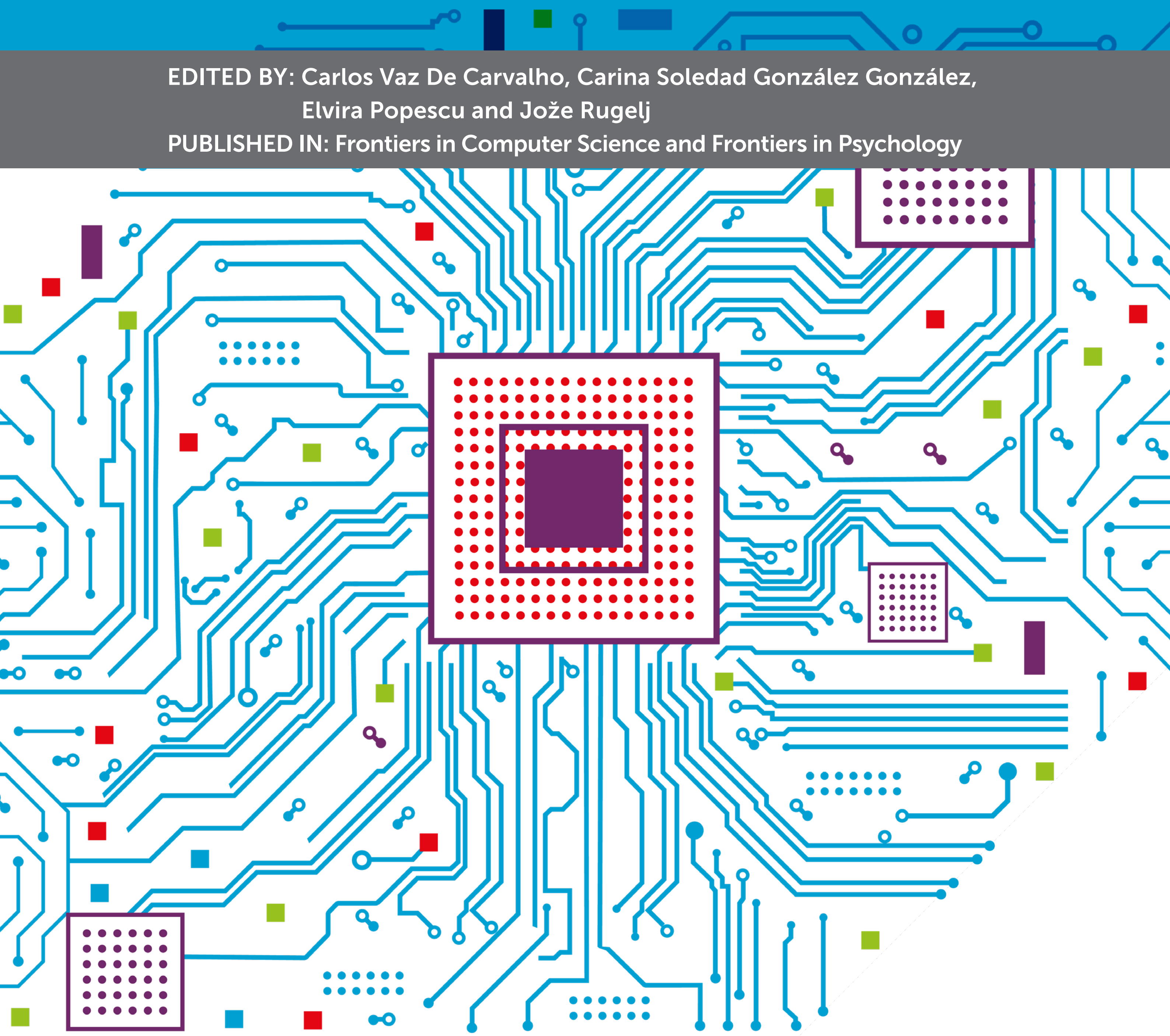
SERIOUS GAMES

EDITED BY: Carlos Vaz De Carvalho, Carina Soledad González González, Elvira Popescu and Jože Rugelj

PUBLISHED IN: Frontiers in Computer Science and Frontiers in Psychology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-944-8
DOI 10.3389/978-2-88966-944-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

SERIOUS GAMES

Topic Editors:
Carlos Vaz De Carvalho, Polytechnic Institute of porto, Portugal
Carina Soledad González González, University of La Laguna, Spain
Elvira Popescu, University of Craiova, Romania
Jože Rugelj, University of Ljubljana, Slovenia

Citation: Vaz De Carvalho, C., González González, C. S., Popescu, E., Rugelj, J., eds. (2021). Serious Games. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-944-8

Table of Contents

frontiers in Computer Science

EDITORIAL
published: 03 May 2021
doi: 10.3389/fcomp.2021.686348

Editorial: Serious Games

Carlos Vaz de Carvalho[1]*, *Carina Soledad González González*[2], *Elvira Popescu*[3] *and Jože Rugelj*[4]

[1] GILT, Instituto Superior de Engenharia Do Porto, Politécnico Do Porto, Porto, Portugal, [2] Department of Computer Engineering and Systems, Universidad de La Laguna, San Cristóbal de La Laguna, Spain, [3] Computers and Information Technology Department, University of Craiova, Craiova, Romania, [4] Faculty of Education, University of Ljubljana, Ljubljana, Slovenia

Keywords: educational games, advergames, awareness raising games, neuroscience of games, games and cognition

OPEN ACCESS

Edited and reviewed by:
Anton Nijholt,
University of Twente, Netherlands

****Correspondence:***
Carlos Vaz de Carvalho
cmc@isep.ipp.pt

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *26 March 2021*
Accepted: *06 April 2021*
Published: *03 May 2021*

Citation:
Vaz de Carvalho C, González González CS, Popescu E and Rugelj J (2021) Editorial: Serious Games. Front. Comput. Sci. 3:686348. doi: 10.3389/fcomp.2021.686348

Editorial on the Research Topic

Serious Games

Games are structured contexts, with clearly defined rules in which players must overcome challenges and face opponents (real or non-player characters) to achieve victory. Games offer incredibly immersive environments where users engage in "learning by doing," receiving immediate feedback on their unsuccessful actions (therefore also learning by error) and getting instant gratification after successful attempts which contributes to their sense of accomplishment and increased self-esteem. Games (and simulations) also offer safe environments to try and learn procedures and actions that, in real life, could be dangerous or potentially damaging if not done correctly. By playing, users can improve their skills and competencies in decision making, strategy, teamwork, social skills, leadership, collaboration, etc.

Serious Games focus precisely on the design, development, use, and application of games for purposes other than entertainment by exploiting the increased user motivation and engagement to promote learning, construct knowledge, sraise awareness or inspire behavior change. As such, the Serious Games domain brought together expertise from very different areas such as cognition, psychology, neurosciences, sociology, technology-enhanced education, evaluation and assessment, multimedia and information technology, interaction and simulation, etc. Serious Game research has been growing in the past few years namely trying to construct supporting theoretical framework and identifying the most effective design, implementation, and validation methodologies and models for each particular area of application and for each particular target group.

This Research Topic contributes to this research effort, while truly showing the multidisciplinary nature and application of Serious Games, by presenting theory, research, practice, and validation in multiple areas. While education and training still represent the main areas of application of Serious Games (and in this Research Topic) we can also see here scientific approaches, experiments and real-life applications in psychology, social sciences, health care and other domains. In sum, we have gathered nine different research works about Serious Games showing their importance and exemplifying potential research paths.

Coelho et al. present an analysis of the pervasive nature of Serious Games, their inherent meaningful gameplay and the usefulness for learning and communication. Several case studies are presented with a special focus on the BEACONING project that aimed to contextualize the teaching and learning process in STEM connecting it with problem-based and location-based game mechanics using real-world interactions and applications.

Smy et al. present a study on the usability of mobile training applications in the Aeronautical industry, in particular for air traffic control. These applications represent alternatives to the equivalent current training practices (a lengthy and expensive process) and the authors conclude that digital applications have the potential to engage future trainees in the air traffic services industry.

Kyrlitsias et al. created an interactive digital reconstruction of an archaeological site and conducted a feasibility study. The results showed high levels of presence and more positive experience by the participants who used the Virtual Reality system although this did not directly translate into greater learning gains.

Othlinghaus-Wulhorst and Hoppe addressed the use of virtual role-playing games to provide an authentic experience of situated learning through different problem-solving and communication strategies. They present a conceptual and technical framework for serious role-playing games for the training of specific social skills in virtual 2D learning environments involving chatbots in dialog-centric settings. They used this framework to create a set of training scenarios that include workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management. The results indicate that the scenarios could be well-suited for real training situations.

Kyrlitsias et al. presented an experimental study that elaborates on the social influence through conformity with a group of virtual agents within an immersive virtual environment. The results showed that participants did, in fact, conform to the agents as they exhibited sympathetic behaviors in the trials.

Christofi et al. compared a perspective-taking immersive Virtual Reality system with and without a number of sensorimotor contingencies (SC) to investigate the effect of the supported SC in promoting empathy and positive attitudes toward drug users. Results demonstrate a strong correlation between closeness to the drug user and empathy in the SC group.

Crepaldi et al. presented the design, implementation, and evaluation of a serious game on inhibition skills in children. Positive correlations between impulsiveness as measured by standard tests and impulsiveness scores in the Serious Game emerged. Furthermore, self-report ratings in the questionnaire showed that the Serious Game was engaging and elicited positive responses from children.

Berenbaum et al. focused on the use of serious games for persons with dementia. Their study showed that, despite declines in cognitive abilities, a tablet based Serious Game could stimulate their emotional and social intelligence and evoke responses in self-awareness, empathy, social, and communication capacities.

Martin-Niedecken et al. present a study on how to boost the attractiveness and effectiveness of an exergame by individualizing it with game adaptations based on physiological parameters like the player's heart rate. The results showed that the formula-based in-exergame adaptation approach was suitable in the presented study population, and that the developed exergame provided an equally reliable in-exergame adaptation and comparable exergame play experiences.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication..

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 05 June 2020
doi: 10.3389/fpsyg.2020.01242

A Virtual Reality Simulation of Drug Users' Everyday Life: The Effect of Supported Sensorimotor Contingencies on Empathy

Maria Christofi[1,2], Despina Michael-Grigoriou[1,2] and Christos Kyrlitsias[1,2]*

[1] GET Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus, [2] Research Centre on Interactive Media Smart Systems and Emerging Technologies – RISE, Nicosia, Cyprus

Perspective taking techniques have been used to transport people into imaginary situations and the lives of others. Virtual Reality provides an immersive way to virtually experience the lives of stigmatized by society members. Through the support of sensorimotor contingencies, people can use natural movements to view and interact with the virtual world around them. In this study, we compared a perspective-taking immersive Virtual Reality system which supports a number of sensorimotor contingencies (SC group) with a perspective-taking desktop system of the same application but without support of any sensorimotor contingencies (NSC group), to investigate the effect of the supported sensorimotor contingencies in promoting empathy and positive attitudes toward drug users. Results demonstrate a strong correlation between closeness to the drug user and empathy in the SC group. In both groups there were a within group significant change in their reported attitudes before and after their exposure. Finally, participants in the SC condition reported significantly higher levels of Place Illusion (PI), body ownership, agency and plausibility of people. Further research is needed to investigate how sensorimotor contingencies can be exploited to the fullest to be used as an effective method to induce empathy and change attitudes toward stigmatized by society people.

Keywords: virtual reality, sensorimotor contingencies, perspective-taking, empathy, attitudes

OPEN ACCESS

Edited by:
Carina Soledad González González,
University of La Laguna, Spain

Reviewed by:
Debaleena Chattopadhyay,
University of Illinois at Chicago, United States
Ke Ma,
Southwest University, China

****Correspondence:***
Despina Michael-Grigoriou
despina.grigoriou@cut.ac.cy

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: *24 December 2019*
Accepted: *12 May 2020*
Published: *05 June 2020*

Citation:
Christofi M, Michael-Grigoriou D and Kyrlitsias C (2020) A Virtual Reality Simulation of Drug Users' Everyday Life: The Effect of Supported Sensorimotor Contingencies on Empathy. Front. Psychol. 11:1242. doi: 10.3389/fpsyg.2020.01242

INTRODUCTION

Members of stigmatized groups are often discriminated against in their workplace, educational settings, health care, and the criminal justice system (Sidanius and Pratto, 2001). A reliable method that has been shown in reducing negative social stereotyping is perspective taking, which can be defined as the cognitive capacity to perceive the world from another person's viewpoint (Davis, 1983; Perdue and Gurtman, 1990; Shechtman and Tanus, 2006; Galinsky et al., 2008; Tomasello, 2009; Zaki and Ochsner, 2012). Furthermore, extensive research by Dan Batson has shown that perspective taking can lead to an increase of empathy, and this can lead to prosocial behaviors toward not only members of stigmatized groups, but toward stigmatized groups as a whole (Batson et al., 1997, 2003; Lamm et al., 2007). Empathy can be defined as the ability to connect emotionally with another individual and understand his point of view (Davis, 1983; Galinsky et al., 2008).

