eg, [43-46]. Introjected regulation (ie, a type of PA motivation indicative of internalization of PA as a behavior) was a correlate of sustained exergaming. Introjected regulation, as described in the SDT [47-49], is motivation from an internal pressure that usually drives short-term behavior change, but does not foster sustainable behavior change [48,50]. It is generally negatively associated with or unrelated to PA levels, although there appears to be sex differences such that girls who report higher introjected regulation also report more PA [49]. The link between behavior regulation and exergaming has been studied in clinical settings and in specific populations such as overweight youth [51,52], but few studies investigate this association in population-based samples of youth using the SDT [53], and it is not fully understood if exercise behavior regulation differs between those who do and do not sustain exergaming. Sustained exergamers may not be highly motivated intrinsically toward traditional PA and are drawn toward exergaming as a PA alternative. They may be interested in developing skills for fitness and weight change, which is fostered through fitness exergames (ie, body weight, body alignment, placement, feeling, and speed captured by motion and sensor-captures and displayed in real time on the screen) [52], or aspects of the advertising or marketing of fitness exergames may play on or increase guilt, particularly among female users. Future research should focus on better understanding the differences between motivation for PA and for exergaming, which in turn lead to exergaming interventions that have a greater impact on PA and sedentary behavior.

No predictors of sustained exergaming were identified in this study. Rather than being a planned behavior, exergaming may be triggered by events in the immediate present such as purchasing a new console or friends coming over, and therefore, has few predictors. It is also possible that the sample size was too small to detect factors associated with sustained exergaming because exergaming is a relatively new area of research, not all relevant predictors may not have been investigated.

Limitations of this study include the small sample size, that self-report data are subject to misclassification, that loss to follow-up may have resulted in selection bias, and that restriction of the sample to francophones may have limited external generalizability. As data were not collected more frequently, it was not possible to confirm whether participants exergamed between data collection cycles.

Conclusions

Exergaming may represent a novel approach to help adolescents remain physically active during a period of life which is notable for sharp declines in PA. Exergaming may be more sustainable if games include components that foster intrinsic PA motivation such as providing more choices in the games offered (ie, with whom the game can be played, whether the setting is collaborative or competitive, how intensely the game is played) or having a coach and/or other social support while learning a new game. Finally, clinicians and practitioners can counsel parents to encourage their children to choose exergaming over more traditional sedentary video games or to exergame as a family.

Acknowledgments

This project was funded by the Canadian Tobacco Control Research Initiative and the Institut national de santé publique du Québec (INSPQ) through a financial contribution from the Québec Ministry of Health and Social Services to the INSPQ. Views expressed in this document do not necessarily reflect those of the Québec Ministry of Health and Social Services. This work is part of the first author’s (EKO) PhD dissertation and coauthors are committee members.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Variables tested in models including response options, coding for analysis, and Cronbach alpha for scales.

[PDF File (Adobe PDF File), 90KB - games_v7i3e13335_app1.pdf]

References

1. Colley R, Garriguet D, Janssen I, Craig CL, Clarke J, Tremblay MS. Physical activity of Canadian children and youth: accelerometer results from the 2007 to 2009 Canadian Health Measures Survey. Health Rep 2011 Mar;22(1):15-23 [FREE Full text] [Medline: 21510586]
2. Rideout V, Foehr UG, Roberts DF. Education Resources Information Center.: ERIC Clearinghouse; 2010. Generation M: Media in the Lives of 8-to 18-Year-Olds URL: https://eric.ed.gov/?id=ED527859 [accessed 2019-04-22] [WebCite Cache ID 77pb0ULzj]
3. Janssen I, LeBlanc AG. Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. Int J Behav Nutr Phys Act 2010;7(1):40-16. [doi: 10.1186/1479-5868-7-40]
4. Warburton D, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. Can Med Assoc J 2006 Mar 14;174(6):801-809 [FREE Full text] [doi: 10.1503/cmaj.051351] [Medline: 16534088]
5. Tremblay M, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8(1):98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
6. Patnode CD, Lytle LA, Erickson DJ, Sirard JR, Barr-Anderson DJ, Story M. Physical activity and sedentary activity patterns among children and adolescents: a latent class analysis approach. J Phys Act Health 2011 May;8(4):457-467 [FREE Full text] [Medline: 21597117]
7. Must A, Tybor DJ. Physical activity and sedentary behavior: a review of longitudinal studies of weight and adiposity in youth. Int J Obes (Lond) 2005 Sep;29(Suppl 2):S84-S96. [Medline: 16385758]
8. Ortega F, Konstabel K, Pasquali E, Ruiz JR, Hurtig-Wennlöf A, Mäestu J, et al. Objectively measured physical activity and sedentary time during childhood, adolescence and young adulthood: a cohort study. PLoS One 2013;8(4):e60871 [FREE Full text] [doi: 10.1371/journal.pone.0060871] [Medline: 23637772]
9. Verloigne M, Van Lippevelde W, Maes L, Yıldırım M, Chinapaw M, Manios Y, et al. Levels of physical activity and sedentary time among 10- to 12-year-old boys and girls across 5 European countries using accelerometers: an observational study within the ENERGY-project. Int J Behav Nutr Phys Act 2012 Mar 31;9:34 [FREE Full text] [doi: 10.1186/1479-5868-9-34] [Medline: 22462550]
10. Madden M, Lenhart A, Duggan M, Cortesi S, Gasser U. Pew Research Center. 2013. Teens and technology 2013 URL: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2013/PIP_TeensandTechnology2013.pdf [accessed 2019-04-22] [WebCite Cache ID 77plVVewg]
11. Fulton J. Circulation. 2010. Abstract P260: Active Video Game Participation in US Youth: Findings from the National Youth Physical Activity and Nutrition Survey, 2010 URL: https://www.ahajournals.org/doi/abs/10.1161/circ.125.suppl_10.ap260 [accessed 2019-04-26] [WebCite Cache ID 77vsIHzHZ]
12. O'Loughlin E, Dugas EN, Sabiston CM, O'Loughlin JL. Prevalence and correlates of exergaming in youth. Pediatrics 2012 Nov;130(5):806-814. [doi: 10.1542/peds.2012-0391] [Medline: 23027171]
13. Kakinami L, O'Loughlin EK, Dugas EN, Sabiston CM, Paradis G, O'Loughlin J. The association between exergaming and physical activity in young adults. J Phys Act Health 2015 Jun;12(6):789-793. [doi: 10.1123/jpah.2013-0447] [Medline: 25133470]
14. Janssen I. Estimating whether replacing time in active outdoor play and sedentary video games with active video games influences youth's mental health. J Adolesc Health 2016 Dec;59(5):517-522. [doi: 10.1016/j.jadohealth.2016.07.007] [Medline: 27567064]
15. Makuck E. Gamespot. 2013. Kinect Sales Reach 24 Million URL: https://www.gamespot.com/articles/kinectsales-reach-24-million/1100-6403766/ [accessed 2019-04-23] [WebCite Cache ID 77qdOcRcL]
16. Best JR. Exergaming in youth: effects on physical and cognitive health. Z Psychol 2013 Apr 01;221(2):72-78 [FREE Full text] [doi: 10.1027/2151-2604/a000137] [Medline: 25097828]
17. Lamboglia CM, da Silva VT, de Vasconcelos Filho JE, Pinheiro MH, Munguba MC, Silva Júnior FV, et al. Exergaming as a strategic tool in the fight against childhood obesity: a systematic review. J Obes 2013;2013:438364 [FREE Full text] [doi: 10.1155/2013/438364] [Medline: 24319594]
18. Sween J, Wallington SF, Sheppard V, Taylor T, Llanos AA, Adams-Campbell LL. The role of exergaming in improving physical activity: a review. J Phys Act Health 2014 May;11(4):864-870 [FREE Full text] [doi: 10.1123/jpah.2011-0425] [Medline: 25078529]
19. Zeng N, Gao Z. Exergaming and obesity in youth: current perspectives. Int J Gen Med 2016;9:275-284 [FREE Full text] [doi: 10.2147/IJGM.S99025] [Medline: 27536158]
20. Graf D, Pratt LV, Hester CN, Short KR. Playing active video games increases energy expenditure in children. Pediatrics 2009 Aug;124(2):534-540. [doi: 10.1542/peds.2008-2851] [Medline: 19596737]
21. Lanningham-Foster L, Jensen TB, Foster RC, Redmond AB, Walker BA, Heinz D, et al. Energy expenditure of sedentary screen time compared with active screen time for children. Pediatrics 2006 Dec;118(6):e1831-e1835. [doi: 10.1542/peds.2006-1087] [Medline: 17142504]
22. Peng W, Lin JH, Crouse J. Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):681-688. [doi: 10.1089/cyber.2010.0578] [Medline: 21668370]

23. Chen J, Wilkosz ME. Efficacy of technology-based interventions for obesity prevention in adolescents: a systematic review. Adolesc Health Med Ther 2014;5:159-170 [FREE Full text] [doi: 10.2147/AHMT.S39969] [Medline: 25177158]
24. Guy S, Ratzki-Leewing A, Gwadry-Sridhar F. Moving beyond the stigma: systematic review of video games and their potential to combat obesity. Int J Hypertens 2011;2011:179124 [FREE Full text] [doi: 10.4061/2011/179124] [Medline: 21629863]
25. Lu A, Kharrazi H, Gharghabi F, Thompson D. A systematic review of health videogames on childhood obesity prevention and intervention. Games Health J 2013 Jun;2(3):131-141 [FREE Full text] [doi: 10.1089/g4h.2013.0025] [Medline: 24353906]
26. Gao Y, Gerling K, Mandryk R, Stanley K. Decreasing sedentary behaviours in pre-adolescents using casual exergames at school. 2014 Oct 19 Presented at: Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play; October 19-21, 2014; Toronto, Ontario, Canada. [doi: 10.1145/2658537.2658693]
27. Li J, Theng YL, Foo S. Effect of exergames on depression: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2016 Jan;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
28. Gao Z, Chen S. Are field-based exergames useful in preventing childhood obesity? A systematic review. Obes Rev 2014 Aug;15(8):676-691. [doi: 10.1111/obr.12164] [Medline: 24602072]
29. LeBlanc A, Chaput JP, McFarlane A, Colley RC, Thivel D, Biddle SJ, et al. Active video games and health indicators in children and youth: a systematic review. PLoS One 2013;8(6):e65351 [FREE Full text] [doi: 10.1371/journal.pone.0065351] [Medline: 23799008]
30. Bethea TC, Berry D, Maloney AE, Sikich L. Pilot study of an active screen time game correlates with improved physical fitness in minority elementary school youth. Games Health J 2012 Feb;1(1):29-36 [FREE Full text] [doi: 10.1089/g4h.2011.0005] [Medline: 26196430]
31. O?Loughlin E. K.., Prevalence and correlates of exergaming in youth. Pediatrics 2012;130(5):806-814.
32. [Ministry of Education]. [National Institute of Public Health of Quebec]. 2003. [School Population Map and Deprivation Indices] URL: http://www.education.gouv.qc.ca/fileadmin/site_web/documents/PSG/statistiques_info_decisionnelle/bulletin_26.pdf [accessed 2019-04-23] [WebCite Cache ID 77qsMkx7o]
33. Sylvestre M, Wellman RJ, O'Loughlin EK, Dugas EN, O'Loughlin J. Gender differences in risk factors for cigarette smoking initiation in childhood. Addict Behav 2017 Dec;72:144-150. [doi: 10.1016/j.addbeh.2017.04.004] [Medline: 28399489]
34. Hagströmer M, Oja P, Sjöström M. The International Physical Activity Questionnaire (IPAQ): a study of concurrent and construct validity. Public Health Nutr 2006 Sep;9(6):755-762. [Medline: 16925881]
35. Sallis J, Prochaska JJ, Taylor WC. A review of correlates of physical activity of children and adolescents. Med Sci Sports Exerc 2000 May;32(5):963-975. [Medline: 10795788]
36. Perez-Chada D, Perez-Lloret S, Videla AJ, Cardinali D, Bergna MA, Fernández-Acquier MA, et al. Sleep disordered breathing and daytime sleepiness are associated with poor academic performance in teenagers. A study using the Pediatric Daytime Sleepiness Scale (PDSS). Sleep 2007 Dec;30(12):1698-1703 [FREE Full text] [doi: 10.1093/sleep/30.12.1698] [Medline: 18246979]
37. Bender R, Lange S. Adjusting for multiple testing--when and how? J Clin Epidemiol 2001 Apr;54(4):343-349. [Medline: 11297884]
38. Simons M, Bernaards C, Slinger J. Active gaming in Dutch adolescents: a descriptive study. Int J Behav Nutr Phys Act 2012 Oct 02;9:118 [FREE Full text] [doi: 10.1186/1479-5868-9-118] [Medline: 23031076]
39. Barnett A, Cerin E, Baranowski T. Active video games for youth: a systematic review. J Phys Act Health 2011 Jul;8(5):724-737. [Medline: 21734319]
40. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010 Jul;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
41. Carmo J, Palmeira A. Can active video games be part of the solution to promote physical activity in youth? A systematic review. Archives of Exercise in Health and Disease 2013;4(1):216-226 [FREE Full text] [doi: 10.4018/ijgcms.2014100105]
42. Kari T. Can exergaming promote physical fitness and physical activity? A systematic review of systematic reviews. Int J Gaming Comput Mediat Simul 2014;6(4):59-77. [doi: 10.4018/ijgcms.2014100105]
43. Edmunds J, Ntoumanis N, Duda JL. Examining exercise dependence symptomatology from a self-determination perspective. J Health Psychol 2006 Nov;11(6):887-903. [doi: 10.1177/1359105306069091] [Medline: 17035261]
44. Edmunds J, Ntoumanis N, Duda J. Testing a self-determination theory-based teaching style intervention in the exercise domain. Eur J Soc Psychol 2008 Mar;38(2):375-388. [doi: 10.1002/ejsp.463]
45. Edmunds J, Ntoumanis N, Duda J. Helping your clients and patients take ownership over their exercise: Fostering exercise adoption, adherence, and associated well-being. ACSMs Health Fit J 2009;13(3):20-25 [FREE Full text]
46. Thøgersen-Ntoumani C, Ntoumanis N. The role of self-determined motivation in the understanding of exercise-related behaviours, cognitions and physical self-evaluations. J Sports Sci 2006 Apr;24(4):393-404. [doi: 10.1080/02640410500131670] [Medline: 16492603]
47. Brötz D. Exercise, physical activity, and Self-Determination Theory: a systematic review. physioscience 2013 Feb 26;9(01):37 [FREE Full text] [doi: 10.1055/s-0032-1330638]

48. Deci E. Overview of self-determination theory: an organismic dialectical perspective. In: Ryan R, editor. Handbook of Self-Determination Research. Rochester, NY, US: University of Rochester Press; 2002:A-33.
49. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012 Jun 22;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
50. Deci E. Self-determination theory. In: Ryan R, editor. Handbook of Theories of Social Psychology. Thousand Oaks, California: Sage publications; 2011:416-433.
51. Fitzgerald D, Trakarnratanakul N, Smyth B, Caulfield B. Effects of a wobble board-based therapeutic exergaming system for balance training on dynamic postural stability and intrinsic motivation levels. J Orthop Sports Phys Ther 2010 Jan;40(1):11-19. [doi: 10.2519/jospt.2010.3121] [Medline: 20044704]
52. Osorio G, Moffat DC, Sykes J. Exergaming, exercise, and gaming: sharing motivations. Games Health J 2012 Jun;1(3):205-210. [doi: 10.1089/g4h.2011.0025] [Medline: 26193438]
53. Staiano AE, Calvert SL. The promise of exergames as tools to measure physical health. Entertain Comput 2011 Jan 01;2(1):17-21 [FREE Full text] [doi: 10.1016/j.entcom.2011.03.008] [Medline: 23378860]

Abbreviations

BMI: body mass index
EE: energy expenditure
INSPQ: Institut national de santé publique du Québec
IPAQ: International Physical Activity Questionnaire
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SDT: self-determination theory

Edited by G Eysenbach; submitted 08.01.19; peer-reviewed by A Staiano, L Gutierrez-Puertas; comments to author 03.02.19; revised version received 31.03.19; accepted 03.04.19; published 31.07.19.

Please cite as:
O'Loughlin EK, Barnett TA, McGrath JJ, Consalvo M, Kakinami L
Factors Associated with Sustained Exergaming: Longitudinal Investigation
JMIR Serious Games 2019;7(3):e13335
URL: https://games.jmir.org/2019/3/e13335/
doi:10.2196/13335
PMID:31368440

Original Paper

Exergaming Improves Executive Functions in Patients With Metabolic Syndrome: Randomized Controlled Trial

Shanshan Wu[1], PhD; Eun-Ah Jo[2], PhD; Hongqing Ji[1], PhD; Kyung-Hee Kim[3], MD; Jung-Jun Park[1], PhD; Bo Hyun Kim[4], MD, PhD; Kyoung Im Cho[5], MD, PhD

[1]Division of Sport Science, Pusan National University, Busan, Republic of Korea
[2]Department of Internal Medicine, Kosin University College of Medicine, Busan, Republic of Korea
[3]Department of Cardiology, Sejong General Hospital, Bucheon, Republic of Korea
[4]Department of Internal Medicine, Pusan National University Hospital and Biomedical Research Institute, Busan, Republic of Korea
[5]Division of Cardiology, Department of Internal Medicine, Kosin University College of Medicine, Busan, Republic of Korea

Corresponding Author:
Kyoung Im Cho, MD, PhD
Division of Cardiology
Department of Internal Medicine
Kosin University College of Medicine
34 Amnam-Dong, Seo-Ku
Busan,
Republic of Korea
Phone: 82 010 4548 6041
Fax: 82 51 990 3005
Email: kyoungim74@gmail.com

Abstract

Background: Recent studies indicate that participation in exercise-related games can improve executive function, attention processing, and visuospatial skills.

Objective: The aim of this study was to investigate whether exercise via exergaming (EXG) can improve executive function in patients with metabolic syndrome (MetS).

Methods: A total of 22 MetS patients were recruited and randomly assigned to an EXG group or a treadmill exercise (TE) group. The reaction time (RT) and electrophysiological signals from the frontal (Fz), central (Cz), and parietal (Pz) cortices were collected during a Stroop task after 12 weeks of exercise.

Results: During the Stroop congruence (facilitation) judgment task, both the EXG and TE groups showed significantly faster RT after 12 weeks of exercise training. For N200 amplitude, the EXG group demonstrated significantly increased electrophysiological signals from the Fz and Cz cortices. These changes were significantly larger in the EXG group than in the TE group. Separately, for the P300 amplitude, the EXG groups presented significantly increased electrophysiological signals from the Fz, Cz, and Pz cortices, whereas the TE group showed significantly increased electrophysiological signals from the Cz and Pz cortices only. During the Stroop incongruence (interference) judgment task, both the EXG and TE groups showed significantly faster RT. For P300 amplitude, the EXG group had significantly increased electrophysiological signals from the Fz and Cz cortices only, whereas the TE group had significantly increased electrophysiological signals from the Fz, Cz, and Pz cortices.

Conclusions: EXG improves executive function in patients with MetS as much as normal aerobic exercise does. In particular, a unique benefit of EXG beyond increased aerobic capacity is the improved selective attention among cognitive functions. Thus, EXG could be recommended to someone who needs to improve their brain responses of concentration and judgment as well as physical fitness.

Trial Registration: ClinicalTrials.gov NCT04015583; https://clinicaltrials.gov/ct2/show/NCT04015583

(JMIR Serious Games 2019;7(3):e13575) doi:10.2196/13575

KEYWORDS
exercise; executive functions; event-related potential; games; metabolic syndrome

Introduction

Background

In recent years, the relationship between cognitive function and metabolic syndrome (MetS) has been widely studied [1,2]. MetS has been found to be associated with a decline in areas related to executive function [3,4] because of multiple risk factors, including hypertension, dyslipidemia, impaired glucose homeostasis, and abdominal obesity. Executive functions include basic cognitive processes such as attentional control, cognitive inhibition, inhibitory control, working memory, and cognitive flexibility [5].

Research on cognitive neuroscience employs Stroop tasks to measure selective attention capacity and skills as well as process speed ability to elucidate the nature of executive functions [6]. Electroencephalographic (EEG) activity using event-related positioning technology has been widely used to measure selective attention capacity and skills, and evaluations of behavioral performance such as reaction time (RT) are commonly used to determine processing speed ability [7,8]. N200 negativity (200-350 milliseconds [ms] poststimulus) is an event-related potential (ERP) indicating the attentional capacity that is usually induced before motion response control and is related to the cognitive processes of stimulus recognition and differentiation [9]. P300 positivity (300-600 ms poststimulus) is another ERP that reflects memory-related neural processing and is involved in categorizing incoming information and updating the context of the working memory (eg, encoding, rehearsal, recognition, and retrieval) [10].

It is well-known that aerobic exercise training provides various beneficial clinical outcomes in metabolic disease patients [11,12]. Its effects on cognitive function, especially executive function, have also been investigated [13]. Furthermore, recent studies reported that both aerobic and resistance exercise training facilitate overall electrophysiological effects (eg, increased P300 amplitudes) and behavior changes (eg, faster RT) in otherwise healthy elderly people [14,15]. In addition, aerobic exercise has been suggested to improve cognitive processes in cortical cognitive control (P300 amplitude) in studies involving chronic stroke patients [16].

Exergaming

Recently, exergaming (EXG, a combination of *exercise* and *gaming*) has attracted much attention as a novel exercise method in terms of improving cognitive function because it utilizes video games that require body movements while simultaneously presenting the user with a cognitively challenging environment [17]. Along with its popular usage for leisure and entertainment, there is a growing interest in the application of EXG to improve clinical outcomes. Recent studies using EXG showed its beneficial effects on cognitive and dual-task functions, which reduced falls in older adults [18] as well as cardiovascular disease risks, such as body fat, serum adipokine levels, and lipid profiles [19]. EXG also promoted improved executive functions and cognitive processing speed in both elderly people and children [20,21]. This growing evidence suggests that EXG offers the benefit of improving both cognitive and physical functions.

Although many previous studies have reported improvements in cognitive function following EXG, it is not clear whether this benefit is due to an exercise effect or video game effect. In addition, all of these studies measured RT instead of ERP using EEG, which limits investigators seeking to illuminate brain activities. Considering that EEG can measure electrical activities in various cortex areas in the brain, it is necessary to investigate ERP using EEG to evaluate executive function. Therefore, we examined the benefits of EXG in comparison with normal exercise and investigated executive function by measuring RT as well as N200 and P300 in 3 cortex areas via Stroop tasks applied in patients with MetS.

Methods

Participants

A total of 22 MetS male and female patients aged between 50 and 80 years participated in this study. MetS was defined according to the modified National Cholesterol Education Program Adult Treatment Panel III definition for South Asians. Briefly, individuals with 3 or more of the following criteria were defined as having MetS: central obesity (waist circumference ≥90 cm for men or ≥85 cm for women); fasting plasma glucose ≥100 mg/dL or current treatment for diabetes mellitus; systolic blood pressure ≥130 mmHg, diastolic blood pressure ≥85 mmHg, or current treatment for hypertension; serum triglyceride level ≥150 mg/dL; and low high-density lipoprotein cholesterol (<40 mg/dL for men or <50 mg/dL for women) [22]. Subjects were asked not to exercise for 24 hours before the experiment. They were also instructed to eat usual meals and to finish meals at least four hours before the experiment, while avoiding alcohol for a day before the experiment and caffeine during the 4 hours before the experiment. All subjects were required to complete a written informed consent form approved by the Institutional Review Board of Kosin University College of Medicine.

The sample size was calculated using a sample size calculation software program (G*Power version 3.1.9.2 for Windows), with an effect size of 0.484, statistical power of .80, and statistical level of significance of .05. The effect size was calculated from previous studies [13,14]. As a result, the sample size for each group was established at 8 patients, so we decided to recruit 11 patients for each group in consideration of a potential 30% (3/11) dropout rate.

Exercise Training Interventions

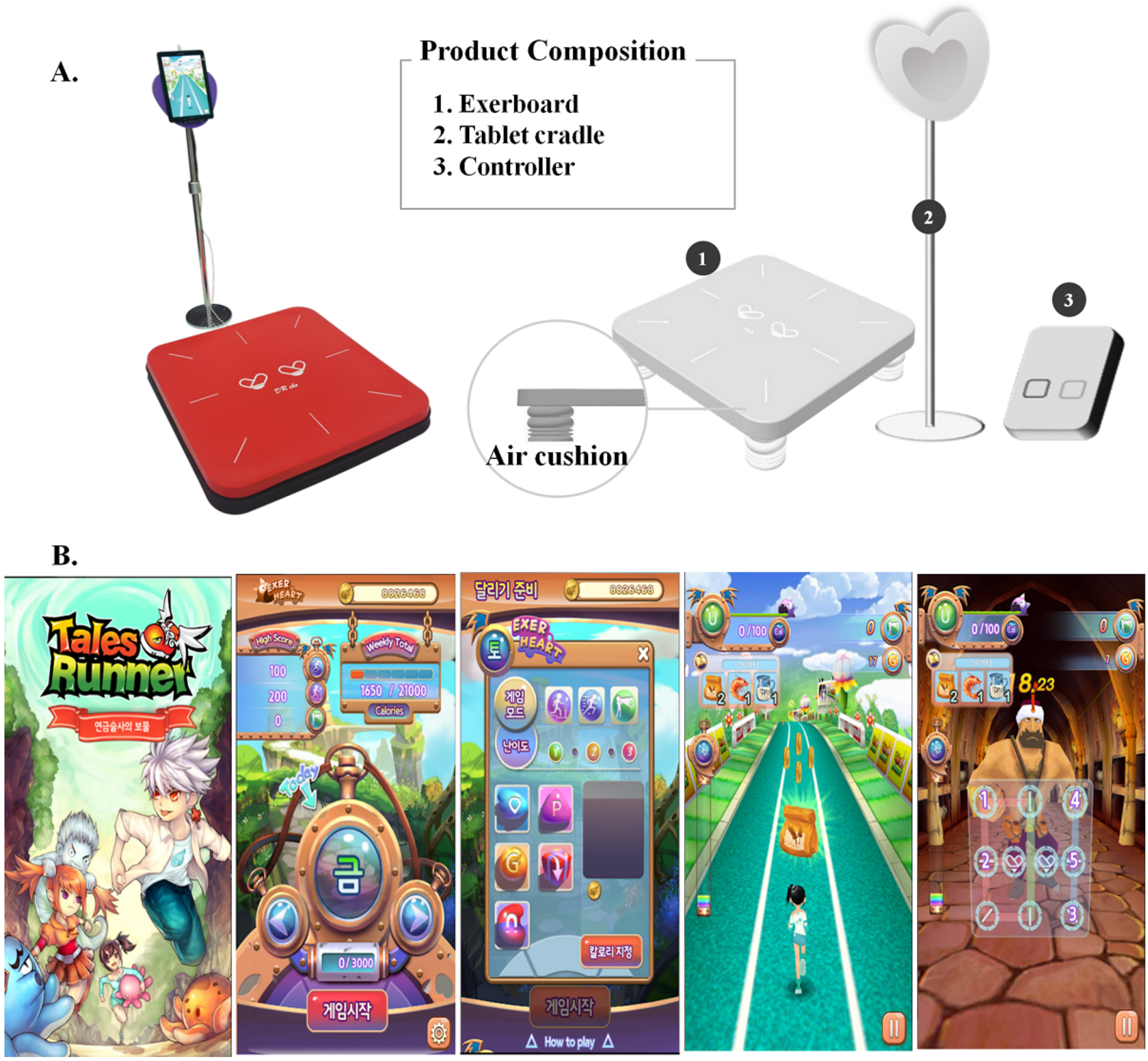
Exercise training was conducted at Kosin University Gospel Hospital. Each participant was instructed to immediately inform the study supervisor if he or she experienced any unusual symptoms during exercise training and to consult a physician if needed. Subjects were excluded from the final analysis if they did not perform more than 80% of the exercise sessions.

All subjects were randomly stratified into either an EXG group or a treadmill exercise (TE) group. Subjects underwent 2 weeks of adaptation and then carried out 12 weeks of exercise training for 60 min per day, 3 days per week, at 60% to 80% of their heart rate reserve (HRR). Each exercise session consisted of 10 min of warm-up, 40 min of main exercise, and 10 min of cooldown.

The EXG group performed exercise using the Exerheart equipment (D&J Humancare) that is composed of a running and jumping mat (730 width × 730 depth × 130 height) and a tablet personal computer placed on a stand (which could be adjusted to any height between 70 and 155 cm; Figure 1). Exerheart is an EXG system developed for in situ running along with the video game called *Alchemist's Treasure* (D&J Humancare; Figure 1). To play this game, the subject has to run or jump on a spot on the mat to move a virtual avatar on the screen of the tablet computer to the front, back, left, and right along with music (Multimedia Appendix 1). The subject can control the speed of their avatar's movement via their running or jumping speed on the mat. Participants in the TE group performed exercise using commercial treadmills (MOTUS). Each subject walked or ran on their treadmill at a comfortable speed.

For both the EXG and TE groups, all subjects' heart rates (HRs) during exercise were monitored using HR monitors (Polar RS400sd) to confirm that the value was within the target HR range. The Karvonen formula [23] was used to calculate the HRR (estimated maximal HR−resting HR) and the target HR during exercise (HRR × given percentage of training intensity + resting HR).

Figure 1. (A) The exergaming group performed exercise using Exerheart devices with permission from D&J Humancare, who is the copyright holder of Exerheart. (B) Features of the video game "Alchemist's Treasure.".



Stroop Test

To assess executive function, a computer-based version of the Stroop task was administered using the Telescan software (LAXTHA Inc). During the task, subjects were presented with a color word appearing in the same color on congruent trials (eg, *blue* printed in blue) and in a different color on incongruent trials (eg, *blue* printed in green) [24]. To provide similar visual content, blue, green, and yellow colors were chosen as stimuli.

Subjects performed the Stroop task twice, before and after exercise training. Subjects sat 1 m from the screen and when the color words appeared on the screen, they clicked the left keyboard for the congruent test and the right keyboard for the incongruent test. Subjects were instructed to respond as quickly and accurately as possible. The rate of measurement targeted for 50%. Each color word (vertical viewing angle: 2 degrees) was presented for 200 ms and a response was accepted within 1500 ms. The interstimulus interval varied randomly between 1500 and 2500 ms.

Electroencephalographic Measurements

EEG activity was recorded during the modified Stroop task by using a computerized polygraph system (type A: a total of 31 channels Poly G-As, LAXTHA). Silver chloride electrodes (LAXTHA) were placed on the frontal (Fz), central (Cz), and parietal (Pz) cortex areas, according to the international 10-20 system. Midline locations were referenced to link earlobe electrodes. Horizontal and vertical electrooculograms were monitored by electrodes placed above and below the left eye and at the outer canthus of both eyes, respectively. The impedance of all electrodes was maintained below 10 kΩ. The bandpass filter of the amplifier was 0.1 to 100 Hz, the sampling rate was 1000 Hz, and a notch filter was established at 60 Hz.

The N200 component was defined as the largest positive peak occurring between 200 and 350 ms poststimulus, whereas the P300 component was defined as the largest positive peak occurring between 300 and 600 ms poststimulus [7]. N200 and P300 amplitudes were measured as the differences between the mean prestimulus baseline and maximum peak amplitude. Telescan's built-in high-pass infinite impulse response filter was used for filtering. Waveforms were digitally smoothed with a low-pass filter using a half power cutoff of 10 Hz before analysis.

Statistical Analysis

Owing to the small sample size of this study, we used nonparametric statistics for data analysis. We used the Wilcoxon signed rank test to examine the changes of each dependent variable after the intervention within each group. The Mann-Whitney U test was employed to compare the delta values between training groups (Δ - EXG group vs Δ - TE group). The effect size of partial eta squared (η2) was reported for significant effects, where the alpha level for all of the tests was set at .05. Data were expressed as mean (SD). All statistical tests were processed using the Statistical Package for the Social Sciences version 24 software program (IBM Corp).

Results

Demographic and physical characteristics for all subjects are provided in Table 1. There were no significant group differences noted during baseline measurements.

Reaction Time

The changes in congruent RT after 12 weeks of exercise training were not significantly different between the EXG and TE groups (Table 2): both showed significantly shortened congruent RT (Figure 2). Similarly, the changes in incongruent RT after 12 weeks of exercise training were not significantly different between the EXG and TE groups, in that both showed significantly shortened incongruent RT (Figure 2).

Table 1. Baseline characteristics of study participants.

Factor	Group		*P* value
	Exergaming (n=11), mean (SD)	Treadmill exercise (n=11), mean (SD)	
Age (years)	64 (10)	60 (7)	.30
Height (cm)	154.24 (5.73)	161.66 (7.1)	.01
Weight (kg)	69.38 (10.54)	71.47 (11.62)	.66
Body mass index (kg/m^2)	29.07 (3.3)	27.31 (3.52)	.24
Waist circumference (cm)	97.36 (10.95)	93.6 (8.78)	.40
Glucose (mg/dl)	123.55 (25.68)	112.36 (29.37)	.35
High-density lipoprotein cholesterol (mmol/L)	46.32 (8.91)	50.59 (10.52)	.32
Low - density lipoprotein cholesterol (mmol/L)	66.19 (22.8)	78.38 (19.52)	.19
Total cholesterol (mmol/L)	134.81 (25.52)	146.67 (12.04)	.18
Triglycerides (mmol/L)	136.82 (78.88)	145.73 (142.74)	.86
Systolic blood pressure (mmHg)	127.09 (16.86)	128.09 (17.39)	.89
Diastolic blood pressure (mmHg)	75.55 (9.43)	77.73 (11.99)	.64

Table 2. Comparison of Stroop task congruent and incongruent reaction times of the exergame group and treadmill exercise group.

Effects, group	Pre, mean (SD)	Post, mean (SD)	Wilcoxon signed rank test[a], *P* value	Chi-square value	Mann-Whitney U test[b], *P* value
Congruent					**.28**
EXG[c] group	1265.91 (383.15)	1012.09 (221.64)	.01	0.053	—[d]
TE[e] group	1124.18 (161.21)	957.82 (138.05)	.01	—	—
Incongruent					**.67**
EXG group	1299.09 (367.48)	984.73 (204.81)	.01	0.008	—
TE group	1171.36 (163.26)	974.55 (120.97)	.01	—	—

[a]Comparison of pre versus post within group.

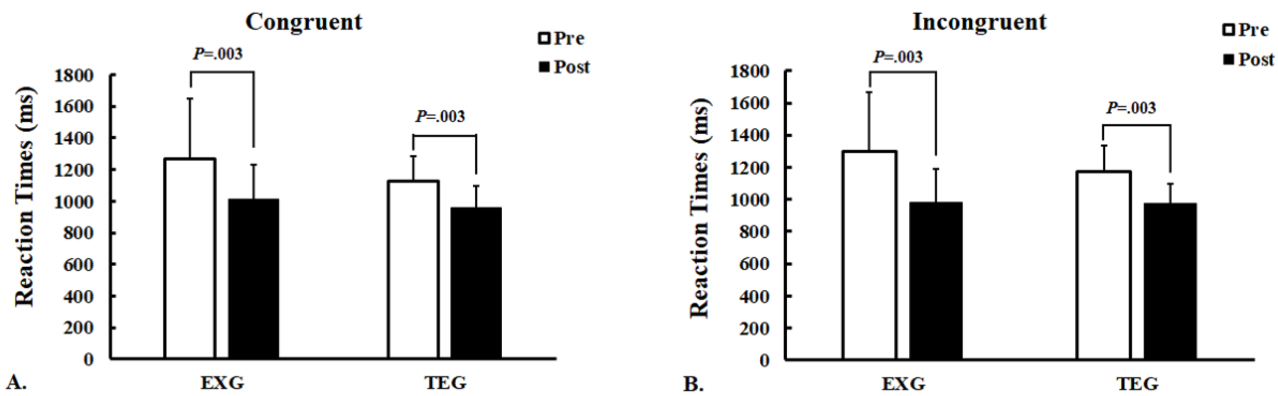
[b]Comparison of the delta values between EXG and TE group.

[c]EXG: exergaming.

[d]Not applicable.

[e]TE: treadmill exercise.

Figure 2. Mean reaction time during the Stroop task in the exergaming and treadmill exercise groups before and after the exercise intervention. (A) Congruent; (B) incongruent. EXG: exergaming; TEG: treadmill exercise group; ms: milliseconds. **Significant difference at P<.01 (Wilcoxon signed-rank test).



Event-Related Potential Data

N200 Amplitude

According to the results in Table 3, after 12 weeks of exercise training, the increases in congruent N200 amplitude on Cz and Pz cortices in the EXG group were significantly greater than in the TE group, but this was not true with regard to Fz. In addition, EXG significantly increased congruent N200 amplitude on Cz, but not for Fz and Pz. On the contrary, the TE group showed no significant changes in congruent N200 amplitude for Fz, Cz, or Pz (Figure 3).

The changes in incongruent N200 amplitude for Fz, Cz, and Pz after 12 weeks of exercise training were not significantly different between the EXG and TE groups (Table 3). Interestingly, EXG did not significantly change the incongruent N200 amplitude for Fz, Cz, or Pz, whereas TE also did not significantly change the incongruent N200 amplitude for Fz, Cz, or Pz (Figure 3). The waveforms of congruent and incongruent N200 amplitudes for Fz, Cz, and Pz for the EXG and TE groups before and after exercise are shown in Figure 3.

Table 3. Comparison of Stroop task congruent and incongruent N200/P300 amplitudes of the exergame group and treadmill exercise group.

Components, effects, and group	Pre, mean (SD)	Post, mean (SD)	Wilcoxon signed rank test[a], *P* value	η2 [b]	Mann-Whitney U test[c], *P* value
N200 amplitude					
Congruent					
Fz[d]					
EXG[e] group	−1.3 (1.95)	−4.1 (3.02)	.09	0.138	.09
TE[f] group	−0.98 (3.14)	−0.41 (3.82)	.33	—[g]	—
Cz[h]					
EXG group	−1.59 (2.54)	−5.13 (2.94)	.03	0.291	.01
TE group	−1.61 (3.89)	−0.97 (4.59)	.42	—	—
Pz[i]					
EXG group	−1.58 (1.89)	−4.06 (2.89)	.06	0.207	.03
TE group	−1.32 (3.62)	−0.89 (3.99)	.53	—	—
Incongruent					
Fz					
EXG group	−2.85 (2.3)	−2.78 (3.02)	.48	0.041	.37
TE group	−1.37 (3.3)	−0.73 (3.18)	.42	—	—
Cz					
EXG group	−3.23 (2.69)	−3.93 (2.92)	.29	0.099	.15
TE group	−1.9 (3.82)	−1.1 (3.65)	.21	—	—
Pz					
EXG group	−2.86 (2.45)	−3.07 (2.62)	.72	0.008	.70
TE group	−1.4 (3.66)	−1.01 (3.52)	.48	—	—
P300 amplitude					
Congruent					
Fz					
EXG group	2.3 (1.94)	7.12 (5.73)	.01	0.008	.70
TE group	3.21 (1.95)	4.82 (3.78)	.01	—	—
Cz					
EXG group	1.92 (1.63)	6.49 (5.28)	.01	0.006	.75
TE group	2.36 (0.93)	5.08 (3.03)	.01	—	—
Pz					
EXG group	1.44 (1.69)	5.2 (5.88)	.01	0.099	.15
TE group	1.74 (1.26)	4.87 (3.64)	.33	—	—
Incongruent					
Fz					
EXG group	2.26 (3.14)	4.48 (3.27)	.09	0.018	.56
TE group	1.93 (2.26)	3.85 (3.04)	.01	—	—
Cz					
EXG group	2.03 (2.8)	4.38 (2.81)	.02	0.002	.85
TE group	1.59 (1.63)	3.74 (2.99)	.02	—	—

Components, effects, and group	Pre, mean (SD)	Post, mean (SD)	Wilcoxon signed rank test[a], *P* value	η2 [b]	Mann-Whitney U test[c], *P* value
Pz					
EXG group	1.58 (2.44)	3.2 (2.73)	.02	0.000	.95
TE group	1.05 (0.92)	3.33 (2.59)	.03	—	—

[a]Comparison of pre versus post within group.

[b]η2: chi-square test value.

[c]Comparison of the delta values between EXG and TE group.

[d]Fz: frontal cortex.

[e]EXG: exergame.

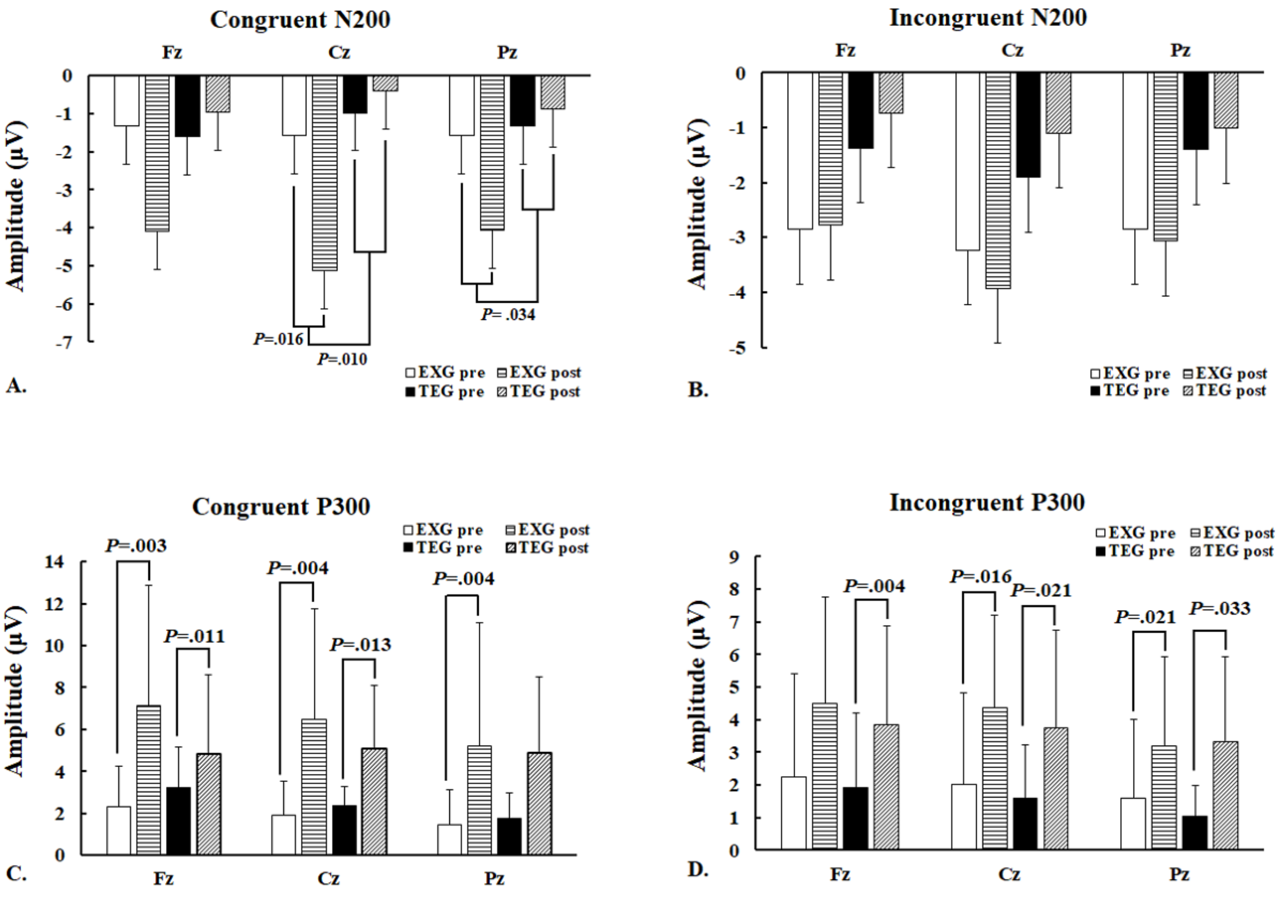
[f]TE: treadmill exercise.

[g]Not applicable.

[h]Cz: central cortex.

[i]Pz: parietal cortex.

Figure 3. N200 and P300 amplitudes (mean [SE]) on 3 electrodes during the Stroop task in the exergaming and treadmill exercise groups before and after 12 weeks of exercise training. (A) Congruent N200 amplitudes; (B) incongruent N200 amplitudes; (C) congruent P300 amplitudes; (D) incongruent P300 amplitudes. Fz: frontal cortex; Cz: central cortex; Pz: parietal cortex; EXG: exergaming group; TEG: treadmill exercise group.