The field of Virtual Reality (VR) offers new ways to induce empathy into people. As Slater and Sanchez-Vives (2016) state, "the primary technological goal of VR is to realize perception through natural sensorimotor contingencies to the best extent possible." People can turn their heads, like in real life, to look around while wearing a head-mounted display (HMD), they can bend over to look under a table, they can reach out their hands to grab objects. With VR, we can show experiences from any point of view, therefore we no longer must rely on people's imagination. We can create experiences that genuinely show people how it is to take the place of somebody else and create a narrative that unfolds around them and therefore people can focus on the events that are happening around them and how they make them feel and go through a more genuine experience of what is like to be that person. In VR, we can also elicit presence, the illusion of being in the virtual world (Place Illusion, PI), and the extent to which the situation and events seemed to be really happening (Plausibility Illusion, Psi) (Slater, 2009).

Bertrand et al.'s (2018) study, highlighted immersive embodied virtual reality (EVR) strategies for empathy. Their strategies included the Body ownership illusion through multisensory and motor perspective taking, which if induced, could modulate bias, mimicry, similarity and emotion after the experience. Also, the agency illusion through embodiment combining voluntary and involuntary actions, which could result in self-attribution of avatar's actions. Another strategy was place and plausibility illusion through sensorimotor stimuli and a highly credible environment, which could result for the users to behave and feel as if they are in the VR environment. Lastly, they mentioned the Proteus Effect (Yee and Bailenson, 2007; Yee et al., 2009) through avatars presenting empathy-related traits and appearances for the reinforcement of positive or negative stereotypes and the modulation of behavior after the experience. It has been shown that the Proteus effect is mediated by the level of embodiment felt by users in relation to their avatar (Ash, 2016) suggesting that EVR can enhance this effect. Bertrand et al.'s (2018) argue that empathy training methods using avatars designed to improve empathy would induce beneficial behavioral changes and improve positive perceptions. They suggested the embodiment of a digital avatar of an outgroup member that presents traits that contradict stereotypes, which we have done in our study.

The use of immersive virtual environments to have an individual take on the perspective of another individual is known as Virtual Reality Perspective-Taking (VRPT) (van Loon et al., 2018). Ahn et al. (2013) conducted three experiments that explored whether embodied experiences via Immersive Virtual Environment Technology (IVET) would elicit greater self-other merging, favorable attitudes, and helping toward persons with disabilities compared to traditional perspective taking. Penn et al. (2010) conducted a 4-condition, between subjects' experiment ($N = 112$), wherein participants were exposed to either a virtual simulation of schizophrenia, a written empathy-set induction of schizophrenia, a combination of both the simulation and written empathy conditions, or a control condition. In Schutte and Stilinović's (2017), randomly assigned participants viewed a documentary featuring a young girl living in a refugee camp either in a virtual reality format or in a control two-dimensional format. Results indicated that the virtual reality experience resulted in greater engagement and a higher level of empathy for the refugee girl compared to the control condition. Chowdhury et al. (2019) study showed that a tracked HMD and a wheelchair interface had significantly larger effects on participants' implicit association toward people with disabilities than a desktop monitor and a gamepad. Hamilton-Giachritsis et al. (2018) study investigated whether embodiment of mothers into the body of a young child and engaging in an interaction with a virtual mother facilitates perspective-taking and empathy. Results showed that experiencing negative maternal behavior increased levels of empathy and participants reported a strong body ownership illusion for the child body that led to cognitive, emotional and physical reactions.

Herrera et al. (2018) conducted two experiments to compare the short and long-term effects of a traditional perspective-taking task and a VR perspective-taking task and to explore the role of technological immersion when it comes to different types of mediated perspective-taking tasks, relating to the homeless. Seinfeld et al. (2018) used virtual reality to allow offenders to virtually embody a victim of domestic abuse. It has been showed by several studies (Groom et al., 2009; Farmer et al., 2012; Maister et al., 2013; Peck et al., 2013; Banakou et al., 2016; Behm-Morawitz et al., 2016; Hasler et al., 2017) that embodiment of light-skinned participants in a dark-skinned virtual body significantly reduced implicit racial bias against dark-skinned people. Finally, a study (Christofi et al., 2018; Stavroulia et al., 2018, 2019) examined the effect of viewing in VR, a drug incident at a school environment from different perspectives and viewpoints in relation to the level of presence achieved and the effects on mood states and empathy of teachers who were immersed in a virtual school where a drug use incident took place.

A survey regarding VR interventions for stigmatized groups has been made by Christofi and Michael-Grigoriou (2017), with additional emphasis on the measurements used for prejudice and empathy. The studies reviewed were grouped according to the social stigma form that the group investigated belonged to, as defined by Goffman (2009), which are; (i) overt or external deformations, (ii) deviation in personal traits, and (iii) tribal stigmas. The survey reviewed the measurements used in these studies for empathy and prejudice. It concluded that the majority of the studies focused on the third form of social stigma, and most specifically on reducing implicit racial bias against dark-skinned people. As far as the measurements used for empathy and prejudice, it was observed that in the reviewed studies they tended to use self-report instruments more and rarely used behavioral observational and neuroscientific methods which could be more accurate than self-reports from the participants.

Drug users were chosen as the social target in this study because they are considered an extreme outgroup that people often struggle to empathize with (Harris and Fiske, 2006), and it is a stigmatized group that hasn't been studied much in this context before. In this study, we compared a VRPT system that supports a number of sensorimotor contingencies (SC condition) and a less immersive perspective-taking system using a desktop computer with a minimum support of sensorimotor contingencies (NSC condition), to investigate their effect in promoting empathy and creating positive attitudes toward drug users.

Following the strategies and suggestions of Bertrand et al.'s (2018), our hypothesis was that participants in the SC condition would be more empathetic after their experience, and have more positive attitudes toward drug users, and feel more emotionally close to them, than participants in the NSC condition.

MATERIALS AND METHODS

Application

An application was designed and developed in which participants were able to virtually experience some scenes of the life of a person who became a drug user, through his point of view, in order to debunk stereotypes surrounding drug users. In total, the participants viewed nine scenes from the life of the drug user. The scenario and scenes they virtually experienced are described in more detail in the following sections.

A number of the scenes, have a mirror on one of the walls (including the first one for familiarization with the VR equipment and the suit), so that the participants could see the avatar they were embodying and progressively see some of the physical changes, the drug consumed by their avatar which in this scenario is cocaine, does to people, which include dilated pupils, increased heart rate, extreme weight loss, bloody nose and cocaine powder running through the nose. These changes were visible to their avatar from the scene after he consumed cocaine in one of the starting scenes (Scene 5 as described below).

The scenario put the participants in the shoes of a man named Mark, a 28-year old man, who recently lost his mother due to cancer, and is living with his wife Amanda. It is said in the narrative, that he is stressed due to his workplace, where his boss assigns him many duties. One day at work, Mark's colleague sitting at the desk next to him, invited him to a party, to which he attends. A man there, offers Mark to try out a drug, in this instance; cocaine, for the first time, to "loosen up and have fun," as the drug dealer tells him. Mark, due to dealing with his mother's death and extensive work stress, obeys, follows that man to the bathroom, does drug use for the first time and becomes addicted to cocaine. Then, participants experience the consequences of this action which include his inability to work, which is shown as a scene at work where his boss yells at him for not doing his job right, a visit to his drug dealer and his changing relationship with his wife, who threatens him in the end of the narrative to leave him if he doesn't seek help for his drug problem. A video illustrating the application can be found in the **Supplementary Video S1**.

Scenes

Scene 1 – Bathroom: You are in the bathroom, standing in front of the sink. There is a mirror in front of you, where you can look at your avatar. Amanda is standing close to you and informs him that she invited your friends for dinner, so that it takes your mind away from work and your mother's loss.

Scene 2 – Dinner: You are in the dinner, and your friends are discussing sports on the table. Then one of your friends, Lisa, mentions that you are quiet tonight, followed by Amanda explaining that it is because of the work stress, among other stuff. In this scene, there is a mirror on the left side of the room, where participants can view their avatar.

Scene 3 – Work #1: You are in your office, and your boss is telling you to prepare some files for him before lunch time because important clients are waiting for them. Your boss leaves and your coworker turns to you and invites you to a party in a few days to release all your stress.

Scene 4 – Party #1: You are in that party, loud music is playing, people dancing, drinking around you. Your coworker is telling you to enjoy yourself, and then a man notices how stressed and serious your character is and tells you to follow him.

Scene 5 – Party #2: You are in the toilets of the house that the party takes place in. In front of you there are two sinks with needles, cocaine powder divided in straight lines and razors. The man stands in front of the second sink, and sniffs cocaine and urges you to do the same (**Figure 1**, top right).

Scene 6 – Drug dealer: You stand in front of your drug dealer who is sitting relaxed at a table where you can see cocaine in lines, razors and a bag and money laying around in packs. The dealer wonders how you can afford all this cocaine, implying that you are a regular and warns you that demand is getting bigger and the prices are getting higher and next time you should bring more money.

Starting from this scene, participants can view the physical changes cocaine had done to their character, which include weight loss (in the starting scenes Mark appears to have a belly so that the change is more noticeable), so in this scene he looks really thin, his eyes are red, he has blood running from his nose, and cocaine power visible on the nose as well.

Scene 7 – Work #2: Your boss yells at you for not showing up to an important meeting the day before (**Figure 1**, top left) and then your coworker tells you to look at yourself in a mirror and that he worries about you.

Scene 8 – Bathroom #2: You locked yourself in the bathroom of your house (the same one you see in the first scene), and standing in front of the mirror, on the sink you can see cocaine powder and a razor (**Figure 1**, bottom left). You can listen to your wife, who is outside the room, wondering why you locked the door and informing you that the food is ready.

Scene 9 – Dining area: You stand in the dining area, your wife is standing by the table and you see your bag opened and cocaine on the table as well as money found by your wife, who is disappointed in you and says she noticed a big withdrawal of money from your shared account and it explains all the symptoms you have been showing lately (**Figure 1**, bottom right). She wonders what your mother would say about all this and urges you to seek some help or she is going to leave you.

The goal of showing the participants the whole journey of this man was to reverse the fundamental attribution error, that people tend to do. According to Ross (1977), the fundamental attribution error is defined as "the tendency for attributors to underestimate the impact of situational factors and to overestimate the role of dispositional factors in controlling behavior." People tend to overestimate the influence of personality and underestimating the influence of situations, when explaining other people's

FIGURE 1 | Workplace scene (top-left), party scene (top-right), bathroom scene, illustrating the physical effects of cocaine use (bottom-left), consequences on the relationship (bottom right).

behavior (Myers, 2010). In this case and with our narrative, we wanted to show the participants, that in most cases situations drive people to drug use. According to the self-medication hypothesis (Khantzian, 1997, 2003), addictive drugs have appeal because during the short term they relieve painful feelings and psychological distress. There is a tendency in society to stigmatize drug users, as people think they use drugs for their enjoyment only. The character in the scenario is seen using drugs as a coping mechanism for the death of his mother and the stress from his work. Additionally, we followed the suggestion of Bertrand et al.'s (2018) for the embodiment of a digital avatar of an outgroup member that presents traits that contradict stereotypes, which we have done in our study. Additional screenshots from the scenes can be found in the **Supplementary Table S1**.

Participants

A total number of 40 participants (n = 40) participated in the study, 21 of them male and 19 female. For the recruitment of the participants we used convenience sampling. Ages of participants varied from 18 to 59 with the majority (42.5%) in the range of 18–24 years old. Furthermore, data on participants' experience in using VR environments, indicated that most of the participants (42.5%) had little experience with VR. A table with the frequencies of the demographics across the two groups can be found in **Supplementary Table S2**.

Experimental Design

A between-groups design was used to conduct the study. The participants were randomly assigned into two groups, the Sensorimotor Contingencies condition (SC) (n = 20) and No Sensorimotor Contingencies (NSC) condition (n = 20), which are explained further in the following sections. The majority (11) of the participants in the NSC group were aged between (18–24), and for the SC group the majority (8) aged between 25 and 29 years old. Regarding occupation, in the NSC group the majority (12) were students and in the SC group (7) were working in the private sector. The study has been approved by the Research's Ethics and Deontology committee of the Cyprus University of Technology. All participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

SC Condition

This condition supports several sensorimotor contingencies for the user. It required the participants to put on a VR HMD allowing stereo viewing in 360 degrees of the surrounding environment and supporting head tracking allowing displacement of the viewing scene in a physical way. Real time tracking of full-body's and hands', at the detail of the fingers', movements was also supported and real time mapping of the participant's movements to the virtual avatar as well. That was achieved by having the participants fitted with a wireless full body motion capture suit and VR data gloves. Wearing both, the participants were virtually embodying the drug user. They were viewing the virtual world in first-person view and could rotate their head to change their point of view. There was no other stimulation than the first-person perspective view over the body.

The participants could not interact with the application in any way or with the other avatars or virtual items in the virtual world. The participants were required to be standing up and could move their body and head around to change their point of view and watch their avatar in the virtual world, which was the drug user, move simultaneously to their real time movements.

NSC Condition

The application of the NSC version was identical with the SC version with the difference being that the participants did not wear any of the equipment mentioned above thus no sensorimotor contingencies were supported. Instead, they were sitting on a chair in front of a desktop computer, viewing the same application, this time, through a flat computer screen and could only turn their point of view with the use of the mouse.

Materials and Technical Setup

The application used for both groups was developed in the Unity software. The models of the avatars in the application were created using the online software Autodesk® Character generator, in which a skeleton was also added to the models as well as facial blend shapes, so that their mouths could move according to the dialogues recorded which were recorded using a Philips audio recorder. These recordings were then edited in the Adobe Audition CS6 software.

For the SC group the virtual environment was displayed through an Oculus Rift CV HMD. This has two 1080 × 1200 pixels OLEDs per eye at 90 Hz display refresh rate, coupled with a positional tracker and built-in headphones. The participants in this group had to wear the Xsens MVN Awinda wireless motion capture trackers for real time body movement tracking which was mapped to their avatar in the virtual world through the Xsens software and the Manus VR – Xsens Edition gloves which offered finger tracking (see **Figure 2**).

For the NSC condition the application was displayed to participants through 1920 × 1080 pixels 24-inch computer screen. A computer mouse was used for the interaction with the application and headphones for the sound (see **Figure 3**).

Procedure

Upon arriving at the laboratory, the participants were randomly assigned to one of the two groups, NSC or SC and were asked to read and sign the consent form. Then they were given the pre-questionnaire. Once this process was completed, participants in the SC group were asked to put on the VR HMD, Xsens Awinda trackers and VR gloves and participants in the NSC group were asked to sit in front of the computer screen and wear the headphones for the sound. They were then briefed about the way they could interact with the application. Then, both groups viewed the application. The intervention lasted approximately 10 min for both groups. Participants completed the same post-intervention questionnaire immediately after viewing the application.

Measures

Two questionnaires were given to the participants in total, which were the same for both groups, NSC and SC.

Pre-questionnaire

The pre-questionnaire was given to the participant before their exposure to the virtual world. Using this questionnaire, we gathered demographic data such as age, gender, occupation, hours spent playing video games, experience in using VR environments and computer programming knowledge.

It also included the following two measures:

FIGURE 2 | One of the participants of the SC condition wearing the Oculus Rift, Xsens Awinda trackers, and Manus VR gloves.

Interpersonal reactivity index (IRI)

Interpersonal reactivity index is a self-report tool that consists of four 7-item subscales answered on a 5-point Likert scale ranging from "Does not describe me well" to "Describes me very well" (Davis, 1983), used to measure individual differences in empathy. The four subscales are: (a) Perspective-Taking scale (PT): the ability to shift perspectives when dealing with other people. (b) Empathic Concern scale (EC): assessment of the degree to which the respondent experiences feelings of warmth, compassion and concern for the observed individual. (c) Personal Distress scale (PD): measures the individual's own feelings of fear, apprehension and discomfort at witnessing the negative experiences of others.

It is important to note, that in the pre-questionnaire, the fantasy scale was not measured, as the participants

FIGURE 3 | One of the participants of the NSC group.

would experience through an application the life of the drug user, not imagine it.

Attitudes toward drug users

This was a 6-item attitudes scale on a 5-point Likert scale (1 = Completely disagree, 5 = Completely agree). It was adapted from the questionnaire used by Batson et al. (1997) and Herrera et al. (2018) to address attitudes toward drug users. Sample questions include "Our society should do more to protect the welfare of drug users" and "For more drug users, it is their fault that they are drug users." Two out of the six questions were reverse coded. Higher scores indicate more positive attitudes toward drug users. This questionnaire was given to the participants before and after their exposure, in order to see if their exposure to the application, can change their attitudes toward drug users.

Post-questionnaire

The questionnaire given to all participants, regardless of their group, after their exposure included questionnaires measuring their level of empathy and personal distress, the Inclusion of the Other in the Self scale, the same questions regarding their attitudes toward drug users as the pre-questionnaire. It also included questions regarding presence [Place Illusion (PI) and Plausibility of the situation (Psi)], Body Ownership and Agency (BOA).

Empathy

Participants were asked to rate the extent to which they felt soft hearted, touched, sympathetic, or compassionate throughout the intervention by using a 7-point Likert scale (1 = Not at All, 7 = Very much). The results of these four questions were used to create an index of empathic concern. This measure was adapted from Batson et al. (1997).

Personal distress

Participants were asked to rate the extent to which they felt uneasy, troubled, distressed, or disturbed throughout the intervention by using a 7-point Likert scale (1 = Not at All, 7 = Very much). The results of these four questions were used to create an index of personal distress. This measure was also adapted from Batson et al. (1997).

IOS

The Inclusion of Other in the Self (IOS) scale measured how close the participants felt to the drug user. Developed by Aron et al. (1992), this scale depicts seven drawings of increasingly overlapping circles, anchored by the first picture of two non-overlapping circles and the seventh picture of two almost completely overlapping circles. Participants had to choose the picture that best represented the extent to which he/she felt connected to the drug user. The pictures are coded from 1 to 7 with the larger numbers indicating a closer relationship with the drug user.

Place illusion (PI) and plausibility of the situation and the virtual people

Place illusion (PI) refers to the sense of "being there" in a virtual world (Slater, 2009). PI was measured using a 5-item questionnaire on a 5-point Likert scale (1 = Not at all, 5 = Very much). Sample questions include "I had a sense of "being there" in the virtual environments" and "During the time of the experience, which was strongest overall, your sense of being in virtual environments, or of being in the real world of the laboratory?"

The Plausibility of the situation (Psi) is the illusion that what is apparently happening is really happening (even though you know that it is not). In the questionnaires, Psi was divided into two aspects; the Psi, and plausibility of the virtual human characters, which is if the participants felt that the virtual people were behaving, moving, reacting as if they were real. They were measured using a 6-itemand 5-item questionnaires, respectively, on a 5-point Likert scale (1 = Not at all, 5 = Very much). Sample questions for the plausibility of situation include "How much did you behave within the scenes as if the situations were real?" and for the plausibility of the virtual people "How much were you thinking things like "I know these people are not real" but then surprisingly finding yourself behaving as if they were?"

Body ownership and agency

According to Tsakiris et al. (2006) body ownership refers to the sense that one's own body is the source of sensations. Gallagher (2000) defined the sense of agency as "the sense that I am the one who is causing or generating an action." They were measured using a 4-item questionnaire on a 5-point Likert scale (1 = Completely agree, 5 = Completely disagree). Sample questions include "During the experience I felt as though I had two bodies" and "During the experience I felt that the virtual body belonged to someone else." The complete questionnaires of all measures can be found in **Supplementary Table S2**.

RESULTS

All results were obtained with using the SPSS Software v.24 by IBM. All datasets generated for this study are included in the **Supplementary Data Sheet S1**.

The means and standard deviations for the outcome variables; IRI, Inclusion of the Other in the Self (IOS), Empathy, Personal Distress, Pre-, and Post-Attitudes toward drug users by condition (NSC and SC) are summarized in **Table 1**.

IRI

An independent-samples *t*-test was conducted to compare the level of empathy of the participants before their exposure in the two groups. Although the mean for the SC group is higher than the NSC group, there was not a statistically significant difference in the IRI scores for the NSC group (M = 71.1, SD = 11.02) and SC group (M = 77.4, SD = 11.08) conditions; $t(38) = -1.802$, $p = 0.079$.

Separately in the subscales, higher scores were shown was all three subscales of IRI for the SC group, but there was also no significant difference in participants' trail-levels levels of empathy on any of the three subscales [empathetic concern (EC), perspective-taking (PT), and personal distress (PD)], of the IRI scales between the two conditions. More specifically, regarding the EC scale, for the NSC group (M = 26.20, SD = 5.22) and SC group (M = 28.45, SD = 4.61) conditions; $t(38) = -1.443$, $p = 0.157$. Next for the PT scale, for the NSC group (M = 25.90, SD = 4.59) and SC group (M = 27.20, SD = 3.95) conditions; $t(38) = -0.959$, $p = 0.344$. Finally for the PD scale, for the NSC group (M = 19.00, SD = 4.06) and SC group (M = 21.75, SD = 5.66) conditions; $t(38) = -1.764$, $p = 0.086$.

TABLE 1 | Means and standard deviations for all outcome variables for both groups.

Measures	Condition			
	NSC		SC	
	Mean	*SD*	Mean	*SD*
IRI	71.1	11.02	77.4	11.08
IOS	2.85	1.89	3.75	1.55
Empathetic concern	19.4	6.87	19.6	5.68
Personal distress	17.8	6.57	19.55	6.19
Pre-attitudes	20.3	3.78	21.6	3.4
Post-attitudes	21.45	3.83	22.7	3.48

SD, Standard deviation.

This suggests that the two conditions were balanced in terms of pre-intervention empathy and they were successfully randomly assigned into these two groups.

Gender differences were found in the SC group regarding IRI. More specifically, there was a significant difference between males (M = 71.11, SD = 11.12) and females (M = 82.54, SD = 8.34); $t(18) = -2.627$, $p = 0.017$. This suggests that females in the SC group had significantly higher levels of empathy than the males in the SC group before their exposure. More separately in the sub-scales, females only in the SC group had significantly higher levels of EC (M = 30.45, SD = 3.47) over the males (M = 26.00, SD = 4.82); $t(18) = -2.401$, $p = 0.027$ as well as levels of PD (M = 24.18, SD = 5.23) over the males (M = 18.77, SD = 4.89); $t(18) = -2.365$, $p = 0.029$. Females also had higher but not significantly different levels of PT (M = 27.90, SD = 3.70) over the males (M = 26.33, SD = 4.30); $t(18) = -0.881$, $p = 0.390$. This aligns with the findings of Davis (1980), where women displayed higher scores than men for each of the four subscales or IRI.

A Pearson product-moment correlation coefficient was computed to estimate the relationship between the IRI total score of the participants that was measured before their exposure and their attitudes toward drug users before and after their exposure. Results indicate significant positive correlation between these variables. More specifically, participants that reported more empathetic before their exposure also had more positive attitudes toward drug users after their exposure (positive medium correlation; $r = 0.353$, $n = 40$, $p = 0.025$) and before their exposure (positive strong correlation; $r = 0.412$, $n = 40$, $p = 0.008$). Separately in the two groups, this correlation was observed to a significant level only in the NSC group for their post-attitudes scores ($r = 0.452$, $n = 20$, $p = 0.045$) where a significant medium positive correlation was found.

Attitudes Toward Drug Users

Both groups showed a significant difference in their reported attitudes toward drug users before and after their exposure. There was a significant difference in the attitudes of the participants in the NSC group for their attitudes before their exposure (M = 20.3, SD = 3.79) and after (M = 21.45, SD = 3.83); $t(19) = -2.529$, $p = 0.020$. Similarly, there was a significant difference in the attitudes of the participants in the SC group for their attitudes before their exposure (M = 21.6, SD = 3.4) and after (M = 22.7, SD = 3.48); $t(19) = -2.125$, $p = 0.047$.

Gender differences were found in the NSC group regarding both pre-attitudes and post-attitudes. More specifically, there was a significant difference regarding the pre-attitudes of males (M = 18.75, SD = 3.57) and females (M = 22.62, SD = 2.92); $t(18) = -2.546$, $p = 0.020$. Also, there was a significant difference

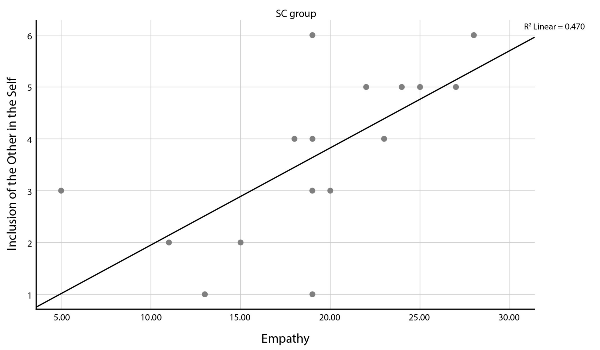


FIGURE 4 | Strong correlation between closeness to the drug user and empathy in the SC group.

regarding the post-attitudes too, of males ($M = 20.08$, $SD = 3.89$) and females ($M = 23.50$, $SD = 2.82$); $t(18) = -2.127$, $p = 0.047$. Both these results suggest that females in the NSC group had more positive attitudes toward drug users both before and after their exposure than the males in the NSC group.