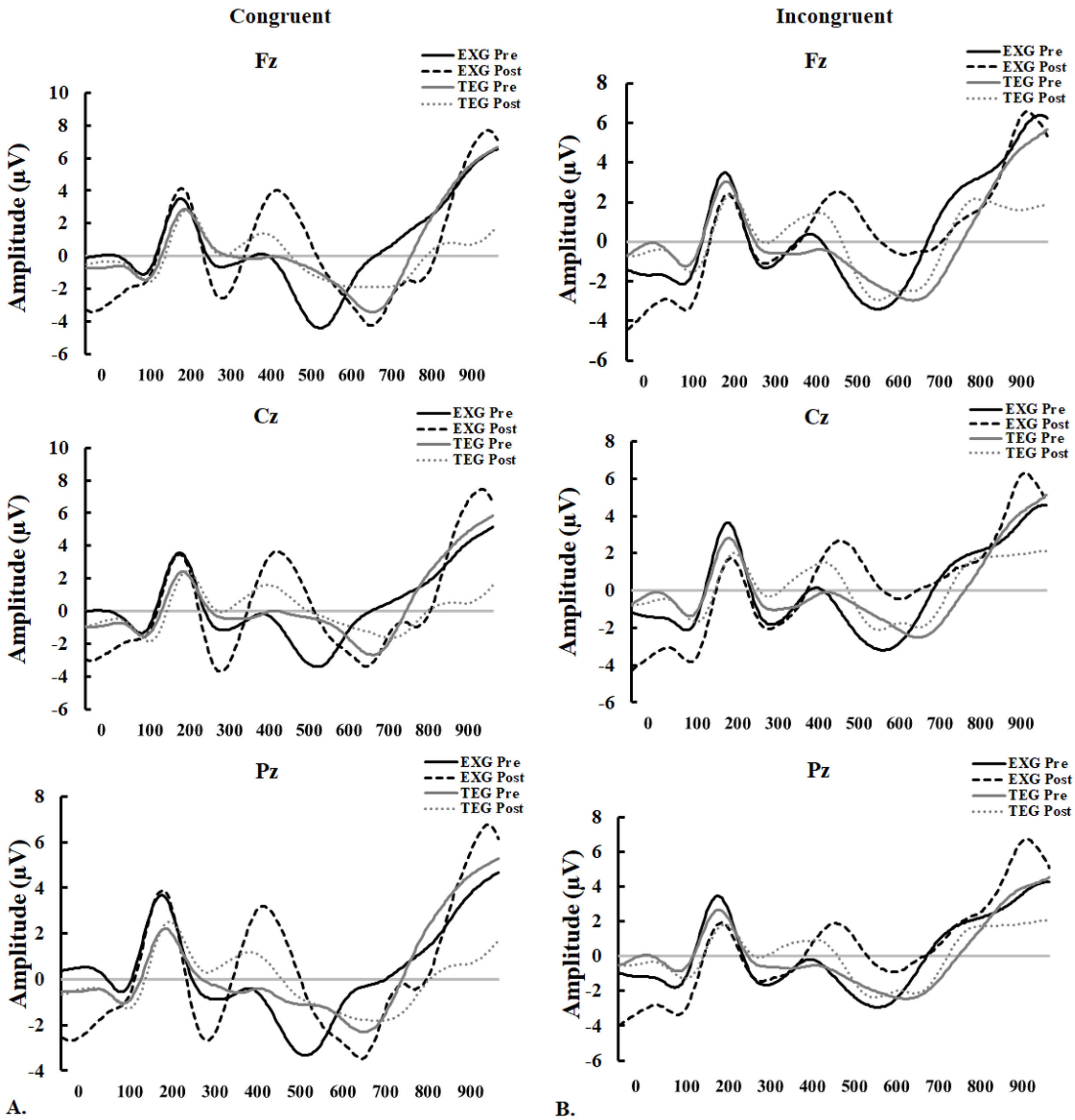


P300 Amplitude

Table 3 shows that the results of the Mann-Whitney U test for changes in the congruent P300 amplitude for Fz, Cz, and Pz after 12 weeks of exercise training were not significantly different between the EXG and TE groups. EXG significantly increased the congruent P300 amplitude for Fz, Cz, and Pz. However, TE significantly increased the congruent P300 amplitude only for Fz and Cz (Figure 3).

There were no significant differences in the changes in the incongruent P300 amplitude for Fz, Cz, and Pz between the EXG and TE groups after 12 weeks of exercise training (Table 3). EXG significantly increased the incongruent P300 amplitude for Cz and Pz, but not for Fz. On the contrary, TE significantly increased the incongruent P300 amplitude for Fz, Cz, and Pz (Figure 3). The waveforms of congruent and incongruent P300 amplitudes for Fz, Cz, and Pz for the EXG and TE groups before and after exercise are shown in Figure 4.

Figure 4. Average event-related potential waveforms of electrodes for mean N200 and P300 amplitudes during the Stroop task in the exergaming and treadmill exercise groups before and after 12 weeks of the exercise training. (A) Congruent; (B) incongruent. Fz: frontal cortex; Cz: central cortex; Pz: parietal cortex; EXG: exergaming group; TEG: treadmill exercise group; ms: milliseconds.



Discussion

Principal Findings

This study was the first to investigate the benefits of EXG in comparison with normal exercise on the behavioral performance and executive function of patients with MetS. We found that 12 weeks of both EXG and TE training similarly and effectively improved behavioral performance and congruent and incongruent memory-related neural processing. However, only EXG training improved congruent selective attention, whereas neither EXG nor TE training affected incongruent selective attention. These results suggest similar overall effects for EXG and normal exercise on behavioral performance and executive function in patients with MetS but that EXG could be more effective than normal exercise for congruent selective attention.

This study showed that both 12 weeks of EXG and TE training effectively improved RT in MetS patients, and these changes were not different between the 2 training protocols. These results suggest that both EXG and normal exercise are able to improve behavioral performance, but EXG does not have more beneficial effects compared with normal exercise. In our previous study, we examined the performance of control tasks using a simple acute aerobic exercise and complex exercise [25]. The results indicated that participants did not have a difference in performance when participating in the simple exercise compared with very skilled complex exercises. In another study, contrary to our results, an acute single-bout study comparing the effects of standard normal exercise with EXG on attention performance in young adults found no significant improvement with EXG [7]. The results of this previous study suggest that, after 20 min of unskilled WiiFit training, the brain requires further information processing capabilities, which may be the source of control requirements and pressure increases, offsetting the potential benefits of the exercise component. However, compared with the WiiFit system, EXG with Exerheart involves running-based aerobic exercise on an air cushion board with game-based contents, such as adventures, racing, and quizzes, which continue to arouse the user's interest in exercise. So, with Exerheart, an individual is running constantly with changing visual stimuli; these repeated effects simultaneously increase physical activity and cognitive function via an interactive virtual reality engagement [20]. Therefore, through EXG and TE, the RT of Stroop task conditions of MetS patients could be shortened, which would promote basic information processing and the executive function of suppression control.

In this study, the congruent and incongruent P300 amplitudes were increased after 12 weeks of both EXG and TE training, with no difference seen between the 2 groups. These results indicate that both EXG and normal exercise improve memory-related neural processing, but the beneficial effects of EXG are more significant than normal exercise. In other words, participation in exercise, regardless of the exercise modality, induces an increase in working memory in executive function. However, our previous study showed that P300 amplitude increased during a control task following futsal relative to seated rest or TE, indicating that complex control of the brain stimulates the executive control network of the cortex [25]. It was found that as one's age increases, the P300 amplitude in the central (Cz) region decreases and the scalp distribution of the P300 amplitude is transferred to the frontal region [26]. Pontifex et al [27] examined P300 components and found that older adults with high cardiorespiratory fitness only exhibited greater P300 amplitudes, whereas Tsai et al [28] revealed that different exercise types have greater P300 amplitudes for older individuals. However, as they pointed out, regardless of the type of older people participating in sports, physical exercise is a lifestyle factor that is crucial to preventing age-related biological degeneration in the frontal-to-parietal areas, thus delaying the cognitive declines associated with later life. Considering that EXG is a kind of aerobic exercise, it stands to reason that EXG could improve not only cardiovascular health but also cognitive plasticity, thereby improving categorization of the incoming information and updating the context of working memory in MetS patients.

We found that neither EXG nor TE training affected the incongruent N200 amplitude. However, the consistent N200 amplitude was increased by EXG training. These findings suggest that, although neither EXG nor normal exercise affected incongruent selective attention, EXG improved congruent selective attention, which suggested that EXG has a more beneficial effect on congruent selective attention compared with normal exercise. The results of many studies on the relationship between exercise and the N200 amplitude indicate that exercise has no significant effect on the N200 amplitude [28,29]. Pontifex et al showed that general decreases in N200 amplitudes across scalp sites were observed during exercise relative to rest [29]. The N200 component plays a key role in the anterior cingulate cortex (ACC), which is part of the potential prefrontal cortex and regulates dopaminergic neurons in cognitive functions, such as working memory, attention, and decision making [30-32]. Therefore, the reduction of N200 amplitude caused by normal aerobic exercise severely limits ACC activity [29]. In light of our N200 amplitude findings, these results suggest that EXG better regulates the activity of ACC in the prefrontal cortex than does aerobic exercise, thereby effectively increasing consistent selective attention.

Recent studies suggest that combining motor and cognitive demands during exercising can improve cognitive function more so than training these domains separately [17,33]. In addition, cognitive video game training can have beneficial effects on memory, attention, and RT in older adults [34,35]. In previous studies, when participants consistently performed exercises in a virtual environment, an increase in the N200 amplitude positively promoted decision making (frontal and central) and visual perception (occipital) [36]. Therefore, EXG positively promotes visual perceptual stimulation in the virtual environment to enhance the selective attention activity associated with the cerebral cortex, thereby strongly promoting executive function. Exercise and video games can each improve brain structure and function [37-40]; thus, their combination can have a complementary effect on brain stimulation and protection.

Our study provides evidence that EXG improves RT and incongruent memory-related neural processing in MetS patients as much as normal aerobic exercise does. In addition, EXG improves congruent selective attention, which was not changed by normal aerobic exercise. Therefore, EXG could provide an innovative way to enjoy aerobic exercise compared with repetitive, conventional exercises.

Limitations

Although this study found significant results, there were some limitations: (1) The sample size in this study was relatively small (2) The age range was relatively large at 50 to 80 years. Considering that, with age, response time and brain activity become slower, we cannot rule out the possibility that age will affect the performance of executive function. However, the mean age was similar in both groups, so this possibility might be low in this study; (3) Finally, the intensity of Exerheart use while playing the *Alchemist's Treasure* game was unable to be controlled in a standard fashion.

Conclusions

The results of this study suggest that EXG enhances brain responses to concentration and judgment, resulting in increased behavioral response among MetS patients comparable with the impact of normal aerobic exercise. Furthermore, the unique advantage of EXG is that it improves selective attention among cognitive functions, unlike normal aerobic exercise. Therefore, EXG could be recommended to some patients who need to improve executive function as well as physical fitness.

Acknowledgments

The authors would like to thank the study participants for their time and effort given while participating in this study. The authors declare that the results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Conflicts of Interest

None declared.

Editorial note: This randomized study was only retrospectively registered, explained by authors with "As the intervention was not related to drug trials, we did not think it was essential to register our study. No other reason but just our mistake." The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low and the study was considered formative, guiding the development of the application. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1

Participants played the video game with Exerheart.

[MP4 File (MP4 Video), 1MB - games_v7i3e13575_app1.mp4]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - games_v7i3e13575_app2.pdf]

References

1. González HM, Tarraf W, Vásquez P, Sanderlin AH, Rosenberg NI, Davis S, et al. Metabolic syndrome and neurocognition among diverse middle-aged and older Hispanics/Latinos: HCHS/SOL results. Diabetes Care 2018;41(7):1501-1509. [doi: 10.2337/dc17-1896] [Medline: 29716895]
2. Shigaeff N, Amaro E, Franco FG, Jacinto AF, Chiochetta G, Cendoroglo MS, et al. Functional magnetic resonance imaging response as an early biomarker of cognitive decline in elderly patients with metabolic syndrome. Arch Gerontol Geriatr 2017 Nov;73:1-7. [doi: 10.1016/j.archger.2017.07.002] [Medline: 28711765]
3. Falkowski J, Atchison T, Debutte-Smith M, Weiner MF, O'Bryant S. Executive functioning and the metabolic syndrome: a project FRONTIER study. Arch Clin Neuropsychol 2014 Feb;29(1):47-53 [FREE Full text] [doi: 10.1093/arclin/act078] [Medline: 24152591]
4. Bokura H, Nagai A, Oguro H, Kobayashi S, Yamaguchi S. The association of metabolic syndrome with executive dysfunction independent of subclinical ischemic brain lesions in Japanese adults. Dement Geriatr Cogn Disord 2010;30(6):479-485 [FREE Full text] [doi: 10.1159/000322057] [Medline: 21252541]
5. Royall DR, Lauterbach EC, Cummings JL, Reeve A, Rummans TA, Kaufer DI, et al. Executive control function: a review of its promise and challenges for clinical research. A report from the committee on research of the American neuropsychiatric association. J Neuropsychiatry Clin Neurosci 2002;14(4):377-405. [doi: 10.1176/jnp.14.4.377] [Medline: 12426407]
6. Badzakova-Trajkov G, Barnett KJ, Waldie KE, Kirk IJ. An ERP investigation of the Stroop task: the role of the cingulate in attentional allocation and conflict resolution. Brain Res 2009 Feb 9;1253:139-148. [doi: 10.1016/j.brainres.2008.11.069] [Medline: 19084509]
7. Patel SH, Azzam PN. Characterization of N200 and P300: selected studies of the event-related potential. Int J Med Sci 2005;2(4):147-154 [FREE Full text] [doi: 10.7150/ijms.2.147] [Medline: 16239953]
8. Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. Nature 1996;381(6582):520-522. [doi: 10.1038/381520a0] [Medline: 8632824]
9. Dennis TA, Chen CC. Neurophysiological mechanisms in the emotional modulation of attention: the interplay between threat sensitivity and attentional control. Biol Psychol 2007 Sep;76(1-2):1-10 [FREE Full text] [doi: 10.1016/j.biopsycho.2007.05.001] [Medline: 17582673]
10. Polich J. Updating P300: an integrative theory of P3a and P3b. Clin Neurophysiol 2007 Oct;118(10):2128-2148 [FREE Full text] [doi: 10.1016/j.clinph.2007.04.019] [Medline: 17573239]
11. Wewege MA, Thom JM, Rye KA, Parmenter BJ. Aerobic, resistance or combined training: a systematic review and meta-analysis of exercise to reduce cardiovascular risk in adults with metabolic syndrome. Atherosclerosis 2018 Jul;274:162-171. [doi: 10.1016/j.atherosclerosis.2018.05.002] [Medline: 29783064]

12. Lemes Í, Turi-Lynch BC, Cavero-Redondo I, Linares SN, Monteiro HL. Aerobic training reduces blood pressure and waist circumference and increases HDL-c in metabolic syndrome: a systematic review and meta-analysis of randomized controlled trials. J Am Soc Hypertens 2018 Aug;12(8):580-588. [doi: 10.1016/j.jash.2018.06.007] [Medline: 29945775]
13. Fong DY, Chi LK, Li F, Chang YK. The benefits of endurance exercise and Tai Chi Chuan for the task-switching aspect of executive function in older adults: an ERP study. Front Aging Neurosci 2014 Oct 28;6:295. [doi: 10.3389/fnagi.2014.00295] [Medline: 25389403]
14. Tsai CL, Pan CY, Chen FC, Tseng YT. Open- and closed-skill exercise interventions produce different neurocognitive effects on executive functions in the elderly: a 6-month randomized, controlled trial. Front Aging Neurosci 2017 Sep 12;9:294. [doi: 10.3389/fnagi.2017.00294] [Medline: 28959200]
15. Tsai CL, Wang WL. Exercise-mode-related changes in task-switching performance in the elderly. Front Behav Neurosci 2015;9:56 [FREE Full text] [doi: 10.3389/fnbeh.2015.00056] [Medline: 25798097]
16. Swatridge K, Regan K, Staines WR, Roy E, Middleton LE. The acute effects of aerobic exercise on cognitive control among people with chronic stroke. J Stroke Cerebrovasc Dis 2017 Dec;26(12):2742-2748. [doi: 10.1016/j.jstrokecerebrovasdis.2017.06.050] [Medline: 28774794]
17. Limperos AM, Schmierbach M. Understanding the relationship between exergame play experiences, enjoyment, and intentions for continued play. Games Health J 2016 Apr;5(2):100-107. [doi: 10.1089/g4h.2015.0042] [Medline: 26871873]
18. Fu AS, Gao KL, Tung AK, Tsang WW, Kwan MM. Effectiveness of exergaming training in reducing risk and incidence of falls in frail older adults with a history of falls. Arch Phys Med Rehabil 2015 Dec;96(12):2096-2102. [doi: 10.1016/j.apmr.2015.08.427] [Medline: 26360975]
19. Amorim MG, de Oliveira MD, Soares DS, da Silva Borges L, Dermargos A, Hatanaka E. Effects of exergaming on cardiovascular risk factors and adipokine levels in women. J Physiol Sci 2018;68(5):671-678 [FREE Full text] [doi: 10.1007/s12576-017-0581-5] [Medline: 29189992]
20. Dimitrova J, Hogan M, Khader P, O'Hora D, Kilmartin L, Walsh JC, et al. Comparing the effects of an acute bout of physical exercise with an acute bout of interactive mental and physical exercise on electrophysiology and executive functioning in younger and older adults. Aging Clin Exp Res 2017 Oct;29(5):959-967. [doi: 10.1007/s40520-016-0683-6] [Medline: 27866346]
21. Benzing V, Schmidt M. Cognitively and physically demanding exergaming to improve executive functions of children with attention deficit hyperactivity disorder: a randomised clinical trial. BMC Pediatr 2017 Jan 10;17(1):8. [doi: 10.1186/s12887-016-0757-9] [Medline: 28068954]
22. Pandit K, Goswami S, Ghosh S, Mukhopadhyay P, Chowdhury S. Metabolic syndrome in south Asians. Indian J Endocrinol Metab 2012;16(1):44-55 [FREE Full text] [doi: 10.4103/2230-8210.91187] [Medline: 22276252]
23. Karoven MJ, Kentala E, Mustala O. The effects of training on heart rate; a longitudinal study. Ann Med Exp Biol Fenn 1957;35(3):307-315. [Medline: 13470504]
24. Ilan AB, Polich J. P300 and response time from a manual Stroop task. Clin Neurophysiol 1999 Feb;110(2):367-373. [doi: 10.1016/S0168-5597(98)00053-7] [Medline: 10210626]
25. Won J, Wu S, Ji H, Smith JC, Park J. Executive function and the P300 after treadmill exercise and futsal in college soccer players. Sports (Basel) 2017;5(4):73 [FREE Full text] [doi: 10.3390/sports5040073] [Medline: 29910433]
26. Picton TW, Stuss DT, Champagne SC, Nelson RF. The effects of age on human event-related potentials. Psychophysiology 1984 May;21(3):312-325. [doi: 10.1111/j.1469-8986.1984.tb02941.x] [Medline: 6739673]
27. Pontifex MB, Hillman CH, Polich J. Age, physical fitness, and attention: P3a and P3b. Psychophysiology 2009 Mar;46(2):379-387 [FREE Full text] [doi: 10.1111/j.1469-8986.2008.00782.x] [Medline: 19170947]
28. Tsai CL, Wang CH, Pan CY, Chen FC, Huang SY, Tseng YT. The effects of different exercise types on visuospatial attention in the elderly. Psychol Sport Exerc 2016 Sep;26:130-138. [doi: 10.1016/j.psychsport.2016.06.013]
29. Pontifex MB, Hillman CH. Neuroelectric and behavioral indices of interference control during acute cycling. Clin Neurophysiol 2007 Mar;118(3):570-580. [doi: 10.1016/j.clinph.2006.09.029] [Medline: 17095295]
30. Buzzell GA, Fedota JR, Roberts DM, McDonald CG. The N2 ERP component as an index of impaired cognitive control in smokers. Neurosci Lett 2014 Mar 20;563:61-65. [doi: 10.1016/j.neulet.2014.01.030] [Medline: 24486891]
31. Seamans JK, Yang CR. The principal features and mechanisms of dopamine modulation in the prefrontal cortex. Prog Neurobiol 2004 Sep;74(1):1-58. [doi: 10.1016/j.pneurobio.2004.05.006] [Medline: 15381316]
32. Tritsch NX, Sabatini BL. Dopaminergic modulation of synaptic transmission in cortex and striatum. Neuron 2012 Oct;76(1):33-50 [FREE Full text] [doi: 10.1016/j.neuron.2012.09.023] [Medline: 23040805]
33. Fabre C, Chamari K, Mucci P, Massé-Biron J, Préfaut C. Improvement of cognitive function by mental and/or individualized aerobic training in healthy elderly subjects. Int J Sports Med 2002 Aug;23(6):415-421. [doi: 10.1055/s-2002-33735] [Medline: 12215960]
34. Lauenroth A, Ioannidis AE, Teichmann B. Influence of combined physical and cognitive training on cognition: a systematic review. BMC Geriatr 2016 Dec 18;16:141 [FREE Full text] [doi: 10.1186/s12877-016-0315-1] [Medline: 27431673]
35. Oswald WD, Gunzelmann T, Rupprecht R, Hagen B. Differential effects of single versus combined cognitive and physical training with older adults: the SimA study in a 5-year perspective. Eur J Ageing 2006;3(4):179 [FREE Full text] [doi: 10.1007/s10433-006-0035-z] [Medline: 28794762]

36. Vogt T, Herpers R, Scherfgen D, Strüder HK, Schneider S. Neuroelectric adaptations to cognitive processing in virtual environments: an exercise-related approach. Exp Brain Res 2015;233(4):1321-1329. [doi: 10.1007/s00221-015-4208-x] [Medline: 25630906]
37. Deslandes A, Moraes H, Ferreira C, Veiga H, Silveira H, Mouta R, et al. Exercise and mental health: many reasons to move. Neuropsychobiology 2009;59(4):191-198 [FREE Full text] [doi: 10.1159/000223730] [Medline: 19521110]
38. Shams TA, Foussias G, Zawadzki JA, Marshe VS, Siddiqui I, Müller DJ, et al. The effects of video games on cognition and brain structure: potential implications for neuropsychiatric disorders. Curr Psychiatry Rep 2015;17(9):71. [doi: 10.1007/s11920-015-0609-6] [Medline: 26216589]
39. Wang P, Zhu XT, Qi Z, Huang S, Li H. Neural basis of enhanced executive function in older video game players: an fMRI study. Front Aging Neurosci 2017;9:382 [FREE Full text] [doi: 10.3389/fnagi.2017.00382] [Medline: 29209202]
40. Zhang X, Ni X, Chen P. Study about the effects of different fitness sports on cognitive function and emotion of the aged. Cell Biochem Biophys 2014;70(3):1591-1596. [doi: 10.1007/s12013-014-0100-8] [Medline: 24997050]

Abbreviations

ACC: anterior cingulate cortex
Cz: central cortex
EEG: electroencephalographic
ERP: event-related potential
EXG: exergaming
Fz: frontal cortex
HR: heart rate
HRR: heart rate reserve
MetS: metabolic syndrome
ms: milliseconds.
Pz: parietal cortex
RT: reaction time
TE: treadmill exercise
TEG: treadmill exercise group

Edited by G Eysenbach; submitted 01.02.19; peer-reviewed by N Skjaeret, M Rosly; comments to author 28.03.19; revised version received 07.05.19; accepted 25.05.19; published 31.07.19.

Please cite as:
Wu S, Jo EA, Ji H, Kim KH, Park JJ, Kim BH, Cho KI
Exergaming Improves Executive Functions in Patients With Metabolic Syndrome: Randomized Controlled Trial
JMIR Serious Games 2019;7(3):e13575
URL: https://games.jmir.org/2019/3/e13575/
doi:10.2196/13575
PMID:31368441

Original Paper

Effects of Full Body Exergaming in Virtual Reality on Cardiovascular and Muscular Parameters: Cross-Sectional Experiment

Boris Feodoroff[1], PhD; Ippokratis Konstantinidis[1], MA; Ingo Froböse[1], Prof Dr

Institute of Movement Therapy and Movement-Oriented Prevention and Rehabilitation, German Sport University Cologne, Cologne, Germany

Corresponding Author:
Boris Feodoroff, PhD
Institute of Movement Therapy and Movement-Oriented Prevention and Rehabilitation
German Sport University Cologne
Nawi-Medi, Ground Floor
Am Sportpark Müngersdorf 6
Cologne, 50933
Germany
Phone: 49 0221 4982 4570
Email: b.feodoroff@dshs-koeln.de

Abstract

Background: In recent years, many studies have associated sedentary behavior in front of screens with health problems in infants, children, and adolescents. Yet options for exergaming—playing video games that require rigorous physical exercise—seem to fall short of the physical activity levels recommended by the World Health Organization.

Objective: The purpose of this study was to investigate the effect of a fully immersive virtual reality (VR)-based training system on cardiovascular and muscular parameters of young adults.

Methods: A cross-sectional experiment design was used to analyze muscle activity (surface electromyography), heart rate, perceived exertion (RPE), cybersickness symptoms, perceived workload, and physical activity enjoyment (PACES) in 33 participants performing two 5-minute flights on a new training device.

Results: Participants' performance of the planking position required to play the game resulted in moderate aerobic intensity (108 [SD 18.69] bpm). Due to the mainly isometric contraction of the dorsal muscle chain (with a mean activation between 20.6% [SD 10.57] and 26.7% [SD 17.39] maximum voluntary isometric contraction), participants described the exercise as a moderate to vigorous activity (RPE 14.6 [SD 1.82]). The majority reported that they enjoyed the exercise (PACES 3.74 [SD 0.16]). However, six participants had to drop out because of cybersickness symptoms and two because of muscle pain due to prior injuries.

Conclusions: Our findings suggest that fully immersive VR training systems can contribute to muscle-strengthening activities for healthy users. However, the dropout rate highlights the need for technological improvements in both software and hardware. In prevention and therapy, movement quality is a fundamental part of providing effective resistance training that benefits health. Exergaming on a regular basis has the potential to develop strong muscles and a healthy back. It is essential that future VR-based training systems take into account the recommendations of sport and exercise science.

(JMIR Serious Games 2019;7(3):e12324) doi:10.2196/12324

KEYWORDS

exergaming; gamification; Immersive virtual reality; exercise; cybersickness; flight simulation

Introduction

Background

There is a growing body of evidence suggesting that a high level of sedentary behavior can harm human health [1-5]. Despite the advantages that come with access to information and rapid communication, sitting in front of screens for hours at a time has been associated with health and psychological problems in infants, children, and adolescents [6]. Screen time and total sitting time have been the most studied exposure variables in epidemiological research [7]. Of these, television viewing has been proposed to be the most deleterious [8-11]. Recently, this field of interest has expanded to examine the impact of computer and video game time on health [12].

The World Health Organization (WHO) has identified the appearance of negative symptoms due to excessive digital gaming as a disorder. Sedentary behavior spent in front of screens has been reported to increase the risk of obesity, high-density lipoprotein dysfunction, and high blood pressure, which are also major risk factors for cardiovascular morbidity [13-15]. Adolescent boys who spend a great deal of time playing video games have been found to have lower bone mineral densities than average [16,17]. Sedentary behavior spent in front of screens is a primary contributor to decreasing physical activity among youths [18-20]. The promotion of exergaming seems to be a promising way to counteract this trend and its negative effects.

Exergaming

The portmanteau word exergaming combines exercise and gaming [21]. The term exergame has many definitions, reflecting the diverse approaches undertaken by various research communities during the past 20 years. Health researchers have introduced several terms similar to exergaming in describing the promotion of physical activity or interactivity during video gaming: physical gaming, exertainment, active video games, and active video training [21]. Bogost [22] refers to exergaming as "the combination of exercise and video games,"—the term video gaming is defined as "the process of gaming in any digital device" [22], but the term exercise allows for a variety of interpretations.

Exercise can refer to the process of becoming more skilled in a set of actions, without necessarily specifying the degree of body movement involved [23]. In this respect, exergaming can refer to the process of training reaction times in a solely sedentary setting, provided that it is planned and structured. Professional players of multiplayer online battle arena games optimize, train, and exercise their eye-hand coordination skills.

Exercise can also describe physical activity involving body movements that do not promote physiological adaptation mechanisms or increase a particular skill level [23]. For instance, Pokémon Go promotes body movement by forcing users to change their physical location. Positive beneficial health behaviors such as a higher level of physical activity, more socialization, and better mood have been associated with Pokémon Go [24].

Using the term exercise interchangeably with physical activity or even physical fitness can, therefore, be misleading when referring to exergaming systems. The proposed definition by Oh and Yang [21] describes exergaming as "playing exergames or any video games that require physical exertion or movements that are more than sedentary activities and also include strength, balance, and flexibility activities." Immersive virtual reality (VR)-based exergaming systems such as VR Boxing and Fastest Fist try to bridge the gap between plain body movements and the planned, structured, and repetitive elements of a training session.

Physical exercise includes an effective stimulation of the muscle adaptation mechanism. WHO recommends that adults between the ages of 18 and 64 years should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week, at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week, or an equivalent combination of moderate- and vigorous-intensity activity. Muscle-strengthening activities should be done for major muscle groups on two or more days per week [25].

Different muscle contraction types can produce different gains in muscle strength and power [26]. The most common contraction types in everyday life are static isometric contractions (holding an object against gravity) and eccentric and concentric contractions, which in combination produce a dynamic motion (such as lifting and lowering an object). The technical capabilities of VR-based motion capture systems can motivate the performance of isometric contractions, for example, by getting users to execute a perfect plank, an exercise known mostly for its ability to improve balance and strengthen core muscles.

Plank Exercise

In 2009, lower back pain ranked third among the most common syndromes and illnesses [27]. Core strength training is the most effective way to help alleviate the symptoms of chronic lower back pain [28]. In the exercise known as the plank, the body is fully stretched with its weight supported entirely by the tips of the toes and the forearms. This primarily activates the core muscles (*musculus erector spinae* and *m. rectus abdominis*) together with the deltoids. Varieties of this strengthening exercise have been introduced with and without the use of external devices [29-31].

Study Objectives

To the best of our knowledge, this is the first qualitative study to examine the muscular training potential of an immersive VR-based full-body exergaming system. The study also analyzed the impact of exergaming on the cardiovascular system and its potential to provide effective endurance training. Understanding its impact on participants' perceived levels of motion sickness, cognitive load, and overall enjoyment could provide insight into the potential and utility of similar approaches.

Methods

Participants

Thirty-three participants (mean age 23.90 [SD 4.58] years) were recruited via social media platforms, email chains, and flyers. The study protocol was approved by the ethics committee of the German Sport University Cologne. Most of the participants were students at the German Sport University Cologne. Data collection occurred between June and August 2017. Participant requirements were defined to ensure that factors such as sex, age, and physical condition would not bias the data:

- Male participants aged younger than 30 years: due to hormonal changes, women tend to be more susceptible to motion sickness than men [32].
- Athletic body: body mass index correlates with body fat percentage, which can influence the signal quality of surface electromyography (sEMG). To get a valid sEMG signal,

participants had to have a BMI <25 kg/m^2, indicating a low body fat percentage.

- Height between 170 and 190 cm: the Icaros device (Figure 1) permits configuration of the distance between the foot holder and arm rest. This distance influences muscle activation of the abdominal and back areas. Therefore, a standardized configuration of the upper torso-to-arm and upper-to-lower-body angle was selected. In addition, the body's center of gravity had to be in line with the device's zero balance point. Based on these requirements, the device's properties allowed a standardized configuration for body heights between 170 and 190 cm.

Figure 1. (A) Torso-arm angle measurement using protractor, (B) Icaros device, (C) participant being familiarized with the device, (D) upper-to-lower limb angle measurement using protractor.

Instrumentation

The experiment protocol required a variety of different systems to be used simultaneously and synchronized for data analysis.

Icaros System

The Icaros VR fitness machine and flight simulator (Icaros GmbH) consists of a device and an attached gyro sensor (Figure 1). The gyro sensor connects via Bluetooth with a smartphone and provides angular velocity data. Body position changes on the sagittal and longitudinal axes of the device are independent of the visual stimuli, allowing a deeper immersion in the VR world through a combination of visual, acoustic, and proprioception sensors.

The device can be rotated around the pivot point with a range of motion between –35° and 35° for roll and –45° to 45° for pitch on the vertical and sagittal axes, respectively.

Head-Mounted Device

The VR system employed the Gear VR headset version SM-R322 (Samsung Electronics) and gamepad for setting in-game options. The Icaros software was installed on a Galaxy S6 (Samsung Electronics) smartphone running Android 6.01. The integrated smartphone accelerometer and gyroscope paired with the head-mounted display allowed users to interact with the visual stimuli in virtual reality while maintaining their body position on the Icaros.

Heart Rate Monitor

Participant heart rates were measured continuously using the RS800 heart rate monitor (Polar Electro).

Muscle Activity

Muscle activity was measured based on sEMG using the TeleMyo 2400T G2 (Noraxon USA).

Questionnaires

The study employed questionnaires measuring sociodemographic and anthropometric data. The Simulator Sickness Questionnaire (SSQ) was used to assess perceived motion sickness and cybersickness symptoms during the flights. Enjoyment of the flight sessions was measured using the Physical Activity Enjoyment Scale (PACES). Results are based on the mean average of the 16-item modified version of the questionnaire ranging from a scale of 1 (strongly disagree) to 5 (strongly agree). A high score indicates a high level of physical activity enjoyment [33]. The Borg scale [34] was used to measure the rate of perceived exertion (RPE). This scale has a high degree of validity for endurance training and physical activity [35,36]. The perceived mental, physical, and emotional demands on participants during the VR experience were assessed using the NASA Task Load Index (NASA-TLX) questionnaire [37]. Five of the six dimensions (frustration, effort, mental demand, physical demand, and temporal demand) ranged from 0 (low) to 100 (high). The performance dimension ranged from 0 (good) to 100 (bad).

Experimental Protocol

All participants were informed about the aim of the study and provided with a written description of the procedure. Participants completed questionnaires collecting sociodemographic and anthropometric data (Table 1). The heart rate monitor system was then configured and participants performed the deep breathing technique [38] for measuring their resting heart rates while seated.

For each test, participant's height was measured, and the distance between the Icaros arm and foot holders was configured to align the body's center of gravity with the pitch axis' zero degree point. The upper-to-lower limb and torso-to-arm angle were configured to 135° and 90°, respectively (seen in Figure 1A and D).

In a 5-minute familiarization session, participants were introduced to the VR in-game tasks. Afterward, eight electrodes were positioned on the recommended sensor locations for the *m. erector spinae* (neck extensors), *m. deltoideus pars clavicularis*, *m. rectus abdominus*, and *m. erector spinae* (lumbar region) as per the Surface EMG for Noninvasive Assessment of Muscles guidelines [39] (Figure 2). Participants were then asked to perform a maximum voluntary isometric contraction (MVIC) for each muscle [40]. Participants completed two consecutive flight sessions, approximately five minutes each, with an interim pause of 15 minutes. The goal of the game was for the participants to navigate their virtualized plane in a first-person view through all 63 rings. Computer software allowed the researchers to see the participant field of view (Figure 3). The speed of the vehicle and the horizon view were standardized to ensure maximal consistency in flight path, speed, and trajectory. During both sessions, muscle activity, heart rate, and device movements were continuously captured. Exertion levels were determined after each successful flight using the RPE scale. After the second flight, participants completed postsession questionnaires (SSQ, NASA-TLX, and PACES).

Table 1. Sociodemographic and anthropometric characteristics of study participants.

Variable	Finishers (n=25), mean (SD)	Motion sickness dropouts (n=6), mean (SD)
Age in years	24.16 (4.82)	25.00 (1.90)
Height (meters)	1.80 (0.63)	1.88 (0.08)
Weight (kg)	77.50 (8.49)	83.00 (6.10)
Body mass index (kg/m^2)	23.80 (2.03)	23.36 (1.18)

Figure 2. Positioning of electromyography electrodes, hardware, and software.

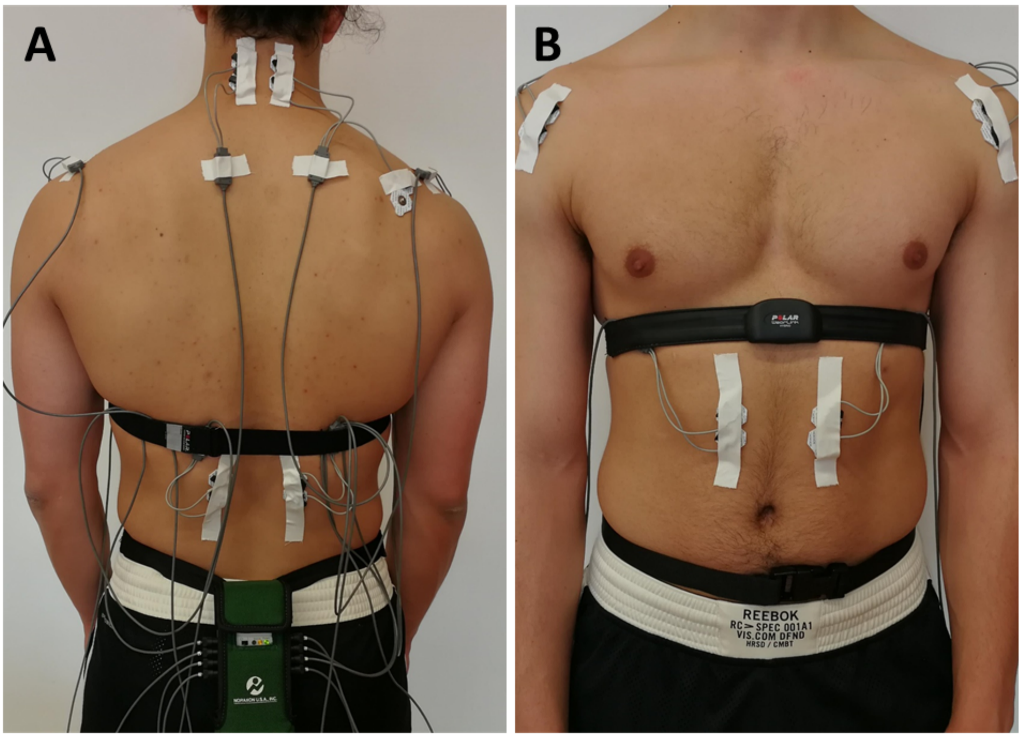
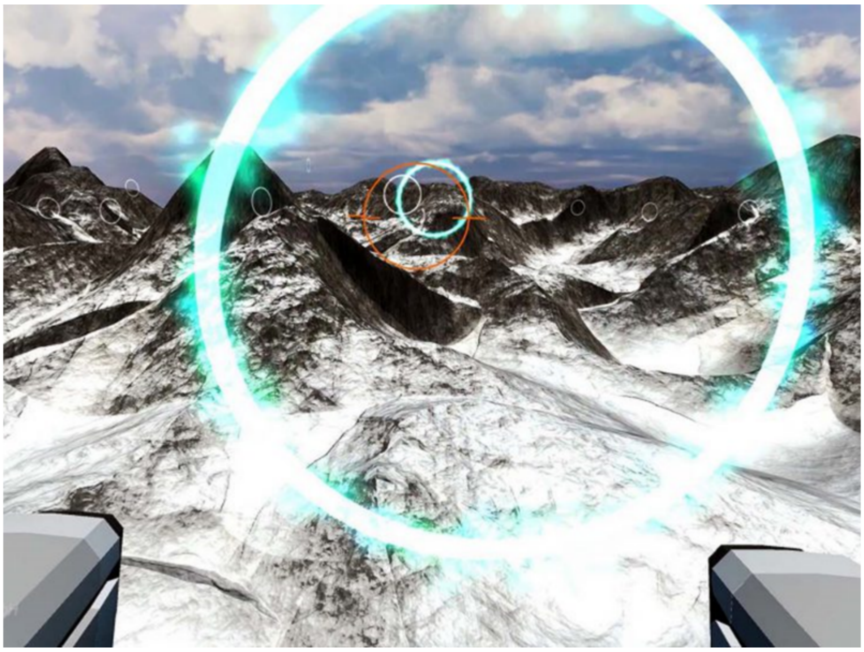


Figure 3. In-game screenshot of the participant field of view.



Outcome Measurements

The captured sEMG signal was smoothed using the root mean square algorithm, and the time interval was set to 300 milliseconds [41]. Afterward, the amplitude of the signal was normalized to the MVIC stack of the participant.

For the purposes of data analysis, a MATLAB script was programmed to provide a time sync of the Icaros device position along with the corresponding muscle activation and heart rate.

Statistical Analysis

The captured data were exported from the MATLAB environment and statistically analyzed (SPSS Statistics 23, IBM Corp).

Power

Power analysis indicated that for an estimated effect size of 0.6 with 80% power and 5% type I error, 25 participants were needed. Flight sessions that had to stop due to muscle pain or cybersickness symptoms were excluded from all data analysis except the SSQ score.

Results

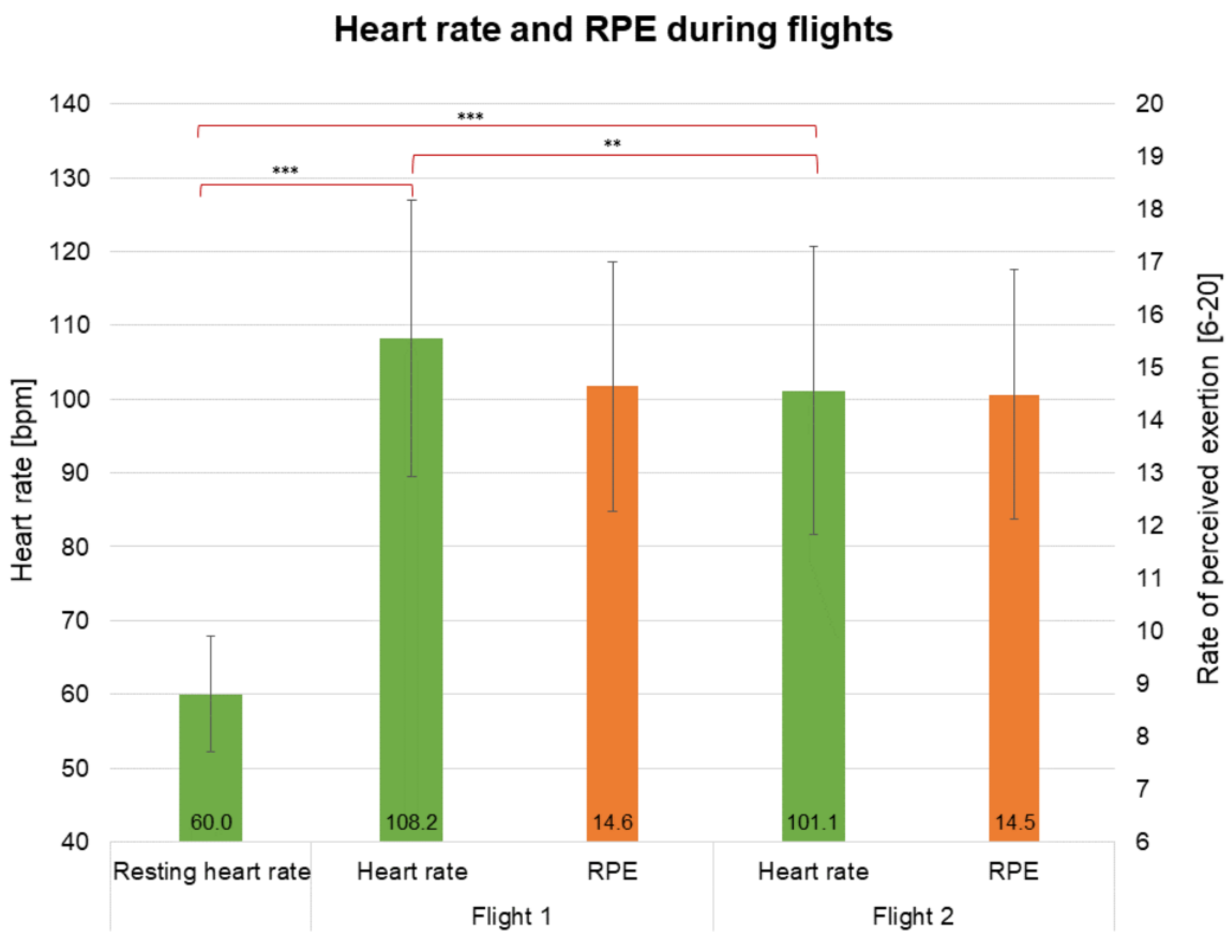
Overview

Six participants had to stop the experiment due to signs of nausea or discomfort indicative of motion sickness. Another two participants stopped the VR flights because of muscle pain due to prior injuries. Post hoc analysis was computed based on a 5% type I error and a total of 25 participants.

Muscle Activity During Flight Sessions

A paired sample mean *t* test resulted in significant differences between the two flights for the sEMG signal captured on the *m. erector spinae* muscles (neck extensors) of the participants (Figure 4). During the first flight, participants had significantly higher muscle activity (mean 24.21 [SD 11.47]) relative to the second flight (mean 20.60 [SD 10.57]), with t_{24}=2.219, *P*=.04, and *d*=0.33. Achieved power was 35%.

Figure 4. Average muscle activation during the two consecutive flights.



Heart Rate and Rate of Perceived Exertion During Flight Sessions

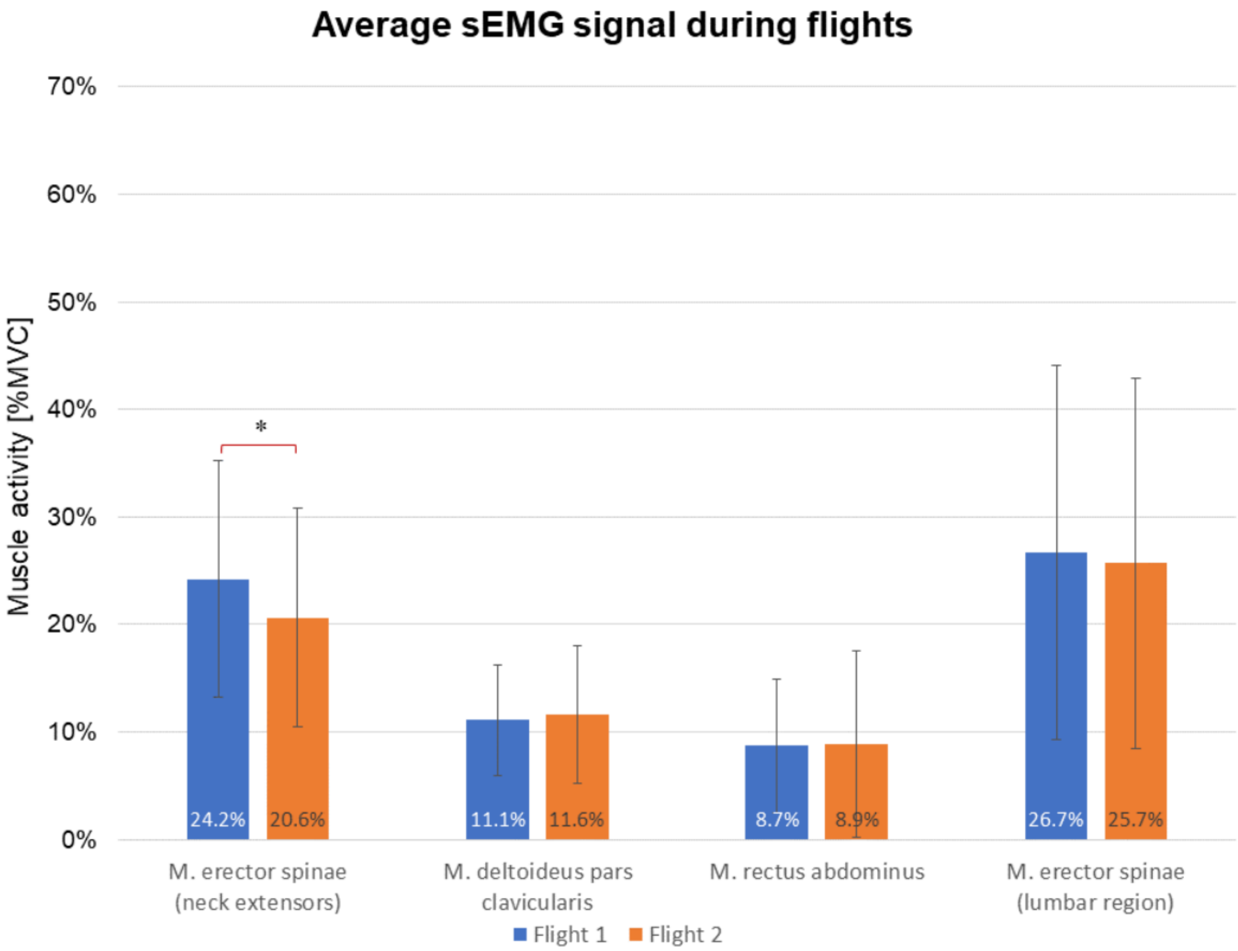
There were significant differences between resting heart rate and heart rate during the flights (Figure 5).