There was not a significant difference between the two groups in their post-attitudes, NSC ($M = 21.45$, $SD = 3.83$) and SC ($M = 22.7$, $SD = 3.48$) conditions; $t(38) = -1.08$, $p = 0.287$.

IOS

There was not a significant difference in the closeness of the participants toward the drug user between the two groups, NSC ($M = 2.85$, $SD = 1.899$) and SC ($M = 3.75$, $SD = 1.552$) conditions; $t(38) = 0.084$, $p = 0.109$. Regarding the SC group, IOS was positively significantly correlated to their reported levels of post-attitudes ($r = 0.512$, $n = 20$, $p = 0.021$) and their reported levels of empathy (**Figure 4**) ($r = 0.686$, $n = 20$, $p = 0.001$). Which means that participants in the SC group who reported being emotionally closer to the drug user, also reported more positive attitudes toward drug users after their exposure, and more empathetic.

Regarding the NSC group, IOS was positively significantly correlated only to their reported levels of BoA ($r = 0.508$, $n = 20$, $p = 0.022$). Which means that participants in the NSC group who reported being emotionally closer to the drug user, also reported higher levels of BOA.

Empathy

Participants empathy was measured after their exposure. The overall reliability of the index was good, Cronbach's alpha = 0.921. There was not a significant difference in the empathy levels between the two groups, NSC ($M = 19.4$, $SD = 6.87$) and SC ($M = 19.6$, $SD = 5.68$) conditions; t (38) = -0.100, $p = 0.921$.

Regarding the NSC group; empathy was positively significantly correlated to their reported levels of BoA ($r = 0.475$, $n = 20$, $p = 0.034$), their reported levels of PVP ($r = 0.564$, $n = 20$, $p = 0.010$) and finally their personal distress levels ($r = 0.630$, $n = 20$, $p = 0.003$). This means that the more empathetic they reported they felt after their exposure, the more distressed they were and reported higher levels of body ownership, agency and plausibility of the virtual people.

Personal Distress

Participants personal distress was measured after their exposure. The overall reliability of the index was good, Cronbach's alpha = 0.889. There was not a significant difference in the personal distress levels between the two groups, NSC ($M = 17.8$ $SD = 6.57$) and SC ($M = 19.55$, $SD = 6.19$) conditions; $t(38) = -0.866$, $p = 0.392$.

Regarding the NSC group, personal distress was positively significantly correlated to their reported levels of PI ($r = 0.457$, $n = 20$, $p = 0.043$), their reported levels of PSI ($r = 0.514$, $n = 20$, $p = 0.020$) and their levels of PVP ($r = 0.513$, $n = 20$, $p = 0.021$). This means that the more distressed they felt, the higher the levels of PI, Psi, and the virtual people.

Factor Analysis

For the questionnaire data regarding the participants experience, factor analysis was carried out to reduce the number of questionnaire variables, which also has the advantage of transforming ordinal variables to continuous ones. Corresponding factor scores were used, and the interpretation of each factor was identified. Subsequent analysis on the derived variables was done using independent samples t-test.

Figure 5 shows the bar chart of the means and standard errors of the derived factor scores for all four variables Place Illusion (PI), Body Ownership and Agency (BOA), Plausibility of the situation (Psi), and Plausibility of the virtual people (PVP).

Place Illusion (PI)

Factor analysis in the questionnaire on the participants reported PI has resulted in a single factor. The factor loadings of the scoring variable PI are shown in **Table 2**. The factor is interpreted as "the illusion of being there." There was a statistically significant difference between the two groups. Participants in the SC condition reported higher levels of presence (0.399 ± 0.169) compared to participants in the NSC condition (-0.399 ± 0.238). An independent samples t-test showed that the above differences are significant [$t(38) = -2.728$, $p = 0.010$].

Body Ownership and Agency (BOA)

Factor analysis in the questionnaire on the participants reported BOA has resulted in a single factor. The factor loadings of the scoring variable BOA are shown in **Table 3**. The factor is interpreted as "the sense that one's own body is the source of sensations and the sense that I am the one who is causing or generating an action." There was a statistically significant difference between the two groups. Participants in the SC condition reported higher levels of BOA (0.541 ± 0.147) compared to participants in the NSC condition (-0.541 ± 0.223). An independent samples t-test showed that the above differences are significant [$t(38) = -4.047$, $p = 0.000$].

Plausibility of the Situation (Psi)

Factor analysis in the questionnaire on the participants reported Psi has resulted in a single factor. The factor

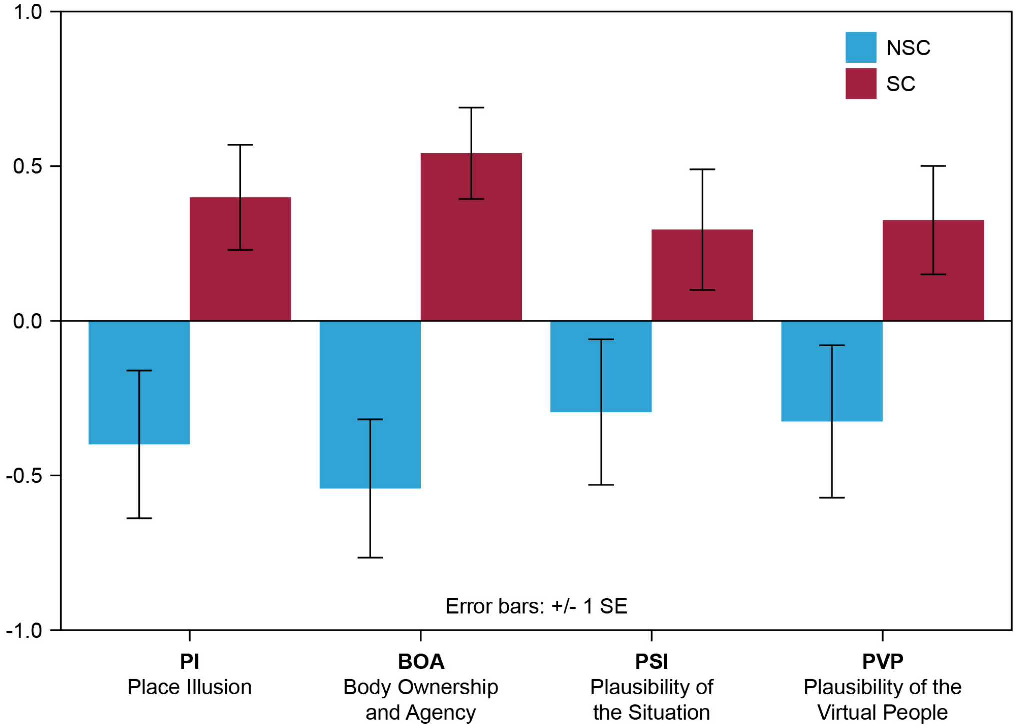


FIGURE 5 | Means and standard errors of the derived factor scores for all four variables PI, BOA, PSI, and PVP.

loadings of the scoring variable PSI are shown in **Table 4**. The factor is interpreted as "the illusion that events are actually happening." Participants in the SC condition rated their sensed illusion more positively (0.294 ± 0.194) compared to participants in the NSC condition (−0.294 ± 0.235). An independent samples *t*-test showed that the above differences are not significant [$t(38) = -1.930$, $p = 0.061$].

Plausibility of the Virtual People (PVP)

Factor analysis in the questionnaire on the participants reported PVP has resulted in a single factor. The factor loadings of the scoring variable PVP are shown in **Table 5**. The factor is interpreted as the "behavioral, physiological, emotional, and thinking responses as if the people were real." There was a statistically significant difference between the two groups. Participants in the SC condition reported higher levels of PVP (0.541 ± 0.147) compared to participants in the NSC condition (−0.541 ± 0.223). An independent samples *t*-test showed that the above differences are significant [$t(38) = -4.047$, $p = 0.000$].

TABLE 2 | Factor analysis for Place Illusion, resulting in a single factor F1 and the corresponding scoring coefficients of the factor score yp1.

Variable	Factor loadings	Scoring coefficients
	F1	yp1
There	0.886	0.257
Real	0.808	0.234
Visited	0.787	0.228
Lab	0.815	0.236
Overwhelm	0.854	0.247

TABLE 3 | Factor analysis for Body Ownership and Agency, resulting in a single factor F1 and the corresponding scoring coefficients of the factor score yboa1.

Variable	Factor loadings	Scoring coefficients
	F1	yboa1
Mybody	0.924	0.393
Twobodies	0.426	0.181
Agency	0.892	0.379
Otherbody	−0.721	−0.307

TABLE 4 | Factor analysis for Plausibility of the situation, resulting in a single factor F1 and the corresponding scoring coefficients of the factor score ypsi1.

Variable	Factor loadings	Scoring coefficients
	F1	ypsi1
Behavereal	0.911	0.196
Emotionreal	0.881	0.213
Thoughtsreal	0.823	0.199
Behaveasifreal	0.821	0.199
Physicalreal	0.807	0.196
Experiencereal	0.833	0.202

TABLE 5 | Factor analysis for Plausibility of the virtual people, resulting in a single factor F1 and the corresponding scoring coefficients of the factor score ypvp1.

Variable	Factor loadings	Scoring coefficients
	F1	ypvp1
Mybody	0.924	0.393
Twobodies	0.426	0.181
Agency	0.892	0.379
Otherbody	−0.721	−0.307

DISCUSSION

The aim of this study was to investigate the role of supported sensorimotor contingencies when attempting to promote empathy and positive attitudes toward drug users, by comparing a VR system that offered perception through sensorimotor contingencies (SC condition) and system that didn't offer that (NSC condition), by showing the participants scenes from the life of a drug user.

Although the SC group had higher means in all the measures, as seen in **Table 1**, they were not significantly different to those of the NSC group. This aligns with the results of Herrera et al. (2018) that showed that the differences between mediated perspective-taking tasks, regardless of how immersive they are, were not significant in terms of IOS, empathy or personal distress. Both groups showed a significant difference in their reported attitudes toward drug users before and after their exposure which means that both interventions succeeded in eliciting more positive attitudes toward drug users to the participants. Participants in the VR group reported significantly higher levels of PI, BOA, and PVP but not the situation. The results could be explained by the fact that most of the participants (60%) had non or little experience with VR and maybe this new experience and technology distracted them from experiencing fully the scenario, resulting in similar levels of empathy and personal distress. The SC groups' higher levels of empathy and body ownership could explain the correlation between their closeness to the drug user and empathy as well as their positive attitudes toward drug users. This is in line with the study by Maister et al. (2013) that showed the more intense the participants' illusion of ownership over a dark-skinned rubber hand, the more positive their implicit racial attitudes. Additionally, the experience lasted approximately 6 min, and this perhaps was not enough time for the participants to truly experience what it is like to be a drug user.

Furthermore, this study aimed to show the participants, that in most cases, people end up doing drug use because of bad situations that are going through in life, and they find "relief" and an "escape" in drug use. Considering that, the scenes and scenario created, wanted to reflect the journey a person goes in, before drug use and after, and the consequences it has in his health, life, family and workspace. Future studies should focus more on the negative aspects that drug use has on drug users and their life. Future studies should also investigate the role of agency and interactivity plays in empathy and attitudes toward stigmatized people. Interactivity has been found to increase empathy (Vorderer et al., 2001), and it can be utilized in VR applications. As many companies are releasing VR headsets every year with way more features and needing less extra equipment to allow hands and even body tracking, it would be interesting to compare in future studies, the difference between low vs. high cost VR and tracking equipment in the supporting sensorimotor contingencies. Examples of that could include the comparison of motion tracking technologies like the more affordable Microsoft Kinect and a more expensive specialized motion tracking suit like the one used by the participants of the SC group in the current study, or even different cost HMD's which offer less or more tracking possibilities.

The findings of this study provide more evidence that show that VRPT, can be used for to induce empathy and prosocial behavior of the participants toward members of stigmatized groups. Further research is needed to investigate how perception through sensorimotor contingencies can be exploited to the fullest to be used as an effective method to induce empathy and change attitudes toward stigmatized by society people.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The study has been approved by the Research's Ethics and Deontology Committee of the Cyprus University of Technology. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this manuscript.

AUTHOR CONTRIBUTIONS

MC made substantial contribution in the conception and design of the study, development of the VR application, data collection, analysis and interpretation of the data, and drafting the article. DM-G contributed in the design and conception of the study, interpretation of the data, revising critically the manuscript, and supervised and coordinated all the steps of the study. CK contributed partially in the development of the VR application and substantially in data analysis.

FUNDING

This research has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant

Agreement No. 739578 (RISE – Call: H2020-WIDESPREAD-01-2016-2017-TeamingPhase2) and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

ACKNOWLEDGMENTS

We would like to thank Dr. Domna Banakou for the discussion and exchanging ideas on the experimental design and methods.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01242/full#supplementary-material

TABLE S1 | Additional views from the various scenes.

TABLE S2 | Participant's demographic data and questionnaires.

DATA SHEET S1 | All datasets generated for this study.

VIDEO S1 | A video illustrating the application.

REFERENCES

Ahn, S. J., Le, A. M. T., and Bailenson, J. (2013). The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychol.* 16, 7–38. doi: 10.1080/15213269.2012.755877

Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *J. Pers. Soc. Psychol.* 63:596. doi: 10.1037/0022-3514.63.4.596

Ash, E. (2016). Priming or proteus effect? Examining the effects of avatar race on in-game behavior and post-play aggressive cognition and affect in video games. *Games Cult.* 11, 422–440. doi: 10.1177/1555412014568870

Banakou, D., Hanumanthu, P. D., and Slater, M. (2016). Virtual embodiment of white people in a black virtual body leads to a sustained reduction in their implicit racial bias. *Front. Hum. Neurosci.* 10:601.

Batson, C. D., Early, S., and Salvarani, G. (1997). Perspective taking: imagining how another feels versus imaging how you would feel. *Pers. Soc. Psychol. Bull.* 23, 751–758. doi: 10.1177/0146167297237008

Batson, C. D., Lishner, D. A., Carpenter, A., Dulin, L., Harjusola-Webb, S., Stocks, E. L., et al. (2003). As you would have them do unto you": does imagining yourself in the other's place stimulate moral action?. *Pers. Soc. Psychol. Bull.* 29, 1190–1201. doi: 10.1177/0146167203254600

Behm-Morawitz, E., Pennell, H., and Speno, A. G. (2016). The effects of virtual racial embodiment in a gaming app on reducing prejudice. *Commun. Monogr.* 83, 396–418. doi: 10.1080/03637751.2015.1128556

Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., and Zenasni, F. (2018). Learning empathy through virtual reality: multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Front. Robot. AI* 5:26. doi: 10.3389/frobt.2018.00026

Chowdhury, T. I., Ferdous, S. M. S., and Quarles, J. (2019). VR disability simulation reduces implicit bias towards persons with disabilities. *IEEE Trans. Vis. Comput. Graph.* doi: 10.1109/TVCG.2019.2958332 [Epub ahead of print].

Christofi, M., and Michael-Grigoriou, D. (2017). "Virtual reality for inducing empathy and reducing prejudice towards stigmatized groups: a survey," in *Proceedings of the 2017 23rd International Conference on Virtual System & Multimedia (VSMM)*, (Piscataway, NJ: IEEE), 1–8. doi: 10.1109/VSMM.2017.8346252

Christofi, M., Baka, E., Stavroulia, K. E., Michael-Grigoriou, D., Lanitis, A., and Magnenat-Thalmann, N. (2018). "Studying levels of presence in a virtual environment simulating drug use in schools: effect on different character perspectives," in *Proceedings of the ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*, (Geneve: The Eurographics Association), 163–170. doi: 10.2312/egve.20181329

Davis, M. H. (1980). Individual differences in empathy: a multidimensional approach. *Dissert. Abstr. Int.* 40:3480.

Davis, M. H. (1983). Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* 44:113. doi: 10.1037/0022-3514.44.1.113

Farmer, H., Tajadura-Jiménez, A., and Tsakiris, M. (2012). Beyond the colour of my skin: how skin colour affects the sense of body-ownership. *Conscious Cogn.* 21, 1242–1256. doi: 10.1016/j.concog.2012.04.011

Galinsky, A. D., Maddux, W. W., Gilin, D., and White, J. B. (2008). Why it pays to get inside the head of your opponent. *Psychol. Sci.* 19, 378–384. doi: 10.1111/j.1467-9280.2008.02096.x

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends Cogn. Sci.* 4, 14–21. doi: 10.1016/S1364-6613(99)01417-5

Goffman, E. (2009). *Stigma: Notes on the Management of Spoiled Identity*. New York, NY: Simon and Schuster.

Groom, V., Bailenson, J. N., and Nass, C. (2009). The influence of racial embodiment on racial bias in immersive virtual environments. *Soc. Influence* 4, 231–248. doi: 10.1080/15534510802643750

Hamilton-Giachritsis, C., Banakou, D., Quiroga, M. G., Giachritsis, C., and Slater, M. (2018). Reducing risk and improving maternal perspective-taking and empathy using virtual embodiment. *Sci. Rep.* 8, 1–10. doi: 10.1038/s41598-018-21036-2

Harris, L. T., and Fiske, S. T. (2006). Dehumanizing the lowest of the low: neuroimaging responses to extreme out-groups. *Psychol. Sci.* 17, 847–853. doi: 10.1111/j.1467-9280.2006.01793.x

Hasler, B. S., Spanlang, B., and Slater, M. (2017). Virtual race transformation reverses racial in-group bias. *PLoS One* 12:e0174965. doi: 10.1371/journal.pone.0174965

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., and Zaki, J. (2018). Building long-term empathy: a large-scale comparison of traditional and virtual reality perspective-taking. *PLoS One* 13:e0204494. doi: 10.1371/journal.pone.0204494

Khantzian, E. J. (1997). The self-medication hypothesis of substance use disorders: a reconsideration and recent applications. *Harv. Rev. Psychiatry* 4, 231–244. doi: 10.3109/10673229709030550

Khantzian, E. J. (2003). The self-medication hypothesis revisited: the dually diagnosed patient. *Prim. Psychiatry* 10, 47–54.

Lamm, C., Batson, C. D., and Decety, J. (2007). The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosci.* 19, 42–58. doi: 10.1162/jocn.2007.19.1.42

Maister, L., Sebanz, N., Knoblich, G., and Tsakiris, M. (2013). Experiencing ownership over a dark-skinned body reduces implicit racial bias. *Cognition* 128, 170–178. doi: 10.1016/j.cognition.2013.04.002

Myers, D. G. (2010). *Psychology*, 9th Edn. New York, NY: Worth.

Peck, T. C., Seinfeld, S., Aglioti, S. M., and Slater, M. (2013). Putting yourself in the skin of a black avatar reduces implicit racial bias. *Conscious Cogn.* 22, 779–787. doi: 10.1016/j.concog.2013.04.016

Penn, D. L., Ivory, J. D., and Judge, A. (2010). The virtual doppelganger: effects of a virtual reality simulator on perceptions of schizophrenia. *J. Nerv. Ment. Dis.* 198, 437–443. doi: 10.1097/NMD.0b013e3181e07d66

Perdue, C. W., and Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *J. Exp. Soc. Psychol.* 26, 199–216. doi: 10.1016/0022-1031(90)90035-K

Ross, L. (1977). "The intuitive psychologist and his shortcomings: distortions in the attribution process," in *Advances in Experimental Social Psychology*, Vol. 10, ed. M. P. Zanna, (Cambridge, MA: Academic Press), 173–220. doi: 10.1016/S0065-2601(08)60357-3

Schutte, N. S., and Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motiv. Emot.* 41, 708–712. doi: 10.3389/fpsyg.2019.01521

Seinfeld, S., Arroyo-Palacios, J., Iruretagoyena, G., Hortensius, R., Zapata, L. E., Borland, D., et al. (2018). Offenders become the victim in virtual reality: impact

of changing perspective in domestic violence. *Sci. Rep.* 8:2692. doi: 10.1038/s41598-018-19987-7
Shechtman, Z., and Tanus, H. (2006). Counseling groups for Arab adolescents in an intergroup conflict in Israel: report of an outcome study. *Peace Conflict* 12, 119–137. doi: 10.1207/s15327949pac1202_2
Sidanius, J., and Pratto, F. (2001). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge: Cambridge University Press, doi: 10.1017/CBO9781139175043
Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philos. Trans. R. Soc. B* 364, 3549–3557. doi: 10.1098/rstb.2009.0138
Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074
Stavroulia, K. E., Baka, E., Christofi, M., Michael-Grigoriou, D., Magnenat-Thalmann, N., and Lanitis, A. (2018). "A virtual reality environment simulating drug use in schools: effect on emotions and mood states," in *Proceedings of the International Conference on Information, Communication Technologies in Education, ICICTE 2018*, Chania,
Stavroulia, K. E., Christofi, M., Baka, E., Michael-Grigoriou, D., Magnenat-Thalmann, N., and Lanitis, A. (2019). Assessing the emotional impact of virtual reality-based teacher training. *Int. J. Inform. Learn. Technol.* 36, 192–217. doi: 10.1108/IJILT-11-2018-0127
Tomasello, M. (2009). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard university press.
Tsakiris, M., Prabhu, G., and Haggard, P. (2006). Having a body versus moving your body: how agency structures body-ownership. *Conscious Cogn.* 15, 423–432. doi: 10.1016/j.concog.2005.09.004
van Loon, A., Bailenson, J., Zaki, J., Bostick, J., and Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLoS One* 13:e0202442. doi: 10.1371/journal.pone.0202442
Vorderer, P., Knobloch, S., and Schramm, H. (2001). Does entertainment suffer from interactivity? The impact of watching an interactive TV movie on viewers' experience of entertainment. *Media Psychol.* 3, 343–363. doi: 10.1207/S1532785XMEP0304_03
Yee, N., and Bailenson, J. (2007). The Proteus effect: the effect of transformed self-representation on behavior. *Hum. Commun. Res.* 33, 271–290. doi: 10.1111/j.1468-2958.2007.00299.x
Yee, N., Bailenson, J. N., and Ducheneaut, N. (2009). The Proteus effect: implications of transformed digital self-representation on online and offline behavior. *Commun. Res.* 36, 285–312. doi: 10.1177/0093650208330254
Zaki, J., and Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nat. Neurosci.* 15:675. doi: 10.1038/nn.3085

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 21 July 2020
doi: 10.3389/fcomp.2020.00022

Digital Training in the Aeronautical Industry: Measuring the Usability of Two Mobile Applications

Phillip James Smy[1]*, Iain Donald[1]*, Kenneth Scott-Brown[2] and Ruth E. Falconer[1]

[1] School of Design and Informatics, Abertay University, Dundee, United Kingdom, [2] School of Applied Sciences, Abertay University, Dundee, United Kingdom

The air traffic control industry is highly regulated, with stringent processes and procedures to ensure that IP (Intellectual Property) and workplaces are kept secure. The training of air traffic controllers (ATCs) and other roles relating to air traffic services is a lengthy and expensive process. The rate in which trainees can be trained is projected to fall significantly short of the demand for staff to work in the air traffic industry. This paper focuses on two prototype mobile training applications—Location Indicators (LI) and the Aircraft Control Positions Operator (ACPO) Starter Pack. LI and the ACPO Starter Pack have been produced to explore how air traffic control training could be improved and supported using digital applications. Each application explores a key learning area for trainees in the air traffic control industry and presents an alternative to the equivalent training that is currently in use. The two prototypes that have been designed focus on producing a succinct user experience alongside gamified elements to improve engagement. As part of this paper, usability testing has been undertaken with LI and the ACPO Starter Pack. A total of nine usability tests have been undertaken at four different locations. These usability tests consisted of participants from differing demographics, varying experience with the current training and differing amounts of time with both applications. The System Usability Scale (SUS) was adapted and used to quantify participant's reactions to the usability of each application. Usability scores for both applications were collected and then averaged to produce an overall score for each application. We can conclude from both usability scores and qualitative feedback that digital applications have the potential to engage future trainees in the air traffic services industry.

Keywords: user experience, gamification, digital training, air traffic control, engagement, aerospace, mobile apps, e-learning

OPEN ACCESS

Edited by:
Elvira Popescu, University of Craiova, Romania

Reviewed by:
Chien-Sing Lee, Sunway University, Malaysia
Jorge Luis Bacca Acosta, Fundación Universitaria Konrad Lorenz, Colombia

****Correspondence:***
Phillip James Smy
p.smy@abertay.ac.uk
Iain Donald
i.donald@abertay.ac.uk

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *18 February 2020*
Accepted: *03 June 2020*
Published: *21 July 2020*

Citation:
Smy PJ, Donald I, Scott-Brown K and Falconer RE (2020) Digital Training in the Aeronautical Industry: Measuring the Usability of Two Mobile Applications. Front. Comput. Sci. 2:22. doi: 10.3389/fcomp.2020.00022

INTRODUCTION

It can take 3 years to train ATCs (Humphreys, 2017; Noyes, 2018). Rigorous pre-testing means only 1% of potential applicants go onto the training phase (BBC, 2018a,b). Experienced ATCs tend to act as teachers (National Research Council, 1997) which further contributes to a shortage of ATCs. With experienced ATCs being assigned to teaching roles, ATCs retiring at a younger age (Federal Aviation Administration, 2017), a low pass rate of ATC applicants and the amount of flights expecting to rise to an extra 500,000 per by year 2030 year (pre-Covid 19), there is an increased demand to both recruit and train ATCs (BBC, 2018a,b). Digital alternatives to training can provide learning which is flexible, with less dependency on instructors (Pollard and Hillage, 2011). Issues pertaining to training time, the costs associated with training and the effectiveness of training, could all be alleviated through the widespread adoption of digital training (Strother, 2002).