The one-way analysis of variance with repeated measures and a Greenhouse Geisser correction showed significant differences in the three heart rate measurements $F_{1.56,37.55}$=161.29, *P*<.001, and η^2=0.87. A Bonferroni post hoc test indicated significant differences between the resting heart rate and the heart rate during flight 1 (*P*<.001, *d*=0.87; –48.18, 95% CI –56.51 to –39.86) and during flight 2 (*P*<.001, *d*=0.83; –41.07, 95% CI –49.51 to –32.64).

Average heart rate during flight 1 was significantly higher (*P*=.005, *d*=0.37; –7.11, 95% CI 1.98 to –12.23) than during flight 2.

Results from the paired sample *t* test showed no significant differences between participant RPE scores at the end of the first flight (mean 14.64 [SD 1.82]) and the second flight (mean 14.48 [SD 2.37]). Achieved power was 98%.

Figure 5. Heart rate and rate of perceived exertion during the two flights.



Movements During Flight Sessions

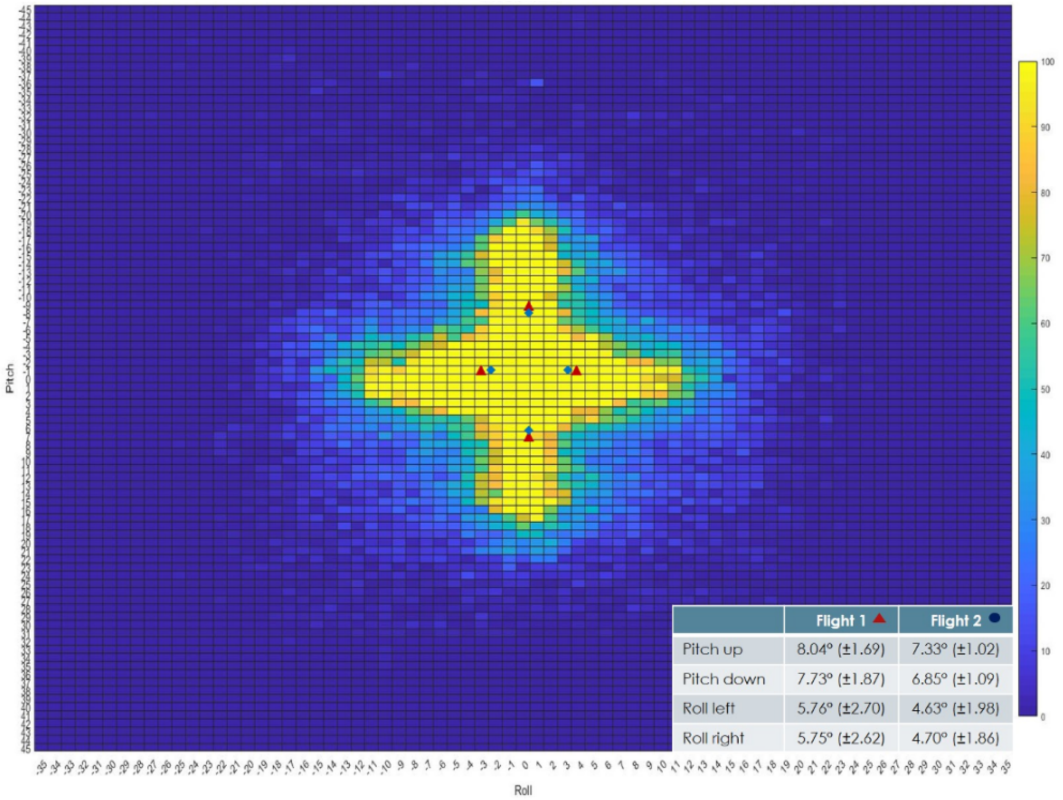
Participants were able to steer pitch and roll by changing their body's center of gravity. A paired *t* test showed that the average upward movements (Figure 6) on the pitch axis during flight 1 (mean 8.04° [SD 1.69]) were significantly larger, with t_{24}=–2.370, *P*=.03, and *d*=0.51, than those during flight 2 (mean 7.33° [SD 1.02]). Achieved power was 69%.

Downward pitch movements also showed a significant difference, with t_{24}=3.013, *P*=.06, and *d*=0.57 between flight 1 (mean 7.73° [SD 1.87]) and flight 2 (mean 6.85° [SD 1.09]). Achieved power was 78%.

The average degrees of roll were similar in both directions. Rolling left showed a significant difference, with t_{24}=3.103, *P*=.05, and *d*=0.48 between flight 1 (mean 5.76° [SD 2.70]) and flight 2 (mean 4.63° [SD 1.98]). Achieved power was 63%.

Rolling right also showed significantly higher average degrees for flight 1 (mean 5.75° [SD 2.62]) than for flight 2 (mean 4.7° [SD 1.86]), with t_{24}=–2.714, *P*=.012, and *d*=0.46. Achieved power was 60%.

Figure 6. Heat plot of the Icaros device's movements for both flights. Symbols depict the average positions throughout the flights on each axis.



Questionnaires

Postsession questionnaires included the SSQ, NASA-TLX, and PACES. Table 2 illustrates the average values of the calculated dimensions and scores.

Levene test indicated unequal variances for all but the nausea score. Therefore, the degrees of freedom were adjusted from 29 to 5.26 for oculomotor, from 29 to 5.31 for disorientation, and from 29 to 5.35 for the total score. Due to the different sample sizes, a weighted effect size based on Hedges' G formula replaced the Cohen *d* effect size.

The independent *t* test indicated a significantly higher score on all four SSQ subcategories. Participants who stopped the experiment due to motion sickness symptoms showed significantly higher oculomotor, disorientation, nausea, and total score values than the finishers (Table 2).

Participants found the sessions to be very physically demanding and exerted much effort when completing the session task. The task was not perceived as mentally demanding; its pace was not perceived as hurried or rushed (temporal demand). Participants felt like they achieved very good results (performance) and didn't feel particularly insecure, discouraged, irritated, stressed, or annoyed (frustration). Despite the high physical demand scores and the moderate to vigorous RPE scores, participants rated the flights as highly enjoyable (PACES score 3.74 [SD 0.16]).

Table 2. Postsession questionnaire scores by dimension.

Questionnaire and dimension	Finishers	Motion sickness dropouts	Significance *P* value	Effect size g
SSQ[a]				
Oculomotor	22.44 (15.85)	87.17 (48.11)	.02	2.62
Disorientation	23.39 (29.72)	146.16 (82.93)	.01	2.80
Nausea	27.86 (27.52)	124.02 (34.66)	<.001	3.24
Total score	28.27 (21.51)	130.90 (56.67)	.006	3.35
NASA-TLX[b]				
Mental demand	45.04 (24.61)	—[c]	—	—
Physical demand	76.76 (15.73)	—	—	—
Temporal demand	43.04 (18.73)	—	—	—
Performance	74.88 (16.84)	—	—	—
Effort	74.52 (10.98)	—	—	—
Frustration	33.24 (24.78)	—	—	—
PACES[d]	3.74 (0.16)	—	—	—

[a]SSQ: Simulator Sickness Questionnaire.
[b]NASA-TLX: NASA Task load Index.
[c]Not available.
[d]PACES: Physical Activity Enjoyment Scale.

Discussion

Principal Findings

Muscle Activity

For isometric contractions, there is a good correlation between muscle activity and generated force [42]. The electrical signal measured on muscles is proportional to the muscle force, although there is debate about the linearity of this relationship [43]. Generally, though, MVIC reflects maximal muscle force. The MVIC normalized muscle activity signal can thus be used as a control mechanism on the effectiveness of muscle strengthening exercises. Depending on the generated muscle force, the effects can vary between mobilization and strength improvement with muscle growth [44].

Muscle activity during flights indicates the effects on muscle blood flow stimulation for all muscle groups. Kibler [45] notes that a minimum of 30% MVIC is required for effective strength endurance training; percentages below 30% provide mere muscle mobilization. In both flights, muscle activity for all assessed muscles ranged between 11% and 26% MVIC.

Dorsal muscle chain activity (neck extensors and lumbar region of *m. erector spinae*) shows higher levels of activation, with values reaching or crossing the 30% threshold in some cases. Dorsal muscle chain exercises (eg, stabilization exercise) could assist in alleviating low back pain in the lumbosacral area [28].

Participants position on the Icaros device resembles the well-established plank exercise. The only difference is the device's shin holders, which provide additional support for users (Figure 1). This explains the rather low activation of the *m. rectus abdominus* and *m. deltoideus pars clavicularis* during both flights (Figure 4). Other studies reported EMG values of 44% to 65% on the rectus abdominus during the performance of plank exercises with unstable devices (such as the Swiss Ball) [29,46].

To achieve the most immersive experience, the virtual horizon was adjusted to a 45° angle, which required the participants to tilt their heads back. This position requires an almost constant isometric muscle contraction of the neck extensors. The neck extensors' average muscle activation was relatively high compared with that of the *m. rectus abdominus*, the *m. deltoideuspars clavicularis*, and the *m. erector spinae* (lumbar region). In part, the high muscle activation was due to the extra weight (approx. 300 grams) of the head-mounted display. During flight 1, participants' average muscle activity of the neck extensors was significantly higher than flight 2 (*P*=.04). Because of increased familiarization with the VR in-game elements, participants undertook fewer head movements during the second flight, which significantly lowered neck extensor muscle activation. The second flight did not show any other significant differences regarding muscle activity.

The similarity between the average muscle activations seems to indicate the potential reliability of the Icaros VR system. Being able to reproduce the physiological muscle activation in a consecutive flight suggests that familiarization effects do not reduce muscle activity straight away. Nevertheless, the long-term outcome of repetitive flights cannot be assessed based on this study's design. The sample size consisted mainly of fit, young sport students whose athleticism distinguishes them from the general public.

Since the second flight was the same as the first one in terms of the virtualized in-game world and properties (eg, flying speed), differences between the two flights are attributable solely to participants' movements.

The heat plot in Figure 6 shows the differences in the range of motion between the first and second flights. Standardizing the body balance point for all participants ensured that body height and weight would not bias muscle activity or device movements.

During the first flight, participants rolled by an average of 5.76° to the left and 5.75° to the right. Pitch movements reached 8.04° downward and 7.73° upward. Flight 2 resulted in a significantly smaller range of motion across all four directions, with effect size varying between 0.46 and 0.57. The smaller range of motion during the second flight can be attributed to familiarization with the in-game elements and device properties during the first flight.

Improved intermuscular coordination from reintroducing and sustaining body balance on the device after pitch and roll movements could also explain the significantly smaller range of motion. Interestingly, the significantly smaller range of motion did not correspond to significantly lower muscle activation during the second flight (see Figure 5). The measured muscle activation seems to be necessary for maintaining the plank position and coping with in-game goals. Compared with other exergaming systems, this represents a step forward. In some other fitness games, for instance, intended exercises can be avoided by simply shaking the remote controller quickly up and down.

Due to the mostly static position of the participants on the device, the effects on their cardiovascular system were expected to be low. Their resting heart rate (60.04 [SD 7.90] bpm) was significantly lower than their heart rate during the first (108.22 [SD 18.72] bpm) and the second (101.12 [SD 19.53] bpm) flights. Despite the significant differences between the two, the effect size was small to medium (d=0.37). The heart rate stayed within a lower aerobic intensity level, and the changes did not reflect a reduction in the intensity zone [47]. Because muscular activity showed no significant differences for all muscles but one, significant changes in heart rate should be attributed to psychological factors rather than to physiological ones. In his meta-analysis, Howard [48] investigated how excitement could potentially account as one factor for the success of VR rehabilitation programs. Furthermore, Bun et al [49] observed increased excitement during users' first contact with VR systems. Such an aroused psychological state could further stimulate the sympathetic nervous system, increasing users' heart rates.

Although heart rate alone is an insufficient metric for measuring energy expenditure and physiological effect [50], the observed heart rate values do seem to point to the positive effects of a low-intensity aerobic training zone. However, it must be borne in mind that the participants needed around 6 minutes per flight session, during which they mostly remained in a static isometric position that undercut the effects of aerobic cardiovascular training.

For muscle strength endurance training, long periods of muscle tension are connected with significantly higher myofibrillar protein synthesis [51]. However, long sessions of isometric exercises are also connected with high blood pressure levels, which are a risk factor for people with undiagnosed hypertension, heart disease, or other cardiovascular problems [52,53]. Such hypertensive participants can develop increased blood pressure and may need a longer period to recuperate than healthy individuals [54].

The reported RPE values of 14.56 (SD 2.09) for both flights indicate a moderate to vigorous activity level. The question of whether the perceived intensity level is due to physical arousal or due to cognitive workload can be further analyzed by looking at the reported NASA-TLX values, in which mental demand is lower than physical demand. Despite the perceived medium to high exertion levels, the reported enjoyment (PACES) during the VR sessions was high (3.74 [SD 0.18]), indicating that the VR system was fun to use. The perceived exertion, task workload, and enjoyment level are good indicators of a system's potential to provide a good combination of fun and physical activity.

Motion Sickness and Cybersickness

Cybersickness can be described as a visually induced motion sickness that is common in immersive VR sessions [55]. Although the SSQ was not explicitly developed for assessing cybersickness, it is widely used in assessing the outcome of experience with VR environments [56-58]. The 6 participants with the strongest motion sickness symptoms had to stop the experiment. After the symptoms faded, they completed the accompanying questionnaire. They showed significantly higher scores on the SSQ than participants who were able to finish the flights (Table 2), with large effect sizes on all 4 subcategories.

The (forced) planking position on the Icaros was never before applied in a fully immersive VR system. It introduces high immersiveness by keeping the users on the transversal plane, simulating the in-game flying position. Changing of the visual stimuli as the device shifts on the axes is a unique feature of the system. Flights with the Icaros system seem to lead to relatively high SSQ scores in general [59]. Here, three properties come into play: the technical elements of the head-mounted display, the participants' posture, and their postural sway.

The system used for our study does not represent the state of the art for head-mounted display technology, which can be found in devices such as Oculus Rift (Facebook Technologies LLC) and Vive Pro (HTC Corporation). Screen flickering caused by high latency and a low refresh rate (60 Hz) of systems like ours has been associated with symptoms of motion sickness [60,61]. Interestingly, it has been proposed that increased realism and immersion through improved optic stimulus may exacerbate motion sickness symptoms [55] as the internal focus is steered toward the sensory mismatches between the visual and vestibular systems.

Gallaghar et al [55] describes the vital role of the vestibular system in causing motion sickness symptoms. The most established cybersickness theories identify multisensory signal conflicts caused mainly by visual and vestibular signal mismatches as the root of motion sickness symptoms. On the Icaros device, posture and movement affect visual stimuli and perceived self-motion, which should lead to fewer visual-vestibular signal discrepancies.

Riccio and Stoffregen [62] propose that postural instability causes motion sickness symptoms. Smart et al [63] point out that postural sway could be related to higher perceived levels of visually induced motion sickness. On the Icaros device, the user's body posture resembles the plank position but undergoes fast changes on pitch and roll axes, which could explain the perceived motion sickness. As users tried to maneuver through the rings, unpredictable changes in the device's direction may have further influenced the perceived motion sickness [60]. Moreover, the vection produced by user head movements as they aimed for the next virtual ring may have also contributed to motion sickness [64,65].

Limitations and Strengths

There are some limitations that warrant discussion. First, requirements regarding the physiological status of participants and the low fat-to-body ratio were very strict. While these requirements ensured high-quality and reliable EMG activation signals, they also restricted the transferability of the results to populations with different body compositions. Furthermore, the exclusive use of male participants for a consistent perception of cybersickness symptoms rendered the study unable to assess the physiological and psychological outcomes on female participants [32]. A third limitation was the assessment of training level intensity based on heart rate alone. Training heart rates can vary among individuals due to variances between actual maximal heart rates and the ones resulting from age-based formulas [50,66]. To eliminate the effects of these variances, we recommend the use of additional tools such as spirometry for future studies measuring physiological effort and energy expenditure.

This paper tries to investigate the combination of a fully immersive VR system with physical activity through a sport-science–based approach. To the best of our knowledge, this is the first qualitative study that examines the muscular training potential of an immersive VR-based full-body exergaming system.

We use established measuring procedures to identify muscular activity in terms of muscle innervation and potential training effects. Our approach points out future development steps of similar VR exergaming systems, assisting developers to optimize their efforts in promoting physical activity.

Conclusion

The Icaros flight session appears to provide little to no cardiovascular benefit. Based on WHO guidelines for resistance training, muscle activity during the sessions occasionally met the lower threshold for effective muscle activation. The lower back's muscle activation corresponds to plank variations with instability devices (eg, planks on exercise balls) [29].

By contrast, Icaros can provide improved muscle strength, especially for the dorsal muscle chain. Differentiated muscle activation can be achieved in virtualized worlds by, say, requiring the user to spend more time in a pitch down position, thereby shifting the body's center of gravity on the pitch axis. But achieving the 30 minutes of daily cardiovascular activity recommended by the WHO requires dynamic instead of isometric movements.

Performing a plank position without external control of one's body posture can result in hyperlordosis, an excessive extension of the lumbar region. Added stress on the visual and vestibular systems is associated with higher immersiveness but can lead to proprioceptive signals receiving less attention by the user. Spending time in flexed and awkward positions has been associated with low back pain syndrome [67]. It is crucial, therefore, that users' movements do not amplify unwelcome body positions, especially during shifts in the body's center of gravity.

Motivating the public to engage in physical activity is probably one of the most important and difficult tasks of the health sector. Gamification of physical exercise can help not only to motivate physical activity but also to promote social contact and interaction.

Future full-body and fully immersive concepts should focus on increasing dynamic muscle activation while considering user susceptibility to cybersickness and motion sickness. VR systems can give the public a refreshing and enjoyable form of physical activity.

In prevention and therapy, movement quality is a fundamental component of effective resistance training to benefit health. It is crucial, therefore, that future VR-based training systems follow the recommendations of sport and exercise science.

Acknowledgments

This research was supported by the Institute of Movement Therapy and Movement-oriented Prevention and Rehabilitation of the German Sport University. We would like to thank the Icaros company for providing us with the necessary hardware and software. No financial agreements of any kind were conducted between the company and the researchers.

Conflicts of Interest

None declared.

References

1. Dunstan DW, Howard B, Healy GN, Owen N. Too much sitting—a health hazard. Diabetes Res Clin Pract 2012 Sep;97(3):368-376. [doi: 10.1016/j.diabres.2012.05.020] [Medline: 22682948]

2. Grøntved A, Hu FB. Television viewing and risk of type 2 diabetes, cardiovascular disease, and all-cause mortality: a meta-analysis. JAMA 2011 Jun 15;305(23):2448-2455 [FREE Full text] [doi: 10.1001/jama.2011.812] [Medline: 21673296]
3. Katzmarzyk PT, Church TS, Craig CL, Bouchard C. Sitting time and mortality from all causes, cardiovascular disease, and cancer. Med Sci Sports Exerc 2009 May;41(5):998-1005. [doi: 10.1249/MSS.0b013e3181930355] [Medline: 19346988]
4. van Uffelen JGZ, Wong J, Chau JY, van der Ploeg HP, Riphagen I, Gilson ND, et al. Occupational sitting and health risks: a systematic review. Am J Prev Med 2010 Oct;39(4):379-388. [doi: 10.1016/j.amepre.2010.05.024] [Medline: 20837291]
5. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary behaviors and subsequent health outcomes in adults a systematic review of longitudinal studies, 1996-2011. Am J Prev Med 2011 Aug;41(2):207-215. [doi: 10.1016/j.amepre.2011.05.004] [Medline: 21767729]
6. Lissak G. Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study. Environ Res 2018 Jul;164:149-157. [doi: 10.1016/j.envres.2018.01.015] [Medline: 29499467]
7. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. Diabetologia 2012 Nov;55(11):2895-2905. [doi: 10.1007/s00125-012-2677-z] [Medline: 22890825]
8. Chau JY, Grunseit A, Midthjell K, Holmen J, Holmen TL, Bauman AE, et al. Cross-sectional associations of total sitting and leisure screen time with cardiometabolic risk in adults. Results from the HUNT Study, Norway. J Sci Med Sport 2014 Jan;17(1):78-84. [doi: 10.1016/j.jsams.2013.03.004] [Medline: 23619159]
9. Owen N, Salmon J, Koohsari MJ, Turrell G, Giles-Corti B. Sedentary behaviour and health: mapping environmental and social contexts to underpin chronic disease prevention. Br J Sports Med 2014 Feb;48(3):174-177. [doi: 10.1136/bjsports-2013-093107] [Medline: 24415410]
10. Rosenberg DE, Lee I, Young DR, Prohaska TR, Owen N, Buchner DM. Novel strategies for sedentary behavior research. Med Sci Sports Exerc 2015 Jun;47(6):1311-1315 [FREE Full text] [doi: 10.1249/MSS.0000000000000520] [Medline: 25222817]
11. Kim Y, Wilkens LR, Park S, Goodman MT, Monroe KR, Kolonel LN. Association between various sedentary behaviours and all-cause, cardiovascular disease and cancer mortality: the Multiethnic Cohort Study. Int J Epidemiol 2013 Aug;42(4):1040-1056 [FREE Full text] [doi: 10.1093/ije/dyt108] [Medline: 24062293]
12. Biddle SJH, García BE, Wiesner G. Sedentary behaviour and adiposity in youth: a systematic review of reviews and analysis of causality. Int J Behav Nutr Phys Act 2017 Dec 28;14(1):43 [FREE Full text] [doi: 10.1186/s12966-017-0497-8] [Medline: 28351363]
13. Merghani A, Malhotra A, Sharma S. The U-shaped relationship between exercise and cardiac morbidity. Trends Cardiovasc Med 2016 Apr;26(3):232-240. [doi: 10.1016/j.tcm.2015.06.005] [Medline: 26187713]
14. Goldfield GS, Kenny GP, Hadjiyannakis S, Phillips P, Alberga AS, Saunders TJ, et al. Video game playing is independently associated with blood pressure and lipids in overweight and obese adolescents. PLoS One 2011 Nov;6(11):e26643 [FREE Full text] [doi: 10.1371/journal.pone.0026643] [Medline: 22069461]
15. Martinez-Gomez D, Rey-López JP, Chillón P, Gómez-Martínez S, Vicente-Rodríguez G, Martín-Matillas M, AVENA Study Group. Excessive TV viewing and cardiovascular disease risk factors in adolescents. The AVENA cross-sectional study. BMC Public Health 2010 May 25;10:274 [FREE Full text] [doi: 10.1186/1471-2458-10-274] [Medline: 20500845]
16. Winther A, Ahmed LA, Furberg A, Grimnes G, Jorde R, Nilsen OA, et al. Leisure time computer use and adolescent bone health—findings from the Tromsø Study, Fit Futures: a cross-sectional study. BMJ Open 2015 Apr 22;5(6):e006665 [FREE Full text] [doi: 10.1136/bmjopen-2014-006665] [Medline: 26063563]
17. Shao H, Xu S, Zhang J, Zheng J, Chen J, Huang Y, et al. Association between duration of playing video games and bone mineral density in Chinese adolescents. J Clin Densitom 2015 Apr;18(2):198-202. [doi: 10.1016/j.jocd.2015.02.007] [Medline: 25937308]
18. Roberts JD, Rodkey L, Ray R, Knight B, Saelens BE. Electronic media time and sedentary behaviors in children: findings from the Built Environment and Active Play Study in the Washington DC area. Prev Med Rep 2017 Jun;6:149-156. [doi: 10.1016/j.pmedr.2017.02.021] [Medline: 28316911]
19. Lieberman D, Chamberlin B, Medina E, Franklin B, Sanner B, Vafiadis D, Power of Play: Innovations in Getting Active Summit Planning Committee. The power of play: Innovations in Getting Active Summit 2011: a science panel proceedings report from the American Heart Association. Circulation 2011 May 31;123(21):2507-2516. [doi: 10.1161/CIR.0b013e318219661d] [Medline: 21518980]
20. Kohorst MA, Warad DM, Nageswara RAA, Rodriguez V. Obesity, sedentary lifestyle, and video games: the new thrombophilia cocktail in adolescents. Pediatr Blood Cancer 2018 Jul;65(7):e27041. [doi: 10.1002/pbc.27041] [Medline: 29528184]
21. Oh Y, Yang S. Defining exergames and exergaming. 2010 Presented at: Meaningful Play Conference 2010; 2010; East Lansing.
22. Bogost I. Persuasive Games: The Expressive Power of Videogames. Cambridge: MIT Press; 2010.
23. Caspersen C, Powell K, Christenson G. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]

24. Marquet O, Alberico C, Adlakha D, Hipp JA. Examining motivations to play Pokémon GO and their influence on perceived outcomes and physical activity. JMIR Serious Games 2017 Oct 24;5(4):e21 [FREE Full text] [doi: 10.2196/games.8048] [Medline: 29066423]
25. Global Recommendations on Physical Activity for Health. Geneva: World Health Organization; 2010.
26. Kanehisa H, Miyashita M. Effect of isometric and isokinetic muscle training on static strength and dynamic power. Eur J Appl Physiol Occup Physiol 1983;50(3):365-371. [Medline: 6683160]
27. Robert Koch-Institut. 2015. Bericht Gesundheit in Deutschland 2015 URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GesInDtld/gesundheit_in_deutschland_2015.pdf?__blob=publicationFile [accessed 2019-08-18]
28. Chang W, Lin H, Lai P. Core strength training for patients with chronic low back pain. J Phys Ther Sci 2015 Mar;27(3):619-622 [FREE Full text] [doi: 10.1589/jpts.27.619] [Medline: 25931693]
29. Snarr RL, Esco MR. Electromyographical comparison of plank variations performed with and without instability devices. J Strength Cond Res 2014 Nov;28(11):3298-3305. [doi: 10.1519/JSC.0000000000000521] [Medline: 25264667]
30. Kim S, Kang M, Kim E, Jung I, Seo E, Oh J. Comparison of EMG activity on abdominal muscles during plank exercise with unilateral and bilateral additional isometric hip adduction. J Electromyogr Kinesiol 2016 Oct;30:9-14. [doi: 10.1016/j.jelekin.2016.05.003] [Medline: 27213781]
31. Schoenfeld BJ, Contreras B, Tiryaki-Sonmez G, Willardson JM, Fontana F. An electromyographic comparison of a modified version of the plank with a long lever and posterior tilt versus the traditional plank exercise. Sports Biomech 2014 Sep;13(3):296-306. [doi: 10.1080/14763141.2014.942355] [Medline: 25325773]
32. Golding JF. Motion sickness susceptibility. Auton Neurosci 2006 Oct 30;129(1-2):67-76. [doi: 10.1016/j.autneu.2006.07.019] [Medline: 16931173]
33. Motl RW, Dishman RK, Saunders R, Dowda M, Felton G, Pate RR. Measuring enjoyment of physical activity in adolescent girls. Am J Prev Med 2001 Aug;21(2):110-117. [Medline: 11457630]
34. Borg GA. Psychophysical bases of perceived exertion. Med Sci Sports Exerc 1982;14(5):377-381. [Medline: 7154893]
35. Ceci R, Hassmén P. Self-monitored exercise at three different RPE intensities in treadmill vs field running. Med Sci Sports Exerc 1991 Jun;23(6):732-738. [Medline: 1886482]
36. Russell WD, Weeks DL. Attentional style in ratings of perceived exertion during physical exercise. Percept Mot Skills 2017 May 30;78(3):779-783. [doi: 10.1177/003151259407800319]
37. Hart SG. Nasa-Task Load Index (NASA-TLX): 20 years later. Proc Hum Factors Erg Soc Ann Meet 2006 Oct 01;50(9):904-908 [FREE Full text] [doi: 10.1177/154193120605000909]
38. Agelink MW, Malessa R, Baumann B, Majewski T, Akila F, Zeit T, et al. Standardized tests of heart rate variability: normal ranges obtained from 309 healthy humans, and effects of age, gender, and heart rate. Clin Autonomic Res 2001 Apr;11(2):99-108. [doi: 10.1007/BF02322053]
39. Recommendations for surface electromyography.: Surface ElectroMyoGraphy for the Non-Invasive Assessment of Muscles URL: http://seniam.org/ [accessed 2019-07-15] [WebCite Cache ID 72hT8gokr]
40. Konrad P. EMG-Fibel: Eine Praxisorintierte Einführung in die Kinesiologische Elektromyopraphie. Scottsdale: Noraxon; 2011.
41. Möller D, Backes K. [Von der Biomechanik zur klinischen Anwendung]. Manuelletherapie 2017 May 19;21(02):88-95. [doi: 10.1055/s-0043-105164]
42. Hof A. The relationship between electromyogram and muscle force. Sportverletz Sportschaden 1997;11(3):79-86. [Medline: 9351163]
43. Schwartz M, editor. EMG Methods for Evaluating Muscle and Nerve Function. London: INTECH Open Access Publisher; 2012.
44. Gottlob A. Differenziertes Krafttraining: mit Schwerpunkt Wirbelsäule. 1 Aufl. Munich: Urban & Fischer; 2002.
45. Kibler W, Sciascia A, Uhl T, Tambay N, Cunningham T. Electromyographic analysis of specific exercises for scapular control in early phases of shoulder rehabilitation. Am J Sports Med 2008 Sep;36(9):1789-1798. [doi: 10.1177/0363546508316281] [Medline: 18469224]
46. Czaprowski D, Afeltowicz A, Gebicka A, Pawlowska P, Kedra A, Barrios C, et al. Abdominal muscle EMG-activity during bridge exercises on stable and unstable surfaces. Phys Ther Sport 2014;15(3):162-168. [Medline: 24268641]
47. Stephen S, Espen T. Intervals, thresholds, and long slow distance: the role of intensity and duration in endurance training. Sportscience 2009;13:32-53.
48. Howard MC. A meta-analysis and systematic literature review of virtual reality rehabilitation programs. Comput Hum Behav 2017 May;70:317-327. [doi: 10.1016/j.chb.2017.01.013]
49. Bun P, Gorski F, Grajewski D, Wichniarek R, Zawadzki P. Low-cost devices used in virtual reality exposure therapy. Procedia Comput Sci 2017;104:445-451. [doi: 10.1016/j.procs.2017.01.158]
50. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156 [FREE Full text] [Medline: 11153730]

51. Burd NA, Andrews RJ, West DWD, Little JP, Cochran AJR, Hector AJ, et al. Muscle time under tension during resistance exercise stimulates differential muscle protein sub-fractional synthetic responses in men. J Physiol 2012 Jan 15;590(2):351-362 [FREE Full text] [doi: 10.1113/jphysiol.2011.221200] [Medline: 22106173]
52. Sharman JE, La Gerche A, Coombes JS. Exercise and cardiovascular risk in patients with hypertension. Am J Hypertens 2015 Feb;28(2):147-158. [doi: 10.1093/ajh/hpu191] [Medline: 25305061]
53. Chrysant SG. Current evidence on the hemodynamic and blood pressure effects of isometric exercise in normotensive and hypertensive persons. J Clin Hypertens (Greenwich) 2010 Sep;12(9):721-726 [FREE Full text] [doi: 10.1111/j.1751-7176.2010.00328.x] [Medline: 20883233]
54. de Sousa Nery S, Gomides RS, da Silva GV, de Moraes Forjaz CL, Mion D, Tinucci T. Intra-arterial blood pressure response in hypertensive subjects during low- and high-intensity resistance exercise. Clinics (Sao Paulo) 2010 Mar;65(3):271-277 [FREE Full text] [doi: 10.1590/S1807-59322010000300006] [Medline: 20360917]
55. Gallagher M, Ferrè ER. Cybersickness: a multisensory integration perspective. Multisens Res 2018 Jan 01;31(7):645-674. [doi: 10.1163/22134808-20181293] [Medline: 31264611]
56. Chen S, Liu M, Cook J, Bass S, Lo SK. Sedentary lifestyle as a risk factor for low back pain: a systematic review. Int Arch Occup Environ Health 2009 Jul;82(7):797-806. [doi: 10.1007/s00420-009-0410-0] [Medline: 19301029]
57. Brunnström K, Wang K, Tavakoli S, Andrén B. Symptoms analysis of 3D TV viewing based on Simulator Sickness Questionnaires. Qual User Exp 2016 Dec 8;2(1). [doi: 10.1007/s41233-016-0003-0]
58. Walch M, Frommel J, Rogers K, Schüssel F, Hock P, Dobbelstein D, et al. Evaluating VR driving simulation from a player experience perspective. 2017 Presented at: Proceedings of the 20 CHI Conference Extended Abstracts on Human Factors in Computing Systems; 2017; Denver p. 2982-2989.
59. Chen C, Jeng M, Fung C, Doong J, Chuang T. Psychological benefits of virtual reality for patients in rehabilitation therapy. J Sport Rehabil 2009 May;18(2):258-268. [Medline: 19561368]
60. Kolasinski E. Simulator sickness in virtual environments. Alexandria: Army Research Institute for the Behavioral and Social Sciences; 1995. Simulator sickness in virtual environments URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a295861.pdf [accessed 2019-07-15]
61. Lavalle SM. Virtual Reality. 2017. URL: http://vr.cs.uiuc.edu/ [accessed 2019-07-15]
62. Riccio GE, Stoffregen TA. An ecological theory of motion sickness and postural instability. Ecol Psychol 1991 Sep;3(3):195-240. [doi: 10.1207/s15326969eco0303_2]
63. Smart LJ, Otten EW, Strang AJ, Littman EM, Cook HE. Influence of complexity and coupling of optic flow on visually induced motion sickness. Ecol Psychol 2014 Oct 28;26(4):301-324. [doi: 10.1080/10407413.2014.958029]
64. Palmisano S, Mursic R, Kim J. Vection and cybersickness generated by head-and-display motion in the Oculus Rift. Displays 2017 Jan;46:1-8. [doi: 10.1016/j.displa.2016.11.001]
65. Ash A, Palmisano S, Kim J. Vection in depth during consistent and inconsistent multisensory stimulation. Perception 2011;40(2):155-174. [doi: 10.1068/p6837] [Medline: 21650090]
66. Robergs RA, Landwehr R. The surprising history of the "HRmax=220-age" equation. J Exer Physiol Online 2002;5(2):1-10.
67. Heneweer H, Staes F, Aufdemkampe G, van Rijn M, Vanhees L. Physical activity and low back pain: a systematic review of recent literature. Eur Spine J 2011 Jun;20(6):826-845 [FREE Full text] [doi: 10.1007/s00586-010-1680-7] [Medline: 21221663]

Abbreviations

MVIC: maximum voluntary isometric contraction
NASA-TLX: NASA Task Load Index
PACES: Physical Activity Enjoyment Scale
RPE: rate of perceived exertion
sEMG: surface electromyography
SSQ: Simulator Sickness Questionnaire
VR: virtual reality
WHO: World Health Organization

Edited by G Eysenbach; submitted 27.09.18; peer-reviewed by W Meekes, J Salisbury; comments to author 01.04.19; revised version received 27.05.19; accepted 28.06.19; published 28.08.19.

Please cite as:
Feodoroff B, Konstantinidis I, Froböse I
Effects of Full Body Exergaming in Virtual Reality on Cardiovascular and Muscular Parameters: Cross-Sectional Experiment
JMIR Serious Games 2019;7(3):e12324
URL: http://games.jmir.org/2019/3/e12324/
doi:10.2196/12324
PMID:31464194

Original Paper

Usability, Acceptability, Feasibility, and Effectiveness of a Gamified Mobile Health Intervention (Triumf) for Pediatric Patients: Qualitative Study

Riin Tark[1,2], MA; Mait Metelitsa[1,2], MA; Kirsti Akkermann[1], PhD; Kadri Saks[3], MD; Sirje Mikkel[4], MD; Kadri Haljas[2], PhD

[1]Institute of Psychology, University of Tartu, Tartu, Estonia

[2]Triumf Research OU, Tartu, Estonia

[3]Department of Oncology and Hematology, Clinic of Pediatrics, Tallinn Children's Hospital, Tallinn, Estonia

[4]Department of Hematology and Bone Marrow Transplantation, Clinic of Hematology and Oncology, Tartu University Hospital, Tartu, Estonia

Corresponding Author:
Riin Tark, MA
Triumf Research OU
Kannikese 33-1
Tartu, 50408
Estonia
Phone: 372 5078577
Email: tarkriin@gmail.com

Abstract

Background: Mental disorders are notably prevalent in children with chronic illnesses, whereas a lack of access to psychological support might lead to potential mental health problems or disruptions in treatment. Digitally delivered psychological interventions have shown promising results as a supportive treatment measure for improving health outcomes during chronic illness.

Objective: This study aimed to evaluate the usability, acceptability, and feasibility of providing psychological and treatment support in a clinical setting via a mobile game environment. In addition, the study aimed to evaluate the preliminary effectiveness of the mobile health game.

Methods: Patients aged 7 to 14 years with less than a year from their diagnosis were eligible to participate in the study. In total, 15 patients were invited to participate by their doctor. A total of 9 patients (age range: 7-12 years; mean age 9.1 years) completed the 60-day-long study in which the Triumf mobile health game was delivered as a digital intervention. In an engaging game environment, patients were offered psychological and treatment support, cognitive challenges, and disease-specific information. The fully digital intervention was followed by a qualitative interview conducted by a trained psychologist. The results of the interview were analyzed in conjunction with patient specific in-game qualitative data. Ethical approval was obtained to conduct the study.

Results: Patients positively perceived the game, resulting in high usability and acceptability evaluations. Participants unanimously described the game as easy to use and engaging in terms of gamified activities, while also providing beneficial and trustworthy information. Furthermore, the overall positive evaluation was emphasized by an observed tendency to carry on gaming post study culmination (67%, 10/15). Psychological support and mini games were the most often used components of the game, simultaneously the participants also highlighted the education module as one of the most preferred. On average, the patients sought and received psychological support or education on 66.6 occasions during the 60-day intervention. Participants spent the most time collecting items from the city environment (on average 15.6 days, SD 8.1), indicative of exploratory behavior, based on the quantitative in-game collected data. During the intervention period, we observed a statistically significant decrease in general health problems (P=.003) and saw a trend toward a decrease in depression and anxiety symptoms.

Conclusions: This study demonstrated that a game environment could be a promising medium for delivering comprehensive supportive care to pediatric patients with cancer alongside standard treatment, with potential application across a variety of chronic conditions. Importantly, the results indicate that the study protocol was feasible with modifications to randomized controlled trials, and the game could be considered applicable in a clinical context. By giving an empirical evaluation of delivering psychological support via the game environment, our work stands to inform future mobile health interventions.

(*JMIR Serious Games 2019;7(3):e13776*) doi:10.2196/13776

KEYWORDS

psychological stress; coping skills; psychological feedback; mobile app; mHealth; mental health; chronic illness; cancer; pediatrics

Introduction

Background

The prevalence of chronic illnesses in children is on the rise, with up to a quarter of children in the general population having at least 1 chronic illness [1], including pediatric cancer. The situation is further complicated as up to 60% of chronically ill children have at least 1 co-occurring mental disorder [2], compared with the 10% to 20% prevalence in the general population [3,4]. Indeed, it has been well established that chronic illnesses predispose children for higher risk of developing mental disorders such as anxiety, depression, and behavioral problems [2,5].

Psychological support, although generally seen as an integral part of comprehensive and effective care [6,7], has not yet been unified across hospitals [8], leading to large differences in access to psychological care across regions. Differences in the availability and quality of psychological support increase the likelihood of leaving psychological problems unattended, which in turn may have long-term negative health effects and interfere with the treatment process of the underlying chronic illness [9]. Furthermore, untreated psychological problems may affect treatment compliance in the pediatric care setting [10,11] and may carry on into adulthood [12]. This attests to the need for the further understanding of psychological problems co-occurring with chronic illnesses, both at the level of risk factors and disease development but also in the search for effective novel intervention strategies. Timely and accessible evidence-based psychological support in the pediatric care setting might be a crucial factor in achieving desirable treatment outcomes.

The factors leading to mental burden among chronically ill patients are largely universal across different conditions, encompassing, for example, changes in daily routines, stressful states related to treatment procedures, and psychological uncertainty [2,13]. Relatedly, psychological problems are comparable between general and chronically ill populations, involving mostly symptoms associated with anxiety, depression, and behavioral problems [2,5,14]. Hence, various traditional intervention strategies aimed at reducing mental burden (eg, psychoeducational programs, solution-focused brief therapy, cognitive behavioral therapy [CBT], and mindfulness-based interventions) have been used interchangeably between chronically ill patients and patients without chronic illness [4,11,15-18].

In addition to traditional interventions, digital tools have been shown to be a promising avenue in delivering psychological support to patients. Currently, there are only a few mobile health (mHealth) apps for children with cancer that are publicly available, including: Pain Squad for pain management [19]; Re-Mission2 for patient empowerment [20]; and Super K where kids can fight cancer cells according to the game description [21]. However, these solutions have not integrated psychological care for patients and are primarily used for pain monitoring or empowering.

Rathbone et al [22] reviewed mHealth apps that used CBT principles for several psychological conditions and concluded that although apps can be effective tools in a health setting, the need for further studies that evaluate the effectiveness of various mHealth solutions is imminent [23]. Similarly, another recent review brought out the mental health app efficacy evaluation as a concern, showing that 38% of app store descriptions included phrases related to claims of effectiveness, whereas only less than 3% provided scientific evidence for such declarations [24]. Furthermore, previous findings also highlight the need to improve user engagement [25]. As an example of user engagement, the gamified version of smartCAT solution was found to be more effective than the nongamified version in delivering brief CBT treatment [26], therefore suggesting that gamification could be effective in achieving desired mHealth platform effectiveness targets. If the solution is effective, it can facilitate desired behavior changes [27] and ultimately lead to improved health outcomes.

Against this background, this study aims to assess the usability, acceptability, and feasibility of a gamified mHealth intervention, Triumf, both as a whole and by its individual constituent components. Furthermore, this paper explores the preliminary effectiveness findings of the mHealth game, Triumf, whereas further analysis with a pretest-posttest design and without randomization of the participants is published as a master thesis [28].

Mobile Health Game, Triumf

A newly developed digital health intervention, called Triumf, aims to reduce the negative psychological changes associated with chronic illness through an mHealth game. For clarity purposes, *intervention* or (mobile) *game* will also be used interchangeably to refer to the Triumf digital health intervention from hereon in. Importantly, the game has been designed in cooperation with pediatric patients with cancer, their parents, and care teams to determine illness-related burdening factors to develop a solution with maximum relevance to patients. Creating a game environment in collaboration with the key stakeholders has allowed building an intervention that approaches children in a way that is familiar to them. In addition, patients may be more empowered to comply with treatment if the intervention is delivered in a way that approaches the treatment process from a new angle [29]. By delivering care through a safe and familiar game environment, it is possible to fill in the gap between attractiveness and effectiveness, a common challenge in digital health solutions. Furthermore, as children in general are increasingly using their mobile devices for day-to day socializing and free-time activities, a new and still unused avenue for interacting, supporting, and educating pediatric care patients has been opened.

Together with the cooperative input from the partners and previous research literature, the intervention has been designed to serve as a basis to deliver effective psychological support that is universally applicable across different chronic illnesses, while maintaining the platform-level flexibility to consider disease- and region-specific differences. Thus, the overall gameplay, mechanics, and setup of the game and its principles are also applicable to other chronic illnesses, such as diabetes and asthma. The aim of the intervention was to offer patients psychological support and information about their health condition and emotions that would (1) help children to better understand their new health condition, (2) offer children external support to promote internal motivation to better cope with the illness, (3) offer children cognitive challenge and distraction, but also activity-based learning of healthy behaviors, (4) profile and screen the psychological well-being of children and offer psychoeducation and coping techniques accordingly, and thereby (5) support the formation of better self-understanding and constructive health behaviors, such as physical activity and diet. In short, through the continuous screening and support of various aspects, the intervention seeks to identify, prevent, and lessen the potential psychological problems and support behavioral change.

Triumf intervention covers various aspects of comprehensive care, and the game user experience is personalized and dynamic. The game learns and adapts with the user and offers individual and targeted experience. The intervention follows a predetermined structure in onboarding where the players are guided through the game narrative. Further gameplay, that is, accessing the educational module, entertainment games, and other elements of the intervention, is determined by the in-game choices made by the player. Furthermore, provision of psychological support is dynamically dependent on the individual emotional state of the patient.

Main Theoretical Background of the Game

An inferior understanding of emotions has been found to be a risk factor for developing psychopathology [30] and poor coping or adaptation to illness [13]. Furthermore, future health of the chronically ill individuals is related to general health behaviors. Thus, a shift toward healthier behaviors is crucial [31]. The interactive gamified setting in the Triumf intervention puts theories of emotions [32,33], coping [13], behavior, and behavior change [34] to practical use and presents as an educational module of the game. The intervention also consists of several mini games that include games related to the application of the in-game learned information and cognitive challenges, as well as entertainment games that offer cognitive distraction. The emphasis of the game is on the storyline—saving the Triumfland City by finding one's inner superpowers and taming the Disease Monster—to achieve player engagement and connectedness, also bringing personal meaning to the game [35].

In addition, the game combines the Self-Determination Theory (SDT) and the Player Experience of Need Satisfaction model [36] to increase autonomy and the player's competency experience in the game. SDT suggests that when the 3 core needs, autonomy, competence, and relatedness, are satisfied, they promote psychological health and intrinsic motivation [37]. It is argued that enhancing SDT needs in a game results in motivation to play [36]. As described by Tark [28], the Triumf intervention could enhance the following: (1) autonomy through offering noncontrolling guidelines and flexibility in choosing the flow of tasks and goals and by using in-game rewards as feedback instead of behavior-controlling mechanisms [36]; (2) competence through broadening knowledge about the illness and the importance of treatment adherence by using rewards and praise for successfully completed health-related actions and by keeping the players optimally challenged (eg, the possibility to choose the difficulty level in mini games) [36]; and (3) relatedness by creating an environment the player can relate to (eg, inclusion of illness-related but also regular child activities) and play against the game (eg, interacting with and helping city kids and playing tic-tac-toe against artificial intelligence). The visual representation of the 3D game, Triumf, is presented in Figure 1 and Multimedia Appendix 1, and individual modules and their theoretical background are presented in Table 1.