Training ATCs is not only lengthy but also costly with each trainee costing ∼£600,000 to be trained (Savill, 2009). This training process has been identified to be "no longer engaging the current generation of trainees" (Humphreys, 2017). With the ever-increasing demand for ATCs and the current training needing to adapt to meet the learning expectations of the newer generation of air traffic controller trainees, the disruption of current air traffic controller training is desperately needed. Whilst not being specific to air traffic controller training, the so called "millennial" generation is frequently found to prefer learning via technology (Oblinger, 2003). Millennials are defined as individuals who have been born between 1979 and 1994 (Myers and Sadaghiani, 2010) and by 2020 millennials will consist of 35% of the workforce (ManpowerGroup, 2016).

To address these problems, aeronautical companies have tried to recruit more individuals to apply to be ATC's (BBC, 2018b). For example, one air traffic service company has created mini-games to determine whether an individual has the correct cognitive skills to become an ATC (NATS(a), n.d.). By playing these mini-games, individuals can quickly identify if they have the required skill set to be an ATC. The FAA is also looking to recruit, with thousands of ATCs being expected to be trained over the next decade to ensure safety and efficient air travel (Lane, 2018). Whilst recruitment is critical of air traffic organisations around the world, there approach to technology is hampering validation rates.

New technologies and procedures are difficult to develop due to the large number of development steps needed before they can become operational within an air traffic company (Haissig, 2013). These include the creation of new procedures, terminology, and a lengthy testing phase because of the safety critical nature of the industry. Technology is widely underutilised in the air traffic services industry, with the lack of technology adoption not only impacting recruitment and validation, but also safety (The Economist, 2016). Furthermore, supporting ATC trainees by providing self-paced digital learning tools can negate the need for trainees to make their own training tools and provide learning which is consistent and effective.

A large amount of companies at present are using one or several digital products to improve their business, whether that is tools to provide a better customer experience or data analysis (Schwertner, 2017). Digital transformation is difficult with Forbes finding that only 7% of 274 executives felt that they had "successfully used digital technologies to evolve their organization into truly digital businesses" (McKendrick, 2015). Whilst digital transformation is a challenge to businesses, it can be a powerful method to improve a company's productivity (Matt et al., 2015). By providing more effective training, costs can be decreased by reducing training time.

E-learning can reduce costs and foster behaviour change if it's well-received. Engaging e-learning content can be produced by utilising gamification techniques (Seixas et al., 2016) and applying formal user experience design practice (Lain and Aston, 2004). Gamification is a term used to describe the implementation of gaming techniques to non-gaming applications (Thom et al., 2012). The primary goal of Gamification is typically to help engage and motivate individuals to complete tasks. Using a multitude of different techniques such as scoring systems, leader boards, achievements, and rewards, the user is further engaged with the material. In training, the concept of the quiz is the most easily gamified interactive element with motivators such as point rewards serving to motivate the player and quizzes delivering the primary learning the player undertakes. By utilising gamification, both engagement and learning retention can be dramatically increased (Burke, 2014; Darejeh and Salim, 2016). Gamification techniques enhance self-learning tools with gamified workplaces being "self-improving, self-learning entities" (Oprescu et al., 2014).

UX (User Experience) is a modern field of design and is deeply associated with other areas of design and science such as HCI (Human Computer Interaction) and Ergonomics (Law et al., 2009). UX focuses on the accomplishment of a user tasks. Buley (2013) explains that "good" UX is reducing the friction between the task and the tool or application. Focusing on UX design often increases engagement, resulting in an increase in the time the user interacts with the application (Hart et al., 2012). Better user experiences are typically more engaging and younger generations are more receptive to activities such as video games and other digital media (Metz, 2011). By creating training applications which possess a high-quality user experience, we can engage individuals with their given training. With a younger generation of ATC trainees and a large sum of crucial information needing to be learned and retained, UX practice can play a key part in creating effective training programs.

This research is based on a joint academic-industry project to explore how ATC training could be improved by taking advantage of UX principles and gamification in digitalisation of training. The intention of the project was to conceptualise and prototype a range of potential solutions to assist with initial training of new recruits through to up-skilling packages for experienced staff. Thirty different concepts were created to explore different ways to improve training. These included an interactive application to teach ATCs visual scanning patterns, a learning tool to help in radar scenarios and an application to assist in the learning of the phonetic alphabet. From this pool, two digital prototypes were produced and then tested to determine if gamified and UX centric digital training applications would be beneficial to the air traffic industry. Data gathered from the usability tests will give us an insight into the utility of the app in providing flexible, portable and engaging learning tools.

LITERATURE REVIEW

This literature review we will examine the existing literature on UX, Gamification, and usability. We will also explore ATC

TABLE 1 | Studies which utilise the SUS to assess the usability of mobile applications.

Background	SUS score	Adjective rating	References
A mobile application designed to support injured service members	78.0	Good	Little J. et al., 2017
A mobile quiz application used in lectures to promote student participation and engagement	74.25	Good	Wang et al., 2008
A mobile application for recording symptoms and physical activity exertion	75.4	Good	Heinonen et al., 2012
An AI based context-aware mobile learning system designed to provide real-time training and support for medical cleaning staff	90.6	Excellent	Tortorella and Kinshuk, 2017
A mobile application designed to help with the training and practice of mindfulness	81.63	Excellent	Plaza García et al., 2017
A mobile game to teach users how to thwart phishing attacks	84	Excellent	Arachchilage et al., 2013
A mobile game-based application for myoelectric prosthesis training	75.2	Good	Winslow et al., 2018
A classroom management app	77.88 and 78.2	Good	Armstrong and Wilkinson, 2016

training and existing digital training applications for employees of air traffic control companies.

User Experience (UX) and Usability

In 2009 Law et al., concluded from a survey of 275 researchers and practitioners working within industry and academia that the definition of User Experience is dynamic and very much dependent on the context, they concluded that UX has origins in HCI (Human-Computer Interaction) and UCD (User-Centered Design) (Law et al., 2009). Don Norman coined the term User Experience (UX), defined as "a philosophy based on the needs and interest of the user, with an emphasis on making products usable and understandable." UCD is ultimately for the benefit of the user, creating experiences more pleasurable and usable (Norman, 1988).

McCarthy and Wright argue that "we don't just use technology, we live it" and that interacting with technology "involves us emotionally, intellectually, and sensually" (McCarthy and Wright, 2004). Interacting with technology is not a passive experience, it is an experience that engages users on an emotional and intellectual level. They further state that those "who design, use, and evaluate interactive systems need to be able to understand and analyse people's felt experience with technology." Having an understanding of an individual's experiences with a user experience is therefore paramount to building something which is usable for users. By conducting usability tests, we can gain an insight into the emotions our participants feel when interacting with the applications produced.

Usability, "the capability to be used by humans easily and effectively" (Shackel and Richardson, 1991) can be measured in respect of: ergonomics and user comfort; effectiveness, learnability, flexible, and attitude. Dowell and Long define usability as "the level of behavioural resource costs demanded of the user." where the process of work is both a physical and mental behaviours and working is a cost on those behaviour (Dowell and Long, 1998). In a long term study using computer games, usability defects negatively correlated with user engagement measured by playtime and playability (Febretti and Garzotto, 2009). The study of usability and user experience is the focus on creating experiences and services more suited toward the user. How the user thinks and feels is at the centre of user experience and is related to the core usability factors previously mentioned. Focusing on UCD, usability and UX principles when designing experiences is important to engage users. Usability has been assessed via a questionnaire with responses linked to a System Usability Scale (SUS) (Brooke, 1996). This results in a score between 0 and 100 with score ranges linked to usability adjectives e.g., poor through to excellent. **Table 1** outlines studies focusing on mobile training applications that have used SUS to assess the usability of the applications.

Air Traffic Controller Training

ATC training initially involves external training (also described as college-based training) in which learning content is presented through practical and theory-based classes. After a year to 18 months, trainees are then sent to do operational training. Finally, after operation training, trainees are given their ATC license and can begin operating as an ATC (NATS(c), n.d.).

Cordeil et al. (2016) outline the challenges of air traffic workers and digital immersive solutions to support air traffic staff. Air traffic staff operate in an incredibly complex environment, having to manage and analyse a large amount of critical data in real-time. They suggest potential immersive solutions such as Augmented Paper Strips to enhance the traditional stripboard method of tracking flights as well as a VR (Virtual Reality) immersive environment designed for long distance collaboration between air traffic controller experts (Cordeil et al., 2016).

Perry et al. (2017) also frames the challenge of ATC training as the application of complex information at a fast rate. Their hypothesis states that a digital learning tool may increase student learning and success rates. The paper outlines a digital tool called "Swivl" which allows students to refer back to simulations and videos they had previously seen in a classroom setting (Perry et al., 2017). In an associated research article Little A. et al. (2017) outlines that trainees in the field of air traffic control have no way to learn what is taught via simulations outside of the classroom.

In 1997, a panel was formed by the FAA (Federal Aviation Authority) to investigate Human Factors in Air Traffic Control. They outlined problems with ATC training including the fact that many trained ATCs do not have sufficient time to teach trainees or receive refresher training themselves. The panel recommended *Flight to the Future: Human Factors in Air Traffic Control* to use computer simulations to improve controller training (National Research Council, 1997).

ATC trainings' current problems have existed for a long period of time. These challenges include the complexity of the required learning for ATCs, full-time controllers being utilized as educators and the difficult environment in which ATCs need to work in. Reviewing the research suggests that ATC training consists of the use of computer simulation in combination with procedural based traditional teaching programmes. Digital learning tools are less utilised, especially the use of digital tools within self-learning scenarios. Research suggests that digital learning can strengthen a students' learning experience and enhance student engagement (Sousa et al., 2017). The current state of air traffic controller training consists of very few digital tools and digitally focused training programs, yet both have the ability to facilitate learning. In the forthcoming sections, we highlight two applications that have been created to evaluate the use of digital tools for learning.

METHODOLOGY

The two prototypes have been designed with an emphasis on producing a succinct user experience alongside gamified elements to improve and measure engagement. The questions the two prototypes address are:

- What elements of the applications contribute to higher or lower usability scores?
- What changes can be made to both applications to improve their usability?
- Are both applications perceived to be fit for purpose to train the air traffic services workforce?

The study began with the investigation of the needs of the target users. By shadowing staff members working at an air traffic services organisation, we understood the challenges the target user faced. Initial concepts were influenced by the problems encountered by staff members. Workshops were undertaken to gain a further insight into user's needs. The workshops found that training is often a stressful experience for trainees and that there is room for improvement regarding the self-learning material given to trainee ATCs. Training material for example is often cumbersome with trainees given large books to review and comprehend. After conducting a game jam (Kultima, 2015) to explore preferred digital platforms/products e.g., PC, VR, and Augmented Reality (AR) applications, mobile applications were identified as being the platform of choice. Two concepts were prototyped—Location Indicators (LI) and The Aircraft Control Positions Operator (ACPO) Starter Pack and evaluated with respect to UX and Usability.

Development Process

LI and the ACPO Starter Pack were designed using an agile approach allowing for iterative development. Both app designs were created through a UCD approach. Agile is an iterative approach to software development with the aim of delivering quality software in smaller increments (Fowler and Highsmith, 2001). By delivering functional software to the customer in a shorter time, the direction of a software project can change frequently. This is especially important in the case of game projects due to the complex nature of the medium. Game developers must balance the user needs and development complexity with good "game feel" (Swink, 2009). Utilising a specific agile approach, such as Scrum, has become a common approach in the games industry. Scrum is an agile methodology for software development which utilises sprints to produce smaller increments of work (Schwaber and Beedle, 2002). Sprints typically involve a set number of tasks to be completed in a short space of time, usually a week or 2 weeks. This approach is very similar to a "lifecycle model for interaction design" (Preece et al., 2002).

1. Identify needs/establish requirements
2. Re(design)
3. Build an interactive version
4. Evaluate

This is a repeating process which focuses on user needs, creating prototypes frequently and then redesigning and reiterating based on user evaluation. This model is similar to an agile approach when developing prototypes, and for the development of both LI and the ACPO Starter Pack, this process was used to focus on meeting user needs.

Both LI and the ACPO Starter Pack were developed iteratively in short increments, taking into account feedback from trainee ATCs. This is effective as Abras et al. (2004) argues it "is only through feedback collected in an interactive iterative process involving users that products can be refined." The involvement of the trainee/employee is central to facilitating a UCD approach. Abras et al. (2004) defines UCD as a "broad term to describe design processes in which end-users influence how a design takes shape." The prototypes of both the ACPO Starter Pack and LI allowed for the usability study to be built into the process and therefore is a key pillar of the research of this paper. The creation of prototypes is necessary as Preece et al. (2002) points out in *Interaction Design: Beyond Human-Computer Interaction*, "For users to evaluate the design of an interactive product effectively, designers must prototype their ideas."

Rams outlines ten principles which were considered central when designing the two applications (de Jong et al., 2017). The Location Indicators design has been made to be as simple as possible with the core interactions being to swipe left and right, this design decision was made in line with Rams principle of "good design is as little design as possible" (Lovell, 2011). This is echoed in the aesthetics of the application, with very little colour being used and minimal screen clutter. This provides contrast when green and red colours are used as the main indicators if the player remember or did not remember the location. One of Rams principles is that "Good design makes a product useful" which is something echoed in his work as a designer who valued "honest design" which focuses on simplicity (de Jong et al., 2017). This is a core principle in the design of LI and the ACPO Starter Pack.

Design of Location Indicators Mobile Application

LI focuses on teaching ICAO (International Civil Aviation Organization) Codes. ICAO Codes are four-letter codes that are designated to each airport around the world. ICAO Codes are referenced when ATCs and pilots communicate with each other via radio to signify the airports they are departing from or arriving to. ICAO Codes are classified geographically with certain letters signifying different regions, countries and continents. The first letter identifies the continent, the second letter identifies the country whilst the remaining two letters are used to identify the airport. Larger countries such as the USA, Australia, and Canada are the exception to this rule and have only one letter to represent a country. In some instances, the third letter may identify a certain region in a country. For example, the ICAO Code EGBB signifies Birmingham Airport. The "E" represents the continent, Europe, the "G" represents the country, the United Kingdom and the "B" represents a certain region within the country. Airports and airfields which are in relatively close proximity to Birmingham such as Derby and Leicester also have the letter "B" as their third letter. This convention is also present with airfields located in London and Scotland with Scottish airfields being represented with "P" as the third letter and London airfields being represented with a "L" as a third letter. It is important to note that this convention is not a rule that is followed comprehensively throughout ICAO Codes and is merely a guide (Benedikz, 2019). This demonstrates the complexity in learning these codes for any prospective ATC.

Learning ICAO Codes is required for ATCs and is a challenging aspects of initial ATC training. To learn these codes, trainees employ various training methods to learn and retain these codes. The main method is group learning by testing their fellow trainee's knowledge of ICAO codes, with the use of flashcards. Trainees use this method consistently, but anecdotal evidence suggests that this method may cause issues due to misinformation. Flashcards are typically self-made by trainees and there is the opportunity for flashcards to be made inaccurately with either spelling or letter placement errors being made or ICAO codes being designated to the wrong airport.

Due to the large number of airports and other aeronautical facilities that possess ICAO codes, it is impossible for a single individual to learn all the codes that exist globally. ATCs regularly are assigned to different sectors of airspace (NATS(b), n.d.), ATCs therefore are required to learn the ICAO codes for all the airfields and aviation facilities for that sector (NATS(b), n.d.). This creates confusion due to trainees unsure if they are required to learn certain codes and their locations. To learn different regions, flashcards are often made by trainees to learn the codes for the airports and aviation facilities in their assigned sector. Flashcards are cumbersome to carry around, even if the trainee is learning codes from a certain region there are still many flashcards that a trainee is required to carry around. A portable learning tool which can be flexible enough to provide sector specific training to trainees seemed to be an obvious potential solution.

LI was designed to be a portable, flexible alternative which guaranteed consistency (Wobbrock et al., 2009). LI was designed to mimic a trainee using flashcards to learn different ICAO codes and their corresponding locations thus providing familiarity with the learning method. LI further utilises gamification by providing scoring systems to each play session, this allows for instant feedback to be provided to the user. The feedback shows the user immediately what codes they got wrong. This should provide motivation to LI users as they are able to identify what codes they need to study. The user experience is designed to be succinct and allows for play sessions with differing codes and for differing amounts of time.

LI has been created for iOS devices primarily but due to technical simplicity of the application and being created within the game engine Unity (Unity, n.d.), the application is able to be played on Android devices or ported to other platforms. Unity is an incredibly popular game engine that has an extensive knowledge base, contributed to by both industry developers, as well as the creators of the game engine. It is a fast and flexible tool which allows for the development of mobile prototypes and smaller scale software projects due to its component system. Whilst standard libraries are usable, Unity is widely used by the games industry and is a more user-friendly game engine in comparison to standard libraries. Due to its usability, Unity is suitable for both game designers and developers. For creating a refined user experience via rapid prototyping, as well as the having the ability to pivot to different platforms, Unity was the clear choice when choosing the software to create LI with. Being playable on Smartphones, trainees have a portable training tool that can fit into their pocket and played anywhere and at any time.

It is essential that ATCs continue to retain the knowledge of ICAO codes. LI is designed to function as a retention tool for ATCs. ATCs are obligated to retain the knowledge of the ICAO codes of the airports within the sectors they are assigned to. This knowledge retention process is supported by the different regions the user can choose from within LI. Users can choose different regions which contain ICAO codes from a specific sector. LI provides the option to use LI for differing amounts of time. This means users can use any incidental time they may have working to retain knowledge of specific ICAO codes. The usage of mobile phones is widespread with 85% of people owning mobile phones (Deloitte, 2017). LI takes advantage of this by providing a mobile experience that is portable and allows users to play for differing amounts of time.

User Flow

LI first leads the user through an onboarding and tutorial section in which users are introduced to the game's interactions. Once the tutorial is completed, the game puts the player into the title screen (see **Figure 1A**) wherein they can start the game by pressing START or go into the settings page (see **Figure 1B**) by swiping

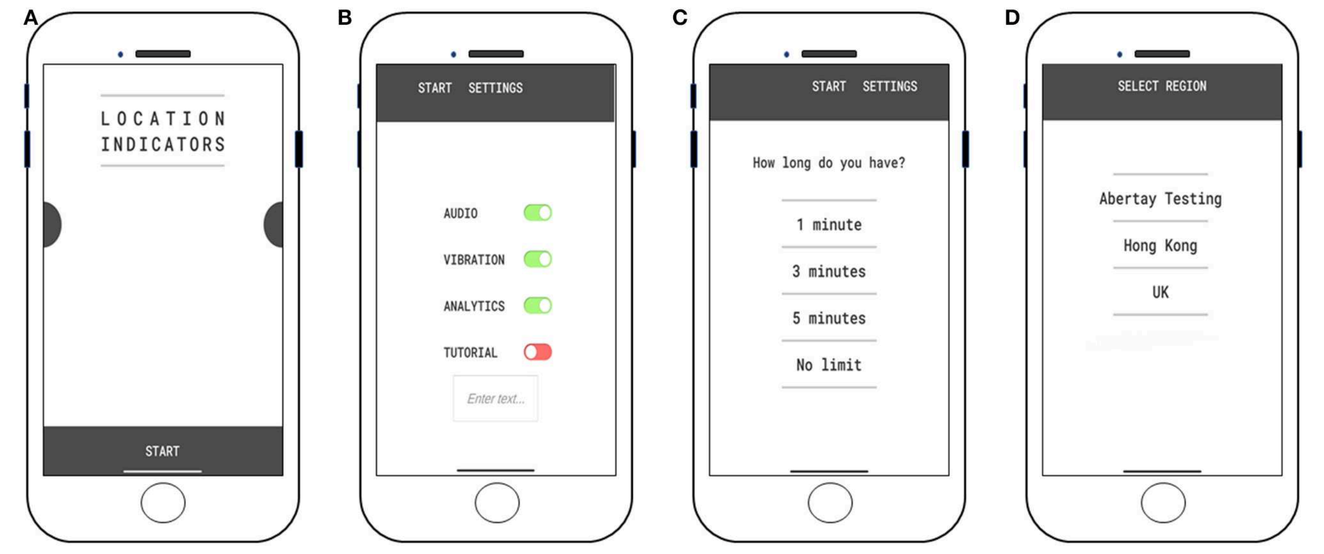


FIGURE 1 | (A) Title screen of Location Indicators. **(B)** Settings screen of Location Indicators. **(C)** "Play time" screen of Location Indicators. **(D)** "Select region" screen of Location Indicators.

FIGURE 2 | "ICAO Code" screen of Location Indicators.

left. In the settings menu the player can turn the audio and vibration on and off and can also turn the tutorial back on. By turning the tutorial back on, the player can replay the tutorial when they next start the game.

Users can start interacting with the gameplay by defining how long they want to play for (see **Figure 1C**). Users can either play the game for 1, 3 or 5 min sessions or request to play without any limit by selecting "no limit." After selecting their desired playtime they have a selection of regions (see **Figure 1D**) to choose from, defined by the database manager. Once the user has defined their play time and region they are taken into the game. Users are presented with ICAO codes in the middle of their screen (see **Figure 2**) and can either swipe up to end their session (see **Figure 3A**) or swipe down to reveal the answer (see **Figure 3B**). After revealing the answer, the player can swipe left to mark that they do remember the code (see **Figure 4**) or swipe right to signify that they don't remember the code (see **Figure 5**). After the time user defined in the "play time" screen has run out or the user manually ends their session by swiping up and press the end button, they are displayed with the results of their session (see **Figure 6**). The results display what codes the user remembered or did not remember as well as their accuracy displayed as a percentage. After seeing their results, the user can press END and return to the title screen, go into another game or enter the settings menu. The whole experience is designed to be as simple as possible with the interactions and gameplay itself being very simple to learn. LI seeks to provide a more effective training tool for trainees ATCs to learn ICAO codes.

Design of the ACPO Starter Pack Mobile Application

Aircraft Control Positions Operators (ACPOs) are individuals who act as "pseudo-pilots" within simulators. An ACPOs job is to assist in the training of ATCs within simulators. ACPOs give trainee ATCs the opportunity to gain experience interacting with individuals who possess the same knowledge as pilots. ACPOs are therefore required to act in the simulators the same way that pilots would in the real world. ACPOs should give the same response to trainees ATCs when communicating. As ATC training simulates high risk scenarios, ACPOs need to act appropriately in response to those scenarios. Working with ACPOs, ATCs have someone to send commands to and communicate with, the core aspect of an ATCs job.

ACPOs are required to learn specific aspects of what a pilot is required to learn. ACPOs learn the basic requirements of their job

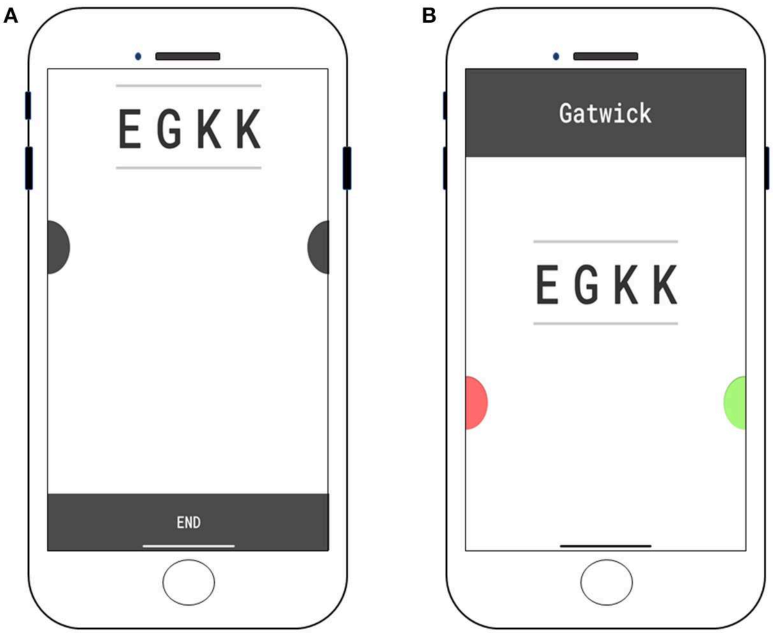


FIGURE 3 | (A) "Swipe Up" screen of Location Indicators. **(B)** "Reveal" screen of Location Indicators.

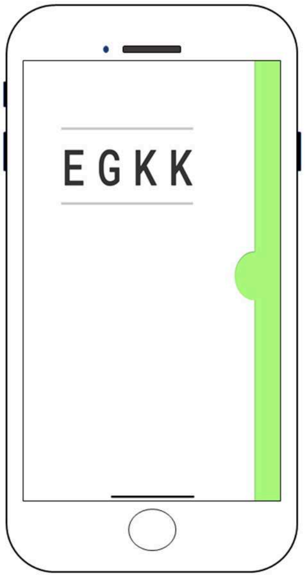


FIGURE 4 | "I do remember" screen of Location Indicators.