Figure 1. Visual representation of the intervention, displaying a screenshot: (a) from the introduction to the game (storyline), (b) from customization, and (c) from the obstacle course mini-game.

Table 1. Overview of the intervention modules.

Module	Rationale	Description
Screening module		
Mental state	Profiles and screens to create preconditions for support [38]	Questions to form the player's profile are prompted during the onboarding of the game. After the onboarding, the question how one feels at the moment is prompted daily. Two or more questions per day about symptoms of depression, anxiety, attention problems, and general health (well-being) are prompted depending on the player's profile. In addition, well-being questions are accessible to the players throughout the game, ie, more than 2 questions per day can be answered by the player.
Motivation and attitudes	Offers more specific psychological targets [36,39]	SDT[a] questions about general attitude toward health, autonomy, competence, and relatedness are monitored once a week.
Educational module		
	Offers relevant information about the illness, treatment rationale, potential side effects, and hospital environment so that the child could be motivated and an informed participant in his or her treatment process [39,40]	After the onboarding, educational module is accessible to the player throughout the game. The topics are not presented in a predetermined order, ie, they appear based on the choices made by the player. The educational module presents each topic at 2 levels. First, the general introductory overview of the topic at hand is presented to the player. After the completion of the introductory level, more challenging in-depth educational description about the topic is presented to the player. This is followed by self-control questions that allow for the assessment and feedback about the acquisition of new information.
Support module		
Psychological support	On the basis of the child's profile, the module offers psychoeducation and coping techniques [4,11,15-18,41]	Symptoms of depression, anxiety, attention problems, and general health questions all have 3 possible answers—sometimes, often, and rarely—based on which psychoeducation, psychological techniques, or praise is offered.
Health behavior change	Motivates children to learn and engage in health-promoting behaviors through an educational module and content-relevant mini games (eg, a mini game that reminds the child to keep oneself well hydrated) [27,34]	Progressing in certain mini games requires applying information learned in educational module, which supports the motivation to engage in healthy behaviors and facilitates the consolidation of the acquired information.
Emotion regulation	Helps children learn about identifying and regulating emotions [32,33,42,43]	Educational module includes information about 6 basic emotions, which the player also has to recognize in city kids. The circumplex model helps player to become better at identifying and relating to various emotional states. To be more specific, circumplex model, based affective state-space, gives the player a mental model to understand and relate to different specific emotional states, thus possibly facilitating cognitive top-town emotional regulation.
Coping	Helps children better understand their new health situation [13]	Educational module information about one's health situation and treatment procedures helps to normalize the daily challenges, wherein the support module offers ways how to cope with those challenges.
Mini games module		
Storyline	Engages and fosters learning [35]	The storyline is introduced to the player during the onboarding.
Activities common among children	Offers regular activities experience through entertaining games (eg, football), as children may be excluded from their social environment	Mini games are always accessible to the player after the onboarding.
Cognitive distraction and cognitive challenge	Offers distraction and challenges through mini games such as puzzles, tic-tac-toe, and memory game	Mini games are always accessible to the player after the onboarding.

[a]SDT: Self-Determination Theory.

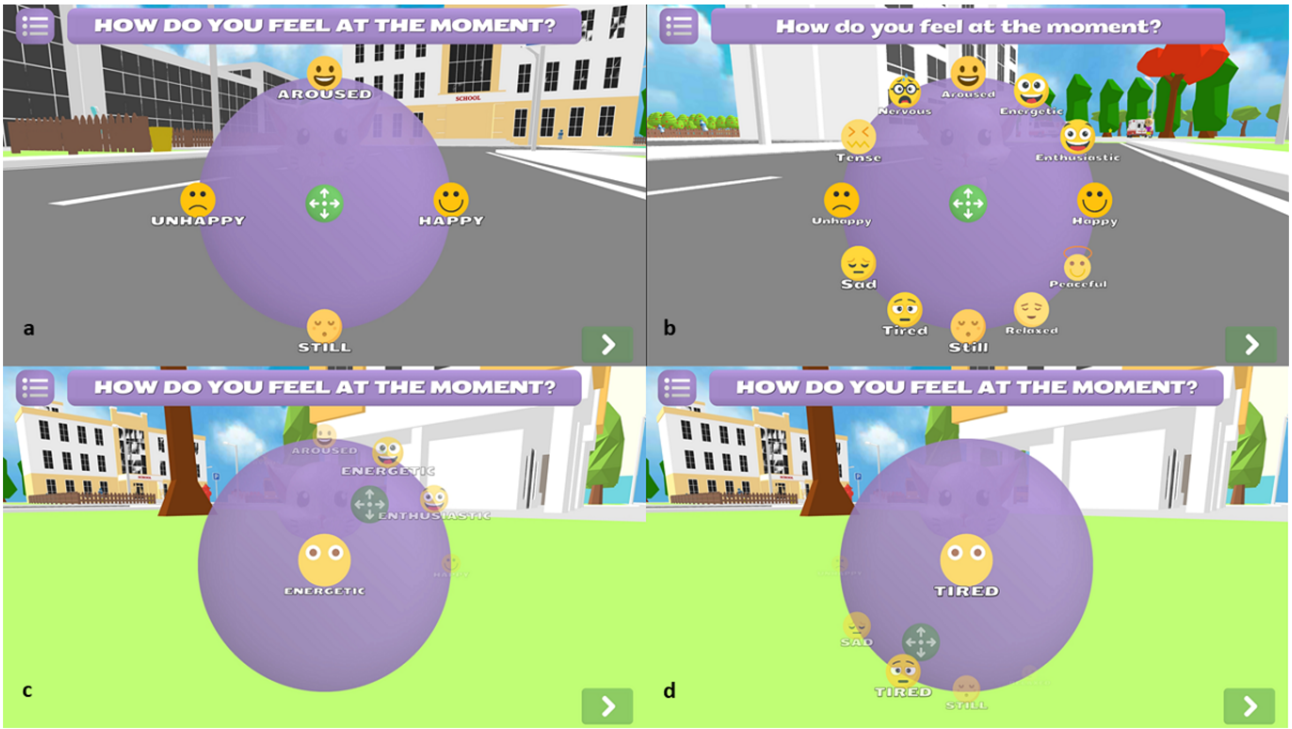
Screening, Support, and Educational Module

The rationale of psychological support provided in the game is based on continuous screening and monitoring of the player (patient). Questions about symptoms of depression, anxiety, attention, and general health are included in the game, wherein the interval and number of questions is based on the player's profile. Questions addressing symptoms of depression, anxiety, and attention problems are based on DSM-5 diagnostic criteria; general health items include questions about hygiene, physical health, healthy eating, and sleep. The profile is created during the onboarding and is based on the answers given to 8 questions about emotional and attention state [32]. The player, based on the profile, is then categorized in one of the following: in good psychological functioning (minimum amount of screening and psychological support), in need of psychological support for emotional problems (more frequent screening and psychological support), and in need of psychological support in more than one area (more frequent screening and psychological support). All support questions have 3 possible answers (sometimes, often, and rarely) and after each answer, psychoeducation, psychological coping techniques, or praise is offered. The content of the psychological support constitutes the gamified versions of established evidence-based therapeutic methods, including mindfulness as well as CBT-based techniques, relaxation methods, and breathing exercises—all of which are commonly and successfully used in the context of chronic illnesses [11,15,16,41].

In addition, questions about how one feels at the moment, how one slept the previous night, and one’s motivation and attitude toward health (SDT) are screened to complement the monitoring of the players’ well-being. A circumplex model [42] is used to probe the player’s emotional state by letting the player interactively indicate his or her emotional state by making a choice between specific emotional states that are situated in the affective state-space circumplex created by the interaction of core affect dimensions arousal (still-aroused) and valence (happy-unhappy). Specific emotional states have been indicated by the corresponding emoji figures (Figure 2). A question about how one slept the previous night compared with the average is prompted daily, as problems with sleep are common among chronically ill patients with accompanying psychological problems [33]. SDT questions about the general attitude toward health, autonomy, competence, and relatedness (health care climate; modified Treatment Self-Regulation Questionnaire [34]) are asked once a week to learn more about the needs and internalization of in-game learnt behaviors [39].

As one of the aims of the intervention is to induce behavioral change, it is essential to educate patients to support better coping, enhance resilience, and make informed decisions regarding their health and well-being [35]. By following the educational module, children learn about emotions, the illness itself, the treatment and its side effects, the care team, and social interactions with friends and family that may be affected by long-term hospital stay. The module is organized by levels that correspond to the Bloom taxonomy [40], which is often used as an underlying theoretical basis in pedagogics. Level 1 sets out to give information in a descriptive manner to facilitate initial learning. Level 2 gives further details on these topics. After reading the detailed material, several questions are asked to facilitate an understanding of the topic. Higher stages of the educational game include interactive modules to facilitate the carryover of new information from the semantic level to real-life related instrumental behaviors (ie, applying knowledge in content-relevant mini games, eg, identifying emotions of the citizens).

Figure 2. Presentation of specific emotional states in emotional state-space circumplex created by core affect dimensions valence and arousal: (a) displays appearance of the prompted question, (b) displays all possible answers, whereas (c) and d) display the appearance of answers when moving towards sectors of the circumplex.



Objectives of the Study

This study has 2 main objectives. First, to evaluate the usability, acceptability, and potential preliminary effectiveness of the intervention among pediatric cancer patients. Second, to assess the feasibility of the study protocol of administering the intervention without the randomization of patients.

Of note, the usability and acceptability assessments were collected both at the level of the whole game and at the level of constituent components. Furthermore, the preliminary assessment of possible beneficial effects on the well-being of pediatric patients with cancer was carried out.

To fulfill the main study aims, the following analytical steps were carried out: (1) assessment of the relations between participants' game behavior, medical treatment, self-reported well-being, motivation and attitudes, sleep, and evaluations on the intervention, (2) assessment of the perceived usefulness, ease of use, and enjoyment of the intervention as a whole and at component level, (3) general evaluation of cooperation and process with hospitals, and (4) general evaluation of applicability of the game in clinical context.

Methods

Participants

All pediatric patients with cancer aged between 7 and 14 years with a new or recurrent diagnosis of cancer, diagnosed no more than 1 year ago, were eligible to participate and were invited to the study (Figure 3). Participants were recruited by their medical doctors from Tallinn Children's Hospital and Tartu University Hospital within a 6-month period starting from June 2018. Of the invited 15 children, 10 agreed (10/15, 67%) to participate in the study, with 1 participant withdrawing during the intervention owing to unfamiliarity with the game interface and without willingness to familiarize oneself. Thus, 90% of participants completed the study, and the sample used for analysis consisted of 9 pediatric patients with cancer with the average age of 9.1 years (SD 1.5; range 7-12), including 4 girls (44%) and 7 with Estonian as their native language (77%). Together with the withdrawn patient, 6 patients in total declined or withdrew their participation, with a mean age of 10.5 years (SD 2.4; range 7-14), 2 of them being girls (33%) and four (67%) having Estonian as their native language. Patients had various reasons for declining participation, where not being interested in the study or in mobile games was reported most frequently (4/6, 67%). However, the sample used for the final analysis did not statistically differ from the patients who decided not to participate or withdrew from the study (Table 2).

Figure 3. Participant flow diagram.

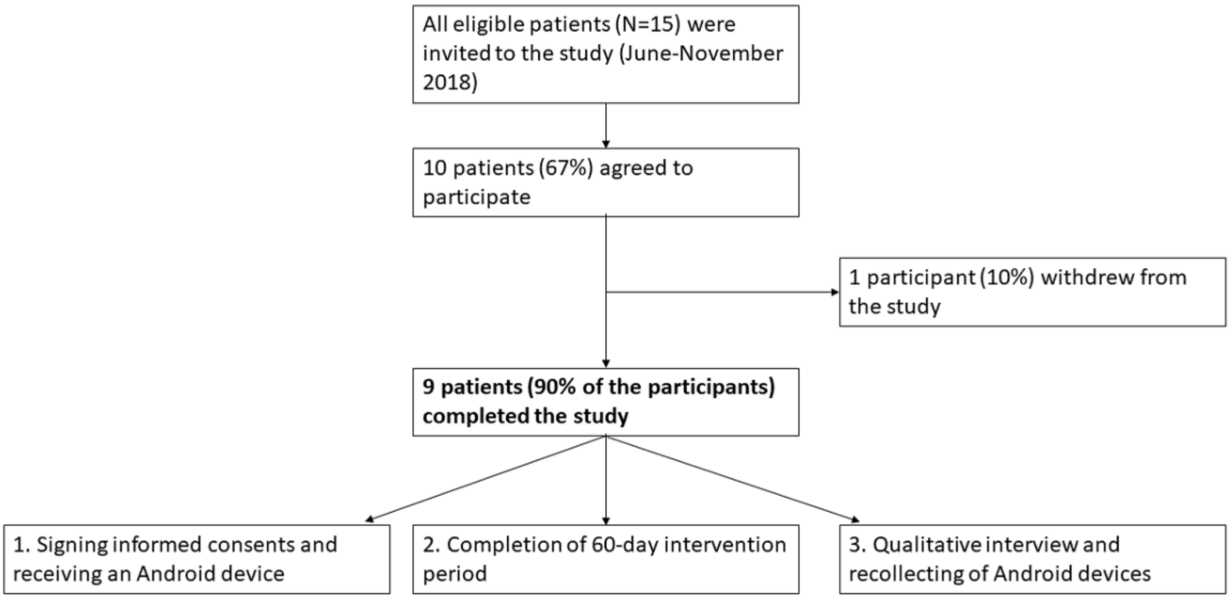


Table 2. Sample characteristics.

Variable	Patients who declined participation	Analytic sample	*P* value of declined patients vs analytic sample
Sex, n (%)			
Girls	2 (33)	4 (44)	.67
Boys	4 (67)	5 (56)	.67
Age (years), mean (SD)	10.5 (2.4)	9.1 (1.5)	.18
Language, n (%)			
Estonian	4 (67)	7 (78)	.63
Russian	2 (33)	2 (22)	.63
Diagnosis category, n (%)			
CNS[a] tumor	N/A[b]	1 (11)	N/A
Leukemia	N/A	6 (67)	N/A
Other	N/A	2 (22)	N/A
Treatment status, n (%)			
Newly diagnosed	N/A	1 (11)	N/A
On treatment	N/A	6 (67)	N/A
Recurrent	N/A	2 (22)	N/A

[a]CNS: central nervous system.

[b]N/A: not applicable.

Each child and their parent gave written informed consent to participate in the study. The Research Ethics Committee of the University of Tartu (decision 283/T-32) approved the study protocol. Recruiting and participation in the study processes did not have deviations from the study protocol.

Procedure

The study protocol involved a predetermined 60-day intervention period with a suggested 10 min of gameplay per day. The game was available in both Estonian and Russian. Android-operating smartphones were prepared and provided for the study period by the research team, and all participants had the option to continue the intervention following study completion using their own Android devices. Updates to the game were provided every other week to improve user engagement through new content. Technical support and information about game updates were offered to participants primarily via email.

Medical doctors recruited the patients to the study based on the eligibility criteria. Subsequently, information about the treatment and a health status for each participant were obtained from the medical doctors. This information included a general evaluation of the participant's medical status during the study period, diagnosis and time from the diagnosis, notification of medical treatments that could affect mood, affect cognitive functions, or cause fatigue, time spent at the hospital according to the treatment plan, and outside planned treatment. Time from diagnosis, treatment that could affect participant's capabilities, and total time spent at the hospital and at home were explored in the analysis. There were no deviations from the initial study protocol.

Assessment

In-Game Assessment

Game activity and self-reported well-being, sleep quality, motivation, and attitudes were included in the analysis. The average scores for symptoms of depression, anxiety, attention problems, and general health across the study period were obtained as well-being indicators to observe possible relations with the SDT score (motivation and attitudes), game behavior, medical treatment, and intervention evaluation (derived from the qualitative interview). In addition, average scores in the beginning (first week average) and in the end (last week average) of the study period were compared to test the potential preliminary effectiveness of the intervention, with lower scores indicative of more problems.

Presentation of the sleep and SDT questions on a visual red-green gradient scale (slider) was coded on a 7-point scale such that higher scores indicate better than average sleep or better general attitude toward health, perceived autonomy, competence, and health care climate. A study period average and the first and last week average sleep score and total score of SDT (average of the questions answered) were calculated.

The progress in educational module was not included in the analysis owing to technical limitations despite being accessible to participants. The full play time from opening to closing of the intervention was also not observed owing to technological restrictions, and therefore, alternative variables indicating game activity were included in the analysis. These included total time spent in all mini games, total amount of collectables gathered (indication of exploratory behavior), total amount of interactions with citizens initiated (helping citizen maintain hydration and dental hygiene), and number of days when the game was used.

Evaluations on the Intervention and Study Process (Qualitative Interview)

A semistructured qualitative interview in the native language of the patient was held after completion of the 60-day study period. Trained psychologists conducted an oral qualitative interview with each participant and one of their parents. The interview was a structured evaluation of the usability and acceptability of the game following Marsac et al [44] and covered the following topics: (1) general attitude toward the game, including negative feedback, (2) participant's subjective evaluation of the game (ie, whether it was beneficial, improved subjective well-being, helped to better follow the treatment plan, offered distraction from treatment, and was empowering), (3) evaluation of the intervention's usability (whether it was easy to use, would they play more, were the guidelines clear, was there too much text, and did they learn something new), (4) acceptability of the game (whether it was fun to use, were the visuals nice, did the game have nice quests and mini games, nice characters, did it contain beneficial information, was the information accurate in their opinion, would they recommend the game to other children, and were the characters like oneself or like other children with cancer), and (5) evaluation on the study processes in general. There were no deviations from study protocol regarding the semistructured qualitative interviews. Questions regarding the general attitude toward and subjective evaluations of the game, as well as the evaluation of the study process, were asked in an undirected and neutral manner (eg, *What did you like about this game?*, *What you did not like?*, and *Do you think the game was beneficial?*). For usability and acceptability evaluations, participants were instructed to give yes/no answers to the neutrally read out statements (eg, *Triumf game was easy to use* and *Triumf game had too much text to read*). In addition, parents were asked to evaluate the general process of participating in the study. A summary of the interview answers, captured verbatim, was forwarded to the research personnel for analysis.

The aforementioned questions, to be included in the analysis, were coded as follows: clear *yes* answers as 1, expressions *yes, liked a little* or similar as 0.5, and clear *no* or *not really* as 0. Negatively formulated questions were reversed, which means higher scores indicated higher evaluations. General commentary on the game and evaluation of the study process were analyzed qualitatively.

From the subjective evaluations on the game, only 1 question (whether the game was beneficial) was answered by every participant (100%) and was rated beneficial unanimously. The remaining questions had missing values on several occasions, owing to some of the concepts (eg, *benefitting one's well-being*) being difficult to understand by the young participants, and were thus excluded from further analysis. Usability questions were answered by all participants. From 8 acceptability questions, all participants answered only 2 questions, owing to difficulties in understanding by younger participants. One participant gave acceptability answers for half of the questions (ie, for 4) and was removed from further analysis concerning the acceptability aspect. To evaluate subjective evaluations, usability, and acceptability, the remaining data set was analyzed as described below.

Statistical Analyses

International Business Machines SPSS version 20.0 [45] and R free software environment [46] were used for data analysis. Owing to this study's exploratory nature, aiming to provide relevant research questions for further similar studies, we have reported tendencies as measure averages and used small group size–based statistical comparisons and correlations. Tendencies and statistically significant results based on our small sample provide valuable insights that could be meaningful and therefore could be more stringently tested in future studies that employ larger sample sizes.

In detail, a Pearson correlation was used to evaluate possible associations between age, total time of all mini games played, exploring around the city environment (total of collected items), helping city kids to maintain healthy behaviors (total of interactions), number of days when the intervention was used, treatment that could affect mood (in days), time spent at home (in days), time spent at the hospital (in days), time from diagnosis (in months), SDT average score, study period average depressive symptoms, study period average anxiety symptoms, study period average attention problems, study period average self-reported general health score, average sleep score, score of questions answered in the qualitative interview (questions answered yes divided by total number of questions answered), subscore of usability from the qualitative interview, and subscore of acceptability from the qualitative interview. Goodman and Kruskal *gamma* was used for exploring relations between sex and the aforementioned variables, as well as treatment complexity and noted variables. A paired sample (dependent) *t* test was used to evaluate group differences of continuous variables and a chi-squared test was used for evaluating differences in dichotomous variables. Covariates were not included in the analyses. A conventional cut-off point for 2-tailed significance ($P<.05$) was used.

Results

Sample Characteristics

Sample characteristics are presented in Table 2. The average time from diagnosis to recruiting was 5.3 months (SD 4.1, range 0-12 months). Participants received medical treatment that could potentially influence emotional state and cognitive functions on 12.7 days (SD 18.5, range 0-60), spent 17.2 days at the hospital (SD 15.8, range 0-37), and 42.8 days at home (SD 17.8, range 23-60). During the study period, 8 participants out of 9 (89%) received chemo or hormone therapy, and 1 participant received the last treatment right before the study period. Out of the 8 participants, 4 (44%) had a more complex treatment regimen (eg, presence of infection in addition to main diagnosis). The group of participants who had a more complex treatment versus the group who received regular treatment presented a statistically significant difference in total time spent playing mini games (*gamma*=−.70, P=.02), that is, participants with more complex treatment spent less time playing mini games. No systematic psychological support in the sample was received during the study period.

In-Game Reported Well-Being Overview

Within the intervention environment, participants were divided into 3 categories on the basis of their profile of existing problems and strengths, resulting in 8 participants falling into the good psychological functioning category and 1 participant into the need for psychological support in more than 1 area category. On the basis of this, the majority of participants (8/9, 89%) received psychological screening and support approximately 2 times a day, with the exception of 1 patient, who received support more often. The participant who was categorized into the greater psychological support group had more depressive and anxiety symptoms, the lowest SDT score, and gave the lowest scores during the qualitative interview, but spent less time playing the game compared with the other group.

General comparisons (at the level of average scores) between the beginning and the end of the study period showed that there were no statistically significant changes in the in-game self-reported depression, anxiety, and attention problem symptoms, although a very slight improvement in depressive symptoms (from 1.15 to 1.19) and anxiety symptoms (from 1.46 to 1.50) was observed, accompanied with a very slight decrease in attention problems (from 1.31 to 1.29). For self-reported general health problems, there was a statistically significant change resulting in less problems (t_7=−4.4; *P*=.003).

In-Game Reported Sleep

The average sleep score of the sample was 5.1 (SD 0.6) from a total of 7, indicating the best sleep quality compared with the average. Comparisons between the beginning (first week average) and the end (last week average) of the study period showed that sleep quality improved from the average score of 5.05 to 5.53, although the results were not statistically significant.

Motivation and Attitudes

Out of the 14 SDT questions, on average, 8.1 questions were answered (range 1-12) and the average SDT score for all participants was 5.4 (range 3.7-7). Comparisons between the beginning (first week average) and the end (last week average) of the study period showed that SDT scores remained stable over the course of the intervention ranging from 6.13 in the beginning of the study to 6.35 in the end. The lower average score of 5.4 for all participants was a result of greater variability across the study period.

Game Activity

On the basis of the quantitative in-game data, participants used the game on average on 20.2 days (SD 9.4), answered on average 66.6 (SD 51.6) support questions, and received support accordingly. Of specific modules, an obstacle course mini game was used on an average of 8.6 days (SD 4.3), a memory game on 6.1 days (SD 3.8), and a medication labyrinth mini game on 7.0 days (SD 3.9). Participants collected stars on 15.6 days (SD 8.1), water on 16.4 days (SD 9.2), and toothbrushes on 15.7 days (SD 8.8); and helped other citizens to stay hydrated and take care of dental health on 13.6 days (SD 7.6). The total average time spent playing mini games was 25.4 min (SD 19.6), total average amount of all collectables was 199.3 (SD 121.6), and helping citizens was initiated on an average of 74.2 times (SD 67.9).

Several statistically significant associations between the in-game data, medical treatment, and evaluations on intervention are presented in Multimedia Appendix 2. Symptoms of anxiety were positively correlated with usability evaluations (*r*=0.71; *P*=.03). General health problems and exploring around the city collecting different items was negatively correlated (*r*=−0.71; *P*=.03). In addition, less general health problems were reported when participants spent more time at the hospital (*r*=0.73; *P*=.03). Sleep score was positively correlated with acceptability evaluation of the intervention (*r*=0.67; *P*=.049).

There was a positive correlation between the SDT total score and time from diagnosing (*r*=0.70; *P*=.04). In addition, participant's sex was negatively related to SDT score (*gamma*=−.80; *P*<.001), that is, boys had lower SDT scores.

Age and usability evaluation for the intervention were positively correlated (*r*=0.72; *P*=.03). Participant's sex was positively related to the total amount of collectables (*gamma*=.60; *P*=.046), which means that boys collected more items. In addition, the more time participants spent at home during the study period, the more they explored around in the city collecting different items (*r*=0.87; *P*=.002). There were also statistically significant correlations between time from diagnosis and days spent at home (*r*=0.68; *P*=.04) and between game-related data (eg, the amount of collectables was related to helping citizens, *r*=0.71; *P*=.03).

Evaluations on Intervention

Quantified usability and acceptability evaluations are presented in Table 3. All participants evaluated that the game was easy to use (100%) and 7 out of 9 (78%) would play it again. About 78% (7 out of 9) concluded that the instructions of the game were clear and that they learned something from the intervention, whereas 56% (5 out of 9) thought that there was too much to read in the educational module.

Table 3. Usability and acceptability evaluations for the intervention.

Evaluations	Value, n (%)
Usability	
Number of participants in analysis	9 (100)
Triumf game was easy to use	9 (100)
Triumf game instructions were not confusing	7 (78)
There was not too much to read in Triumf game	4 (44)
I learned something new from Triumf game	7 (78)
I would play Triumf game more	7 (78)
Acceptability	
Number of participants in analysis	8 (100)
I liked Triumf game visuals	7.5 (94)
I liked Triumf game activities	8 (100)
I liked Triumf game characters	7.5 (94)
Triumf game contained beneficial information	8 (100)
Triumf game information was trustworthy	8 (100)

Acceptability evaluation was based on the total of 8 participants' answers. All of them (100%) concluded that they liked Triumf game activities and that the intervention contained beneficial information and that the intervention was trustworthy; 94% (7.5 out of 8) liked the visuals and the characters of the intervention.

The qualitatively analyzed subjective free form feedback included both positive and negative feedback. Interviews showed that each participant liked the different modules of the game the most. Specifically mentioned by the participants were the features of the memory game (1/9, 11%), obstacle course (2/9, 22%), medical labyrinth mini game (1/9, 11%), collecting stars (29, 22%), helping citizens (1/9, 11%), the well-being questions (1/9, 11%), educational module (3/9, 33%), and characters in general (5/9, 56%). The game in general was liked by 67% (6/9) of the participants. Interviews indicated that all the study participants used educational module. From negative aspects, different kind of preferences regarding the intervention were presented, for example, possibilities to access more buildings and use collectables in more advanced ways. The commonly reported critique indicated that the game was perceived to be too short for the 60-day intervention and that it was more interesting to play the game in the beginning of the study period until approximately half of the study period. Parents in general evaluated that the study process was smooth, and they were given sufficient amount of information.

When offered to continue playing the game following the study period, on their personal Android device, 6 of the participants out of 9 (67%) wished to do so. The reasons for the 3 participants not continuing were lack of a personal Android device (1 participant, 33%) and not interested (2 participants, 67%).

Discussion

Principal Findings

To the authors' best knowledge, this was the first study to evaluate the usability, acceptability, potential preliminary effectiveness, and feasibility of a personalized digital health intervention that uses a game environment to deliver psychoeducation, coping techniques, and treatment support. This approach allowed the researchers to use gamified, personalized therapeutics to offer comprehensive supportive care to pediatric patients. Previous studies have repeatedly stressed that it is highly important to integrate pediatric oncology-psychology research and standard of care [47], as well as the comprehensive supportive care of other chronic illnesses [6]. This study gave support to the prospect of delivering supportive care through a digital mHealth game in addition to traditional methodology.

The main findings of the study showed that the patients positively perceived the game, specifically with regard to their engagement, liking of the intervention, and learnings from it. The quantitative data showed that the mini games and the support module were used the most, whereas the qualitative findings also indicated the use of the educational module by all patients. However, on the basis of the qualitative interview evaluations, all patients expressed their own opinions about the most favorite parts, showing that all participants found something valuable to them which, in turn, offered more personal content. In general, our findings are supported by previous literature which has found mHealth solutions to be well suited for children, most of whom are savvy technology users [23].

Out of the specific modules we found, on the basis of consistent monitoring during the intervention, the well-being of patients improved when considering general health, but did not change significantly for depression, anxiety, and attention problems.

The evidence of no statistically significant change in mental health aspects could be explained by the generally stable mental state of the patients during the study period. Consistent monitoring is crucial for prevention and early detection of psychological problems as, for example, the intensity of the treatment and the change of health may influence the mental well-being. Furthermore, it could be expected that mental distress is experienced, even when the overall health status has improved, as noted in previous findings showing that psychological problems could last into adulthood [12]. The evidence that children receiving more complex treatment played fewer mini games requires further investigation of the potential different game behavior in this group of patients.

We also found that lower anxiety levels resulted in higher intervention evaluations, which could indicate that feeling less anxious during the intervention, and the qualitative interview may influence the child's view and opinion about the intervention. Furthermore, longer stays at the hospital resulted in less general health problems, which could indicate that receiving treatment at the hospital, and therefore receiving greater monitoring by medical personnel and timely adjustment in diagnosis, could result in less-reported general health problems. Unexpectedly, better sleep was only related to higher evaluations on the intervention but not with mental or physical health status. However, the importance of sleep should still be emphasized, and sleep quality should still be continuously monitored to observe whether sleep disruptions indicate short-term or long-term problems [48,49].

General attitudes toward health showed that a higher SDT score was related to more time from diagnosis, indicating more positive attitudes and motivation toward health. However, boys had lower motivation and attitudes compared with girls and thus may need more support to reach desired health outcomes. Focusing on an SDT-based approach could be suggested, as previous research indicates that components of SDT are associated with improved self-care among chronically ill patients, and thus positively related to treatment compliance, quality of life, and other health-related outcomes [50]. In addition, using general theories such as SDT as guiding theories for Triumf game development efforts, many of which having been widely used to explain the facilitation of motivated health behavior in wide variety of previous studies [51], enables us to adapt the intervention for different chronic conditions. SDT theory is also associated with the concept of mental toughness that refers to an individual's capacity to be consistently successful in coping with difficult life circumstances [52], one of the goals of Triumf intervention.

We also found that higher evaluations on the usability of the intervention were related to higher age, which could indicate that it was easier for older patients to understand the guidelines and text in the game. The youngest patients gave feedback that they did not understand some of the words used in the educational module. A revision of the text has been included in updates following the study, although already now the text is presented in levels, which allows younger patients to access more simple explanations compared with older ones who are able to access further details. Boys and those who spent more time at home engaged more in the exploratory use of the game. This could mean that patients feel more secure in the hospital setting and at home look for more support or that boys are more curious about the possibilities of the game.

The usability and acceptability results indicated high usability in general, and very high acceptability, with only too much reading being specifically brought out with regard to usability. In previous literature, a shortened text has been used [44], or alternatively, a more engaging way of presenting information could be implemented. The usability and acceptability findings are in accordance with previous literature that has highlighted that gamification is more engaging for users than intervention without gamification [26] and that using mHealth solutions is a valuable resource to deliver psychological techniques [22]. In general, the choice to collect feedback by constituent components of the game proved to be a valuable source of information, informing future studies about the presence of both strengths and *areas of improvement* at the same time.

The study protocol was generally feasible and was followed without deviations throughout the study period, although some amendments are necessary. Feedback from parents and medical doctors did not bring any modifications to the study protocol from the co-operation and communication perspective. As the questions of qualitative interview appeared difficult to understand for younger participants, a revision is needed. A simplification of the wording or more optimal amount of questions should be considered. In addition, transforming the oral interview format into a digital form could be more optimal for larger sample sizes. On the basis of the refusal and completion rates and necessary changes to the study protocol and the intervention itself, it could be concluded that the protocol is feasible for the randomized controlled trial, with minor modifications, and the game could be considered applicable in a clinical setting.

Strengths

This study has several major strengths. First, the rationale behind the intervention has been clearly supported by previous research, which is central in understanding the included elements and therefore allows for the selection and design of gamified components that are potentially most effective [53]. Second, the intervention was built in a digital environment only, which indicates players' engagement without external social encouragement and creates possibility to evaluate the cost-effectiveness of digital tools [53]. Moreover, a novel technological methodology to monitor and support patients in a timely manner was used. This interventional algorithm took individual responses and tailored the proposed support components accordingly. This study was conducted in 2 different hospitals, allowing to evaluate its effect independently of the hospital treatment context. Furthermore, it was observed that patients liked different modules of the game the most, suggesting that the personalized way of delivering psychological support is a preferred method, as we showed that all participants were able to use beneficial components of their liking. Taken together, our work informs future studies and contributes toward the development of effective mHealth interventions by giving the empirical evaluation of delivering psychological support in a health-focused digital game environment.

Limitations and Further Directions

Several patients and their parents stated that the game was interesting for approximately half of the study period (30 days), followed by a decrease in engagement. A decrease in the intensity of use of mHealth solutions over time has been recognized [23], although this could be related to the intervention design, especially from the motivation and engagement perspective. On the basis of the above, the intervention period could be shortened in the randomized controlled study or the game could undergo significant updates throughout the 60-day intervention period to keep patients engaged with the intervention (a process that is already in place to an extensive degree). It is possible that a shorter intervention time may be sufficient to induce behavioral change as habit formation time is very individual [54]. For example, studies on improvement of physical activity using health apps have shown that shorter interventions are more effective, although findings on intervention effects over time are still scarce [55]. Taken together, a change to the study protocol includes the reconsideration of the length of the intervention.

As this study involved patients aged 7 to 12 years, future studies are needed to evaluate the game among other age groups (eg, 5-7 years and 12-16 years). The game was evaluated in the context of pediatric cancer; thus, the intervention could also be evaluated among children with other chronic illnesses.

In addition, it was found that the 1 individual who needed psychological support the most ended up using the intervention the least. There have been observations that more serious psychological problems may interfere with engagement with the game [26]. Considering that digital intervention studies have been conducted on patients with mild to moderate symptoms [53], it should be investigated further. To continue, amendments to the game should be made to accommodate the needs of those individuals who need psychological support the most. Thus, it was concluded that the intervention should be delivered in 2 steps. During the initial stage, only educational module and fun components of the game would be accessible. Subsequently, psychological intervention would follow. Through this 2-step approach, new information would be given in different stages of the intervention. This might result in higher compliance through reducing the initial load of information. The abovementioned modifications to the intervention would be implemented before the next study.

Although data on general health attitudes were collected, it was proposed that these findings would need reassessment, as reported general health problem results might have been dependent on whether prompted questions were about general health behaviors or related to the current health situation. Separate measures of the current health situation and general perceived health condition and behaviors should be considered.

Conclusions

In conclusion, this study shows that delivering comprehensive supportive care through a game environment to pediatric patients is a feasible intervention strategy and is accepted by the patients and applicable in clinical context. This study showed that a game environment is a safe and engaging way of collecting real-time comprehensive data that can be used for personalized support.

Acknowledgments

We would like to thank psychologists Ada Urm and Olga Korkkinen for conducting qualitative interviews. We are thankful to the medical personnel at Tartu University Hospital and Tallinn Children’s Hospital for their support in conducting the clinical trials. We are grateful for all study participants and their families. We would like to acknowledge the contribution of Triumf Health team members who participated in the development of the digital health intervention involving Georgi Merdzhanov, Leandro Oscar Gil, Jose Manuel Prieto Garay, Risto Haljas, Sven Parkel, Anna Suarez Figueiredo, Nguyen Hong Vân Anh, Kertu Tombak, and Kaari Kink.

Conflicts of Interest

KH, RT, and MM are part of Triumf Health start-up and have been involved in the creation of the digital health intervention strategy. RT was one of the trained psychologists to conduct qualitative interviews.

Multimedia Appendix 1
Visual representation of the current Triumf mobile health game (version 1.0.5.2) displaying 4 screenshots of the mHealth intervention.
[PDF File (Adobe PDF File)368 KB - games_v7i3e13776_app1.pdf]

Multimedia Appendix 2
Correlations with confidence intervals.
[PDF File (Adobe PDF File)56 KB - games_v7i3e13776_app2.pdf]

References

1. van Cleave J, Gortmaker SL, Perrin JM. Dynamics of obesity and chronic health conditions among children and youth. J Am Med Assoc 2010 Feb 17;303(7):623-630. [doi: 10.1001/jama.2010.104] [Medline: 20159870]

2. Butler A, van Lieshout RJ, Lipman EL, MacMillan HL, Gonzalez A, Gorter JW, et al. Mental disorder in children with physical conditions: a pilot study. BMJ Open 2018 Jan 3;8(1):e019011 [FREE Full text] [doi: 10.1136/bmjopen-2017-019011] [Medline: 29301763]
3. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. Lancet 2011 Oct 22;378(9801):1515-1525. [doi: 10.1016/S0140-6736(11)60827-1] [Medline: 22008427]
4. Tark R, Jõeveer K, Tali A, Haljas K. Tervisepsühholoogia: psühholoogi roll vähiravis. Eesti Arst 2018 Jun;97(6):319-325 [FREE Full text]
5. Hysing M, Elgen I, Gillberg C, Lie SA, Lundervold AJ. Chronic physical illness and mental health in children. Results from a large-scale population study. J Child Psychol Psychiatry 2007 Aug;48(8):785-792. [doi: 10.1111/j.1469-7610.2007.01755.x] [Medline: 17683450]
6. American Diabetes Association. Abridged for primary care providers. Clin Diabetes 2018 Jan;36(1):14-37 [FREE Full text] [doi: 10.2337/cd17-0119] [Medline: 29382975]
7. Kirch R, Reaman G, Feudtner C, Wiener L, Schwartz LA, Sung L, et al. Advancing a comprehensive cancer care agenda for children and their families: Institute of Medicine Workshop highlights and next steps. CA Cancer J Clin 2016 Sep;66(5):398-407 [FREE Full text] [doi: 10.3322/caac.21347] [Medline: 27145249]
8. Feudtner C, Womer J, Augustin R, Remke S, Wolfe J, Friebert S, et al. Pediatric palliative care programs in children's hospitals: a cross-sectional national survey. Pediatrics 2013 Dec;132(6):1063-1070. [doi: 10.1542/peds.2013-1286] [Medline: 24190689]
9. Prince M, Patel V, Saxena S, Maj M, Maselko J, Phillips MR, et al. No health without mental health. Lancet 2007 Sep 8;370(9590):859-877. [doi: 10.1016/S0140-6736(07)61238-0] [Medline: 17804063]
10. 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]
11. Pai AL, McGrady M. Systematic review and meta-analysis of psychological interventions to promote treatment adherence in children, adolescents, and young adults with chronic illness. J Pediatr Psychol 2014 Sep;39(8):918-931 [FREE Full text] [doi: 10.1093/jpepsy/jsu038] [Medline: 24952359]
12. Secinti E, Thompson EJ, Richards M, Gaysina D. Research review: childhood chronic physical illness and adult emotional health - a systematic review and meta-analysis. J Child Psychol Psychiatry 2017 Jul;58(7):753-769. [doi: 10.1111/jcpp.12727] [Medline: 28449285]
13. Compas BE, Jaser SS, Dunn MJ, Rodriguez EM. Coping with chronic illness in childhood and adolescence. Annu Rev Clin Psychol 2012;8:455-480 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032511-143108] [Medline: 22224836]
14. Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci 2009;11(1):7-20 [FREE Full text] [Medline: 19432384]
15. Beale IL. Scholarly literature review: efficacy of psychological interventions for pediatric chronic illnesses. J Pediatr Psychol 2006 Jun;31(5):437-451. [doi: 10.1093/jpepsy/jsj079] [Medline: 16162841]
16. Bennett S, Shafran R, Coughtrey A, Walker S, Heyman I. Psychological interventions for mental health disorders in children with chronic physical illness: a systematic review. Arch Dis Child 2015 Apr;100(4):308-316 [FREE Full text] [doi: 10.1136/archdischild-2014-307474] [Medline: 25784736]
17. Bond C, Woods K, Humphrey N, Symes W, Green L. Practitioner review: the effectiveness of solution focused brief therapy with children and families: a systematic and critical evaluation of the literature from 1990-2010. J Child Psychol Psychiatry 2013 Jul;54(7):707-723. [doi: 10.1111/jcpp.12058] [Medline: 23452301]
18. Wiener L, Pao M, Kazak A, Kupst M, Patenaude A, Arceci R, editors. Pediatric Psycho-Oncology: A Quick Reference on the Psychosocial Dimensions of Cancer Symptom Management. Second Edition. New York: Oxford University Press; 2015.
19. App Store - Apple. 2014. Pain Squad: The Hospital for Sick Children URL: https://itunes.apple.com/us/app/pain-squad/id929781246?mt=8 [accessed 2019-04-26] [WebCite Cache ID 77v8mNkpF]
20. Google Play. 2015. Re-Mission2: Nanobot's Revenge URL: https://play.google.com/store/apps/details?id=air.com.rm2.nbr [accessed 2019-04-26] [WebCite Cache ID 77v964Awp]
21. Google Play. 2018. Super K URL: https://play.google.com/store/apps/details?id=com.BNdemo.SuperK [accessed 2019-04-26] [WebCite Cache ID 77v9WFcck]
22. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
23. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]
24. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]

25. Ernsting C, Dombrowski SU, Oedekoven M, Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 5;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
26. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. JMIR Serious Games 2018 May 10;6(2):e9 [FREE Full text] [doi: 10.2196/games.8902] [Medline: 29748165]
27. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 1;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
28. Tark R. DSpace: University of Tartu. 2019. The Effect of Triumf Mobile Health Game on Psychological Well-Being and Health-Related Quality of Life Among Pediatric Cancer Patients: A Pilot Study URL: https://dspace.ut.ee/handle/10062/65009 [accessed 2019-08-15]
29. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
30. McLaughlin KA, Hatzenbuehler ML, Mennin DS, Nolen-Hoeksema S. Emotion dysregulation and adolescent psychopathology: a prospective study. Behav Res Ther 2011 Sep;49(9):544-554 [FREE Full text] [doi: 10.1016/j.brat.2011.06.003] [Medline: 21718967]
31. Ford JS, Barnett M, Werk R. Health behaviors of childhood cancer survivors. Children (Basel) 2014 Oct 22;1(3):355-373 [FREE Full text] [doi: 10.3390/children1030355] [Medline: 27417484]
32. Ekman P. Facial expressions of emotion: new findings, new questions. Psychol Sci 2017 Apr 25;3(1):34-38. [doi: 10.1111/j.1467-9280.1992.tb00253.x]
33. Russell JA. Mixed emotions viewed from the psychological constructionist perspective. Emotion Rev 2017 Jan 30;9(2):111-117. [doi: 10.1177/1754073916639658]
34. Sheeran P, Klein WM, Rothman AJ. Health behavior change: moving from observation to intervention. Annu Rev Psychol 2017 Jan 3;68:573-600. [doi: 10.1146/annurev-psych-010416-044007] [Medline: 27618942]
35. 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]
36. 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]
37. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037/0022-3514.34] [Medline: 11392867]
38. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry 2001 Nov;40(11):1337-1345. [doi: 10.1097/00004583-200111000-00015] [Medline: 11699809]
39. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Psychol 2008;10:2-5 [FREE Full text]
40. Anderson LW, Anderson DR, Blooms BS. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Boston: Allyn & Bacon; 2001.
41. Scholten L, Willemen AM, Grootenhuis MA, Maurice-Stam H, Schuengel C, Last BF. A cognitive behavioral based group intervention for children with a chronic illness and their parents: a multicentre randomized controlled trial. BMC Pediatr 2011 Jul 14;11:65 [FREE Full text] [doi: 10.1186/1471-2431-11-65] [Medline: 21756299]
42. Yik M, Russell JA, Steiger JH. A 12-point circumplex structure of core affect. Emotion 2011 Aug;11(4):705-731. [doi: 10.1037/a0023980] [Medline: 21707162]
43. Suveg C, Zeman J. Emotion regulation in children with anxiety disorders. J Clin Child Adolesc Psychol 2004 Dec;33(4):750-759. [doi: 10.1207/s15374424jccp3304_10] [Medline: 15498742]
44. Marsac ML, Winston FK, Hildenbrand AK, Kohser KL, March S, Kenardy J, et al. Systematic, theoretically-grounded development and feasibility testing of an innovative, preventive web-based game for children exposed to acute trauma. Clin Pract Pediatr Psychol 2015;3(1):12-24 [FREE Full text] [doi: 10.1037/cpp0000080] [Medline: 25844276]
45. IBM - United States. 2019. IBM SPSS Software URL: https://www.ibm.com/analytics/spss-statistics-software [accessed 2019-02-20] [WebCite Cache ID 76KMFLKEx]
46. The R Project for Statistical Computing. 2019. What is R?: Introduction to R URL: https://www.r-project.org/about.html [accessed 2019-02-20] [WebCite Cache ID 76KMZaBTw]
47. Armstrong FD, Reaman GH, Children's Oncology Group. Psychological research in childhood cancer: the children's oncology group perspective. J Pediatr Psychol 2005;30(1):89-97. [doi: 10.1093/jpepsy/jsi020] [Medline: 15610989]
48. Sivertsen B, Hysing M, Elgen I, Stormark KM, Lundervold AJ. Chronicity of sleep problems in children with chronic illness: a longitudinal population-based study. Child Adolesc Psychiatry Ment Health 2009 Aug 27;3(1):22 [FREE Full text] [doi: 10.1186/1753-2000-3-22] [Medline: 19712458]

49. Dewald JF, Meijer AM, Oort FJ, Kerkhof GA, Bögels SM. The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: a meta-analytic review. Sleep Med Rev 2010 Jun;14(3):179-189. [doi: 10.1016/j.smrv.2009.10.004] [Medline: 20093054]
50. Williams GC, Patrick H, Niemiec CP, Williams LK, Divine G, Lafata JE, et al. Reducing the health risks of diabetes: how self-determination theory may help improve medication adherence and quality of life. Diabetes Educ 2009;35(3):484-492 [FREE Full text] [doi: 10.1177/0145721709333856] [Medline: 19325022]
51. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol 2008;49(3):182-185. [doi: 10.1037/a0012801]
52. Bahmani DS, Hatzinger M, Gerber M, Lemola S, Clough PJ, Perren S, et al. The origins of mental toughness - prosocial behavior and low internalizing and externalizing problems at age 5 predict higher mental toughness scores at age 14. Front Psychol 2016;7:1221 [FREE Full text] [doi: 10.3389/fpsyg.2016.01221] [Medline: 27605919]
53. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
54. Lally P, van Jaarsveld CH, Potts HW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
55. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]

Abbreviations

CBT: cognitive behavioral therapy
mHealth: mobile health
SDT: Self-Determination Theory

Edited by N Zary, G Eysenbach; submitted 21.02.19; peer-reviewed by D Fedele, S Bennett; comments to author 31.03.19; revised version received 18.05.19; accepted 05.08.19; published 30.09.19.

Please cite as:
Tark R, Metelitsa M, Akkermann K, Saks K, Mikkel S, Haljas K
Usability, Acceptability, Feasibility, and Effectiveness of a Gamified Mobile Health Intervention (Triumf) for Pediatric Patients: Qualitative Study
JMIR Serious Games 2019;7(3):e13776
URL: https://games.jmir.org/2019/3/e13776
doi:10.2196/13776
PMID:31573904

Original Paper

Impact of Using a 3D Visual Metaphor Serious Game to Teach History-Taking Content to Medical Students: Longitudinal Mixed Methods Pilot Study

Hussain Alyami[1,2], MBChB, FRANZCP, PGDCBT; Mohammed Alawami[3], MBChB, FRACP; Mataroria Lyndon[1], MBChB, MPH, PhD; Mohsen Alyami[1], BSc, MSc; Christin Coomarasamy[4], MPhil; Marcus Henning[1], MA, MBUS, PhD; Andrew Hill[1], MBChB, MD, EdD, FRACS; Frederick Sundram[1], MBChB, MSc, MA, PhD, FRCPsych

[1]South Auckland Clinical Campus, Faculty of Medical and Health Sciences, The University of Auckland, Auckland, New Zealand

[2]College of Medicine, Taif University, Taif, Saudi Arabia

[3]Auckland City Hospital, Auckland, New Zealand

[4]Middlemore Hospital, Auckland, New Zealand

Corresponding Author:
Hussain Alyami, MBChB, FRANZCP, PGDCBT
South Auckland Clinical Campus
Faculty of Medical and Health Sciences
The University of Auckland
Middlemore Hospital
Private Bag 93311
Auckland, 1640
New Zealand
Phone: 00966 541118828
Email: haly282@aucklanduni.ac.nz

Abstract

Background: History taking is a key component of clinical practice; however, this skill is often poorly performed by students and doctors.

Objective: The study aimed to determine whether Metaphoria, a 3D serious game (SG), is superior to another electronic medium (PDF text file) in learning the history-taking content of a single organ system (cardiac).

Methods: In 2015, a longitudinal mixed methods (quantitative and qualitative) pilot study was conducted over multiple sampling time points (10 weeks) on a group of undergraduate medical students at The University of Auckland Medical School, New Zealand. Assessors involved in the study were blinded to group allocation. From an initial sample of 83, a total of 46 medical students were recruited. Participants were assigned to either a PDF group (n=19) or a game group (n=27). In total, 1 participant left the PDF group after allocation was revealed and was excluded. A total of 24 students in the game group and 14 students in the PDF group completed follow-up 7 weeks later. Using an iterative design process for over a year, with input from a variety of clinical disciplines, a cardiac history-taking game and PDF file were designed and informed by Cognitive Load Theory. Each group completed its intervention in 40 min. A total of 3 levels of Kirkpatrick training evaluation model were examined using validated questionnaires: affective (perception and satisfaction), cognitive (knowledge gains and cognitive load), and behavioral attitudes (Objective Structured Clinical Exam) as well as qualitative assessment. A priori hypotheses were formulated before data collection.

Results: Compared with baseline, both groups showed significant improvement in knowledge and self-efficacy longitudinally ($P<.001$). Apart from the game group having a statistically significant difference in terms of satisfaction ($P<.001$), there were no significant differences between groups in knowledge gain, self-efficacy, cognitive load, ease of use, acceptability, or objective structured clinical examination scores. However, qualitative findings indicated that the game was more engaging and enjoyable, and it served as a visual aid compared with the PDF file.

Conclusions: Students favored learning through utilization of an SG with regard to cardiac history taking. This may be relevant to other areas of medicine, and this highlights the importance of innovative methods of teaching the next generation of medical students.

(JMIR Serious Games 2019;7(3):e13748) doi:10.2196/13748

KEYWORDS

video games; instructional technology; memory; retention; metaphor; learning; clinical competence

Introduction

Background

Since the time of Hippocrates, clinical history taking has remained a cornerstone of medicine [1]. Although history taking is the most common medical procedure performed by doctors [2], research has shown a deficit in this essential skill among medical students [3-6], interns, and, more worryingly, general practitioners [7]. History taking is more valuable than physical examination in reaching a diagnosis in around 80% of medical outpatient referrals [8,9]. When adequately performed, history taking is associated with improved patient physiological and psychological health outcomes [10], satisfaction [11,12], and compliance [13,14]. Although the literature shows that this essential clinical skill can be learned, it is inadequately taught [3,5,15].

A fine balance between the history-taking components, that is, history-taking content (HTC) and history-taking process (HTP), needs to be achieved for optimal effectiveness. HTC is commonly referred to as data or information gathering, and it is concerned with the *what* part of the medical interview, eliciting specific information about the patient's symptomatology from the presenting complaint through to social and occupational histories [1]. On the other hand, HTP, commonly known as communication skills, is the method by which this content is elicited; thus, it is more concerned with the *how* of the medical interview [1]. There is evidence to suggest that medical students are distracted when trying to remember HTC, which impairs their communication skills [16,17]. Despite the importance of teaching HTC, this remains an understudied area, with a recent systematic review showing only 6 studies of educational interventions that targeted this essential part of the medical interview; this review also found that technologically enhanced interventions improved teaching of HTC [18].

Objectives

The literature shows that medical students use multiple learning styles, including visual, auditory, and kinesthetic [19]; therefore, we propose that a visual metaphor system could prove to be of significant value to enhance the teaching of HTC. A visual metaphor is defined as "a graphic structure that uses the shape and elements of a familiar natural or man-made artifact or of an easily recognizable activity or story, to organize content meaningfully and use the associations with the metaphor to convey additional meaning about the content" [20:233]. The use of visual metaphors in medicine dates back to antiquity, when Aristotle explained how blood is contained in the heart and associated vessels and compared this with a "vase" [21]. Currently, visual metaphors continue to be used successfully to teach some of the more complicated and abstract scientific principles, for example, Emil Fischer's Lock-and-Key model of enzyme-substrate interaction in physiology [22]. Until recently, metaphor was considered a decorative component of speech; it is now considered "to pervade all forms of knowledge" [23:86]. A number of benefits have been reported for using a visual metaphor system as a learning and knowledge sharing tool, including improved audience engagement, attention, memory, and comprehension [24]. Visual metaphors can provide a means of transferring a complex concept by using simple visual symbols that facilitate the connection between thoughts and feelings [25]. Visual metaphors, when created and presented well, enable high levels of content comprehension, retention, and recall, and they are easier to construct and interpret, compared with mind maps [20]. They also assist learners to incorporate newly learned material with previous knowledge [26,27]. Serious games (SGs) offer a promising modality of delivering visual metaphors, as they have been shown to provide learners with an "anchor" for knowledge [28]. Fabricatore described how metaphors could be embedded in SGs to enhance learning, especially if they are designed to being at the heart of gameplay rather than serving as a mere decorative component [29]. There are potential benefits of SGs, which utilize visual metaphorical elements, in medical education [30]; however, the application and assessment of metaphor use in SG design are limited [31]. SGs are defined as games "that engage the user and contribute to the achievement of a defined purpose other than pure entertainment" [30: 5]. Although adopting visual metaphors in SG design is in its infancy, the SGs' literature suggests they can improve engagement and facilitate learning of the intended teaching points [28,32,33]. However, SGs are considered complex, and they need to be designed while taking into consideration the limits of human cognitive capacity [34]. Cognitive Load Theory (CLT) offers a wide range of instructional design principles [35]. CLT has been used in the design of SGs to manage the limited cognitive capacity of the learner [36], and it highlights the limited capacity of working memory (WM) to process from 3 to 7 items at any given time, compared with the potentially unlimited capacity of long-term memory [37]. CLT aims to reduce the cognitive load imposed by instructional design (extraneous load) on WM to improve learning (germane load). CLT design principles mostly target the extraneous cognitive load imposed by poorly designed material [38,39]. These design principles include using worked examples and minimizing redundant information and split attention [38]. The aim of this study was to evaluate the impact of an SG on learning HTC in medical students during their first clinical exposure. To our knowledge, this is the first visual metaphor–enhanced SG implemented for teaching HTC to medical students. To evaluate this SG's impact, Kirkpatrick's training evaluation model was adopted, as it is the most widely used method in evaluating training programs [40] and is commonly used to assess medical educational interventions [41-43]. Kirkpatrick's model was modified by Freeth et al [43] and adopted by the Best Evidence Medical Education Collaboration to assess medical teaching interventions [44]. The domains included are the following: affective (cognitive load, material difficulty level, and perceptions and satisfaction), cognitive (knowledge gains), and behavioral attitudes (objective structured clinical exam). The fourth level in Kirkpatrick 's model (outcomes related to the impact of the interventions on patients' care) is not commonly measured in educational settings [45] and is beyond the scope of this study.

Methods

Overview

This pilot study used a mixed methods approach by combining a repeated-measures nonrandomized quasi-experimental design with a qualitative component. The study was approved by University of Auckland Human Participants Ethics Committee, reference number 015567. Med Metaphoria game design and development were funded by The New Zealand Health Innovation Hub, in collaboration with Counties Manukau District Health Board, who provided the first author with a clinical fellowship in Health Systems Innovation and Improvement.

Recruitment

The University of Auckland has a 6-year undergraduate medical program, predominantly comprising basic sciences (Years 1-3) and then clinical practice (Years 4-6). Our sample was a preclinical population of year-3 medical students at the South Auckland Clinical Campus of the University of Auckland, who were attending a 10-week clinical methods module at the end of 2015. This is a foundation course for teaching clinical skills. All 83 students on the rotation were invited to participate in this study. Recruitment strategies involved both verbal and poster invitations. Students were preallocated, at the medical school administration level, to 1 of 2 teaching days (Wednesday or Thursday), independent of the investigators.

Interventions

To minimize potential contamination, the control group (PDF file group) was allocated to Wednesday, followed by the intervention group (Visual Metaphor Game; VMG group) on Thursday. This was done to minimize contamination in keeping with medical education recommendations [46], which suggested restricting access to the intended educational intervention. Both groups had the same face-to-face cardiac history teaching (whole morning session) delivered on the day of the interventions, followed by a one-off 40-min session, using either a PDF file or the VMG. Both groups used iPads to access content, whereby the same content was covered but differed in instructional design, as the PDF file was textual, and the game used visual metaphors. A visual metaphor–enhanced game (Med Metaphoria) was developed as a 3D SG that contains elements of HTC (Figure 1 and Multimedia Appendix 1). The 3D game was codesigned by the lead author in collaboration with clinicians, medical educationalists, code developers, and students. We followed a participatory iterative agile game development process [47], using the I's development framework, proposed by Annetta et al for the design of SGs [48] (Figure 2). This "I's" framework included the following elements: Identity, Immersion, Interactivity, Increasing complexity, Informed teaching, and Instructional. Game development (versions 1 and 2) incorporated student and clinician feedback by using a think-aloud protocol, originally developed by Ericsson and Simon [47,49] and later adopted by SGs' design literature [50] until the final iteration (version 3) was developed (Figure 3). Although there are no significant differences between 2D and 3D educational games in terms of learning gains [51], 3D design was chosen as the main game design modality, despite a higher associated cost for the following reasons. First, a key teaching point in the history-taking visual metaphor is to avoid "tunnel vision" history-taking, which is similar to looking at a 2D pyramid, focusing on 1 causative side or system and ignoring the other 3D pyramid sides and systems. Therefore, 3D depth of the visual metaphor was essential to convey this point, which is commonly used in the game literature and called "spatial metaphor" [52]. Second, the literature has shown that health sciences students' preference for 3D over 2D games [53]. Finally, the game literature has found that 3D games facilitate a more enhanced sense of presence than 2D games [53]. Overall, 2D graphics were also used if there was no spatial need for metaphors.

Figure 1. In-game snapshot showing the symptoms and signs relevant to the cardiovascular system.

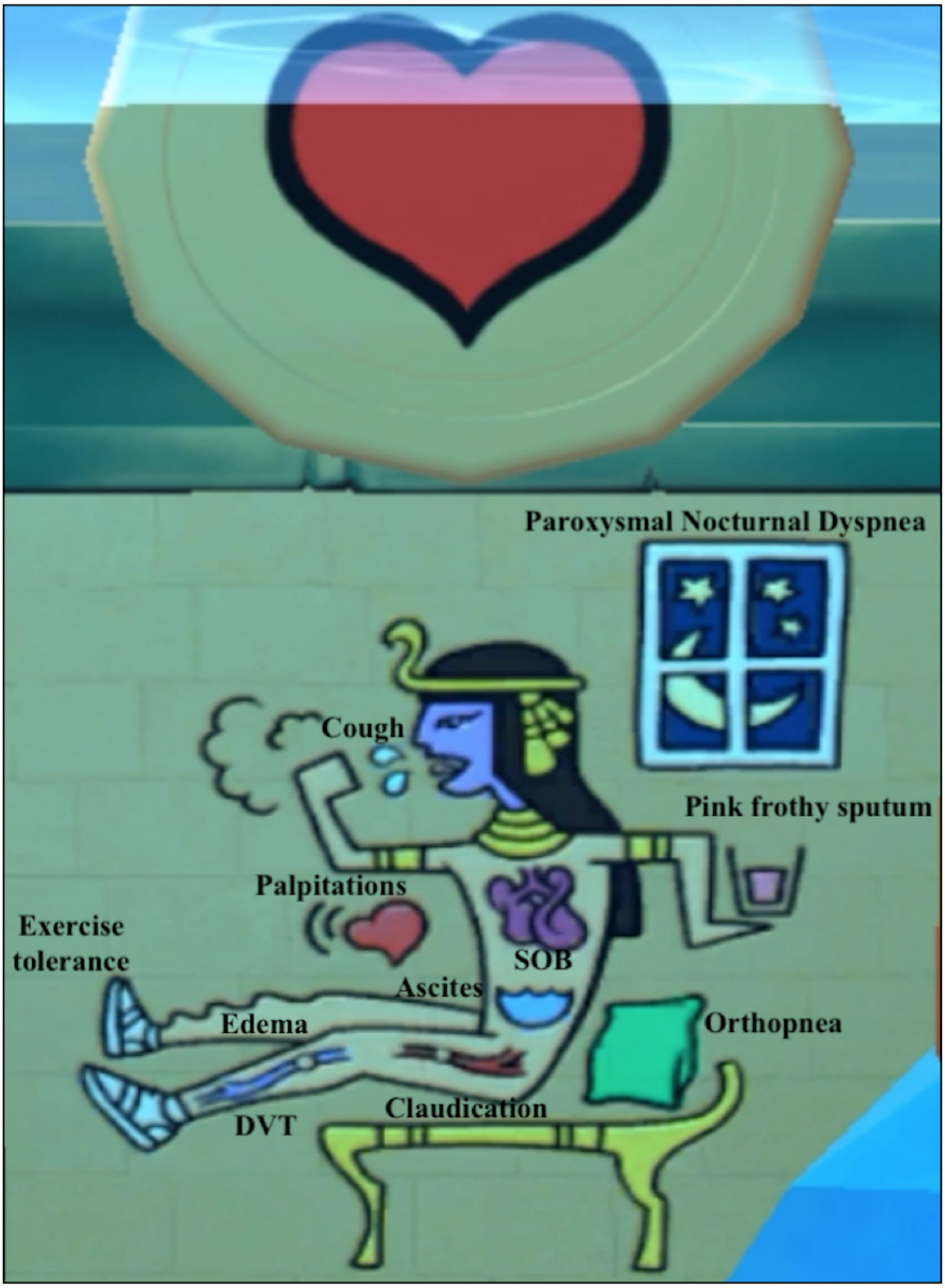


Figure 2. Visual metaphor and game design process in each game version.

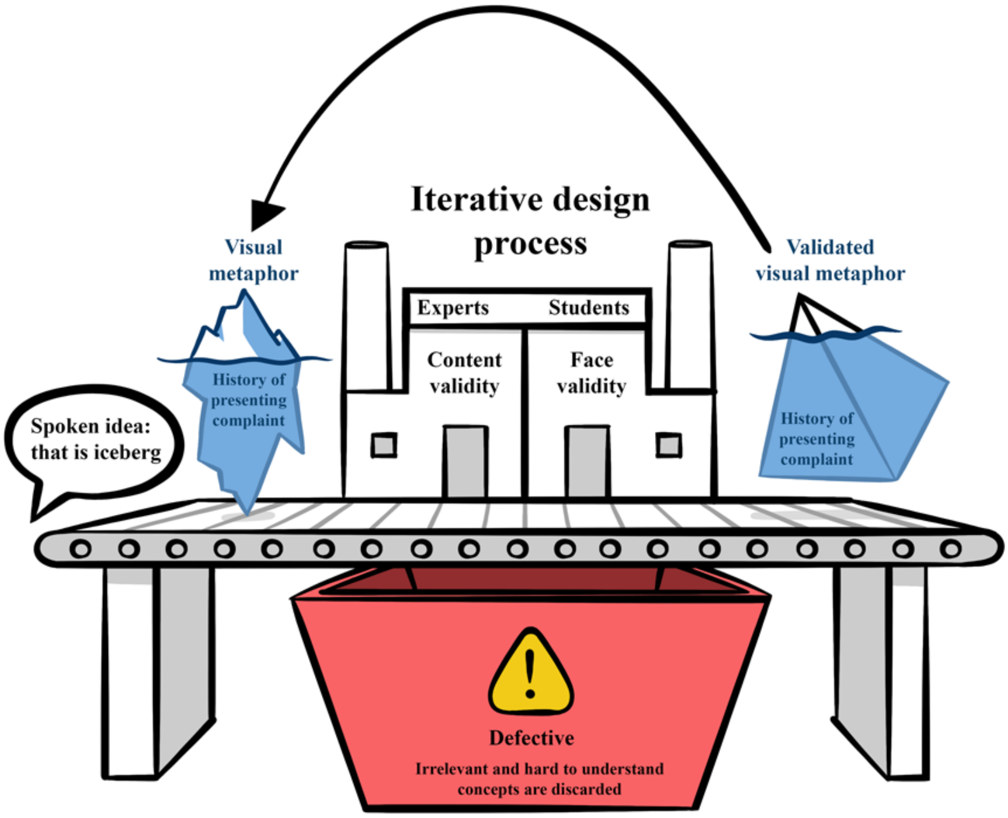
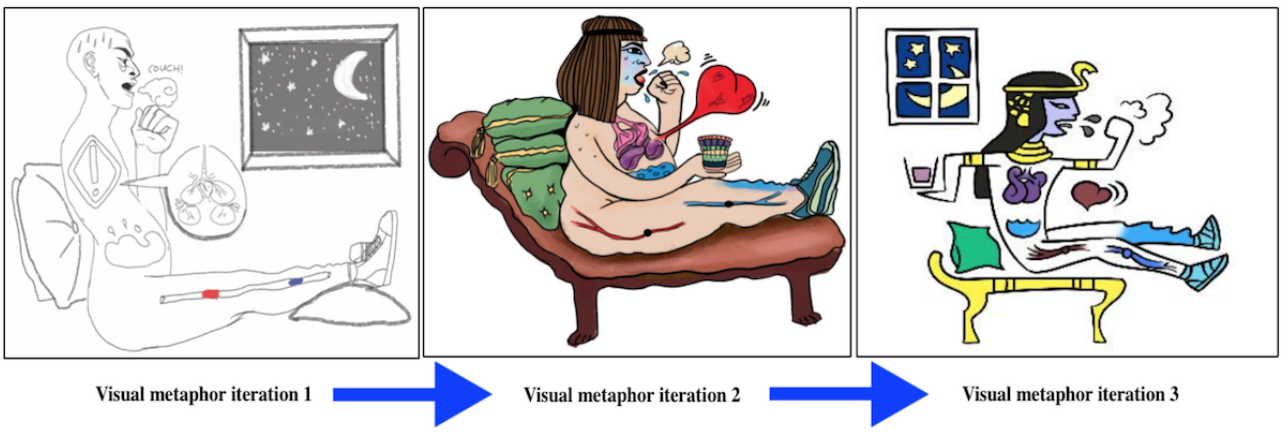
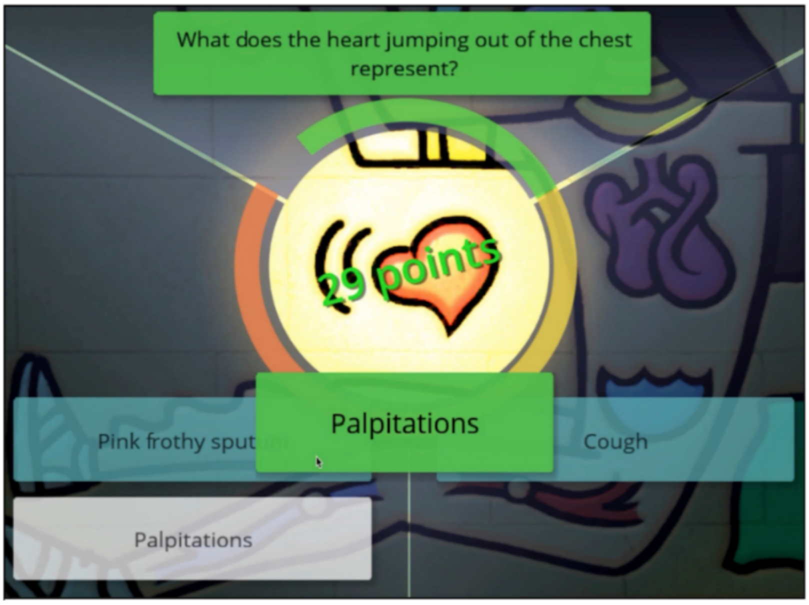


Figure 3. Iterative visual metaphor design.



The main characters in the game were the medical students in the submarine (the rescuer and students had the option of choosing their own personalized name) and John the Archaeologist (the patient). The game was a point-and-click adventure game, taking place in a 3D fantasy world where John was inadvertently locked up in a pyramid after deciphering a set of visual metaphors wrongly. The game rules included deciphering the visual metaphors correctly by associating them with the correct multiple-choice answer to rescue John, who is complaining of chest pain. The quicker this was done, the higher the score and rank achieved. Each section of the history taking was a separate part of the game journey. Audio visual and textual feedback was provided, in addition to static and animated game narrative and cut scenes. Students had to decipher the visual metaphors by choosing the right answer out of 3 or 4 multiple-choice answers per question. Although question order was constant, answer positions were randomly ordered for each question. The game had 3 difficulty levels, with the beginner level allowing 30 seconds per question, using the full textual question, followed by the intermediate level, which allows 20 seconds, with questions partially visible, and the advanced level, allowing 10 seconds, which only shows the visual item with no textual question. The total number of seconds per difficulty level was divided into thirds, and students were ranked into 1 of 3 ranks on the basis of their time to answer each question correctly: Consultant (first third), Resident (second third), or Intern (last third). Students also scored points corresponding to their speed in choosing the right answer (Figure 4).

Figure 4. In-game snapshot showing the visual metaphor in the background, question and multiple choice options. The number of points (29) represents that it took the player in the beginner level one second to decipher the visual metaphor and answer the question correctly (out of the 30 seconds allocated per question).



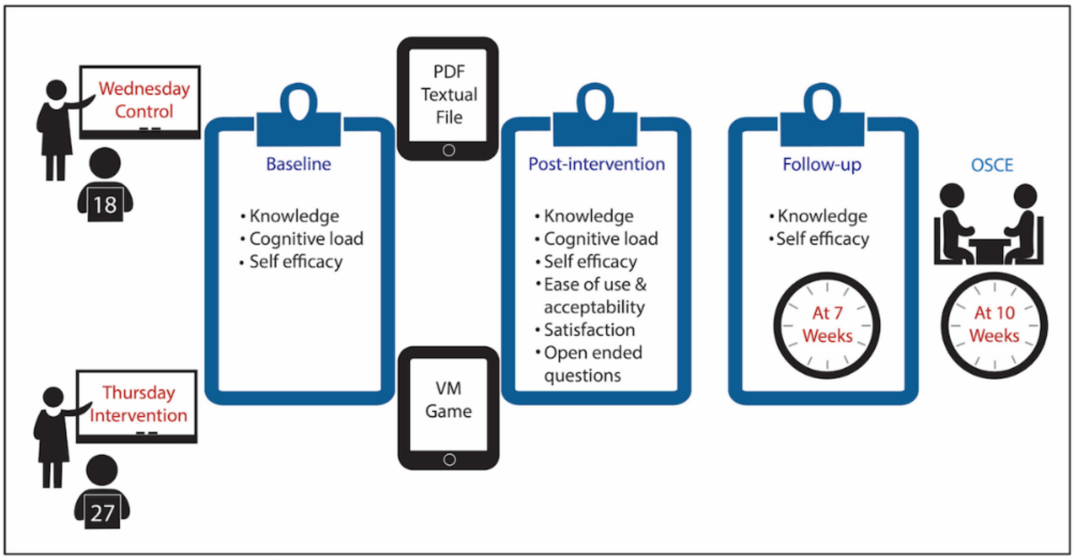
Development and Validation of Teaching Interventions and History-Taking Content Measures

Both PDF files and the VMG were developed after consulting with the recommended reading list the students received, which included a clinical skills module booklet and a recommended clinical skills textbook [54]. Content validity and instructional design of both interventions were carried out through iterative consultation with multiple junior and senior doctors, from medical and surgical disciplines, including Internal Medicine, Cardiology, Surgery, and Psychiatry. Face validity was established through iterative discussions with multiple medical students across the medical school program.

Measures

For both groups, knowledge about HTC was assessed using a paper-based context-rich open-ended question. This was asked immediately before intervention (baseline), immediately after intervention (postintervention), and then again 7 weeks after intervention (follow-up). These tests were designed to assess baseline HTC knowledge and immediate and delayed recall (see Figure 5 for study design). This open-ended question was deemed more appropriate than using closed multiple-choice questions [55]. It read “what are the cardiac symptoms you would ask about as part of the history of presenting complaint in a patient presenting with chest pain?” The model answer template had a total of 25 symptoms. Answers that included lay or medical terms were accepted if they were consistent with the symptom meaning. Marking was conducted by a senior cardiology registrar, who was blinded to both groups, out of a total score of 25 based on predefined marking criteria (Multimedia Appendix 2).

Figure 5. Study design and sampling time points.



Pre, post, and follow-up questionnaires were designed to assess the 3 levels of Kirkpatrick's training evaluation model. These key points about the measures used are summarized in Table 1. The Mental Effort Scale [56] was used to assess students' perceived history-taking cognitive load, whereas history-taking difficulty was measured by using the Difficulty Level Scale [57]. The Mental Effort Scale is a subjective rating scale for cognitive load, with a Cronbach coefficient alpha of .90 (internal reliability) [56]. This scale has been studied widely in the literature, and it has shown good psychometric properties [37]. It has been shown to be comparable to objective physiological measures and has good validity and reliability [58], and it is more sensitive and simpler than objective physiologic measures [37]. The adapted 5-point Likert scale is commonly used to assess cognitive load [57], where the question "In taking a history, I invest..." is answered on a 5-point Likert scale. As a one-off administration of the scale was shown to over or underestimate scores [59,60], it was administered at 2 timelines (baseline and postintervention). The perceived task difficulty level was also measured, as it is an important but different construct to cognitive load [61]. According to Gog and Paas, tasks perceived to be highly difficult might lead to using low mental effort in solving them [61]. This scale was administered to answer the question "I experienced history-taking as..."

Despite the known criticism of single-item scales [62], such as cognitive load and perceived task difficulty scales, research has shown these scales to have comparable properties with multi-item scales [63]. The Self-Efficacy Scale [64] was used to measure student perceived self-confidence in history taking. Medical students' self-efficacy has been shown to influence academic performance [65-67]. Bandura's seminal work on self-efficacy theory highlighted that competency to perform a certain task is not limited to only knowledge and skill acquisition but also in how the participant believes in their efficacy of reaching this goal [68]. In the context of SGs, Gee highlighted how they could increase self-efficacy [69]. The General Self-Efficacy (GSE) Scale devised by Schwarzer and Jerusalem [64] was used to measure student perceived self-confidence in history taking. GSE refers to the global confidence in one's coping ability across multiple situations [64]. This scale has been used extensively in research, with an internal consistency of 0.75 to 0.91, with a reliability of 0.67 [70]. However, it is considered a nondomain-specific scale, which goes against Bandura's domain-specific conceptualization of self-efficacy [71]. Therefore, GSE scale adaptation was necessary to make it history-taking domain specific to fit with Bandura's self-efficacy theory [71]. Given that self-efficacy is dynamic, researchers have advised to measure this construct longitudinally rather than cross-sectionally [72]. Although some researchers have measured self-efficacy at 2 time points, medical educationalists have recommended doing so more than twice [71]. Given that both educational interventions were delivered using a technologically enhanced medium (an iPad) in this study, the Technology Acceptance Model (TAM) was used [73]. The TAM's Ease of Use and Acceptability questionnaire has been previously used to investigate medical students' [74] and physicians' intentions to use technology [75,76]. Owing to its robustness, simplicity, and adaptive nature, it has become one of the most widely used models in measuring technology acceptance [77].

SGs that involve instructional design optimization have been shown to improve satisfaction levels [78]. Therefore, a satisfaction questionnaire was designed as it could serve as an indirect academic performance measure [79]. These questions were drawn from other relevant medical education literature evaluating technologically enhanced educational interventions [80,81]. The questionnaire face and content validation process followed an iterative design process similar to game design (Figure 2). Immediately after the intervention, 3 open-ended questions were used to assess students' perceptions and suggestions of what they learned, enjoyed, and did not enjoy, as well as suggestions for further improvement.

History-taking clinical skills were objectively assessed at 10 weeks postintervention by comparing end-of-year objective structured clinical examination (OSCE) results for the history-taking station. The history-taking station was designed independent of the investigators, and it targeted assessment of abdominal pain; however, the overall history-taking structure and pain questions were similar.

Table 1. Measures used and relevant information.

Scale name and Construct	Items	1 is	5 is	Timelines
Mental effort				
Cognitive load	1	Very low	Very high	Baseline and postintervention
Perceived difficulty				
Task difficulty	1	Not difficult at all	Very high difficulty	Baseline and postintervention
General Self-Efficacy				
Self-confidence in history taking	10	Strongly disagree	Strongly agree	Baseline, postintervention, and 7 weeks later
Technology acceptance model				
Ease of use	6	Strongly disagree	Strongly agree	Postintervention
Usefulness	6	Strongly disagree	Strongly agree	Postintervention
Satisfaction				
Satisfaction	8	Strongly disagree	Strongly agree	Postintervention
Knowledge				
Open-ended knowledge question	Marks out of 25	—[a]	—	Baseline and postintervention
Clinical skills				
Objective structured clinical examination	Marks out of 25	—	—	10 weeks postintervention

[a]Not applicable.

Statistical Analysis

Descriptive statistics were used to assess baseline demographic characteristics, and mean scores of continuous variables were calculated. Chi-square or Fisher exact tests were used to test for associations between the PDF and VMG groups. To examine changes in self-efficacy and knowledge across baseline, postintervention, and follow-up, a repeated measure analysis of variance was carried out. This was adjusted for confounders, including age, gender, ethnicity, education level, marital status, and specialty preference. Interaction between group and time were also tested and adjusted in the model if it was deemed to be significant (P<.05). All the analyses were carried out using SAS (version 9.3; SAS Institute) software. Arbitrary missing data at the follow-up test were handled using multiple imputation, using the 2-fold fully conditional specification approach proposed by Welch et al, using the predicted mean matching [82].

Qualitative Analysis

Data analysis was conducted using the 6-phase thematic analyses guide developed by Braun and Clark [83].

Authors HA and MA independently followed these 6 steps, and consensus was reached using the constant comparison approach where disagreements were discussed until mutual agreement was reached. The first phase involved entering responses in an Excel spreadsheet where they were read and reread as part of the data familiarization stage, and notes were taken for noticeable patterns. Manual coding was performed through highlighting repeated meanings that represented certain patterns as part of a data driven analysis. After data were coded, broader themes were generated. We used tables of codes and generated themes accordingly. At this stage, themes were reviewed and refined against the coded data level, followed by the whole dataset level. These themes were refined and named to ensure they reflected the meaning of the dataset they represent, highlighting interesting points and providing any potential explanations. Each theme was analyzed separately and in relation to other themes within the overall dataset meaning. Finally, report writing involved writing the meaning of the datasets on the basis of the generated themes supported by data extracts. At this stage, critical explanations about the rationale for such themes in the context of our research question about learning enhancement were sought.

Results

Sample Characteristics

A total of 83 students were eligible for the study, and 46 students agreed to participate (55% response rate). In total, 1 student left the PDF group, leaving 18 participants in this group and 27 in the VMG group. The student demographic characteristics in both groups were similar, including age, gender, specialty preference, and gaming frequency (Table 1). All students participated in the baseline and postintervention measures. Of those, 24 (88%) of the students in the VMG group and 14 (73%) students in the PDF group completed the follow-up test 7 weeks later. The sample characteristics are provided in Table 2.

Table 2. Sample characteristics.

Baseline characteristics	Group		Total	*P* value
	PDF	Game		
Gender, n (%)			.73[a]	
Male	4 (21.1)	8 (29.6)	12 (26.1)	
Female	15 (79)	19 (70.4)	34 (73.9)	
Age (years), n (%)				.68[a]
19-21	10 (52.6)	18 (66.7)	28 (60.9)	
22-25	8 (42.1)	8 (29.6)	16 (34.8)	
>26	1 (5.3)	1 (3.7)	2 (4.4)	
Ethnicity, n (%)				.32[a]
NZ European	5 (26.3)	8 (30.8)	13 (28.9)	
Maori	2 (10.5)	3 (11.5)	5 (11.1)	
Pacific	5 (26.3)	2 (7.7)	7 (15.6)	
Asian	1 (5.3)	6 (23.1)	7 (15.6)	
Other	6 (31.6)	7 (26.9)	13 (28.9)	
Marital status, n (%)				.58[a]
Single	16 (84.2)	18 (69.2)	34 (75.6)	
Couple/De Facto	3 (15.8)	7 (26.9)	10 (22.2)	
Married	0	1 (3.9)	1 (2.2)	
Specialty, n (%)				.75[a]
Medicine	10 (52.6)	14 (53.9)	24 (53.3)	
Surgery	0	2 (7.7)	2 (4.4)	
Subspecialities	2 (10.5)	3 (11.5)	5 (11.1)	
Do not know	7 (36.8)	7 (26.9)	14 (31.1)	
Gaming frequency, n (%)				.88[a]
Once a day	1 (5.3)	3 (11.5)	4 (8.9)	
More than one a day	1 (5.3)	2 (7.7)	3 (6.7)	
Once a week	3 (15.8)	5 (19.2)	8 (17.8)	
More than once a week	0	1 (3.9)	1 (2.2)	
I do not play games	14 (73.7)	15 (57.7)	29 (64.4)	
Education, n (%)				.31[b]
Undergraduate entry	12 (63.2)	20 (76.9)	32 (71.1)	
Master's degree	7 (36.8)	6 (23.1)	13 (28.9)	
Enrolment status, n (%)				.50[a]
Domestic	19 (100)	25 (92.6)	44 (95.7)	
International	0	1 (7.4)	2 (4.4)	

[a]Fisher exact test used.

[b]Chi-square test used.

Knowledge and Self-Efficacy

Although between-group differences in knowledge and self-efficacy at postintervention and 7-week time points compared with baseline were not significant (knowledge *P*=.95 and self-efficacy *P*=.85), the within-group differences were statistically significant (*P*<.001) for both groups (Figures 6 and 7). This significant difference persisted across unadjusted and adjusted models (adjusted by age, gender, ethnicity, education level, marital status, and specialty preference).

Clinical skills (Objective Structured Clinical Examination Scores)

The OSCE took place 10 weeks after the intervention. No significant differences between the intervention and control groups were found (*P*=.60; Table 3).

Figure 6. Knowledge across baseline, postintervention and follow up time points (no statistically significant differences).

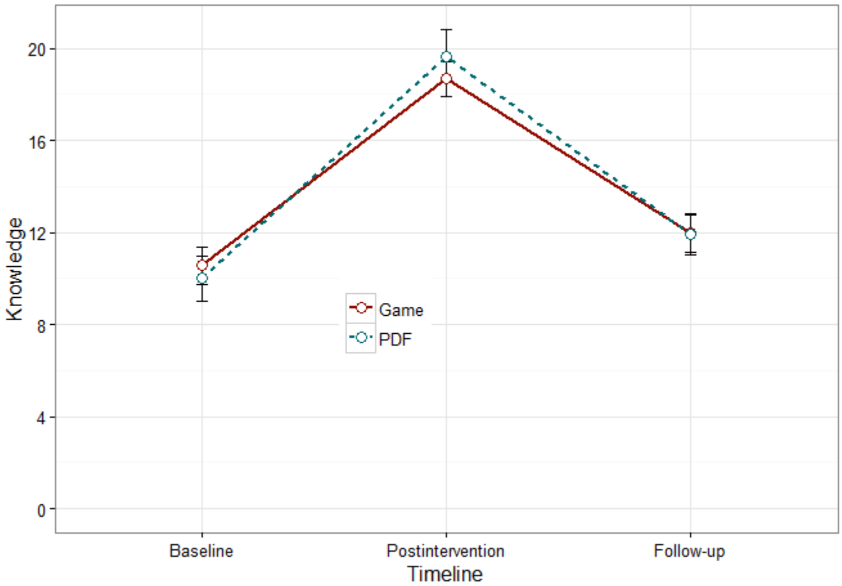


Figure 7. History-taking self-efficacy across baseline, postintervention and follow up time points (no statistically significant differences).

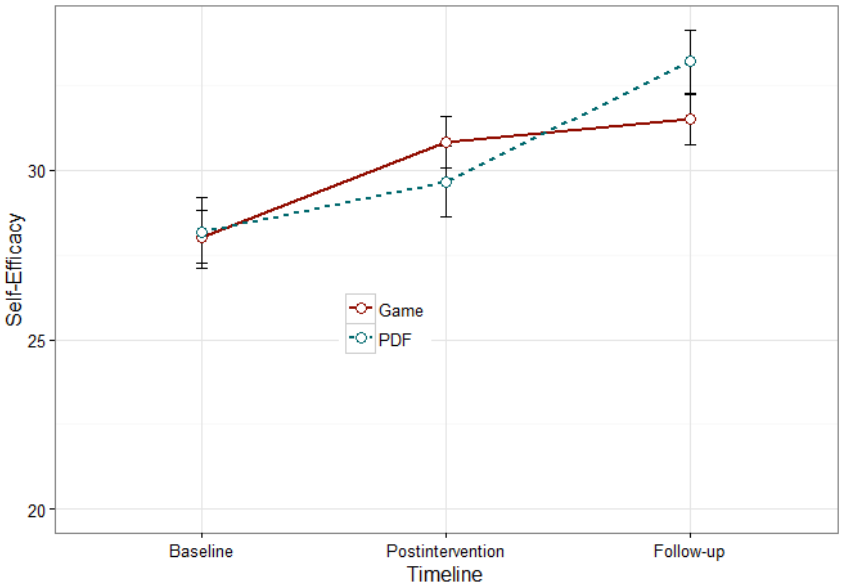


Table 3. History-taking cognitive load, difficulty level, objective structured clinical examination results, ease of use, usefulness, and satisfaction means, SD, and P values.

Outcome, group, and timeline	Mean (SD)	*P* value
Cognitive load		.88[a]
PDF		
Baseline	3.94 (0.64)	
Post	3.89 (0.68)	
Game		
Baseline	3.81 (0.62)	
Post	3.74 (0.66)	
Difficulty level		.68[a]
PDF		
Baseline	2.94 (0.73)	
Post	3.06 (0.64)	
Game		
Baseline	2.96 (0.59)	
Post	3.04 (0.52)	
Satisfaction		<.001
PDF		
Post	23.72 (4.35)	
Game		
Post	29.85 (4.86)	
Usefulness		.30
PDF		
Post	21.39 (4.1)	
Game		
Post	22.59 (3.62)	
Ease of use		.19
PDF		
Post	23.28 (3.68)	
Game		
Post	24.7 (3.36)	
Objective structured clinical examination		.60
PDF		
Post	14.33 (4.52)	
Game		
Post	15.04 (4.28)	

[a]Wilcoxon 2-sample test (nonparametric test).

History-Taking Cognitive Load and Level of Difficulty

There were no statistically significant differences from baseline between the 2 groups in either cognitive load scores (*P*=.88) or level of difficulty scores (*P*=.68; Table 2). Similarly, the ease-of-use and usefulness levels assessed at postintervention did not reach statistical significance (ease of use *P*=.19 and usefulness (*P*=.30; Table 3).

Satisfaction

There was a significant difference in satisfaction levels between the groups, with the VMG group having a statistically higher satisfaction level compared with the PDF group (mean difference of 6.13; 95% CI 3.27-9.00, P<.001; Table 3).

Qualitative Results

The overall findings indicate that the VMG and PDF file presented both HTC and structure well to participants. However, although VMG was viewed as a learning tool that aided understanding, encoding, and recalling history taking in an engaging and fun way, the PDF file was viewed as easy to read but difficult to learn or remember. In addition, the PDF file was seen as an electronic handout, which was no different to other handouts that were boring and nonengaging. A total of 4 major themes emerged from this analysis for both interventions: interventions as learning tools, enjoyment and engagement, generalizability, and suggestions for future improvement (Table 4). Identities are masked and replaced by pseudonyms in the section below to preserve confidentiality.

Table 4. Thematic analysis summary.

Game	Shared main themes	PDF
Memory aid for conceptual, procedural, and metacognitive knowledge development	Learning tool	An easy-to-read but hard-to-learn reference tool
Enhances enjoyment in learning history taking	Affects enjoyment and engagement	Neither engaging nor enjoyable
Text size, spelling, grammar optimization, add skip button, and increase difficulty levels	Design and functionality optimization suggestions	Add pictures, color and animations, as well as enhancing interactivity
Game specific theme: Game is generalizable to other clinical scenarios of different systems	—[a]	—

[a]Not applicable.

Theme 1: Interventions as Learning Tools

Visual Metaphor Game as a Memory Aid for Conceptual, Procedural, and Metacognitive Knowledge Development

A recognized benefit of using visual metaphors is enhancing memory retrieval through activating prior knowledge, thus allowing new knowledge and concepts to be learned [21]. For example, Stacey discussed how visual metaphors enhanced understanding and recall: “The use of symbols made important factors to ask about easy to recall.” Michael described how visual metaphors went beyond being just a memory aid to facilitating the “how” aspects of knowledge in terms of structure and systematic history-taking approach, as they added a “very good systematic process with visual symbols – aids memory,” whereas James shared that visual metaphors “helped realise that a system is important to follow - logical order/sequence.” On a metacognitive level in terms of “thinking about thinking” [84], visual metaphors enable the interpreter to make strong associations among concepts [85]. This enables the students to understand how they think and process information. Tanya stated that she enjoyed “the metaphors - since it has visual association with aspects of history-taking. I like that it is very structured in its approach,” and Sarah highlighted how visual metaphors were “very effective associating concepts with images.”

PDF Viewed as an Easy-to-Read but Hard-to-Learn Reference Tool

Although a majority of the PDF group participants viewed it as a clear and concise list, they found it hard to learn and remember. Although it was seen by William as “easy to read bullet points and concise information,” it was “not interactive, cannot remember just by reading, boring.” Perceiving the PDF file as a reference tool is consistent with the medical education literature, as technology-based handouts were thought of as a reference tool [86,87]. The negative impact of lack of PDF file interactivity on learning outcomes has been observed in a previous study where students in the noninteractive PDF handouts group had significantly reduced educational outcomes compared with other interactive electronic handouts [88]. This may have implications for medical education in general, as information (including difficult to understand concepts) is often presented in textual format, without accompanying images.

Theme 2: Instructional Design Influences Enjoyment and Engagement

Visual Metaphors Enhance Enjoyment in Learning History Taking

Almost all the VMG participants enjoyed the VMG. For example, Jake enjoyed “the results afterwards, i.e., remembering content of history-taking.” Stacey highlighted this by stating:

> *I enjoyed how easy it was to use…..made things fun to follow.*

This highlighted how enjoyment was linked to improved learning the material, as well as the ease of playing the game. This increase in game enjoyment is reported in the SGs literature, as studies have highlighted how students experienced high enjoyment levels while playing SGs [28,89]. Some researchers explained this increased enjoyment by the balance achieved between the learner’s abilities and the challenge posed by the game [90], which is consistent with Stacey’s statement above.

PDF Was Neither Engaging Nor Enjoyable

A majority of the PDF group participants found it boring and nonengaging because of a lack of visuals or interactivity, as Josephine stated, “It was not interactive at all, so it was not fun and I probably won't remember the stuff well.” Sam mentioned that “It was plain and was not necessarily enjoyable but just another handout but electronic.” These observations are supported by the literature, as technologically noninteractive modalities of presenting information, such as textual information, are considered less productive than interactive educational modalities [91]. Incorporating visual and verbal modalities of information is said to improve student understanding of taught material [92], which could explain the above statements.

Theme 3: Visual Metaphor Game Is generalizable to Other Clinical Scenarios of Different Systems

Overall, a majority of the VMG group participants appreciated the teaching value of the VMG and suggested generalizing it to teach other organ systems and clinical scenarios. Martin suggested the following:

> *To have different cases played out but with the same format so we learn about other systems with depth too. I feel that it has definitely helped me a lot with my memory recall, thanks.*

Mike said the following:

> *I think the game is focused on one type of clinical problem, chest pain. It would be nice to include more scenarios*

This did not feature in the PDF group responses. Visual metaphor use with games has been reported to serve as an "anchor" for conceptual knowledge [28], and this could explain the above responses. In the current study, though the history-taking visual metaphor had a cardiac focus, this could be easily applied to other organ systems (eg, respiratory system).

Theme 4: Design and Functionality Suggestions for Improvement

Participants in both groups offered suggestions on how both interventions could be optimized. Although PDF group participants offered several design suggestions, the VMG group had limited design suggestions.

Game Prototype Design and Functionality Suggestions

VMG participants made suggestions as to how the functionality and user interface could be optimized. In terms of design, some participants suggested text size optimization, as Simon commented "make text larger maybe?" Another suggested enlarging the buttons' size. Others suggested spelling and grammatical corrections and adding more levels of difficulty. In terms of functionality, suggestions included adding back and skip buttons, lag improvement, and minimizing repetition. Some participants, such as Harris, commented on the game prototype lagging and requested to "make it faster if possible. Great concept." Stephanie shared a similar opinion that we need to "Make the game respond faster." Lag in the game literature refers to the game delay in response to the actions of the player [93], and this has been shown to negatively impact the gameplay experience [49,93,94]. A possible explanation for the reported lag is the use of data-rich 3D graphics in the game [95].

Angela appreciated the positive impact of repetition on memory and acknowledged that repetition was less enjoyable when they have learned the material. This was captured in her response: "repetition was really good for memory, but after I was able to remember it, it became a little boring, sorry." The use of repetition in SGs is associated with improved learning [96], and the reported boredom after learning content is a good illustration of the flow state proposed by Csikszentmihalyi [95]. In the game literature, a balance needs to be struck between the game task difficulty and the player's ability to reach a state of flow [97]. On the basis of this, hard games could frustrate the player, whereas easy ones could bore the player. Therefore, the literature suggests increasing game difficulty as the user's knowledge improves, to maintain this state of flow. This increase in difficulty levels was suggested by Jasmin, as she wanted "an even more advanced level." However, others enjoyed the current 3 difficulty levels, which is evident in Yvonne's response about enjoying the "different levels to eventually learn history from memory." Finally, given that the reported boredom only occurred after the material was learned, which is the educational purpose of this SG; adding more difficulty levels might be questionable.

PDF Design and Functionality Optimization Suggestions

Most PDF group participants had suggestions to enhance PDF design and functionality through adding pictures, color, and animations, as well as enhancing interactivity.

In terms of design, Patrick suggested "using more animations, colours, pictures making it more interesting and interactive," whereas Patricia suggested adding functions, such as a record button, and a function to write on the file by having a "marker so you can write on and interact with the file to improve memorisation," and Ronald suggested to "have a record button that records the conversation and puts all patients' responses into each section immediately." These suggestions are in line with educational research, which showed that interactive audio visual functionalities enhance learning outcomes [98].

Discussion

Summary of Key Findings in This Study

This quasi-experimental pilot study that utilized both the quantitative and qualitative research methods approach evaluated a visual metaphor–enhanced 3D SG in teaching cardiac HTC to year-3 medical students in comparison to PDF-delivered teaching. More than half the sample indicated a preference to specialize in General Medicine in the future and were not playing games regularly. This study showed that the game is comparable to and as effective as textual information in increasing knowledge gains, enabling self-efficacy, and managing cognitive load, level of difficulty, perceived ease of use, and acceptability at various time points. However, the game was superior to textual information with regard to higher satisfaction scores relative to the PDF group. The reasons for this could be drawn from the qualitative analyses, which showed that the VMG was perceived to be a useful visual memory aid, more enjoyable, and transferrable to other clinical scenarios. On the contrary, the PDF group found the PDF file to be boring and merely another handout that was neither interactive nor engaging.

Quantitative Findings in the Context of the Literature

The quantitative findings were consistent with the current pre and postgraduate medical education SGs literature, which shows that they are at least as effective as traditional teaching methods. Consistent with the higher satisfaction scores in the VMG group and the qualitative findings of students viewing the game as more enjoyable, fun, and as a useful memory aid, 2 systematic reviews reported that SGs have the advantage of increasing students' enjoyment and interest in the topic taught [99,100]. The improvement in satisfaction was consistent with medical education game use among medical students when compared with traditional teaching [81,101]. The main benefit of the new teaching approach was that the VMG was statistically more significant in terms of satisfaction and was more favored in the qualitative analysis. As student satisfaction improvement is on its own an important learning outcome that complements academic achievement [102], this significant satisfaction improvement in the game group is worthy of further consideration. Research has shown that improved satisfaction enhances student academic performance, and its enhancement is strongly advised by educationalists [103,104]. A previous systematic review assessing the impact of educational games on medical students' learning outcomes found potential for improving learning outcomes, but it highlighted the need for more rigorous research [105]. Graafland et al systematically reviewed the literature for the impact of SGs on training health professionals and assessed their validity [106]. They included a total of 30 SGs, 17 of which were designed for educational purposes. However, none of these games were fully validated. The authors suggested validating games before incorporating them into teaching, and such validation needs to include content, face, and concurrent and predictive validity. In 2013, a Cochrane review looking at the use of games as a teaching method for health professionals found insufficient evidence for or against their use [107]. Wang et al conducted a systematic review in 2016 of SGs in medical education, and they identified 42 studies reflecting an increase in the number of SGs. Of these, only 19 studies included an evaluation of SGs, with a majority of these (n=17) associated with significant educational benefits [108]. Out of the 19 studies, only 4 were relevant to a medical student population, which assessed the use of SGs on medical students' knowledge and skill acquisition [81,109-111]. Only 1 of these studies found a significant improvement in knowledge, but this was for immediate recall only, as long-term retention was not assessed [111]. There are several possible reasons why this study showed no significant quantitative differences between the teaching methods apart from satisfaction scores. The duration (40 min) for learning the textual information may be a confounder, as it was regarded as too long by many students, which meant that students started interacting and assessing each other. As assessment drives learning [112], it could be argued that the PDF group assessing each other instead of engaging with the iPad only could have improved their scores independent of the PDF file. The influence of student-student interaction has been found in previous work to be similar to interactions between students and facilitators [81]. SGs are usually played more than once, which is different to the one-off 40-min play session design in this study, and this could also explain the lack of significant findings. In 2013, a meta-analysis of SGs reported how a lack of significant difference between games and traditional teaching could be attributed to the lack of playing the game on multiple occasions, as normally happens [32]. In that study, compared with multiple gameplay sessions, single session gameplay was not more beneficial than traditional teaching methods. Another possible reason for lack of significant difference in scores between the groups is the progression of skill development over time in the absence of formal teaching [3]. Although the VMG group only had 40 min of gameplay, both groups had unlimited access to traditional teaching material from which the PDF file was designed, which could have been a confounding factor in this study. Another possible explanation for this result is that the control group's awareness of its control status, as text PDF file is 1 of 2 interventions where the other is a game, influenced the group's behavior, as the members of the group may have tried to outperform the intervention group. This is a statistical confounder called the John Henry Effect [113], evident by PDF group participants assessing each other during the experiment and highlighted by the fact that a participant left the control group as soon as she knew she saw the PDF file was not the game. Embedding qualitative research alongside quantitative approaches has been suggested as a way of addressing the John Henry Effect [114].

Qualitative Findings in Context of the Literature

The qualitative research in this study was to investigate students' perceptions about the 2 educational approaches. Although both were seen as learning tools, the depth of such learning was better in the VMG group (conceptual, procedural, and metacognitive knowledge levels) compared with the PDF group (conceptual level). Several students touched on educational benefits of visual metaphors, for example, students saw how it enabled them to "associate" visual metaphors with HTCs. This mirrors findings on how visual metaphors work in terms of associating existing knowledge of certain visuals (familiar objects) and linking them with new (unfamiliar) concepts [20,115-117]. This associative function could be because of its visual nature, which, according to Pavio's Dual Coding Theory, creates cognitive associations with textual information, thus enhancing learning [118]. Active cognitive processing marrying new-to-previous knowledge is commonly referred to as "meaningful learning," a term that was pioneered by the educational psychologist, Ausubel, in the 1960s. He stated that "If I had to reduce all of educational psychology to just one principle, I would say this: The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly" [119].

The VMG group's awareness of such associative cognitive processes shows how visual metaphor enabled them to "think about thinking," which is an important metacognitive skill that is essential for self-regulated learning [120]. Another benefit reported by students using visual metaphor was reaching a personalized deeper meaning of history taking. This "Cognitive Elaboration" [117] was observed in some responses alluding to reaching a conceptual understanding of how history taking has multiple interrelated parts, which need to be uncovered systematically and carefully. This was consistent with the visual metaphor literature, as it was seen to provide the learner with a deeper meaning and understanding of the presented material [115].

Game Design Considerations

For the VMG group, most participants suggested developing the game for other organ systems and clinical scenarios suggesting the usefulness of the VMG. Few students commented on the need to address the gameplay lag. This need for immediacy is one of the main features contributing to improved gameplay [121], and this will be addressed in the next iteration of the game. In terms of implications for this study, clinical educational interventions that focus on HTC should continue to explore the value of visual metaphor use delivered through SGs to better teach this crucial skill. First, our SG was delivered as a one-off 40-min session, and according to the literature, more than one session of playing SGs yielded significantly better results compared with traditional teaching [32]. Therefore, the SG could be tested using a bigger sample and randomized study design for multiple sessions, to assess its actual impact compared with traditional teaching, given the overall literature support of SGs' benefits. Second, students' feedback on game optimization is being taken into consideration, as we are at the final stages of developing another game iteration, taking into account students' feedback. It is hoped that this version will be tested in a multicenter randomized controlled trial to assess its impact against traditional teaching. Third, another area that warrants further exploration is the use of visual metaphors in teaching other clinical topics, as students' qualitative feedback showed its perceived generalizability to other history-taking topics apart from cardiac history taking. In terms of cross-cultural impact, SGs could be useful for medical students across the globe, especially with the technically savvy current and future millennial medical students. Although most of the developed SGs are from developed countries, it is expected that other less-developed countries will follow suit [30]. The global benefits of SGs are widely recognized [122], as once they are developed and deployed, they can be accessed from all around the world, and collaboration could be sought to translate SGs to local languages to facilitate medical education efforts across the globe. This is increasingly made possible with higher access to the internet internationally. An example of this exponential increase to mobile phone access has jumped from 10% in 2002 to 80% in 2015 in Kenya, with more access to smartphones for higher education students [123]. Internet is seen by a majority of the 32 underdeveloped countries as beneficial for education [124].

Limitations

This study has several limitations. First, this study included a relatively small sample size, but of note, this was a pilot study, targeting one of the medical school's 3 main teaching sites in Auckland. The response rate was 55%, which is consistent with the SGs literature [125], as some studies response rates were as low as less than 50%. The studies with higher response rates incorporated the interventions as part of the curriculum, which was not possible in this study. Although a randomized study design would have been ideal, it was not feasible, as students were preallocated at the medical school administration level, factoring in students' residential proximity to the clinical teaching campus. This lack of contextual feasibility has been previously cited as one of the barriers to randomized study design adoption in medical education [126,127], and this is reflected in that less than 20% of SGs studies have used a randomized study design [128]. Moreover, participants were not blinded, which is a common issue in educational interventions, such as games, as students can easily identify the intervention. In addition, student-student interaction in the PDF group was not possible to control, which could have influenced their learning; potentially reducing the time limit for the PDF and VMG groups from 40 min may have helped address this issue. Finally, our design included a one-off game session, which differs from the usual gameplay experience of playing it more than once, which has been found to explain the lack of positive findings associated with SGs when compared with traditional teaching methods [32].

Conclusions

In a mixed-method experimental pilot study, we provide evidence that 40 min of playing an SG is as effective as textual information in teaching cardiac history taking to year-3 medical students, with the added value of increasing student satisfaction compared with traditional teaching. The qualitative analysis showed the game was more engaging, fun, enjoyable, and perceived as a useful visual memory aid. The game was as effective for the 3 Kirkpatrick model levels of evaluating educational interventions: affective, cognitive, and behavioral attitudes. Although SGs are in their infancy, their potential educational benefits should be harnessed, considering the current and future digitally adept medical students, especially when motivation and enjoyment of learning is at stake.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A showreel of the Med Metaphoria game.

[MP4 File (MP4 Video)120956 KB - games_v7i3e13748_app1.mp4]

Multimedia Appendix 2

Knowledge test marking criteria.

[PDF File (Adobe PDF File)206 KB - games_v7i3e13748_app2.pdf]

References

1. Nardone DA, Reuler JB, Girard DE. Teaching history-taking: where are we? Yale J Biol Med 1980;53(3):233-250 [FREE Full text] [Medline: 7405275]
2. Lipkin Jr M, Frankel RM, Beckman HB, Charon R, Fein O. Performing the interview. In: Lipkin MJ, Putnam SM, Lazare A, Carroll JG, Frankel RM, Keller A, et al, editors. The Medical Interview. New York: Springer; 1995:65-82.
3. Craig JL. Retention of interviewing skills learned by first-year medical students: a longitudinal study. Med Educ 1992;26(4):276-281. [doi: 10.1111/j.1365-2923.1992.tb00169.x]
4. Maguire GP, Rutter DR. History-taking for medical students. I-deficiencies in performance. Lancet 1976 Sep 11;2(7985):556-558. [doi: 10.1016/s0140-6736(76)91804-3] [Medline: 60632]
5. Pfeiffer C, Madray H, Ardolino A, Willms J. The rise and fall of students' skill in obtaining a medical history. Med Educ 1998 May;32(3):283-288. [doi: 10.1046/j.1365-2923.1998.00222.x] [Medline: 9743783]
6. Platt FW, McMath JC. Clinical hypocompetence: the interview. Ann Intern Med 1979 Dec;91(6):898-902. [doi: 10.7326/0003-4819-91-6-898] [Medline: 517891]
7. Ramsey PG, Curtis J, Paauw DS, Carline JD, Wenrich MD. History-taking and preventive medicine skills among primary care physicians: an assessment using standardized patients. Am J Med 1998 Feb;104(2):152-158. [doi: 10.1016/s0002-9343(97)00310-0] [Medline: 9528734]
8. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med J 1975 May 31;2(5969):486-489 [FREE Full text] [doi: 10.1136/bmj.2.5969.486] [Medline: 1148666]
9. Peterson MC, Holbrook JH, von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med 1992 Feb;156(2):163-165 [FREE Full text] [doi: 10.1097/00006254-199210000-00013] [Medline: 1536065]
10. Stewart MA. Effective physician-patient communication and health outcomes: a review. Can Med Assoc J 1995 May 1;152(9):1423-1433 [FREE Full text] [Medline: 7728691]
11. Robbins JA, Bertakis KD, Helms LJ, Azari R, Callahan EJ, Creten DA. The influence of physician practice behaviors on patient satisfaction. Fam Med 1993 Jan;25(1):17-20. [Medline: 8454118]
12. Kravitz RL, Callahan EJ, Paterniti D, Antonius D, Dunham M, Lewis CE. Prevalence and sources of patients' unmet expectations for care. Ann Intern Med 1996 Nov 1;125(9):730-737. [doi: 10.7326/0003-4819-125-9-199611010-00004] [Medline: 8929006]
13. Harmon G, Lefante J, Krousel-Wood M. Overcoming barriers: the role of providers in improving patient adherence to antihypertensive medications. Curr Opin Cardiol 2006 Jul;21(4):310-315. [doi: 10.1097/01.hco.0000231400.10104.e2] [Medline: 16755199]
14. Zolnierek KB, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care 2009 Aug;47(8):826-834 [FREE Full text] [doi: 10.1097/MLR.0b013e31819a5acc] [Medline: 19584762]
15. Aspegren K, Lønberg-Madsen P. Which basic communication skills in medicine are learnt spontaneously and which need to be taught and trained? Med Teach 2005 Sep;27(6):539-543. [doi: 10.1080/01421590500136501] [Medline: 16199362]
16. van Dalen J, Bartholomeus P, Kerkhofs E, Lulofs R, Van Thiel J, Rethans JJ, et al. Teaching and assessing communication skills in Maastricht: the first twenty years. Med Teach 2001 May;23(3):245-251. [doi: 10.1080/01421590120042991] [Medline: 12098395]
17. Harding SR, D'Eon MF. Using a Lego-based communications simulation to introduce medical students to patient-centered interviewing. Teach Learn Med 2001;13(2):130-135. [doi: 10.1207/S15328015TLM1302_8] [Medline: 11302033]
18. Alyami H, Su'a B, Sundram F, Alyami M, Lyndon M, Yu TC, et al. Teaching medical students history taking content: a systematic review. Am J Educ Res 2016;4(3):227-233 [FREE Full text] [doi: 10.12691/education-4-3-2. 1]
19. Hilliard RI. How do medical students learn: medical student learning styles and factors that affect these learning styles. Teach Learn Med 1995 Jan;7(4):201-210. [doi: 10.1080/10401339509539745]
20. Eppler MJ. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. Inf Vis 2006 Jun 22;5(3):202-210. [doi: 10.1057/palgrave.ivs.9500131]
21. Marcos A. The tension between Aristotle's theories and uses of metaphor. Stud Hist Philos Sci A 1997 Mar;28(1):123-139. [doi: 10.1016/s0039-3681(96)00010-6]
22. Paton RC. Towards a metaphorical biology. Biol Philos 1992 Jul;7(3):279-294. [doi: 10.1007/bf00129971]
23. St Clair RN. Visual metaphor, cultural knowledge, the new rhetoric. In: Reyhner J, Martin J, Lockard L, Gilbert WS, editors. Learn in Beauty: Indigenous Education for a New Century. Flagstaff, Arizona: Northern Arizona University; 2000:85-101.
24. Hanson L. Affective response to learning via 'visual metaphor'. In: Beauchamp DG, Braden RA, Baca JC, editors. Visual Literacy in the Digital Age. Blacksburg, VA: The International Visual Literacy Association; 1993:271-279.
25. Ortony A. Metaphor, language and thought. In: Metaphor and Thought. Second Edition. United Kingdom: Cambridge University Press; 1993:1-16.
26. Roth WM, Bowen GM, McGinn MK. Differences in graph-related practices between high school biology textbooks and scientific ecology journals. J Res Sci Teach 1999 Nov;36(9):977-1019. [doi: 10.1002/(SICI)1098-2736(199911)36:9<977::AID-TEA3>3.0.CO;2-V]

27. Eppler MJ, Burkhard R. Google Books. 2004. Knowledge Visualization: Towards a New Discipline and Its Fields of Application URL: https://books.google.co.in/books/about/Knowledge_Visualization_Towards_a_New_Di.html?id=15gunQAACAAJ&redir_esc=y [accessed 2019-08-27]
28. Rieber LP, Noah D. Games, simulations, and visual metaphors in education: antagonism between enjoyment and learning. Educ Media Int 2008 Jun;45(2):77-92. [doi: 10.1080/09523980802107096]
29. Fabricatore C. Learning Video Games: An Unexploited Synergy. In: Proceedings of the Annual Convention of the Association for Educational Communications and Technology. 2000 Presented at: AECT'00; February 16-19, 2000; Long Beach, CA, USA URL: http://eprints.hud.ac.uk/id/eprint/28000/
30. Susi T, Johannesson M, Backlund P. DiVA portal. 2007. Serious Games – An Overview URL: http://www.diva-portal.org/smash/record.jsf?pid=diva2:2416 [accessed 2017-10-01]
31. Kuipers DA, Terlouw G, Wartena BO, van't Veer JT, Prins JT, Pierie JP. The role of transfer in designing games and simulations for health: systematic review. JMIR Serious Games 2017 Nov 24;5(4):e23 [FREE Full text] [doi: 10.2196/games.7880] [Medline: 29175812]
32. Wouters P, van Nimwegen C, van Oostendorp H, van der Spek ED. A meta-analysis of the cognitive and motivational effects of serious games. J Educ Psychol 2013;105(2):249-265. [doi: 10.1037/a0031311]
33. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
34. Sweller J. Cognitive load during problem solving: effects on learning. Cogn Sci 1988 Jun;12(2):257-285. [doi: 10.1016/0364-0213(88)90023-7]
35. Adams DM, Clark DB. Integrating self-explanation functionality into a complex game environment: keeping gaming in motion. Comput Educ 2014 Apr;73:149-159. [doi: 10.1016/j.compedu.2014.01.002]
36. Sweller J, van Merriënboer JJ, Paas FG. Cognitive architecture and instructional design. Educ Psychol Rev 1998 Sep;10(3):251-296. [doi: 10.1023/A:1022193728205]
37. Sweller J, van Merriënboer JJ, Paas FG. Cognitive architecture and instructional design: 20 years later. Educ Psychol Rev 2019 Jan 22;31(2):261-292 [FREE Full text] [doi: 10.1007/s10648-019-09465-5]
38. Wong A, Marcus N, Sweller J. Instructional animations: more complex to learn from than at first sight? In: Campos P, Graham N, Jorge J, Nunes N, Palanque P, Winckler M, editors. Human-Computer Interaction -- INTERACT 2011. Berlin, Heidelberg: Springer; 2011:552-555.
39. Bates R. A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. Eval Program Plan 2004 Aug;27(3):341-347. [doi: 10.1016/j.evalprogplan.2004.04.011]
40. Curran VR, Fleet L. A review of evaluation outcomes of web-based continuing medical education. Med Educ 2005 Jun;39(6):561-567. [doi: 10.1111/j.1365-2929.2005.02173.x] [Medline: 15910431]
41. Hill AG, Yu TC, Barrow M, Hattie J. A systematic review of resident-as-teacher programmes. Med Educ 2009 Dec;43(12):1129-1140. [doi: 10.1111/j.1365-2923.2009.03523.x] [Medline: 19930503]
42. Yardley S, Dornan T. Kirkpatrick's levels and education 'evidence'. Med Educ 2012 Jan;46(1):97-106. [doi: 10.1111/j.1365-2923.2011.04076.x] [Medline: 22150201]
43. Freeth D, Reeves S, Koppel I, Hammick M, Barr H. University of Nebraska Medical. 2005. Evaluating Interprofessional Education: A Self-Help Guide URL: https://www.unmc.edu/bhecn/_documents/Interprofessional_education_eval_self_help_guide.pdf [accessed 2019-04-25]
44. Walker SL, Fraser BJ. Development and validation of an instrument for assessing distance education learning environments in higher education: the distance education learning environments survey (DELES). Learn Environ Res 2005 Nov;8(3):289-308. [doi: 10.1007/s10984-005-1568-3]
45. O'Hagan AO, Coleman G, O'Connor RV. Software Development Processes for Games: A Systematic Literature Review. In: Proceedings of the European Conference on Software Process Improvement. 2014 Presented at: EuroSPI'14; June 25-27, 2014; Luxembourg p. 182-193. [doi: 10.1007/978-3-662-43896-1_16]
46. Howe A, Keogh-Brown M, Miles S, Bachmann M. Expert consensus on contamination in educational trials elicited by a Delphi exercise. Med Educ 2007 Feb;41(2):196-204. [doi: 10.1111/j.1365-2929.2006.02674.x] [Medline: 17269954]
47. Ericsson KA, Simon HA. Protocol Analysis: Verbal Reports as Data. Cambridge, MA, US: MIT Press; 1984.
48. Annetta LA. The 'I's' have it: a framework for serious educational game design. Rev Gen Psychol 2010 Jun;14(2):105-113. [doi: 10.1037/a0018985]
49. Olsen T, Procci K, Bowers C. Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness. In: Proceedings of the International Conference of Design, User Experience, and Usability. 2011 Presented at: DUXU'11; July 9-14, 2011; Orlando, FL, USA p. 625-634. [doi: 10.1007/978-3-642-21708-1_70]
50. Ak O, Kutlu B. Comparing 2D and 3D game-based learning environments in terms of learning gains and student perceptions. Br J Educ Technol 2015 Aug 17;48(1):129-144. [doi: 10.1111/bjet.12346]
51. Möring SM. CORE. 2013. Games and Metaphor – A Critical Analysis of the Metaphor Discourse in Game Studies URL: https://core.ac.uk/download/pdf/50527885.pdf [accessed 2019-08-27]

52. Chang HY, Poh DY, Wong LL, Yap JY, Yap KY. Student preferences on gaming aspects for a serious game in pharmacy practice education: a cross-sectional study. JMIR Med Educ 2015 May 11;1(1):e2 [FREE Full text] [doi: 10.2196/mededu.3754] [Medline: 27731304]
53. Roettl J, Terlutter R. The same video game in 2D, 3D or virtual reality - how does technology impact game evaluation and brand placements? PLoS One 2018;13(7):e0200724 [FREE Full text] [doi: 10.1371/journal.pone.0200724] [Medline: 30028839]
54. Talley NJ, O'Connor S. Clinical Examination: A Systematic Guide to Physical Diagnosis. New South Wales, Australia: Churchill Livingstone; 2013.
55. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ 2004 Sep;38(9):974-979. [doi: 10.1111/j.1365-2929.2004.01916.x] [Medline: 15327679]
56. Paas FG. Training strategies for attaining transfer of problem-solving skill in statistics: a cognitive-load approach. J Educ Psychol 1992;84(4):429-434. [doi: 10.1037/0022-0663.84.4.429]
57. Klepsch M, Schmitz F, Seufert T. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. Front Psychol 2017;8:1997 [FREE Full text] [doi: 10.3389/fpsyg.2017.01997] [Medline: 29201011]
58. Szulewski A, Gegenfurtner A, Howes DW, Sivilotti ML, van Merriënboer JJ. Measuring physician cognitive load: validity evidence for a physiologic and a psychometric tool. Adv Health Sci Educ Theory Pract 2017 Oct;22(4):951-968. [doi: 10.1007/s10459-016-9725-2] [Medline: 27787677]
59. van Gog T, Kirschner F, Kester L, Paas F. Timing and frequency of mental effort measurement: evidence in favour of repeated measures. Appl Cognit Psychol 2012 Nov 28;26(6):833-839. [doi: 10.1002/acp.2883]
60. Schmeck A, Opfermann M, van Gog T, Paas F, Leutner D. Measuring cognitive load with subjective rating scales during problem solving: differences between immediate and delayed ratings. Instr Sci 2014 Aug 21;43(1):93-114. [doi: 10.1007/s11251-014-9328-3]
61. van Gog T, Paas F. Instructional efficiency: revisiting the original construct in educational research. Educ Psychol 2008 Jan 28;43(1):16-26. [doi: 10.1080/00461520701756248]
62. Brunken R, Plass JL, Leutner D. Direct measurement of cognitive load in multimedia learning. Educ Psychol 2003 Mar;38(1):53-61. [doi: 10.1207/S15326985EP3801_7]
63. Yeo G, Neal A. Subjective cognitive effort: a model of states, traits, and time. J Appl Psychol 2008 May;93(3):617-631. [doi: 10.1037/0021-9010.93.3.617] [Medline: 18457490]
64. Schwarzer R, Jerusalem M. Generalized self-efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A User’s Portfolio Causal and Control Beliefs. Windsor, UK: Nfer-Nelson; 1995:35-37.
65. Artino AR, La Rochelle JS, Durning SJ. Second-year medical students' motivational beliefs, emotions, and achievement. Med Educ 2010 Dec;44(12):1203-1212. [doi: 10.1111/j.1365-2923.2010.03712.x] [Medline: 21091760]
66. Woods JL, Pasold TL, Boateng BA, Hense DJ. Medical student self-efficacy, knowledge and communication in adolescent medicine. Int J Med Educ 2014 Aug 20;5:165-172 [FREE Full text] [doi: 10.5116/ijme.53d3.7b30] [Medline: 25341226]
67. Young HN, Schumacher JB, Moreno MA, Brown RL, Sigrest TD, McIntosh GK, et al. Medical student self-efficacy with family-centered care during bedside rounds. Acad Med 2012 Jun;87(6):767-775 [FREE Full text] [doi: 10.1097/ACM.0b013e318253dcdb] [Medline: 22534602]
68. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Jan;84(2):191-215. [doi: 10.1037/0033-295X.84.2.191]
69. Gee JP. What Video Games Have to Teach Us About Learning and Literacy. New York, USA: St Martin's Griffin Publishing; 2004.
70. Scholz U, Doña BG, Sud S, Schwarzer R. Is general self-efficacy a universal construct? Eur J Psychol Assess 2002 Sep;18(3):242-251. [doi: 10.1027//1015-5759.18.3.242]
71. Klassen RM, Klassen JR. Self-efficacy beliefs of medical students: a critical review. Perspect Med Educ 2018 Apr;7(2):76-82 [FREE Full text] [doi: 10.1007/s40037-018-0411-3] [Medline: 29484552]
72. Campbell J, Tirapelle L, Yates K, Clark R, Inaba K, Green D, et al. The effectiveness of a cognitive task analysis informed curriculum to increase self-efficacy and improve performance for an open cricothyrotomy. J Surg Educ 2011;68(5):403-407. [doi: 10.1016/j.jsurg.2011.05.007] [Medline: 21821221]
73. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
74. Briz-Ponce L, García-Peñalvo FJ. An empirical assessment of a technology acceptance model for apps in medical education. J Med Syst 2015 Nov;39(11):176. [doi: 10.1007/s10916-015-0352-x] [Medline: 26411928]
75. Hu PJ, Chau PY, Sheng OR, Tam KY. Examining the technology acceptance model using physician acceptance of telemedicine technology. J Manage Inform Syst 2015 Dec 2;16(2):91-112. [doi: 10.1080/07421222.1999.11518247]
76. Huang HM, Liaw SS, Lai CM. Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems. Interact Learn Environ 2013 Jul 22;24(1):3-19. [doi: 10.1080/10494820.2013.817436]
77. King WR, He J. A meta-analysis of the technology acceptance model. Inf Manag 2006 Sep;43(6):740-755. [doi: 10.1016/j.im.2006.05.003]

78. Gentry SV, Gauthier A, Ehrstrom BL, Wortley D, Lilienthal A, Car LT, et al. Serious gaming and gamification education in health professions: systematic review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
79. Witowski LL. Semantic Scholar. 2008. The Relationship Between Instructional Delivery Methods and Student Learning Preferences: What Contributes to Student Satisfaction in an Online Learning Environment? URL: https://pdfs.semanticscholar.org/0a7f/026a312295085edb20647fb2d97e5181ade1.pdf [accessed 2019-08-27]
80. Succar T, Zebington G, Billson F, Byth K, Barrie S, McCluskey P, et al. The impact of the virtual ophthalmology clinic on medical students' learning: a randomised controlled trial. Eye (Lond) 2013 Oct;27(10):1151-1157 [FREE Full text] [doi: 10.1038/eye.2013.143] [Medline: 23867718]
81. Sward KA, Richardson S, Kendrick J, Maloney C. Use of a web-based game to teach pediatric content to medical students. Ambul Pediatr 2008;8(6):354-359. [doi: 10.1016/j.ambp.2008.07.007] [Medline: 19084784]
82. Welch CA, Petersen I, Bartlett JW, White IR, Marston L, Morris RW, et al. Evaluation of two-fold fully conditional specification multiple imputation for longitudinal electronic health record data. Stat Med 2014 Sep 20;33(21):3725-3737 [FREE Full text] [doi: 10.1002/sim.6184] [Medline: 24782349]
83. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
84. Pintrich PR. The role of metacognitive knowledge in learning, teaching, and assessing. Theor Pract 2002 Nov;41(4):219-225. [doi: 10.1207/s15430421tip4104_3]
85. Serig D. A conceptual structure of visual metaphor. Stud Art Educ 2015 Dec 18;47(3):229-247. [doi: 10.1080/00393541.2006.11650084]
86. Islam MN, Majumder MD, Ja'afar R, Rahman S. Students' perceptions of 'technology-based' lecture handouts. Malays J Med Sci 2005 Jan;12(1):26-28 [FREE Full text] [Medline: 22605944]
87. Buzzetto-More N. Student perceptions of various e-learning components. Interdiscip J E-Learn Learn Obj 2008 Sep 01;4(17):113-135. [doi: 10.28945/370]
88. McLaughlin JE, Rhoney DH. Comparison of an interactive e-learning preparatory tool and a conventional downloadable handout used within a flipped neurologic pharmacotherapy lecture. Curr Pharm Teach Learn 2015 Jan;7(1):12-19. [doi: 10.1016/j.cptl.2014.09.016]
89. Ke F. Computer games application within alternative classroom goal structures: cognitive, metacognitive, and affective evaluation. Educ Tech Res Dev 2008 Jan 25;56(5-6):539-556. [doi: 10.1007/s11423-008-9086-5]
90. Sherry JL. Flow and media enjoyment. Commun Theory 2004 Nov;14(4):328-347. [doi: 10.1111/j.1468-2885.2004.tb00318.x]
91. Moreno R, Mayer R. Interactive multimodal learning environments. Educ Psychol Rev 2007 Jun 22;19(3):309-326. [doi: 10.1007/s10648-007-9047-2]
92. Fletcher JD, Tobias S. The multimedia principle. In: Mayer RE, editor. The Cambridge Handbook of Multimedia Learning. New York: Cambridge University Press; 2005:117-133.
93. Tseng P, Wang N, Lin R, Chen K. On the Battle Between Lag and Online Gamers. In: Proceedings of the Conference on Communications Quality and Reliability. 2011 Presented at: CQR'11; April 10, 2011; Florida, USA. [doi: 10.1109/CQR.2011.5996093]
94. Loh CS, Sheng Y, Ifenthaler D. Serious games analytics: theoretical framework. In: Loh CS, Sheng Y, Ifenthaler D, editors. Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement. New York, USA: Springer International Publishing; 2015:3-29.
95. Csikszentmihalyi M, Abuhamdeh S, Nakamura J. Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi. Heidelberg: Springer Netherlands; 2014.
96. Drummond D, Hadchouel A, Tesnière A. Serious games for health: three steps forwards. Adv Simul (Lond) 2017;2:3 [FREE Full text] [doi: 10.1186/s41077-017-0036-3] [Medline: 29450004]
97. Harteveld C, Guimarães R, Mayer I, Bidarra R. Balancing Pedagogy, Game and Reality Components Within a Unique Serious Game for Training Levee Inspection. In: Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment. 2007 Presented at: Edutainment'07; June 11-13, 2007; Hong Kong, China p. 128-139. [doi: 10.1007/978-3-540-73011-8_15]
98. Kennedy G. Promoting cognition in multimedia interactivity research. J Interact Learn Res 2004;15(1):43-61 [FREE Full text]
99. Blakely G, Skirton H, Cooper S, Allum P, Nelmes P. Educational gaming in the health sciences: systematic review. J Adv Nurs 2009 Feb;65(2):259-269. [doi: 10.1111/j.1365-2648.2008.04843.x] [Medline: 19032512]
100. Ricciardi F, de Paolis LT. A comprehensive review of serious games in health professions. Int J Comput Games Technol 2014;2014:1-11. [doi: 10.1155/2014/787968]
101. O'Leary S, Diepenhorst L, Churley-Strom R, Magrane D. Educational games in an obstetrics and gynecology core curriculum. Am J Obstet Gynecol 2005 Nov;193(5):1848-1851. [doi: 10.1016/j.ajog.2005.07.059] [Medline: 16260247]
102. Doménech-Betoret F, Abellán-Roselló L, Gómez-Artiga A. Self-efficacy, satisfaction, and academic achievement: the mediator role of students' expectancy-value beliefs. Front Psychol 2017;8:1193 [FREE Full text] [doi: 10.3389/fpsyg.2017.01193] [Medline: 28769839]

103. Artino AR. Motivational beliefs and perceptions of instructional quality: predicting satisfaction with online training. J Comput Assist 2008;24(3):260-270. [doi: 10.1111/j.1365-2729.2007.00258.x]
104. Eom SB, Wen HJ, Ashill N. The determinants of students' perceived learning outcomes and satisfaction in university online education: an empirical investigation*. Decis Sci J Innov Educ 2006 Jul;4(2):215-235. [doi: 10.1111/j.1540-4609.2006.00114.x]
105. Akl EA, Pretorius RW, Sackett K, Erdley WS, Bhoopathi PS, Alfarah Z, et al. The effect of educational games on medical students' learning outcomes: a systematic review: BEME guide no 14. Med Teach 2010 Jan;32(1):16-27. [doi: 10.3109/01421590903473969] [Medline: 20095770]
106. Graafland M, Bemelman WA, Schijven MP. Prospective cohort study on surgeons' response to equipment failure in the laparoscopic environment. Surg Endosc 2014 Sep;28(9):2695-2701. [doi: 10.1007/s00464-014-3530-x] [Medline: 24771195]
107. Akl EA, Sackett KM, Erdley WS, Mustafa RA, Fiander M, Gabriel C, et al. Educational games for health professionals. Cochrane Database Syst Rev 2013 Jan 31(1):CD006411. [doi: 10.1002/14651858.CD006411.pub3] [Medline: 23440807]
108. Wang R, DeMaria S, Goldberg A, Katz D. A systematic review of serious games in training health care professionals. Simul Healthc 2016 Feb;11(1):41-51. [doi: 10.1097/SIH.0000000000000118] [Medline: 26536340]
109. Aagaard E, Teherani A, Irby DM. Effectiveness of the one-minute preceptor model for diagnosing the patient and the learner: proof of concept. Acad Med 2004 Jan;79(1):42-49. [doi: 10.1097/00001888-200401000-00010] [Medline: 14690996]
110. Duque G, Fung S, Mallet L, Posel N, Fleiszer D. Learning while having fun: the use of video gaming to teach geriatric house calls to medical students. J Am Geriatr Soc 2008 Jul;56(7):1328-1332. [doi: 10.1111/j.1532-5415.2008.01759.x] [Medline: 18482292]
111. Boeker M, Andel P, Vach W, Frankenschmidt A. Game-based e-learning is more effective than a conventional instructional method: a randomized controlled trial with third-year medical students. PLoS One 2013;8(12):e82328 [FREE Full text] [doi: 10.1371/journal.pone.0082328] [Medline: 24349257]
112. Wormald BW, Schoeman S, Somasunderam A, Penn M. Assessment drives learning: an unavoidable truth? Anat Sci Educ 2009 Oct;2(5):199-204. [doi: 10.1002/ase.102] [Medline: 19743508]
113. Salkind NJ, editor. John Henry effect. In: Encyclopedia Of Research Design. Newbury Park, California: Sage Publications; 2010.
114. Kocakaya S. An educational dilemma: are educational experiments working? Educ Res Rev 2011;6(1):110-123 [FREE Full text]
115. Feinstein H. Meaning and visual metaphor. Stud Art Educ 1982;23(2):45-55. [doi: 10.2307/1319679]
116. Jeffreys MR. Guided visual metaphor: a creative strategy for teaching nursing diagnosis. Nurs Diagn 1993;4(3):99-106. [doi: 10.1111/j.1744-618X.1993.tb00097.x] [Medline: 7692908]
117. Jeong S. Visual metaphor in advertising: is the persuasive effect attributable to visual argumentation or metaphorical rhetoric? J Market Comm 2008 Feb;14(1):59-73. [doi: 10.1080/14697010701717488]
118. Clark JM, Paivio A. Dual coding theory and education. Educ Psychol Rev 1991 Sep;3(3):149-210. [doi: 10.1007/BF01320076]
119. Ausubel DP. Educational Psychology: A Cognitive View. New York: Rinehart and Winston; 1968.
120. Schraw G, Crippen KJ, Hartley K. Promoting self-regulation in science education: metacognition as part of a broader perspective on learning. Res Sci Educ 2006 Mar;36(1-2):111-139. [doi: 10.1007/s11165-005-3917-8]
121. Klimmt C, Hartmann T. Effectance, self-efficacy, and the motivation to play video games. In: Vorderer P, Bryant J, editors. Playing Video Games: Motives, Responses, and Consequences. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; 2006:133-145.
122. Edgcombe H, Paton C, English M. Enhancing emergency care in low-income countries using mobile technology-based training tools. Arch Dis Child 2016 Dec;101(12):1149-1152 [FREE Full text] [doi: 10.1136/archdischild-2016-310875] [Medline: 27658948]
123. Pew Research Center. 2015. Cell Phones in Africa: Communication Lifeline URL: https://www.pewresearch.org/global/2015/04/15/cell-phones-in-africa-communication-lifeline/ [accessed 2019-07-01]
124. Pew Research Center. 2015. Internet Seen as Positive Influence on Education but Negative on Morality in Emerging and Developing Nations URL: https://www.pewresearch.org/global/2015/03/19/internet-seen-as-positive-influence-on-education-but-negative-influence-on-morality-in-emerging-and-developing-nations/ [accessed 2019-07-01]
125. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
126. Prideaux D. Researching the outcomes of educational interventions: a matter of design. RTCs have important limitations in evaluating educational interventions. Br Med J 2002 Jan 19;324(7330):126-127 [FREE Full text] [doi: 10.1136/bmj.324.7330.126] [Medline: 11799017]
127. Dauphinee WD, Wood-Dauphinee S. The need for evidence in medical education: the development of best evidence medical education as an opportunity to inform, guide, and sustain medical education research. Acad Med 2004 Oct;79(10):925-930. [doi: 10.1097/00001888-200410000-00005] [Medline: 15383347]

128. 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]

Abbreviations

CLT: cognitive load theory
GSE: General Self-Efficacy
HTC: history-taking content
HTP: history-taking process
OSCE: objective structured clinical examination
SG: serious game
TAM: Technology Acceptance Model
VMG: Visual Metaphor Game
WM: working memory

Edited by T Rashid Soron; submitted 18.02.19; peer-reviewed by C Eichenberg, D Elliott; comments to author 28.04.19; revised version received 17.06.19; accepted 17.07.19; published 26.09.19.

Please cite as:
Alyami H, Alawami M, Lyndon M, Alyami M, Coomarasamy C, Henning M, Hill A, Sundram F
Impact of Using a 3D Visual Metaphor Serious Game to Teach History-Taking Content to Medical Students: Longitudinal Mixed Methods Pilot Study
JMIR Serious Games 2019;7(3):e13748
URL: https://games.jmir.org/2019/3/e13748
doi:10.2196/13748
PMID:31573895

Original Paper

Diagnostic Markers of User Experience, Play, and Learning for Digital Serious Games: A Conceptual Framework Study

Jun Wen Tan[1], BSc; Nabil Zary[1,2], PhD

[1]Games for Health Innovations Centre, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[2]Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

Corresponding Author:
Jun Wen Tan, BSc
Games for Health Innovations Centre
Lee Kong Chian School of Medicine
Nanyang Technological University
11 Mandalay Road
Singapore,
Singapore
Phone: 65 96945179
Email: junwen.tan2@gmail.com

Abstract

Background: Serious games for medical education have seen a resurgence in recent years, partly due to the growth of the video game industry and the ability of such games to support learning achievements. However, there is little consensus on what the serious and game components in a serious game are composed of. As a result, electronic learning (e-learning) and medical simulation modules are sometimes mislabeled as serious games. We hypothesize that one of the main reasons is the difficulty for a medical educator to systematically and accurately evaluate key aspects of serious games.

Objective: This study aimed to identify markers that can evaluate serious games and distinguish between serious games, entertainment games, and e-learning.

Methods: Jabareen's eight-phase framework-building procedure was used to identify the core markers of a serious game. The procedure was modified slightly to elicit "diagnostic criteria" as opposed to its original purpose of a conceptual framework. Following the identification of purported markers, the newly developed markers were tested on a series of freely available health care serious games—Dr. Game Surgeon Trouble, Staying Alive, and Touch Surgery—and the results were compared to the published test validity for each game.

Results: Diagnostic criteria for serious games were created, comprising the clusters of User Experience (UX), Play, and Learning. Each cluster was formed from six base markers, a minimum of four of which were required for a cluster to be considered present. These criteria were tested on the three games, and Dr. Game Surgeon Trouble and Staying Alive fit the criteria to be considered a serious game. Touch Surgery did not meet the criteria, but fit the definition of an e-learning module.

Conclusions: The diagnostic criteria appear to accurately distinguish between serious games and mediums commonly misidentified as serious games, such as e-learning modules. However, the diagnostic criteria do not determine if a serious game will be efficacious; they only determine if it is a serious game. Future research should include a much larger sample of games designed specifically for health care purposes.

(JMIR Serious Games 2019;7(3):e14620) doi:10.2196/14620

KEYWORDS

serious games; diagnostic criteria; medical education

Introduction

The past decade has seen a surge of digital games seeking to educate, train, or otherwise inform their players in a broad spectrum of topics [1]. These serious games are often defined as games developed for nonrecreational purposes and are frequently deployed as adjuncts to education and therapy [2,3]. Documented uses of serious games include teaching mathematics to young children, motivating the elderly to exercise, and increasing surgical competencies of medical students and residents [4-7]. Although such games are unable to replace full-time educators, they are effective at reinforcing concepts learned through conventional teaching methods [8].

This reinforcement is particularly notable in fields characterized by intensive, high-volume learning, such as medical education. Perhaps, due to the ubiquity of game playing in the present day, where video games generate astronomical levels of revenue, it is no surprise that educators are seeking to bend the vast potential of gaming toward serious causes like learning [3,9,10]. This shift in paradigm comes at the heels of the once ironclad assumptions linking the notions of play with time-wasting and frivolousness [1,11]. As more serious games begin featuring in institutions across the globe, they receive ever-increasing attention from stakeholders seeking to leverage the benefits of serious games—be it for profit or learning.

However, despite the growing interest in serious games and the vast array of such games being pushed to market, there remains little consensus on core components that afford a serious game its seriousness while also retaining the fun and entertaining elements that characterize recreational games [3,12,13]. Many serious games are affected by a variety of issues that impede the said serious agenda. Although such games are often the result of attempts to gamify an existing teaching method, these often suffer from poor instructional or game design and thus perform poorly as compared to methods they were meant to replace or support [14,15]. Surprisingly, the reverse is also true: When educational elements are inserted into near-completed games as an afterthought, they do not necessarily perform well [16].

For an educational serious game to be effective, the educational content must be robust, appropriate for the target audience, and well-integrated into the game [17,18]. Thus, it comes as no surprise that many serious games are now subject to rigorous validation studies long before implementation.

Recent reviews of the literature suggested that the vast majority of research remains dedicated to testing the efficacies of specific serious games for their intended purposes and, where applicable, the cognitive processes activated by playing these games [3,4,6,19,20]. Often, newly developed games are trailed to a targeted sample, typically with a control group, and game efficacy is determined by whether the game achieved its intended result, such as increases in standardized test scores. This user-centric focus, while meritorious, often does not addressed the factors, mechanisms, or processes that afford serious games efficacy and receptivity by their target audiences. It leaves the following question unanswered: “How do I know this serious game will be both serious and a game?”

Consequently, this has resulted in a gap in the literature with regard to consistent, definitive, and validated diagnostic criteria to serve both as a reference and basis of evaluation for serious games. This has impeded the development of new serious games whose topics have not yet been subject to extensive testing and the objective-appropriate selection of ready-made games already available on the market. The lack of criteria is especially punishing for educational pathways featuring intensive study, such as medical institutions, that may be seeking additional educational strategies to enhance student engagement and quality of learning.

The development and acquisition of serious games are costly, time-consuming endeavors, especially in the context of health care. Often, the initiators of serious game development are health care professionals idealizing games to tackle specific health or educational challenges, while the actual game makers are technical specialists with comprehensive knowledge of game development. The absence of practical guidelines, or diagnostic criteria, risks the production of ineffective serious games which, in turn, is compounded by the poor allocation and utilization of resources and leads to institutions being discouraged from adopting games that might significantly enhance the performance of their students.

Although Yusoff et al [21] and Rooney [22] have proposed two frameworks established for use by game designers and educators, both were designed for use during the game development process, as opposed to the validation of ready-made games.

The framework by Yusoff et al [21] combined learning theory with gaming requirements to ensure games meet learning outcomes [21]. The Learning Activity, built from the intended learning outcomes and game aspects that support learning and engagement, was key to this framework. It acted with the game’s genre and could be modified based on feedback derived from a player’s achievement within the game. However, the framework acts as a guiding tool during the developmental phase of game design and does not readily function as a validation framework. The theoretical bases of the framework also require that users possess familiarity with either or both game development and pedagogy and may not be used easily by prospective game producers unfamiliar with either.

Rooney [22] proposed a triadic interaction of play, pedagogy, and fidelity that together form a framework for serious game design in higher education. They discussed the theoretical underpinnings and key literature and challenges addressed. Although play and pedagogy referred to game play and the pedagogical aspects of learning, fidelity was defined as the extent to which the game emulated the real world, both physically, such as visual displays and behaviors of physics engines, and functionally, as the extent to which the game behaves like the real world in response to player actions. However, Rooney [22] acknowledged difficulties in balancing the framework’s components, in part, due to the multidisciplinary and sometimes competing nature of game design that has thus far prevented reconciliation of the components into a coherent theoretical framework. In addition, no validation of the framework, such as designing a game from the ground up, was provided.

Although several frameworks related to serious games have been proposed, all require above-average competencies in the knowledge of game development and are applicable in stages of game development. While touched upon, both frameworks did not define the base markers comprising serious games, were not validated against the existing ready-made serious games, and were not created with specific relevance to or confirmed to be compatible with medical and medical education games. In this context, none are easily distilled into the base components of a serious game.

Therefore, this study aims to deconstruct serious games into their base markers before validating them and an accompanying diagnostic criteria for evaluating serious games that can be used by nongaming experts. The proposed “diagnostic criteria” would enable the validation of ready-made serious games and act as a guide to ensure newly created serious games are both serious and games. The study’s objectives are hence twofold. The study will first detail the procedure used for the deconstruction of serious games into their base markers. These markers will then be tested on three established serious games to ensure that the diagnostic criteria are able to assess the serious and game aspects of serious games.

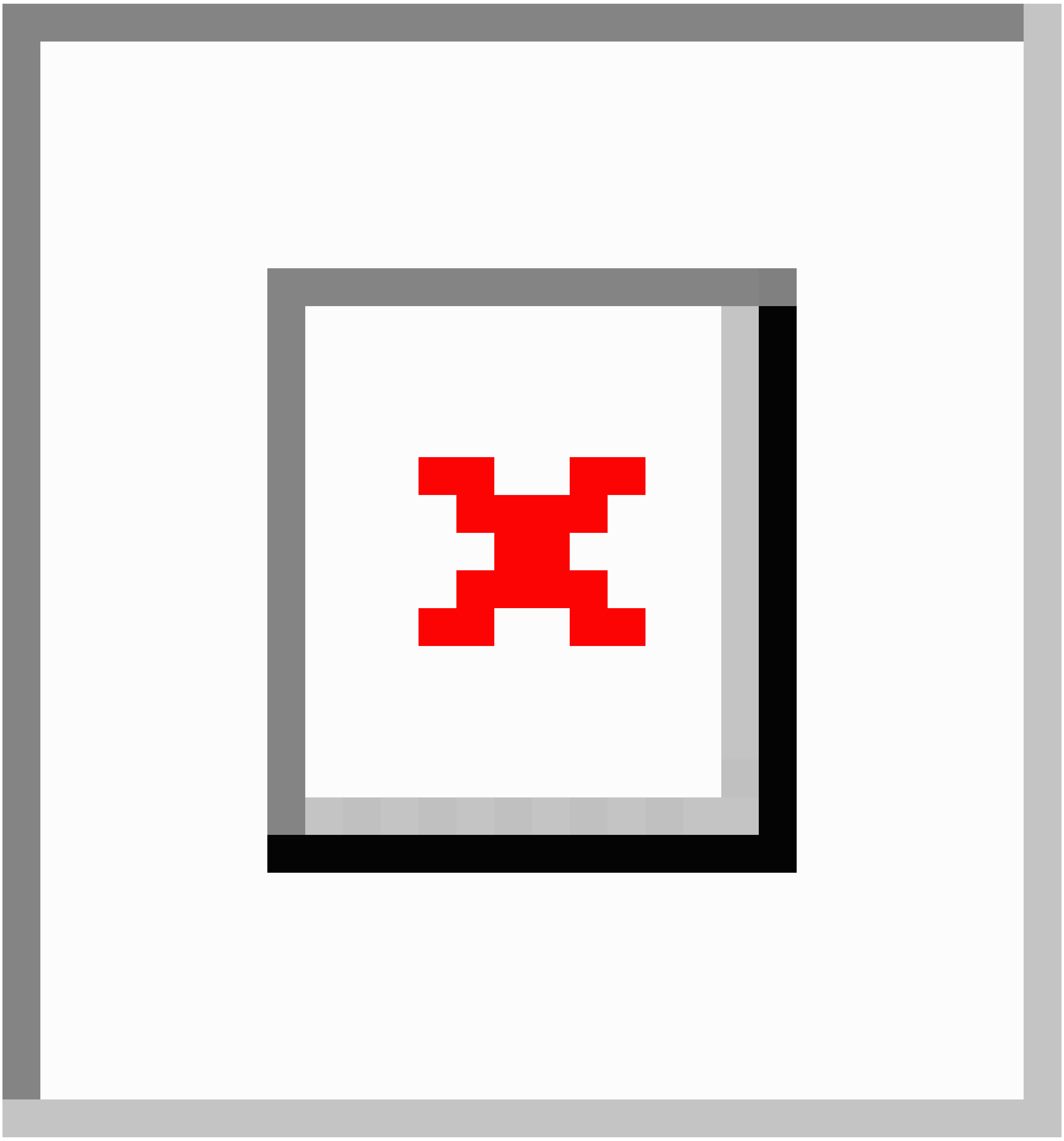
Methods

Deconstruction of Serious Games Into Base Markers

Figure 1 shows the eight-phase procedure by Jabareen [23] for building a conceptual framework that was used for the isolation of base markers.

In phase 1, sources of serious game data were extensively mapped. Sources consisted of multidisciplinary journal articles reporting on evaluative research on serious games, consultations with educators experienced in game-based teaching, and auditing of learning-through-play workshops during international conferences at the Lee Kong Chian School of Medicine in Singapore. In phase 2, the accumulated data were extensively reviewed and categorized according to discipline, representative power, and importance. In phase 3, a qualitative inquiry into the data was conducted. Concepts that surfaced were identified and named even if they competed with or contradicted each other. In phase 4, concepts were deconstructed to identify their main attributes, characteristics, and roles in serious games. Deconstructed concepts were then sorted by their ontological, epistemological, and methodological roles. In phase 5, concepts with similar features were condensed, reducing the total number of concepts to more manageable levels. In phase 6, using an iterative process, the newly integrated concepts were synthesized into a single theoretical framework. Concepts were subject to repeated resynthesis until a sensible general theoretical framework was recognized. In the study, three clusters of interrelated serious games markers were produced instead of a framework. In phase 7, we examined three established serious games to determine if the markers proposed by Jabareen’s process can be confirmed in games via the detection of each cluster’s base markers. Finally, in phase 8, as theoretical frameworks representing multidisciplinary phenomena remain dynamic, the tested framework will eventually be revised in light of new findings and developments as part of future studies undertaken by the authors and the broader serious games scientific community.

Figure 1. Overview of Jabareen's eight phases [23].



Games Used to Validate the Framework

Three serious games for medical education were used in the validation of the clusters and their base markers. All three games had been previously validated as serious games efficacious in their target areas or audiences and are easily accessible by medical students and professionals, as they are freely available from either the Android or iOS mobile app stores or from the developer's website [24-26].

"Touch Surgery" simulates hundreds of different medical procedures in three-dimensional environments and enables users to train in a three-dimensional environment. Upon selection of a procedure, users are guided through all the appropriate steps of an operation step-by-step, following which, players are provided opportunities for rehearsal and self-assessment via multiple-choice questions that focus on cognitive decision making. Of the available modules, the chest tube insertion procedure was chosen due to its relative novelty to the investigators of the study.

"Dr Game, Surgeon Trouble" trains medical residents to recognize and correct responses to equipment failure events during laparoscopic surgery. Player attention is held by a match-three-puzzle minigame unrelated to surgery and must concurrently solve equipment-related problems in a visually embedded laparoscopic tower. The onset of problems is accompanied by signals such as camera blurring or changes in lighting intensity, which occur partially beyond a player's direct focus of attention, emulating a surgical environment. Upon detection of a problem, the player is then moved to a troubleshooting mode where the educational aspect of the game features. At this stage, players interact with the laparoscopic tower to resolve equipment malfunctions and are given a set number of "attempts" at the correct solution. Throughout, the game maintains a cycle of challenges, actions, and direct feedback.

"Staying Alive" teaches the general public and health professionals about the management of a sudden cardiac arrest. Players are presented a three-dimensional environment wherein a man has just collapsed due to the onset of cardiac arrest and is guided through the appropriate measures and techniques required to maximize his chances of survival. Difficulty increases across levels, where level 1 takes place in an indoor office, while level 2 takes place in a sports field.

Validation of the Markers

In the validation phase of the study, each game was played till completion and markers that surfaced were noted using the Serious Games Markers Scoring Protocol (Multimedia Appendices 1 and 2). As the markers serve as a means to detect the base markers of the game, as opposed to how well or how much of a marker is represented, a binary scoring system was utilized. Markers that clearly featured in games were marked as present and their total scores were tallied.

Results

Markers of Serious Games

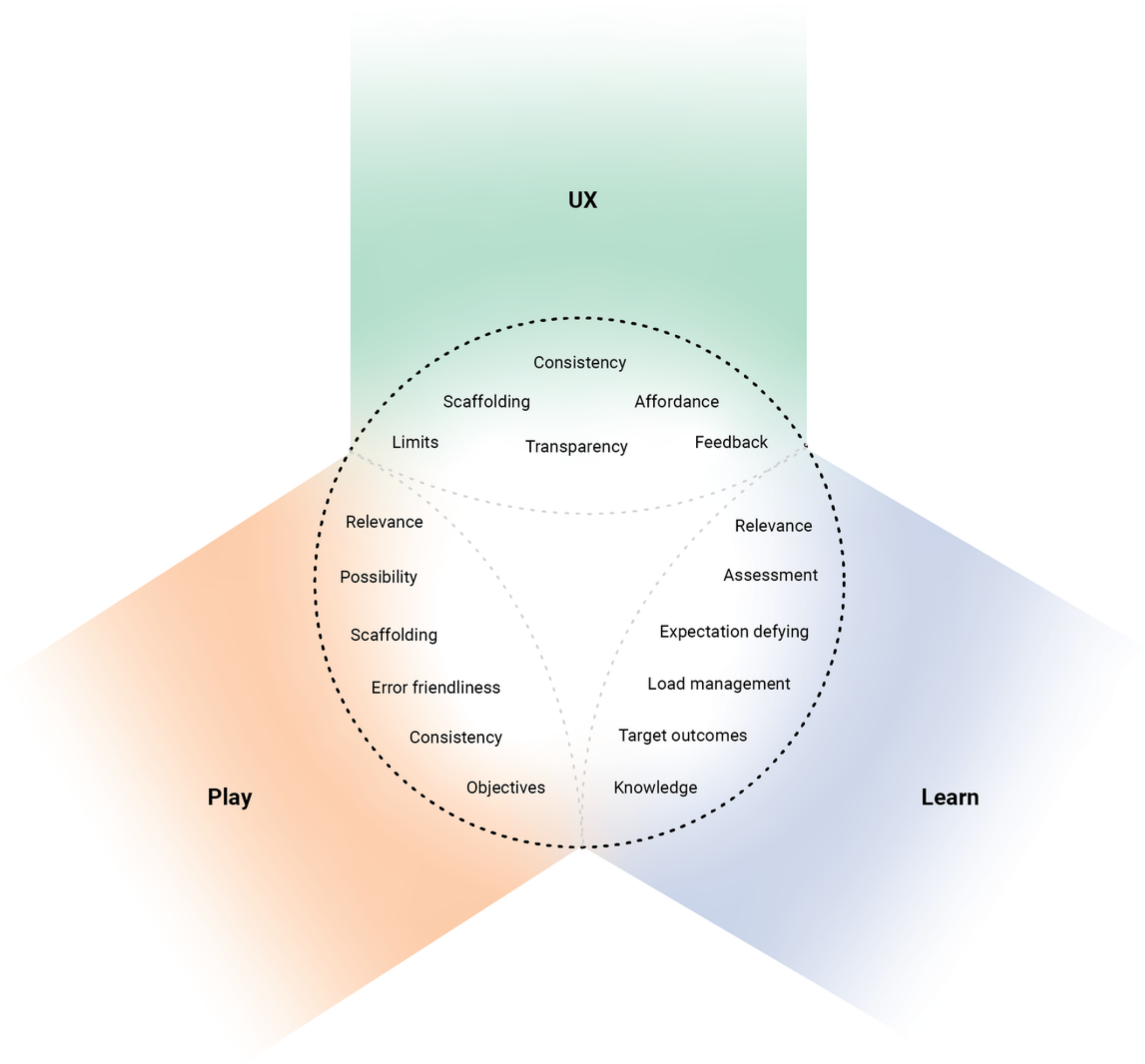
Figure 2 shows the three clusters of markers that, when brought together, form the base composition of a serious game—User Experience (UX), Play, and Learning—that must be present for a serious game to be considered both serious and a game. This is in contrast to recreational games and e-learning, which are respectively missing the clusters of Learning and Play.

UX encompassed the player's cognitive and affective experiences while playing and interacting with the game and found ways in which they were moderated by the game's actual play, professional applicability, and usability through its user interface. The cluster comprised elements detailed in Textbox 1.

The cluster of Play encompassed the game's nonlearning mechanics, genres, style, control mechanism (ie, controller and keyboard), within-game objectives, playable content, and infrastructure (ie, cloud-based, hardware requirements). The cluster comprised markers detailed in Textbox 2.

Learning encompassed the scenes, scenarios, or situations wherein the player is exposed to the knowledge or skills the game intends to impart. They need not be singular, and distinct instances can be spread out or embedded into the game's central mechanics. The cluster comprised markers detailed in Textbox 3.

Figure 2. Overview of the diagnostic criteria of serious games. UX: User Experience.



Textbox 1. Markers comprising the cluster of User Experience.

- Limits: The game attempts to govern how users should play, in accordance with the developer's intent, and discourages styles that circumvent its overall purpose. It also refers to constraining the player within the play area to avoid unintended bugs.
- Feedback: For most of the time, the game responds to actions taken by the player, confirming to the player that the said action had taken place. It typically refers to auditory or visual cues in response to all actions, no matter how inconsequential, such as clicks or bumps when interacting with objects.
- Scaffolding: The game gradually increases the intensity of mechanics that directly influence player impressions and feelings toward it in the moment of play. It also refers to the gradual introduction of new mechanics across time, if such mechanics are available.
- Affordance: Actions or purposes available to the player should be affordable, in that the game makes clear how the presented options are to be used, or such that the player will be able to figure out how options are to be used. Does not apply to puzzle-based games or objectives that require significant cognitive input.
- Transparency: Options available to the player, in both gameplay and the game's user interface, require little to no thought to discern their purpose. Does not apply to mechanics designed to assess levels of competencies and novel mechanics accompanied by dedicated explanations.
- Consistency: Game manages its interactions with the player in a mostly consistent manner, inclusive of predictable outcomes when using external instruments (mice, keyboards, etc) to interact with the game.

Textbox 2. Markers comprising the cluster of Play.

- Relevance: The game's overall themes, style, genre, and design should be professionally applicable to its target audience at the time of publication.
- Objectives: The game possessed explicit overarching or instance-specific purposes for the player to attain.
- Possibility: The game should introduce more substance, usually new game content, events, and features over time. These may include new methods to overcome challenges, and entirely novel materials or themes within the game.
- Consistency: Methods, controls, game mechanics, and rules for gameplay should display consistency most of the time. Does not consider the actual playable content of the game.
- Scaffolding: The effort or cognitive input required by a player to overcome challenges or content presented by the game should gradually increase as play time increases. Play should begin at a manageable level before increasing in difficulty to maintain challenge and flow. If difficulty is moderated by variations of in-game mechanics, then these mechanics should be gradually introduced.
- Error Friendliness: The game allows for player errors to be made, allows consequences of said errors, and does not bar an action or cease missions upon detection of an incorrect choice.

Textbox 3. Markers comprising the cluster of Learning.

- Target Outcomes: The game has clear goals or target outcomes it intends to achieve. It can be operationalized as intent to raise latent or constant awareness of a problem. If targeted at behaviors, it can be operationalized as behavioral modification attempts.
- Knowledge: The game contains the knowledge and skills it intends for the player to take in and utilize.
- Relevance: Knowledge or skills present in the game have been set to a standard suitable for the target audience's learning level and interests, wherein interest is defined as whether the content will be applicable to the said audience.
- Cognitive Load Management: The game supports the player's exposure to new knowledge and allows for the regular intake of new information. The game ideally seeks to maintain players within the zone of proximal development, avoids overloading the player, and ensures that the pace is not too slow as to induce boredom or disinterest.
- Expectation Defying: The game attempts to prevent players from being conditioned to one stimulus to the point that introducing a second, necessary stimulus has no effect. Operationally, the game avoids monotonous and predictable "learning moments" and takes steps to keep players from knowing what will happen next with regard to learning.
- Assessment: The game features a system to assess player learning improvements with regard to the target learning outcomes. It need not be operationalized as traditional scorecards and may feature as achievement or medal systems. "Total Score" features that lack clarity and specifics do not qualify.

Scoring Methodology

Due to the need for a simple validation process, each marker of each cluster is marked on a binary level—they are either present (1) or not present (0)—and must be indisputably present in a game to be considered so. Each cluster (UX, Play, and Learning) contains six markers and will be scored from 0 to 6 depending on the number of markers present, with a score of ≥4 denoting a cluster as present. When all clusters score ≥4, the game is considered a serious game.

The serious game to be validated is to be played from start to finish or, for games designed to end midway due to player error, until five game-ending mistakes are committed or until the player is no longer willing to continue due to the inability to overcome the obstacle wherein the mistakes were committed, whichever comes first. The player is to observe the game for each cluster's markers and record them accordingly.

To facilitate the ease of reporting game scores posttabulation, scores are reported in a UX/Play/Learning format, abbreviated as #U/#P/#L, where # ranges from 0 to 6 depending on the number of markers recorded for the respective clusters.

Scoring the Seriousness of the Games

Game Scores

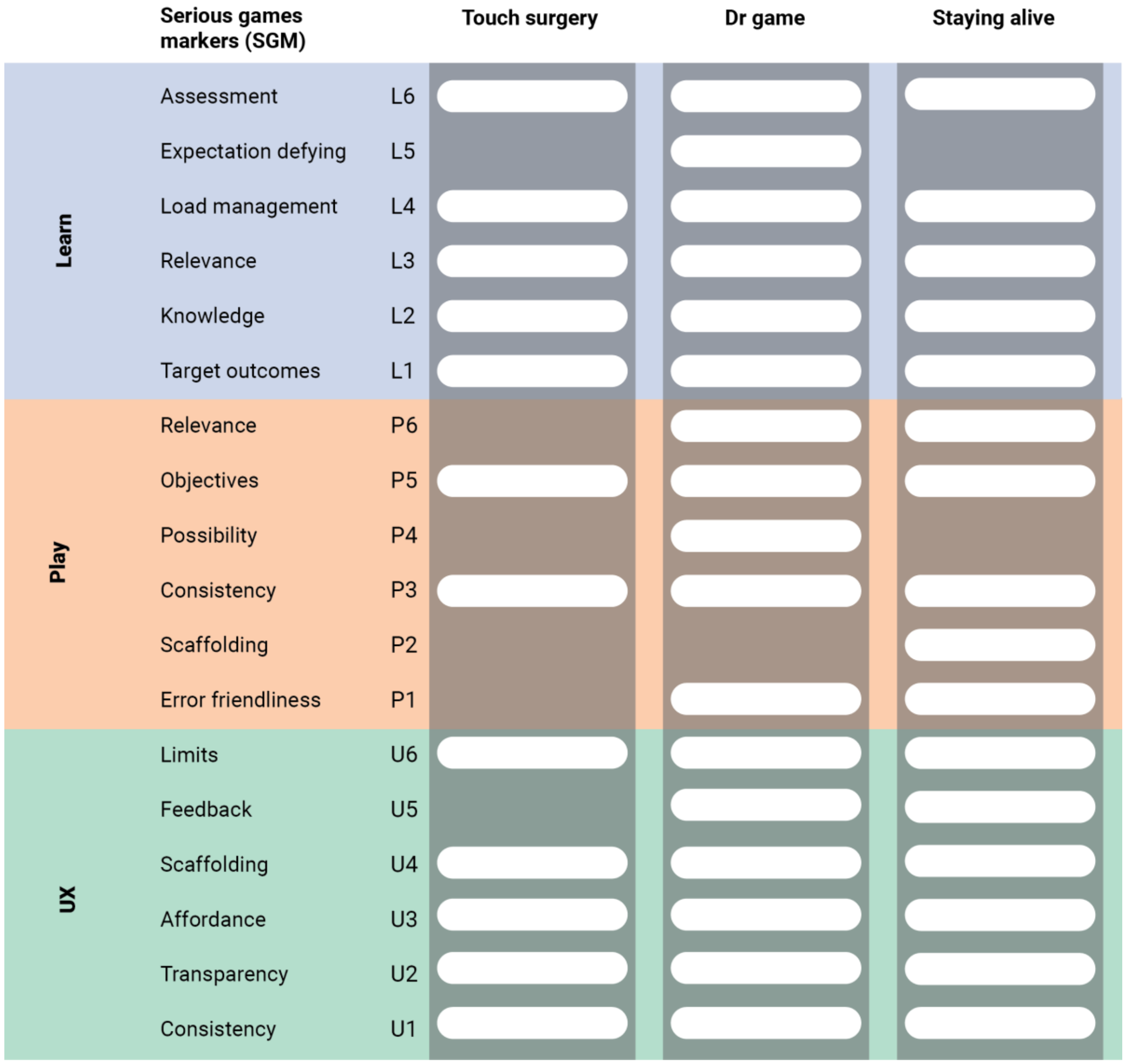
"Touch Surgery" scored 6U/0P/5L. It was found to possess all six markers of UX and five markers for Learning, missing out on Expectation Defying due to the absence of conditioning stimuli in the Learning cluster of the game. However, the game scored 2 points in the cluster of Play, as the "game" component resembled e-learning as opposed to games defined by Alvarez and Djaouti [2].

"Dr. Game Surgeon Trouble" scored 6U/5P/6L. It was found to possess all six markers of UX and Learning, but missed out the Scaffolding component in Play due to the study being unable to confirm if its match-three-puzzle minigame increased in difficulty over time nor did the frequency of problem signaling increasing over time.

"Staying Alive" scored 6U/5P/5L. It was also found to possess all six markers of UX but missed out on the Possibility component in Play because the second level was mostly the same as the first and there was uncertainty about whether there was an increase in difficulty.

Figure 3 summarizes the game scores for the three games.

Figure 3. Markers for the three games evaluated. UX: User Experience.



Discussion

Principal Findings

The diagnostic markers of serious games were developed by an iterative process that began with the identification of knowledge within the existing literature. During this process, the two existing frameworks of serious games proposed by Yusoff et al [21] and Rooney [22] were found to be used during the game development process and not as a validation tool for ready-made games [21,22]. As no such tool existed, this study attempted to close the gap in the literature through the construction and validation of diagnostic markers that included extensive deconstruction and resynthetization of data until a sensible cluster of markers was recognized. The newly developed clusters were then tested on a series of ready-made health care serious games. Although all games were previously validated by at least one study, some appeared to defy the criteria's definition of a serious game.

Haubruck et al [25] conducted a randomized controlled trial that demonstrated the validity and effectivity of "Touch Surgery" as an educational adjunct for chest tube insertion procedures [25]. The study applied the Gameplay/Purpose/Scope (G/P/S) classification model by Djaouti et al [13] to confirm Touch Surgery's status as a serious game. However, the G/P/S model aims to precisely classify a serious game into a subgenre via a system that acknowledges both the "serious" and "game" dimensions. The model's principal goal was to allow educators to find serious games that were useful for their specific causes but were otherwise not designed for or marketed toward their use. A secondary application of the model was to identify entertainment games that could be repurposed for use in education. In both circumstances, the model requires that the game being classified qualifies as one to begin with.

Similarly, notwithstanding the study by Haubruck et al, no citation referring to Touch Surgery as a serious game could be found, including the manufacturer's website. This discrepancy is most likely due to a case of mistaken identity due to the multitude of definitions for what exactly construes a serious game. Djaouti et al [13] noted that serious gaming draws expertise from a broad range of fields such as pedagogy, computer science, and medicine that may not wholly agree with the definition of a serious game [13].

Alvarez and Djaouti defined "serious games" for health care as games that were designed specifically for the serious purpose of providing education via digital devices [2]. The term "serious gaming" was defined as the use of any digital games for health care education and could also refer to nonserious games used for a serious purpose, which they termed "serious diverting" [2]. Therefore, although none of the definitions define what construes a game, it can be inferred that the related software must be games to begin with. A related concept ("gamification") was defined as the use of game elements in real-world applications that typically were not games to begin with [27]. Such characteristics include rewarding users of e-learning with points, badges, or achievements upon completion of a module. As "Touch Surgery" does not demonstrate transparent implementation of game elements, does not offer users challenging problems outside of its assessment module, and was not cited as a serious game by its manufacturer, it is more applicably categorized as an e-learning tool than a serious game, thereby falling in line with the markers and proposed diagnostic criteria, notwithstanding its efficacy as an educational tool.

Conversely, "Dr. Game Surgeon Trouble" appears to fit well with the proposed criteria for a serious game, and its status as one is supported by existing literature. Graafland et al [24] had developed the game specifically to improve the problem recognition and resolution skills of surgical trainees [24,28]. The game has had its construct validity established, and a randomized controlled trial demonstrated favorable outcomes in improving trainee problem recognition and resolution. Moreover, while there is no evidence to suggest that Graafland et al [24] employed existing conceptual frameworks of serious games during the development phase, the game demonstrates the distinct, yet well-integrated game and educational components that fit neatly within common definitions of serious games. Although requiring only a few minutes to complete, "Staying Alive" appeared to fit well within the diagnostic criteria, and its status as a serious game is supported by both the manufacturer and a randomized controlled study by Drummond et al [26].

Validating all three games through either Yusoff's [21] or Rooney's [22] conceptual frameworks would logically require that both first be adapted for use in ready-made games. Such a venture, while plausible, is likely a complex undertaking that requires some degree of familiarity with game development, game playing, and related pedagogies such as learning through play or game-based learning in addition to validating the suitability of the adapted frameworks or diagnostic criteria. Instead, due to the differing goals of each, a more appropriate use would be their original purposes: Yusoff's [21] framework to ensure newly developed serious games can meet their learning outcomes and Rooney's [22] framework to ensure a balance of design and pedagogical elements during game development.

Limitations

The diagnostic markers for serious games were designed with an aim of simplicity and this, in turn, also serves as one of its limitations. In this regard, the markers may only be used to discriminate between serious games and other digital solutions purporting to be serious games, but belonging to other fields like e-learning. These markers, and the associated diagnostic criteria, cannot determine the efficacy of a serious game, as demonstrated by the analysis of "Staying Alive," a serious game that did not perform as well as its creators had hoped as compared to a conventional e-learning alternative.

Another limitation is the lack of readily available, high-quality health care serious games to test the markers. Owing to the specific or controlled uses of health care serious games, many remain noncommercialized or unmaintained and thus difficult to acquire. The study was also unable to utilize serious games that required specialized or customized hardware due to logistical constraints.

Conclusions

The diagnostic markers of serious games presented in this study offer a simpler alternative that may be used by professionals or educators without extensive familiarity with serious games. They allow for the validation of ready-made games on the market and are used to confirm the presence of "seriousness" in any given health care serious game.

Although the markers may also be used during the game development phase to ensure the end product is both serious and a game, further evaluation is required to confirm its validity in this regard. The dynamic nature of diagnostic markers and the accompanying criteria, akin to the those employed by modern medicine, could potentially see new, possibly industry-specific markers being developed from evidence surfacing in future works, which might result in criteria that enable educators or professionals to determine the efficacy of premade serious games while maintaining their simplistic approach to game validation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Serious Games Markers Scoring Protocol (analog).

[DOCX File, 18KB - games_v7i3e14620_app1.docx]

Multimedia Appendix 2

Serious Games Markers Scoring Protocol (digital).

[XLSX File (Microsoft Excel File), 15KB - games_v7i3e14620_app2.xlsx]

References

1. Connolly T, Boyle E, MacArthur E, Hainey T, Boyle J. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 2012 Sep;59(2):661-686. [doi: 10.1016/j.compedu.2012.03.004]
2. Alvarez J, Djaouti D. Introduction Au Serious Games. Limoges, France: Questions Théoriques; 2012.
3. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious Gaming and Gamification Education in Health Professions: Systematic Review. J Med Internet Res 2019 Mar 28;21(3):e12994 [FREE Full text] [doi: 10.2196/12994] [Medline: 30920375]
4. Ravyse W, Seugnet Blignaut A, Leendertz V, Woolner A. Success factors for serious games to enhance learning: a systematic review. Virtual Reality 2016 Sep 20;21(1):31-58. [doi: 10.1007/s10055-016-0298-4]
5. Ke F, Abras T. Games for engaged learning of middle school children with special learning needs. Br J Educ Technol 2012 May 23;44(2):225-242. [doi: 10.1111/j.1467-8535.2012.01326.x]
6. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. Med Educ Online 2018 Dec;23(1):1438718 [FREE Full text] [doi: 10.1080/10872981.2018.1438718] [Medline: 29457760]
7. Chon S, Timmermann F, Dratsch T, Schuelper N, Plum P, Berlth F, et al. Serious Games in Surgical Medical Education: A Virtual Emergency Department as a Tool for Teaching Clinical Reasoning to Medical Students. JMIR Serious Games 2019 Mar 05;7(1):e13028 [FREE Full text] [doi: 10.2196/13028] [Medline: 30835239]
8. Dobrovsky A, Borghoff UM, Hofmann M. An approach to interactive deep reinforcement learning for serious games. 2016 Presented at: IEEE; Oct 16-18, 2016; Wroclaw, Poland p. 85.
9. Heimo OI, Harviainen JT, Kimppa KK, Mäkilä T. Virtual to Virtuous Money: A Virtue Ethics Perspective on Video Game Business Logic. J Bus Ethics 2016 Dec 10;153(1):95-103. [doi: 10.1007/s10551-016-3408-z]
10. Bigdeli S, Kaufman D. Digital games in medical education: Key terms, concepts, and definitions. Med J Islam Repub Iran 2017;31:52 [FREE Full text] [doi: 10.14196/mjiri.31.52] [Medline: 29445681]
11. Ogletree S, Drake R. College Students' Video Game Participation and Perceptions: Gender Differences and Implications. Sex Roles 2007 Mar 21;56(7-8):537-542. [doi: 10.1007/s11199-007-9193-5]
12. Breuer J, Bente G. Why So Serious? On the Relation of Serious Games and Learning. Journal for Computer Game Culture 2010;4(1) [FREE Full text]
13. Djaouti D, Alvarez J, Jessel JP. Classifying Serious Games: The G/P/S Model. In: Felicia P, editor. Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches. Pennsylvania, United States: IGI Global; 2011.
14. Becker K, Parker J. The Guide to Computer Simulations and Games. Indianapolis, IN: John Wiley and Sons; 2011.
15. Ives C. Book review: Instructional-Design Theories and Models, Volume III: Building a Common Knowledge Base (Eds: Charles M Reigeluth & Alison A Carr-Chellman). Educational Technology & Society 2010;13(2):219-221 [FREE Full text]
16. Kapralos B, Fisher S, Clarkson J, van Oostveen R. A course on serious game design and development using an online problem-based learning approach. Interactive Tech & Smart Ed 2015 Jun 15;12(2):116-136. [doi: 10.1108/ITSE-10-2014-0033]
17. Fisch S. Characteristics of effective materials for informal education: a crossmedia comparison of television, magazines,interactive media. In: Rabinowitz M, Blumberg FC, Everson HT, editors. The Design of Instruction and Evaluation: Affordances of Using Media and Technology. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:3-18.
18. Markopoulos P, Bekker M. Interaction design and children. Interacting with Computers 2003 Apr;15(2):141-149. [doi: 10.1016/S0953-5438(03)00004-3]
19. Hays RT. The effectiveness of instructional games: A literature review and discussion. Orlando, FL: Naval Air Warfare Center; 2005.
20. Haoran G, Bazakidi E, Zary N. Serious Games in Health Professions Education: Review of Trends and Learning Efficacy. Yearb Med Inform 2019 Apr 25 [FREE Full text] [doi: 10.1055/s-0039-1677904] [Medline: 31022747]
21. Yusoff A, Crowder R, Gilbert L, Wills G. A Conceptual Framework for Serious Games. : IEEE; 2009 Presented at: 2009 Ninth IEEE International Conference on Advanced Learning Technologies; July 15-17, 2009; Riga, Latvia.
22. Rooney P. A Theoretical Framework for Serious Game Designxploring Pedagogy, Play and Fidelity and Their Implications for the Design Process. International Journal of Game-Based Learning 2012;2(4):41-50. [doi: 10.4018/ijgbl.2012100103]
23. Jabareen Y. Building a Conceptual Framework: Philosophy, Definitions, and Procedure. International Journal of Qualitative Methods 2009 Dec;8(4):49-62. [doi: 10.1177/160940690900800406]

24. Graafland M, Bemelman W, Schijven M. Game-based training improves the surgeon's situational awareness in the operation room: a randomized controlled trial. Surg Endosc 2017 Dec;31(10):4093-4101 [FREE Full text] [doi: 10.1007/s00464-017-5456-6] [Medline: 28281117]
25. Haubruck P, Nickel F, Ober J, Walker T, Bergdolt C, Friedrich M, et al. Evaluation of App-Based Serious Gaming as a Training Method in Teaching Chest Tube Insertion to Medical Students: Randomized Controlled Trial. J Med Internet Res 2018 Dec 21;20(5):e195 [FREE Full text] [doi: 10.2196/jmir.9956] [Medline: 29784634]
26. Drummond D, Delval P, Abdenouri S, Truchot J, Ceccaldi P, Plaisance P, et al. Serious game versus online course for pretraining medical students before a simulation-based mastery learning course on cardiopulmonary resuscitation: A randomised controlled study. Eur J Anaesthesiol 2017 Dec;34(12):836-844. [doi: 10.1097/EJA.0000000000000675] [Medline: 28731928]
27. Stieglitz S. Gamification: Using Game Elements In Serious Contexts (Progress in IS). Basel, Switzerland: Springer; 2016.
28. Graafland M, Bemelman W, Schijven M. Prospective cohort study on surgeons' response to equipment failure in the laparoscopic environment. Surg Endosc 2014 Sep;28(9):2695-2701. [doi: 10.1007/s00464-014-3530-x] [Medline: 24771195]

Abbreviations

UX: User Experience

Edited by G Eysenbach; submitted 06.05.19; peer-reviewed by P Staccini, R Ospina, K Glover; comments to author 27.05.19; revised version received 30.05.19; accepted 09.06.19; published 16.07.19.

Please cite as:
Tan JW, Zary N
Diagnostic Markers of User Experience, Play, and Learning for Digital Serious Games: A Conceptual Framework Study
JMIR Serious Games 2019;7(3):e14620
URL: http://games.jmir.org/2019/3/e14620/
doi:10.2196/14620
PMID:31313660

Original Paper

A Brief Measure of Interpersonal Interaction for 2-Player Serious Games: Questionnaire Validation

Maja Gorsic[1], MSc; Joshua D Clapp[2], PhD; Ali Darzi[1], MSc; Domen Novak[1], PhD

[1]Department of Electrical and Computer Engineering, University of Wyoming, Laramie, WY, United States

[2]Department of Psychology, University of Wyoming, Laramie, WY, United States

Corresponding Author:
Domen Novak, PhD
Department of Electrical and Computer Engineering
University of Wyoming
1000 E University Ave
Laramie, WY, 82071
United States
Phone: 1 307 766 5599
Email: dnovak1@uwyo.edu

Abstract

Background: Competitive and cooperative serious games have become increasingly popular in areas such as rehabilitation and education and have several potential advantages over single-player games. However, they are not suitable for everyone, and the user experience in competitive and cooperative serious games depends on many factors. One important factor is the verbal interaction between players, but the effect of this factor has not been extensively studied because of the lack of a validated measurement tool.

Objective: This paper aimed to validate a brief questionnaire that measures the verbal interaction between 2 players of a serious game. The questionnaire consists of 8 questions pertaining to the amount of conversation, its valence (positive or negative emotion), and its game relatedness.

Methods: The questionnaire was validated with 30 pairs of participants who played a competitive serious game for 10 min while being recorded with cameras. The questionnaire was filled out by both participants, an in-person observer, and 2 members of our research group who watched the videos. Results from these raters were used to develop questionnaire instructions, and the finalized questionnaire was given to 2 additional raters who were trained on 5 videos and then rated the other 25 videos independently.

Results: The questionnaire's interrater reliability is excellent for the amount of conversation and its game relatedness (intraclass correlation coefficients [ICCs] above 0.9). Interrater reliability is fair to good for conversation valence (ICCs between 0.4 and 0.7). We believe that the lower interrater reliability for valence is primarily because of a limited spread of valence values in our sample. Furthermore, questionnaire ratings were significantly correlated with players' personality characteristics (eg, amount of conversation was correlated with extraversion) and pressure/tension experienced in the competitive game.

Conclusions: The validated questionnaire has the potential to be a useful tool for studying user experience in competitive and cooperative serious games. Furthermore, it could be adapted for other applications such as entertainment games. However, it has only been validated with unimpaired university students in a 2-player competitive serious game and should next be validated with different target populations (eg, stroke survivors) and different game designs (eg, cooperative games).

(JMIR Serious Games 2019;7(3):e12788) doi:10.2196/12788

KEYWORDS

attitude to computers; competitive behavior; exercise; motivation; questionnaire design; virtual reality

Introduction

2-Player Serious Games

Although serious games have traditionally involved only a single player, 2-player and multiplayer serious games have become increasingly popular in the last decade. For example, competitive and collaborative serious games can be used for motor rehabilitation [1-3]; for weight loss and general fitness [4-7]; for language therapy [8]; for military training [9]; to teach school subjects such as mathematics [10], language [11], and programming [12]; to teach more informal skills such as recycling [13], energy awareness [14], and sexual risk reduction [15]; and for many other applications. Such games have several potential advantages: compared with single-player games, competitive and cooperative games have been found to result in higher motivation and energy expenditure in rehabilitation and weight loss [3,4,16] as well as higher motivation and faster learning in educational applications [12].

However, not all users benefit equally from competitive and collaborative games. For example, studies in rehabilitation [2,17], weight loss [4], and education [13] have found that people who like competition often do not like cooperation (and vice versa). Furthermore, a person's experience with a competitive or cooperative game depends on factors such as age, gender, personality, and the person's relationship with the other player(s) [2,17-23]. The exact effects of these factors are often unclear and interaction effects are likely—for example, our previous studies on rehabilitation games have found that effects of personality are stronger in pairs of strangers than in pairs of friends [2,17]. To enable more effective deployment and personalization of serious games, these effects critically need to be studied in more detail.

Interpersonal Interaction in 2-Player Serious Games

One factor that strongly affects users' experience with competitive and cooperative serious games is the amount of interaction (both verbal and nonverbal) between players. For example, our previous study found a very strong correlation between self-reported enjoyment and the amount of conversation between 2 players [17], and other studies have found that self-reported enjoyment decreases as interaction elements between players are removed [20]. However, studying the effect of interpersonal interaction on user experience in serious games is difficult because a commonly accepted objective or subjective measure does not exist. In our previous study, we measured the amount of conversation between players using a simple 0 to 3 scale reported by an observer [16], but that measurement was not validated and had several methodological issues (eg, lack of consistent rating guidelines, no ability to rate positive vs negative conversation, and no ability to rate whether 1 player is talking more than the other). Other studies of 2-player serious games have acknowledged that verbal interaction between players is an important factor but did not analyze it because of a lack of a validated measure [1,3,24,25].

The goal of this study was to validate a brief questionnaire that measures the verbal interaction between 2 players of a serious game. The questionnaire consists of 8 questions pertaining to the amount of conversation, its valence (positive or negative emotion), and its game relatedness. The justification for these 3 quantities is as follows:

- Amount of conversation: previous research with unvalidated, ad-hoc measures has shown that the amount of conversation is correlated with subjective experience in competitive serious games [16] and that the amount of interaction differs between participant groups (eg, young and old [23]). Furthermore, although this has not been studied, we believe that the conversation balance (relative amount of talking done by each person in a pair) could provide insights into, for example, cooperation dynamics. Thus, measuring the amount of conversation is expected to provide insight into diverse aspects of competitive and cooperative serious games.
- Valence: providing a positive experience is a critical aspect of serious games for rehabilitation [1-3], physical fitness [7], and education [12], and we reasonably believe that a positive overall experience will result in positive conversational valence. Thus, a measure of conversational valence can provide insights into the players' experience with the serious game.
- Game relatedness: although less well justified than the other 2 quantities, we believe that measuring game relatedness would provide important insights into whether the participants are focused on the serious game. Studies have suggested that participants need to be actively engaged with a serious game (rather than just going through the motions) to maximally benefit from it [26], and a high amount of game-unrelated conversation may indicate that the players are simply chatting rather than actively participating in the game.

The questionnaire is primarily meant to be filled out by an observer who watches the game session either in person or on video, although it could also be filled out by the players themselves. To validate the questionnaire, 30 pairs of participants played a competitive exercise game for 10 min while recorded on video. Results of the questionnaire are compared between the participants, an in-person observer, and several raters who watched the videos after the data collection had been completed.

Methods

In this section, we first present the questionnaire and its rating instructions, then describe the methods used to validate the questionnaire.

Questionnaire Items

Our brief interpersonal interaction questionnaire was developed to evaluate the verbal interaction between 2 game players who are ideally colocated (in the same room). It consists of eight 5-point items that can be answered in less than 5 min by either a player (after gameplay has concluded) or an observer (during or after gameplay). For an observer, the items are as follows:

1. How much did player A talk to player B? (1: little to no talking; 5: nearly constant talking)
2. How much did player B talk to player A? (same scale as above)

3. How balanced was the conversation? In other words, did both players talk about the same amount or did 1 player talk more than the other? (1: both players talked about the same amount; 3: 1 player talked moderately more; and 5: 1 player dominated the conversation)
4. How positive or negative were the things that player A said during the game? (1: very negative; 5: very positive)
5. How positive or negative were the things that player B said during the game? (same scale as above)
6. Were the things that player A said about the game or other unrelated topics? (1: mostly unrelated to the game; 5: mostly related to the game)
7. Were the things that player B said about the game or other unrelated topics? (same scale as above)
8. How would you rate the overall conversation between the players? (1: very negative; 5: very positive)

For a player, these questions were modified to refer to “you” and “the other player” instead of players A and B. Full copies of both versions of the questionnaire are available in Multimedia Appendix 1. For observers, players A and B should be defined within the context of the experiment to avoid confusion; in our study, for example, we defined player A as the person sitting on the left and player B as the person sitting on the right.

Rating Instructions

Administering the questionnaire is relatively easy and should not take more than a few minutes. However, as some situations can be confusing for raters, the following instructions should be read carefully before using the questionnaire to ensure consistent answers.

General: Raters should answer all the questions even if the amount of conversation is limited; answers of “not applicable” should only be permitted if a player says nothing during the entire gameplay session. Raters are allowed to take notes and make preliminary scores during the gameplay interval and may rewind and replay the videos as desired. However, raters have to watch the entire interval before giving their final ratings.

Items 1 and 2 (amount of conversation): If both players talk almost constantly, both should receive a score of 5; conversely, if 1 player talks for approximately half the gameplay session and the other never talks, the silent player should receive a score of 1, whereas the talking player should receive a 3 or 4.

Item 3 (balance): Very high values (4 or 5) should only be used in cases where 1 player is talking frequently and the other is not. For example, if 1 player talks for 5 min and the other player never talks, that would be a 5. However, if 1 player never talks and the other player only says 1 sentence, that should be rated a 2.

Items 4 and 5 (valence): Raters should rate all things spoken, not only those directed at the other person. For example, if a player appears to be talking to themselves, those comments should be considered for these items. Furthermore, raters should rate not only the words but also the tone and facial expression. For example, neutral words (eg, “the game is getting harder”) accompanied by a smile should result in a 4. In cases of sarcasm (eg, “you did such a good job” said insincerely), raters should consider the other player’s reaction—if the other player appears amused by the sarcasm, it should be treated as positive; conversely, if the other player appears annoyed by the sarcasm, it should be treated as negative. Finally, raters should use the extreme values (1 and 5) sparingly—1 should be used when the players are actively antagonizing each other, whereas 5 should be used when the players are actively praising each other or the game.

Items 6 and 7 (game relatedness): The answers 1 and 5 indicate “mostly unrelated to the game” and “mostly related to the game”, respectively. The players do not need to talk 100% about the game to get a 5, and the answers 1 and 5 can be used relatively frequently here. For example, a 5 would correspond to about 90% of the conversation being game related, whereas a 1 would correspond to about 10% of the conversation being game related. Furthermore, metagame discussion (eg, criticizing the game's features or wondering how it is programmed) counts as game related. However, discussion about other games (other than the one being played) does not count as game related.

Item 8 (overall mood): This item is not meant to be an average of items 4 and 5 and does not only include conversation but also includes facial expressions and body language. The answers 1 and 5 can be used more frequently here than on questionnaire items 4 and 5; for example, if the players are smiling and appear to be having a good time, but their conversation is mostly about neutral topics (eg, “oh, the game is getting harder”), raters could answer 4 to items 4 and 5, but 5 to item 8.

Evaluation

Our Serious Game

The validation study was performed with a single 2-player competitive serious game: the game of Pong previously used in our arm rehabilitation studies [16,17]. Each player controls a paddle near the top or bottom of the screen and moves it left or right using their controller. A ball bounces around the game field, and each player’s goal is to intercept the ball so that it does not pass their paddle. If the ball passes a player’s paddle and reaches the top or bottom of the screen, the opponent scores a point. Once the point is scored, the ball moves to the middle of the screen and begins moving in a random direction again after a 1-second pause. Every 60 seconds, the difficulty of the game changes according to a simple adaptation algorithm that changes the ball speed and the size of the 2 paddles depending on the players’ relative score as described in our previous paper [16]. A screenshot of the game is shown in Figure 1.

Both players play the game on the same computer and are seated side by side in front of the same screen. To control the game, we reused the same hardware from our previous rehabilitation study [16]. One player (*participant A*) uses a joystick and tilts it left and right to move their paddle left and right. The other player (*participant B*) uses a Bimeo arm rehabilitation device (Kinestica d.o.o), which consists of 2 acceleration sensors attached to armbands and a spherical handheld module; this module must be tilted left and right to move the player’s paddle left and right. A photo of 2 participants playing the game is shown in Figure 2.

Figure 1. Screenshot of the competitive Pong game. Each player controls one of the 2 paddles. The current game duration, score, ball speed, and time until the next automated difficulty adaptation are shown on the right side of the playing field. Image reused from Gorsic et al with permission [16].

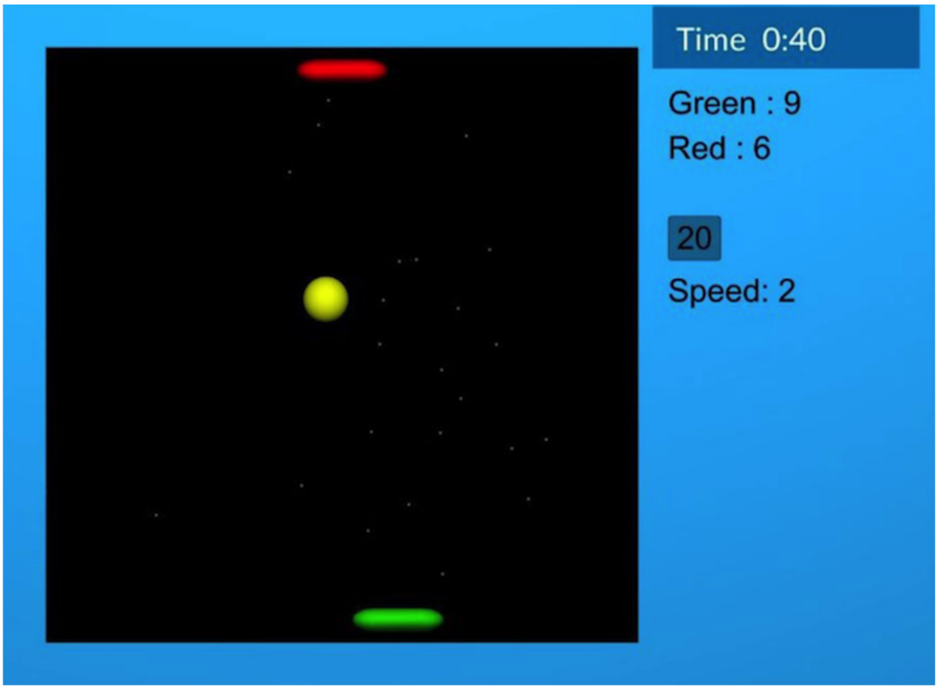
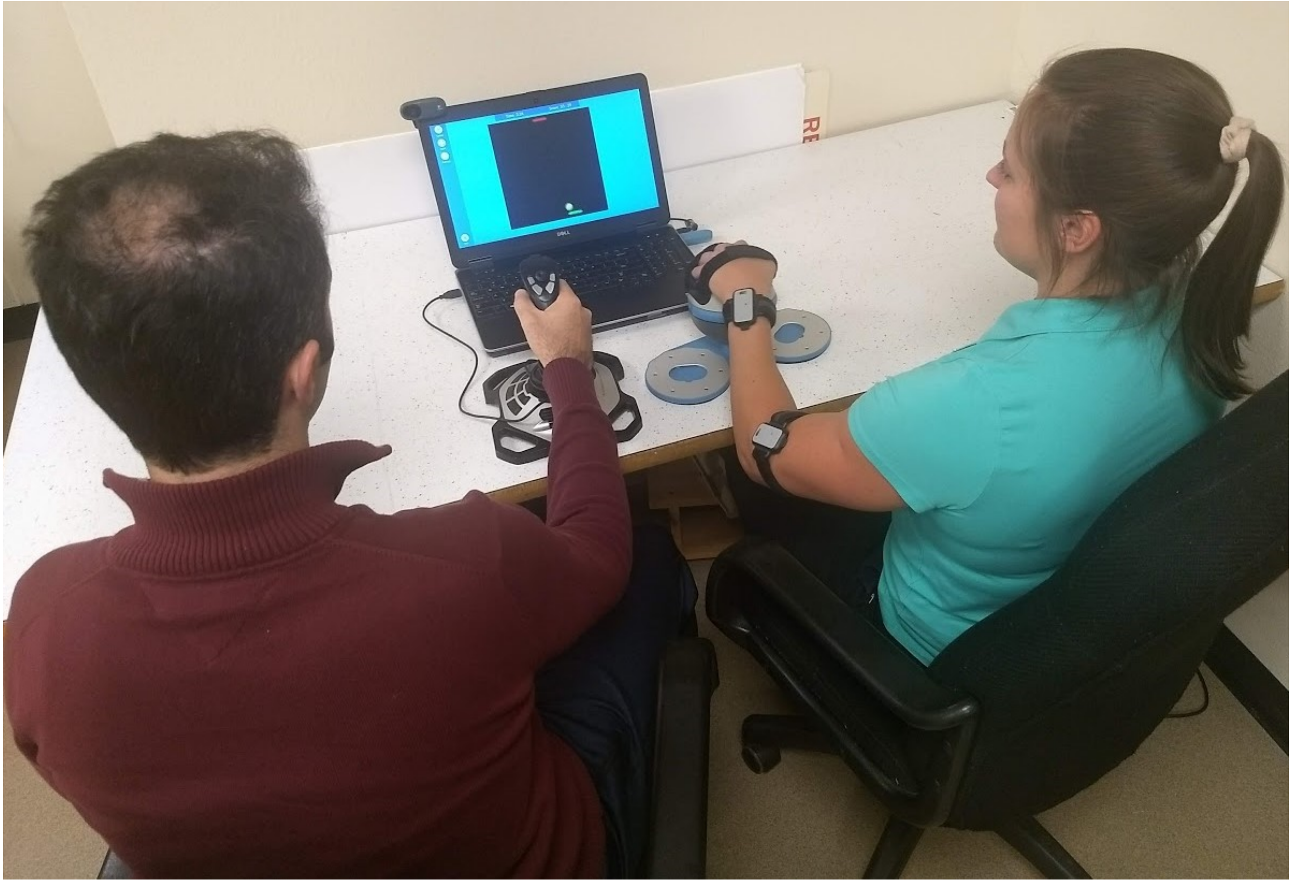


Figure 2. Study setup. Two participants play Pong using a joystick and Bimeo arm rehabilitation device while observed by an experimenter (not shown) and 2 webcams. Participants could use whichever arm they wished to play the game.



Participants

Data collection was carried out in the first half of 2018. Overall, 30 pairs (60 participants) were recruited among students and staff of the University of Wyoming. Participants were not allowed to take part in the study if they had played our specific version of Pong before, although they were not asked if they had played Pong in general. Participants could volunteer for the session alone or in self-selected pairs (eg, 2 friends); if a participant volunteered alone, they were paired with another random single participant. No additional restrictions were placed on allowed pairings. Of the 30 pairs, 22 were same-gender pairs, whereas 8 were mixed-gender pairs; 14 were self-selected pairs, whereas 16 were paired together randomly. There were 42 male and 18 female participants (none self-identified as nonbinary), aged mean 22.3 (SD 5.5) years.

Study Protocol

Each pair of participants took part in a single session. On arrival to the laboratory, participants were told that the purpose of the study was to examine player behavior in a competitive game and that they would be videotaped, although they were not specifically told that the goal was to validate the interpersonal interaction questionnaire. The experiment procedure was explained, the game and questionnaires were demonstrated, and informed consent forms were signed. Participants filled out a brief demographic questionnaire and were then seated in front of the computer and played the game for 10 min without breaks. A total of 2 webcams (1 atop the computer screen and 1 integrated into the screen) were used to take video and audio recordings of the participants. An example image from the 2 webcams is shown in Figure 3. After the 10-min period concluded, both participants filled out the interpersonal interaction questionnaire as well as an 8-item version of the Intrinsic Motivation Inventory (IMI), which measures 4 motivation scales during gameplay: interest/enjoyment, effort/importance, perceived competence, and pressure/tension, each using two 7-point Likert items. The specific version of the IMI was reused from our previous work [16]. Furthermore, they filled out the Ten-Item Personality Inventory, which measures the Big Five personality factors (extraversion, agreeableness, openness, conscientiousness, and neuroticism), each using two 7-point Likert items [27].

At the same time as the participants, the experimenter also filled out the interpersonal interaction questionnaire. The videos were then provided to 2 other raters (members of our research group), who filled out the interpersonal interaction questionnaire by watching the videos. Both raters scored each video, and the rating procedure was done in 2 stages to develop the rating instructions. The experimenter and the 2 video raters first collected and scored the first 15 videos. They then met to compare their answers, rewatched videos where the 3 ratings did not match (at least a 2-point discrepancy between any 2 raters on any item) and developed the rating instructions (*Rating Instructions* section) to address the discrepancies. With the rating instructions at hand, the experimenter and 2 raters adjusted their ratings for the first 15 videos as desired. After that, the remaining 15 pairs were rated independently by the experimenter and 2 video raters using the rating instructions.

Once all 30 pairs had been rated by the original 3 raters, the instructions were finalized, and the videos and the questionnaire were given to 2 external raters (paid US $10 per hour for this study, but otherwise unaffiliated with our research group). A total of 5 videos were selected as *training* videos: the external raters rated them one by one and received feedback about their score and clarifications about the rating procedure after each rated video. Of these 5 videos, the first 2 were considered *easy* by the original raters and had resulted in identical ratings; the other 3 videos were considered more difficult to rate and included either unbalanced conversation, an overall low amount of conversation (making it difficult to rate valence and game relatedness), or sarcasm (making it difficult to rate valence). For items 1-2 and 6-8, the external raters' first ratings (before receiving feedback) were already very similar to those of the original raters for all 5 training videos. For item 3 (balance), 1 external rater had to be reminded that an answer of 1 corresponds to perfectly balanced conversation, as they had assumed that an answer of 3 corresponds to balanced conversation. For items 4 and 5 (valence), the external raters' first ratings were very similar to the original raters' first ratings for the first 3 training videos, but feedback was needed for the *low amount of conversation* and *sarcasm* training videos. After completing the 5 training videos, the external raters rated the other 25 videos independently.

Figure 3. Screenshot of the video from the webcams. Videos from the 2 webcams are automatically synchronized and shown to raters side-by-side as one video. Faces are blurred for anonymity.

Data Analysis

As the primary analysis, intraclass correlation coefficients (ICCs) were calculated between pairs of raters for each interpersonal interaction questionnaire item separately. ICCs are a standard method to assess the consistency of measurements made by different observers who are measuring the same quantity; in our case, they were calculated to determine whether our questionnaire can produce consistent ratings of the same conversation when used by different people, a necessary prerequisite for the use of the questionnaire in research. ICCs provide a more stringent test of consistency than standard Pearson correlations because their estimates account for both the covariation and absolute agreement between raters. They were calculated using a 2-way mixed, single-measures model with absolute agreement [28]. Interpretive benchmarks for ICCs (excellent: 0.75-1.00; good: 0.60-0.74; fair: 0.40-0.59; and poor: below 0.40) are provided by Cicchetti [29].

As the secondary analysis, we also calculated Spearman correlation coefficients between the interpersonal interaction questionnaire, the 4 IMI scales, and the Big Five personality factors. This provides an estimate of how the results of our questionnaire correlate with personality and game experience; as an important future application of the questionnaire is to study correlations between conversation and game experience in 2-player serious games, the secondary analysis provides preliminary data in this regard. For the interpersonal interaction questionnaire, the correlations were calculated separately using the averaged answers of the 3 original raters, the averaged answers of the 2 external raters, and the participants' own answers.

For participants and the 3 original raters, ICCs and correlation coefficients were calculated over all 30 pairs; for external raters, only 25 pairs were included because the 5 training pairs were excluded. We chose to use all 30 pairs for the 3 original raters, as they updated their ratings after the rating instructions had been finalized, although we acknowledge that this may have introduced some bias.

Results

Participants' mean self-reported answers to the Ten-Item Personality Inventory (on scales of 2-14) were 10.6 (SD 1.5) for openness, 10.3 (SD 2.2) for conscientiousness, 8.2 (SD 2.7) for extraversion, 6.3 (SD 2.2) for agreeableness, and 5.8 (SD 2.2) for neuroticism. Their mean self-reported answers to the IMI (on scales of 2-14) were 10.3 (SD 2.5) for interest/enjoyment, 10.6 (SD 2.8) for effort/importance, 9.8 (SD 2.6) for perceived competence, and 7.3 (SD 3.1) for pressure/tension.

ICCs for different pairs of raters are shown in Table 1. Furthermore, Spearman correlation coefficients between the interpersonal interaction questionnaire, the IMI, and the Ten-Item Personality Inventory are listed in Tables 2 and 3.

Table 1. Intraclass correlation coefficients for the 8 questionnaire items and different pairs of raters.

Questionnaire item	Participants A-B	In-person[a]: video 1	In-person: video 2	Videos 1 and 2[b]	Externals 1 and 2[c]
1. How much A talks to B	0.82	0.92	0.91	0.92	0.94
2. How much B talks to A	0.75	0.93	0.93	0.91	0.94
3. Balance	0.27	0.63	0.47	0.45	0.63
4. Valence A to B	0.58	0.55	0.15	0.37	0.49
5. Valence B to A	0.65	0.59	0.42	0.48	0.65
6. Game relatedness A to B	0.51	0.93	0.93	0.97	0.90
7. Game relatedness B to A	0.75	0.98	0.93	0.94	0.97
8. Overall valence	0.52	0.60	0.69	0.65	0.68

[a]In-person: in-person observer (experimenter).

[b]Videos 1 and 2: video raters used to develop the questionnaire instructions.

[c]Externals 1 and 2: external video raters used for the validation of completed questionnaire.

Table 2. Spearman correlation coefficients for correlations between the interpersonal interaction questionnaire items and the Intrinsic Motivation Inventory scales. If a questionnaire item does not appear in the table, no correlations had a *P* value below .10 for that item.

Rater and Questionnaire Item	Interest/enjoyment		Effort/importance		Competence		Pressure/tension	
	CC[a]	*P* value	CC	*P* value	CC	*P* value	CC	*P* value
Self-rating								
How much participant talks to other one	0.11	.39	-0.16	.22	0.07	.58	*-0.30*[b]	*.02*
How much participant was talked to	0.11	.42	*-0.22*	*.097*	0.05	.70	*-0.38*	*.002*
Game-relatedness of participant's statements	*0.31*	*.02*	*0.26*	*.05*	-0.09	.51	0.10	.44
Overall valence	*0.25*	*.06*	-0.02	.90	0.09	.52	*-0.36*	*.004*
Average of internal raters								
How much participant talks to other one	0.08	.53	-0.08	.55	0.18	.16	*-0.41*	*.001*
How much participant was talked to	0.04	.79	-0.01	.91	0.13	.33	*-0.41*	*.001*
Game-relatedness of participant's statements	0.17	.23	0.01	.94	*-0.29*	*.03*	*0.34*	*.01*
Overall valence	*0.22*	*.097*	-0.04	.76	0.02	.90	*-0.36*	*.005*
Average of external raters								
How much participant talks to other one	0.11	.44	-0.07	.62	*0.25*	*.08*	*-0.32*	*.02*
How much participant was talked to	0.05	.71	-0.03	.81	0.05	.75	*-0.37*	*.008*
Game-relatedness of participant's statements	0.13	.41	0.00	>.99	*-0.38*	*.01*	*0.33*	*.03*
Overall valence	0.19	.19	0.08	.56	0.09	.52	-0.19	.19

[a]CC: correlation coefficient.

[b]Correlations with a *P* value below .10 are in italic.

Table 3. Spearman correlation coefficients for correlations between the interpersonal interaction questionnaire items and the Ten-Item Personality Inventory scales. If a questionnaire item does not appear in the table, no correlations had a *P* value below 0.1 for that item.

Rater and Questionnaire Item	Openness		Conscientiousness		Extraversion		Agreeableness		Neuroticism	
	CC[a]	*P* value	CC	*P* value	CC	*P* value	CC	*P* value	CC	*P* value
Self-rating										
How much participant talks to other one	0.15	.25	0.03	.83	*0.29* [b]	*.02*	-0.08	.52	*-0.32*	*.01*
How much participant was talked to	*0.21*	*.10*	0.02	.88	*0.21*	*.10*	-0.17	.19	*-0.31*	*.02*
Valence of participant's statements	0.11	.42	*0.22*	*.097*	0.03	.83	*-0.22*	*.10*	*-0.29*	*.03*
Game-relatedness of participant's statements	*-0.22*	*.10*	0.09	.50	*-0.24*	*.08*	-0.14	.29	0.10	.45
Overall valence	0.20	.13	0.05	.70	0.18	.18	-0.12	.38	*-0.32*	*.01*
Average of internal raters										
How much participant talks to other one	*0.29*	*.03*	0.05	.69	*0.30*	*.02*	-0.14	.27	*-0.34*	*.007*
How much participant was talked to	*0.21*	*.10*	0.08	.53	*0.27*	*.03*	-0.10	.47	*-0.26*	*.04*
Valence of participant's statements	0.19	.15	0.01	.97	-0.18	.18	-0.12	.38	0.05	.68
Game-relatedness of participant's statements	*-0.30*	*.03*	-0.03	.85	*-0.28*	*.04*	-0.07	.63	0.21	.13
Overall valence	0.14	.29	0.11	.42	0.17	.19	-0.19	.15	*-0.32*	*.01*
Average of external raters										
How much participant talks to other one	*0.38*	*.006*	0.05	.74	*0.37*	*.009*	-0.20	.16	*-0.47*	*.001*
How much participant was talked to	0.17	.25	0.05	.72	*0.31*	*.03*	-0.02	.92	-0.16	.27
Valence of participant's statements	0.12	.43	-0.12	.43	-0.11	.46	-0.13	.40	0.03	.84
Game-relatedness of participant's statements	*-0.30*	*.049*	0.06	.68	*-0.26*	*.09*	-0.07	.64	0.20	.19
Overall valence	0.20	.17	0.12	.40	*0.31*	*.03*	0.00	.98	*-0.24*	*.10*

[a]CC: correlation coefficient.

[b]Correlations with a *P* value below .10 are in italic.

Discussion

In this section, we first discuss the systematic differences between participants A and B observed on items 1-2 and 4-7, then discuss the results for different items of the questionnaire. We then discuss the correlations between the interpersonal interaction questionnaire, IMI, and personality scales. Finally, we discuss the need for further validation of our questionnaire with different populations and different game designs and briefly discuss possible alternative approaches to measuring conversation in serious games.

Systematic Differences Between Participants

From an observer's perspective, there should be little difference between the *A* participants and *B* participants, and items 1 and 2 should thus yield essentially the same ICC as should items 4 and 5 as well as 6 and 7. Although many ICCs were the same for these items, there were also notable differences—for example, ICCs for items 4 and 5 differ by as much as 0.27. We performed a follow-up analysis to determine whether the *A* participants were significantly different from the *B* participants but found no significant differences for age, gender, personality, or IMI. Thus, we believe that the difference in ICCs may be because of systematic differences in the study setup: different input devices (Bimeo for A and joystick for B) and different paddle positions on the screen (top for A and bottom for B). This possibility is supported by participants' conversation (several commented on the differences between the Bimeo and joystick) and suggests that the reliability of the questionnaire may depend on the hardware used; however, the differences in ICCs may also be simply because of statistical noise, and this should be explored further.

Amount of Conversation and Game Relatedness

On the basis of the ICCs for questions 1, 2, 6, and 7, we can conclude that (at least for this group of participants) observers can use our questionnaire to measure both the amount of conversation and its game relatedness very consistently, with both the internal and external raters exhibiting ICCs over .9. The most important clarifications for these items were what to do in cases of unbalanced conversation (1 person talks more than the other), what constitutes a 1 or 5, and whether metagame discussion should count as game related.

In addition, the raters sometimes asked how to compare coherent conversation (eg, focusing on a single topic for an extended amount of time) with incoherent conversation (eg, frequent grunts, "oops"es, "ohhh"s, and similar exclamations but few full sentences). In the end, we did not include explicit instructions regarding conversation coherence, although this could be addressed in a future update of the questionnaire—potentially even with additional items to measure this aspect. Another addition to the questionnaire could be a measurement of who takes the lead on the conversation: we observed several cases where both players talked about the same, but all periods of conversation were initiated by the same player even if the other player then contributed equally.

Finally, the participants were able to provide reasonably consistent ratings of their gameplay session, with ICCs between participants ranging from 0.51 to 0.82 for questions 1, 2, 6, and 7. The lower ICCs compared with observers can likely be attributed to the fact that participants are focused on playing the game throughout the session and thus do not keep track of their conversation to the degree that observers do.

Conversation Balance

The item about conversation balance (item 3) exhibited relatively low ICCs compared with items 1 and 2. We believe that one major reason for this was the lack of actual imbalance in the dataset: most of the answers to this item were 1 or 2, whereas 4 occurred in only 1 pair, and 5 never occurred. Thus, the low ICCs for this item are likely because of a limited range of values. However, it is worth noting that the ICC between the participants was especially poor (0.27). As with items 1 and 2, this worse ICC is likely because of the difficulty of keeping track of conversation while playing the game.

Our opinion on this item is mixed. On the one hand, it is similar to items 1 and 2, and the same information should ideally be obtainable as the difference between those 2 items. However, in our dataset, there were several cases where raters gave the same answers to items 1 and 2, then indicated some conversational imbalance in item 3. In most of these cases, the raters' reasoning was that there was not enough of an imbalance to warrant different answers for items 1 and 2 but enough of an imbalance to be noted in item 3. Thus, item 3 may be more sensitive to small imbalances than items 1 and 2. In the future, this item could be validated further by artificially introducing imbalanced conversations (thus determining if the lower ICCs in this study were because of a limited range of self-reported values). Alternatively, it may be possible to simply omit this item and modify items 1 and 2 so that they are more sensitive to small imbalances.

Valence

The 2 items related to individual participants' valence (4 and 5) exhibited relatively low ICCs: although most ICCs were between 0.4 and 0.6 (in the *fair* range), one was as low as 0.15. As with the balance item, we believe that these low ICCs were primarily because of a limited spread of values in our sample. Most pairs were rated as 3 or 4 on these 2 items, only 1 pair was rated as 2 on either item by any rater, and no pairs were rated 1. As a result of this limited spread, the ICCs are low despite good matches between raters: for example, the ICC between the 2 external raters is .49 for item 4, but those 2 raters gave the same answer to that item for 21 of the 25 independently rated pairs and never disagreed by more than 1 point on the 5-point scale.

The narrow spread of valence values is to be expected from a laboratory study, as participants do not wish to exhibit negative behavior when they know they are being observed and recorded on video. We believe that such negative behavior can be easily observed in real-world serious game environments, as several studies have documented very negative responses to competitive serious games [4,25]. Similarly, although few pairs were rated a *5* with regard to valence, we believe that this is also realistic for a laboratory study—a real-world answer of *5* would correspond to, for example, a therapist actively praising and verbally supporting a patient during exercise. The questionnaire could be validated for such extreme negative and positive valence ratings using, for example, actors, but we ultimately elected to simply acknowledge this limited evaluation, as we believe that the questionnaire is nonetheless valid and useful for cases of extremely positive or negative conversational valence.

On both items 4 and 5, the ICC between participants is higher than all other ICCs. We believe that this is because the participants have a better insight into their own valence than the raters do; although the raters have to determine valence based only on facial expressions and conversation, participants are largely aware of their own internal emotional processes. Furthermore, although raters sometimes had difficulty differentiating between honest statements, good-natured ribbing, and sarcasm or insults, the meaning of each sentence was likely clearer to the participants. In addition, ICCs involving video rater 2 were lower than the other ICCs. After a follow-up analysis, we believe that this is because video rater 2 was more likely than the other raters to rate pairs a *5* on items 4 and 5. We believe that this can be avoided in the future by more clearly emphasizing that extreme values should be used sparingly (as stated in the Rating instructions).

The item about overall valence (item 8) exhibited better ICCs than items 4 and 5—between 0.6 and 0.7. The reason for this difference is not entirely clear, although we believe that it is because this item had a somewhat greater spread compared with items 4 and 5. For example, there were several cases where raters gave a score of *4* on items 4 and 5 but a score of *5* on this item, with the justification "neither participant's behavior was very positive on their own, but the overall mood was very positive." We therefore believe that item 8 does provide useful data on its own and is not simply an average of items 4 and 5.

Finally, 1 way to potentially improve ICCs for valence items would be to provide more detailed instructions on how to analyze nonverbal behavior. Although our rating instructions included some examples on how to combine verbal and nonverbal behavior for purposes of rating valence, the raters commented that this was not always an easy task, and additional instructions may help produce more consistent ratings. In a longer version of the questionnaire, we could potentially even include separate items for verbal and nonverbal valence.

Correlations With Intrinsic Motivation Inventory and Ten-Item Personality Inventory

Multiple correlations were observed between our questionnaire and the IMI (Table 2), confirming that conversation can provide insight into participants' motivation in competitive serious games. Most notably, pressure/tension was negatively correlated with the amount of conversation as rated by both participants and internal and external raters, confirming preliminary findings obtained with unvalidated questionnaires [16]. Interestingly, however, correlations with the other 3 IMI scales differed among raters. First, a correlation between game relatedness and perceived competence was observed for internal and external raters but not the participants. Second, a correlation between overall valence and interest/enjoyment was observed for participants and internal raters but not external raters. Finally, correlations between game relatedness, interest/enjoyment, and effort/importance were observed only for participants.

For the Ten-Item Personality Inventory, all 3 groups' (participants, internal raters, and external raters) ratings about the amount of conversation were correlated with openness, extraversion, and (negatively) neuroticism. Furthermore, all 3 groups' ratings of overall valence were negatively correlated with neuroticism. This is not an unexpected result but does show that the amount and valence of conversation can differentiate between different groups of people, as preliminarily observed with unvalidated measures [23]. Interestingly, game relatedness ratings were negatively correlated with openness and extraversion, indicating that extraverted participants were more likely to chat with the other participants about other topics, whereas introverted participants were less likely to talk unless it is related to the game. However, similarly to the IMI, some correlations were only observed for some raters. For example, correlations between our questionnaire and agreeableness or conscientiousness were only observed for participants (but not observers); on the other hand, some correlations were only observed for observers but not the participants.

These differences in significant correlations between participants, internal raters and external raters are likely because of a mixture of statistical noise, perception biases, and additional insights. For example, as participants likely have a better insight into their own valence, their valence ratings are more likely to be correlated with personality and interest/enjoyment than the observers' ratings. On the contrary, the fact that some correlations were not found for participants' ratings (but were found for observers) may be because of differences in self-perception and perception of others. Although we cannot determine the reasons for these differences in detail, we believe that they should be considered when deciding whether to administer the questionnaire to participants. Furthermore, administering the questionnaire to participants and observers simultaneously may even allow explicit study of perception and bias in the context of competitive serious games.

Further Validation: Target Populations for Serious Games

As the immediate next step, the questionnaire should be used in applied studies with actual target populations for serious games. This is a critical step for 2 reasons. First, participants who play serious games with an actual goal (eg, learning new skills) will likely exhibit a wider range of conversational valence than participants in a laboratory experiment, allowing better validation of the valence items. Second, the positive results observed in our study were obtained with a population of young unimpaired university students and are not guaranteed to generalize with other populations.

On the basis of our previous experience with serious games, we believe that our questionnaire would be directly usable with obese adolescents and adults (a common target population for exercise games [4,5]) as well as with older adults who use serious games to socialize [20] or maintain their cognitive abilities but do not have major cognitive impairments. Although such populations may, for example, talk less than the students evaluated in our study, the conversation would likely still be accurately measured by the questionnaire. Similarly, the questionnaire could still be used to measure the amount of conversation and game relatedness in people with, for example, chronic depression or reduced emotional expressivity, although valence ratings may be less reliable in such populations. However, the questionnaire may be significantly less reliable in populations with communication disorders (seen in, eg, cooperative games for language therapy [8]) or other cognitive impairments (seen in, eg, motor rehabilitation after stroke or traumatic brain injury, depending on the injury location). In such populations, all items of the questionnaire may be unreliable, and this should be evaluated in follow-up studies.

Further Validation: Cooperative Games and Games With More Than 2 Players

Although results of our validation are promising, the questionnaire has only been tested with a 2-player competitive game. Cooperative serious games may involve different verbal interaction patterns that may reduce the reliability of the questionnaire. We believe that the questionnaire is general enough to apply to cooperative games, although mean values of different items may change—for example, because players may need to plan their actions for optimal cooperation, the game relatedness of the conversation may increase. However, we acknowledge that this needs to be verified with different cooperative game designs. A future version of the questionnaire could even include items that are specific to cooperative game designs, such as identifying leaders and followers based on the conversation.

The questionnaire also has not been tested with games for more than 2 players and includes items that refer to specific players. We believe that it could be easily expanded for 3- or 4-player serious games (suggested for, eg, language therapy [8]), but that significant modifications would need to be made for group games (seen in, eg, weight loss [4]). On the contrary, although the questionnaire was developed and validated for 2-player serious games played on a single computer, we believe that minimal modifications would be needed for Web-based gameplay (eg, telerehabilitation), entertainment games, or even for nongame tasks.

Alternative Conversation Measures

Finally, although our questionnaire is designed to be brief and usable by both players and observers, we acknowledge that alternative measures may be able to obtain a more detailed or objective picture of the interpersonal interaction. One possible alternative would be to have video raters count the number of conversation instances as well as estimate each instance's speaker, length, valence, and game relatedness. Although this would be time consuming and likely only feasible in offline analysis, it may provide additional details. Alternatively, automated audio analysis could be used to estimate the amount of conversation as, for example, the mean sound level recorded by each player's microphone or the percentage of time that each player's microphone sound level exceeds a certain threshold. Although this would likely not allow analysis of valence or game relatedness, it would provide a very objective measurement.

Conclusions

Our brief measure of interpersonal interaction allows players and observers of a 2-player competitive serious game to rate the players' amount of conversation as well as the conversation's valence (positive or negative emotional content) and game-relatedness using a total of eight 5-point items. The amount of conversation and its game relatedness can be rated reliably, with ICCs over .9 for pairs of trained raters. Valence is more difficult to rate reliably, with ICCs between .5 and .7, but we believe that this is because of the limited range of valence values in our data (neutral to moderately positive) and that the brief measure could nonetheless be used to rate very negative or very positive conversations.

The questionnaire can be used to study user experience in competitive and cooperative serious games, which are becoming increasingly popular in fields such as rehabilitation and education. User experience with such games is known to depend on factors such as the player's relationship with their coplayer, and a validated measure of interpersonal interaction will enable a better understanding of these factors, potentially leading to more efficient deployment of competitive and cooperative serious games. Furthermore, the questionnaire could be adapted for other applications such as entertainment games. However, we acknowledge that it has only been validated with healthy university students in a 2-player competitive serious game, and it should be further validated with different target populations for serious games (eg, stroke survivors) and with other game designs (eg, 2-player cooperation and group games).

Acknowledgments

This research is supported by the National Science Foundation under grant number 1717705 and the National Institute of General Medical Sciences of the National Institutes of Health under grant number P20GM103432.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full interpersonal interaction questionnaire used in our study. This includes separate versions for observers and participants.

[PDF File (Adobe PDF File), 46KB - games_v7i3e12788_app1.pdf]

References

1. Ballester BR, Bermúdez BS, Verschure PF. Including social interaction in stroke VR-based motor rehabilitation enhances performance: a pilot study. Presence: Teleoperators Virtual Environ 2012 Nov;21(4):490-501. [doi: 10.1162/PRES_a_00129]
2. Novak D, Nagle A, Keller U, Riener R. Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. J Neuroeng Rehabil 2014 Apr 16;11:64 [FREE Full text] [doi: 10.1186/1743-0003-11-64] [Medline: 24739255]
3. Mace M, Kinany N, Rinne P, Rayner A, Bentley P, Burdet E. Balancing the playing field: collaborative gaming for physical training. J Neuroeng Rehabil 2017 Nov 20;14(1):116 [FREE Full text] [doi: 10.1186/s12984-017-0319-x] [Medline: 29151360]
4. Marker AM, Staiano AE. Better together: outcomes of cooperation versus competition in social exergaming. Games Health J 2015 Feb;4(1):25-30 [FREE Full text] [doi: 10.1089/g4h.2014.0066] [Medline: 26181677]
5. Staiano AE, Abraham AA, Calvert SL. Motivating effects of cooperative exergame play for overweight and obese adolescents. J Diabetes Sci Technol 2012 Jul 1;6(4):812-819 [FREE Full text] [doi: 10.1177/193229681200600412] [Medline: 22920807]
6. 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 Jan 28;4(1):e1 [FREE Full text] [doi: 10.2196/games.5358] [Medline: 26821955]
7. Li J, Erdt M, Chen L, Cao Y, Lee SQ, Theng YL. The social effects of exergames on older adults: systematic review and metric analysis. J Med Internet Res 2018 Dec 28;20(6):e10486 [FREE Full text] [doi: 10.2196/10486] [Medline: 29954727]
8. Grechuta K, Rubio BR, Duff A, Oller ED, Pulvermüller F, Verschure PF. Intensive language-action therapy in virtual reality for a rehabilitation gaming system. J Pain Manag 2016;9(3):243-254 [FREE Full text]
9. Pasquier P, Mérat S, Malgras B, Petit L, Queran X, Bay C, et al. A serious game for massive training and assessment of french soldiers involved in forward combat casualty care (3S-SC1): development and deployment. JMIR Serious Games 2016 May 18;4(1):e5 [FREE Full text] [doi: 10.2196/games.5340] [Medline: 27194369]

10. Pareto L, Haake M, Lindström P, Sjödén B, Gulz A. A teachable-agent-based game affording collaboration and competition: evaluating math comprehension and motivation. Educ Technol Res Dev 2012 Apr 21;60(5):723-751. [doi: 10.1007/s11423-012-9246-5]
11. Wong LH, Hsu TC, Jizhen S, Iviza B. How flexible grouping affects the collaborative patterns in a mobile-assisted chinese character learning game. J Educ Techno Soc 2013;16(2):174-187 [FREE Full text]
12. Cagiltay NE, Ozcelik E, Ozcelik NS. The effect of competition on learning in games. Comput Educ 2015 Sep;87(3):35-41 [FREE Full text] [doi: 10.1016/j.compedu.2015.04.001]
13. Ibánez JJ, Wang AI. Learning recycling from playing a Kinect game. Int J Game-Based Learn 2015;5(3):25-44. [doi: 10.4018/IJGBL.2015070103]
14. Madeira RN, Silva A, Santos C, Teixeira B, Romão T, Dias AE, et al. LEY!: Persuasive Pervasive Gaming on Domestic Energy Consumption-Awareness. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology. 2011 Presented at: ACE'11; November 8-11, 2011; Lisbon, Portugal. [doi: 10.1145/2071423.2071512]
15. Aebersold M, Villarruel A, Tschannen D, Valladares A, Yaksich J, Yeagley E, et al. Using a virtual environment to deliver evidence-based interventions: the facilitator's experience. JMIR Serious Games 2015 Jul 21;3(2):e5 [FREE Full text] [doi: 10.2196/games.4293] [Medline: 26199045]
16. Gorsic M, Cikajlo I, Goljar N, Novak D. A multisession evaluation of an adaptive competitive arm rehabilitation game. J Neuroeng Rehabil 2017 Dec 6;14(1):128 [FREE Full text] [doi: 10.1186/s12984-017-0336-9] [Medline: 29208017]
17. Gorsic M, Cikajlo I, Novak D. Competitive and cooperative arm rehabilitation games played by a patient and unimpaired person: effects on motivation and exercise intensity. J Neuroeng Rehabil 2017 Mar 23;14(1):23 [FREE Full text] [doi: 10.1186/s12984-017-0231-4] [Medline: 28330504]
18. Schmierbach M, Xu Q, Oeldorf-Hirsch A, Dardis FE. Electronic friend or virtual foe: exploring the role of competitive and cooperative multiplayer video game modes in fostering enjoyment. Media Psychol 2012 Jul;15(3):356-371. [doi: 10.1080/15213269.2012.702603]
19. Wittchen M, Krimmel A, Kohler M, Hertel G. The two sides of competition: competition-induced effort and affect during intergroup versus interindividual competition. Br J Psychol 2013 Aug;104(3):320-338. [doi: 10.1111/j.2044-8295.2012.02123.x] [Medline: 23848384]
20. Gajadhar BJ, Nap HH, de Kort YA, Ijsselsteijn WA. Out of Sight, Out of Mind: Co-Player Effects on Seniors' Player Experience. In: Proceedings of the 3rd International Conference on Fun and Games. 2010 Presented at: Fun and Games'10; September 15-17, 2010; Leuven, Belgium p. 74-83. [doi: 10.1145/1823818.1823826]
21. Fang X, Zhu M. Extraversion and Computer Game Play: Who Plays What Games? In: Proceedings of the Conference on Human-Computer Interaction. Users and Applications. 2011 Presented at: HCI'11; July 9-14, 2011; Orlando, Florida, USA p. 659-667. [doi: 10.1007/978-3-642-21619-0_78]
22. Peng W, Hsieh G. The influence of competition, cooperation, and player relationship in a motor performance centered computer game. Comput Hum Behav 2012 Nov;28(6):2100-2106. [doi: 10.1016/j.chb.2012.06.014]
23. Rice M, Tan WP, Ong J, Yau LJ, Wan M, Ng J. The Dynamics of Younger and Older Adult's Paired Behavior When Playing an Interactive Silhouette Game. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 1081-1090. [doi: 10.1145/2470654.2466138]
24. Baur K, Wolf P, Riener R, Duarte JE. Making Neurorehabilitation Fun: Multiplayer Training via Damping Forces Balancing Differences in Skill Levels. In: Proceedings of the International Conference on Rehabilitation Robotics. 2017 Presented at: IEEE'17; July 17-20, 2017; London, UK p. 876-881. [doi: 10.1109/ICORR.2017.8009359]
25. Alankus G, Lazar A, May M, Kelleher C. Towards Customizable Games for Stroke Rehabilitation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI'10; April 10-15, 2010; Atlanta, Georgia, USA p. 2113-2122. [doi: 10.1145/1753326.1753649]
26. 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]
27. Gosling SD, Rentfrow PJ, Swann Jr WB. A very brief measure of the Big-Five personality domains. J Res Pers 2003 Dec;37(6):504-528. [doi: 10.1016/S0092-6566(03)00046-1]
28. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
29. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994 Dec;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]

Abbreviations

ICC: intraclass correlation coefficient
IMI: Intrinsic Motivation Inventory

Edited by G Eysenbach; submitted 10.11.18; peer-reviewed by B Ballester, J Duarte, S Brigitte; comments to author 31.03.19; revised version received 26.05.19; accepted 18.06.19; published 23.07.19.

Please cite as:
Gorsic M, Clapp JD, Darzi A, Novak D
A Brief Measure of Interpersonal Interaction for 2-Player Serious Games: Questionnaire Validation
JMIR Serious Games 2019;7(3):e12788
URL: https://games.jmir.org/2019/3/e12788/
doi:10.2196/12788
PMID:31339107

Corrigenda and Addenda

Title Correction: Younger Adolescents' Perceptions of Physical Activity, Exergaming, and Virtual Reality: Qualitative Intervention Development Study

Nuša Faric[1], BSc, MSc; Eleanor Yorke[1], BSc, MSc; Laura Varnes[1], BSc, MSc; Katie Newby[2], BSc, MSc, PhD; Henry WW Potts[3], BA, MSc, PhD; Lee Smith[4], BSc, MSc, PhD; Adrian Hon[5], BA, MA; Andrew Steptoe[1], BA, MA, DPhil, DSC; Abigail Fisher[1], BSc, PhD

[1]Department of Behavioural Science & Health, University College London, London, United Kingdom
[2]Faculty Research Centre for Advances in Behavioural Science, Coventry University, Coventry, United Kingdom
[3]Institute of Health Informatics, University College London, London, United Kingdom
[4]Cambridge Centre for Sport and Exercise Sciences, Anglia Ruskin University, Cambridge, United Kingdom
[5]Six to Start, London, United Kingdom

Corresponding Author:
Nuša Faric, BSc, MSc
Department of Behavioural Science & Health
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 02067694466 ext 41914
Email: nfaric@gmail.com

Related Article:

Correction of: http://games.jmir.org/2019/2/e11960/

(JMIR Serious Games 2019;7(3):e15657) doi:10.2196/15657

The title of this paper (JMIR Serious Games 2019;7(2):e11960) has been changed from "Younger Adolescents' Perceptions of Physical Activity, Exergaming, and Virtual Reality: Qualitative Intervention Study" to "Younger Adolescents' Perceptions of Physical Activity, Exergaming, and Virtual Reality: Qualitative Intervention Development Study".

The corrections will appear in the online version of the paper on the JMIR website on September 13, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Submitted 26.07.19; this is a non–peer-reviewed article;accepted 31.07.19; published 13.09.19.

Please cite as:
Faric N, Yorke E, Varnes L, Newby K, Potts HWW, Smith L, Hon A, Steptoe A, Fisher A
Title Correction: Younger Adolescents' Perceptions of Physical Activity, Exergaming, and Virtual Reality: Qualitative Intervention Development Study
JMIR Serious Games 2019;7(3):e15657
URL: http://games.jmir.org/2019/3/e15657/
doi:10.2196/15657
PMID:31538952

Corrigenda and Addenda

Figure Correction: Exploring Efficacy of a Serious Game (Tobbstop) for Smoking Cessation During Pregnancy: Randomized Controlled Trial

Francesc X Marin-Gomez[1,2,3,4], MD; Rocio Garcia-Moreno Marchán[5], MD; Anabel Mayos-Fernandez[5], MSN; Gemma Flores-Mateo[6,7], MD; Esther Granado-Font[6,8,9], SRA; Maria Luisa Barrera Uriarte[6,10], SRA; Jordi Duch[6,11], PhD; Cristina Rey-Reñones[6,8], PhD

[1]Servei d'Atenció Primària d'Osona, Gerència Territorial de la Catalunya Central, Institut Català de la Salut, Vic, Spain

[2]Unitat de Suport a la Recerca Catalunya Central, Institut Universitari d'Investigació en Atenció Primària Jordi Gol, Sant Fruitós de Bages, Spain

[3]Health Promotion in Rural Areas Research Group, Institut Català de la Salut, Sant Fruitós de Bages, Spain

[4]Digital Care Research Group, Universitat de Vic–Universitat Central de Catalunya, Centre for Health and Social Care Research, Vic, Spain

[5]Sexual and Reproductive Health Unit, Servei d'Atenció Primària d'Osona, Institut Català de la Salut, Vic, Spain

[6]Grup de Recerca en Tecnologies de la Informació en Atenció Primaria, Unitat de Suport a la Recerca Tarragona-Reus, Institut Universitari d'Investigació en Atenció Primària Jordi Gol, Reus, Spain

[7]Unitat d'Anàlisi i Qualitat, Xarxa Sanitària i Social Santa Tecla, Tarragona, Spain

[8]Departament d'Infermeria, Facultat d'Infermeria, Universitat Rovira i Virgili, Tarragona, Spain

[9]Centre d'Atenció Primària Horts de Miró (Reus-4), Gerència d'Àmbit d'Atenció Primària Camp de Tarragona, Institut Català de la Salut, Tarragona, Spain

[10]Centre d'Atenció Primària La Granja (Tarragona-2), Gerència d'Àmbit d'Atenció Primària Camp de Tarragona, Institut Català de la Salut, Torreforta,Tarragona, Spain

[11]Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Tarragona, Spain

Corresponding Author:
Francesc X Marin-Gomez, MD
Servei d'Atenció Primària d'Osona
Gerència Territorial de la Catalunya Central
Institut Català de la Salut
Pl Divina Pastora N7
Vic, 08500
Spain
Phone: 34 938890222
Email: xavy.marin@gmail.com

Related Article:

Correction of: http://games.jmir.org/2019/1/e12835/

(JMIR Serious Games 2019;7(3):e14381) doi:10.2196/14381

The manuscript "Exploring Efficacy of a Serious Game (Tobbstop) for Smoking Cessation During Pregnancy: Randomized Controlled Trial" (JMIR Serious Games 2019;7(1):e12835) was erroneously published with a duplicate of Figure 2 in place of Figure 3 due to a display error on the publisher's website. The correct version of Figure 3 can be seen below, and now also appears in the appropriate place in the corrected manuscript.

The correction will appear in the online version of the paper on the JMIR website on July 11, 2019, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article also has been resubmitted to those repositories.

Figure 3. Screenshots of Tobbstop.

Submitted 13.04.19; this is a non–peer-reviewed article;accepted 13.06.19; published 11.07.19.

Please cite as:
Marin-Gomez FX, Garcia-Moreno Marchán R, Mayos-Fernandez A, Flores-Mateo G, Granado-Font E, Barrera Uriarte ML, Duch J, Rey-Reñones C
Figure Correction: Exploring Efficacy of a Serious Game (Tobbstop) for Smoking Cessation During Pregnancy: Randomized Controlled Trial
JMIR Serious Games 2019;7(3):e14381
URL: http://games.jmir.org/2019/3/e14381/
doi:10.2196/14381
PMID:31298219

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