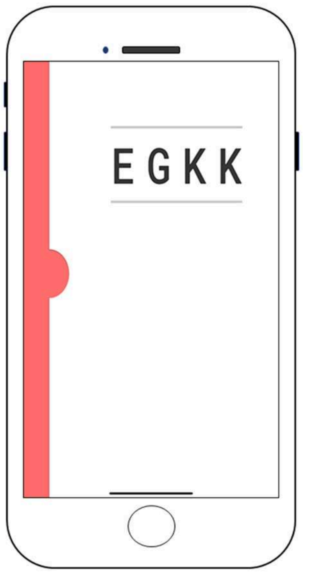


FIGURE 5 | "I don't remember" screen of Location Indicators.

by reading and referencing a document called the ACPO Starter Pack. The ACPO Starter Pack comes in the form of a document, with the majority of the content being written word with very few visuals. The current ACPO Starter Pack is seen as non-engaging and hard to digest, therefore a digital gamified alternative was identified as a potentially better tool to engage ACPOs.

The ACPO Starter Pack is a digitised version of the existing starter pack. The ACPO Starter Pack is a mobile application that has been developed for Apple iPad devices. The app contains information regarding the usage of the Phonetic Alphabet (see **Figure 7**), background information on how aircraft navigate the skies, what happens when aircraft travel across different sectors (see **Figure 8**) and how airspace blocks work (see **Figure 9**).

Each page covers a different topic, contain text on relevant subject areas as well as images and animations. The ACPO Starter Pack presents a steep learning challenge to trainees due

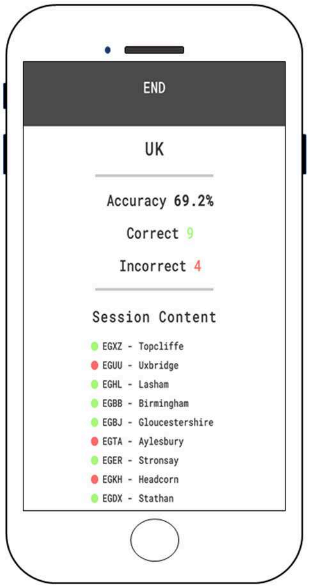


FIGURE 6 | Results screen of Location Indicators.

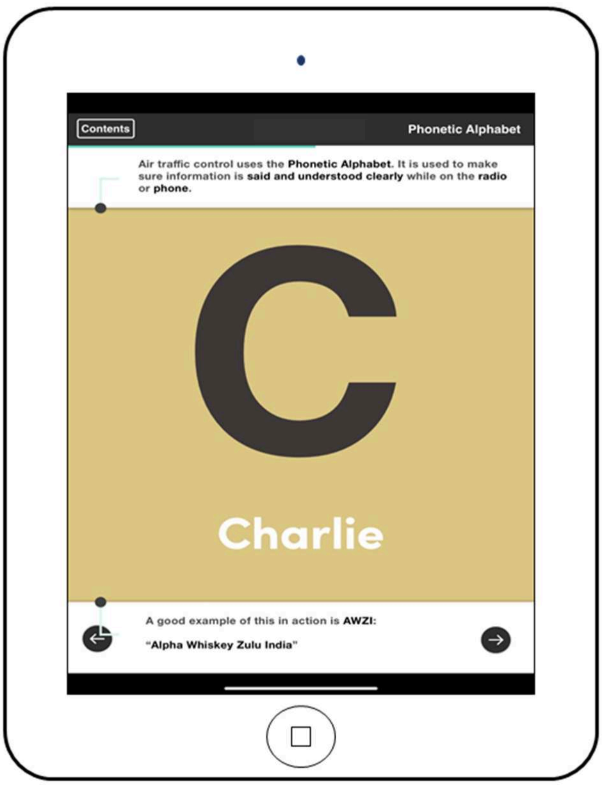


FIGURE 7 | "Phonetic Alphabet" screen of the ACPO starter pack.

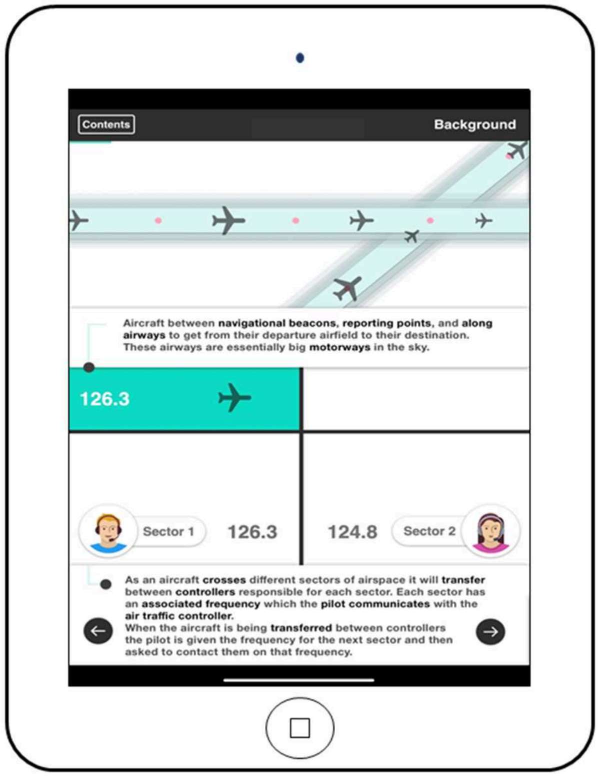


FIGURE 8 | "Background" screen of the ACPO starter pack.

to the complex information that trainees are required to learn. The ACPO Starter Pack application is therefore designed with a focus on the user's experience, with animations and visuals being utilised to explain complex concepts. Users can navigate each screen via two arrow icons at the bottom of the screen, which either go forward or back a page when interacted with. Users can also select the "Contents" button which brings them to the beginning contents page where they can select any page to go to.

The decision to include relevant animations, images, icons and other resources as alternatives to the standard material was mainly to present the learning content in a more engaging way but also to break up blocks of text. To ensure material was accurate and was consistent with the learning content in the traditional training, an expert in the company that the research team is collaborating with, reviewed the material. Material was iterated on to ensure that it was equivalent to the content of the traditional training tools. The prototype was created over a period of ∼3 months and iterated on for and after informal user tests with trainers over several more months. Design decisions were again influenced by the works of Dieter Rams and their 10 principles for good design previously referenced in the "Development Process" section.

The ACPO Starter Pack was created in the software program Principle. Principle is a design tool which facilitates the creation of interactive user interfaces, allowing for the quick creation of multi-screen apps. Principle is incredibly lightweight allowing for rapid prototyping and the implementation of new content and alteration of existing content based on user feedback. It allows for instant deployment to iOS devices and can be shared to multiple stakeholders quickly for review. Using principle allowed

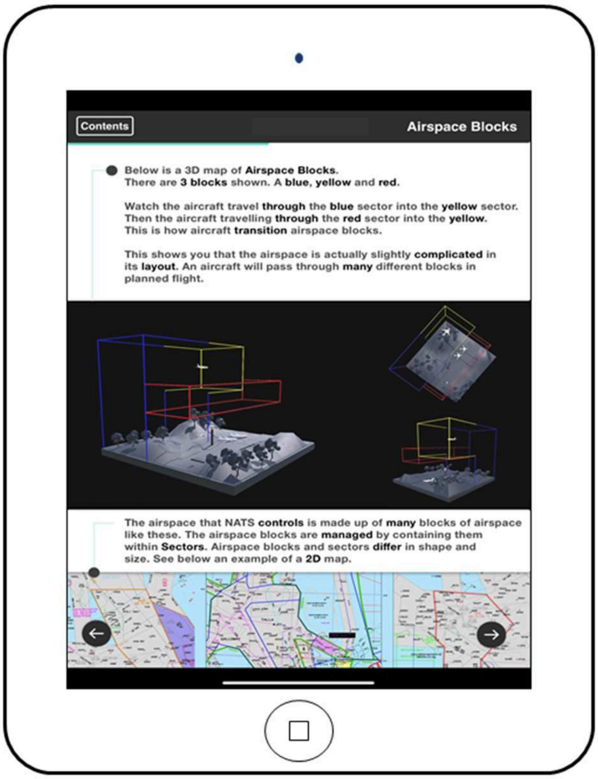


FIGURE 9 | "Airspace Blocks" screen of the ACPO starter pack.

us to garner feedback by being able to quickly present the ACPO Starter Pack to testing participants through iOS devices easily.

The ACPO Starter Pack has been designed to be a more engaging alternative to the existing paper-based starter pack. The application was also designed with younger trainees in mind with a potential preference for digital technology over more traditional methods, and designed to better engage and teach more effectively with animations and images being used to teach complex concepts easier.

Design of Usability Study

The study was designed to obtain both qualitative and quantitative data focused on the usability of the two applications. Qualitative data was gathered from participants to support the quantitative usability scores and gather potential changes to both applications to improve their usability. Feedback on the usability of each application were recorded via a SUS questionnaire, recorded observations during usability tests and individual conversations with the research team. One usability test session included a group discussion. By gathering quantitative and qualitative data we can produce thematic analyses for each usability session to gain an understanding of the overall usability of each application.

A total of nine usability test sessions have been undertaken at four different locations. In total, four sessions had five participants, four sessions had four participants, and one session had three participants with thirty-nine participants in total. Participants were either undergraduates or working within the ATC sector. The undergraduate participants comprised 100% of six of the usability studies, accounting for 25 of the 39 participants. Sixteen participants of the 25 graduates who participated in the usability testing were enrolled in a game design and production course. Three of the participants were enrolled in a computer games technology course, one participant was enrolled in a health course, one participant was enrolled in a business course with the rest of the student participants enrolled in computing based courses. The participants who were employed within the aeronautical industry had experience ranging from 4 months to 22 years and accounted for 14 out of the 39 participants. The aeronautical industry employees represented a group of individuals who have previously undertaken self-learning with the traditional training equivalents of the two applications. The undergraduate students represent individuals who would be prospective applicants to work in the air traffic services industry, this user group was necessary to test due to the wide demographic of potential ATC trainees. Having qualitative results regarding the usability of the two applications from both user groups was extremely valuable. The undergraduate students all had no prior knowledge of air traffic control training except for one participant who had experience in air traffic services before enrolling (see **Table A1** in the Supplementary Material).

Usability testing sessions involved a brief introduction to the project and the two applications, with the researcher briefing participants on the subject matter for a short introduction, no information was given about the design of the application nor how to operate the applications. Participants were given 5–10 min "playtime" with an application in which they were encouraged to explore the two applications and to experience all of the content within both applications. In regard to playtime, the participants of five of the usability sessions were made up of individuals who spent a considerable amount of time with the applications as part of a longitudinal study. After interacting with the applications, participants were given usability questionnaires to fill out and told not to discuss any of their initial feelings about the two application. We purposefully instructed participants to not share their thoughts with other participants or with the researchers present due to not wanting to impact other participant's opinions. After the usability questionnaires were completed and the researcher gathered additional feedback from participants through conversation. This process was repeated with the other application. For the participants who had spent a considerable amount of time with the two application as part of a longitudinal study, usability questionnaires were given out at the end of their time using both applications. These users also gave feedback on a one to one basis at the end of their time using both applications.

Participants were given two copies of the usability questionnaire document, one for the ACPO Starter Pack and one for LI. The two copies were identical and the two documents asked the same questions. The usability document asked the participants for demographic data, asking for age and name with length of employment, course and year studying being gathered in the one to one or group discussion sessions with participants and the researcher. The questionnaire was comprised of ten questions based on a Likert scale of a value from 1 to 5 being given to each statement. The odd numbered statements were positive attitude questions whilst the odd were negative attitude. Calculating the SUS involves subtracting 1 from the user's answers to odd statements and subtract the value of the answers of even statements by 5. After this process, responses will range from 0 to 4, after adding responses and multiplying by 2.5, we achieved our average SUS scores (Usability.gov, n.d.). The results of each usability assessment allowed a usability score to be associated to each participant for each application. With this quantitative data we can produce a succinct and aggregate usability value for each application. The SUS is a globally recognised usability scale which provides a reliable method of assessing usability whilst being scalable and easy to use (Brooke, 1996). The SUS is beneficial to studies with smaller sample sizes, which makes it ideal for this study (Usability.gov, n.d.).

Participants were given open-ended questions as the final section of the usability questionnaires. These open-ended questions served to generate more qualitative data on the usability and design of each application and to gather potential design changes to both applications. The answers to these open-ended questions also added to the thematic analysis alongside the collected feedback from participants after completing the usability questionnaire. Qualitative data gathered from discussion between researchers and participants as well as the usability questionnaires would support usability scores and give more insight to each application's usability scores.

After completing the usability questionnaires, participants were interviewed within a group or on a one to one basis with the researcher. Questions were asked in order to gain more informal qualitative feedback on the design and usability of the two applications. Participants were given prompts so that qualitative feedback could be collected for thematic analysis. Participants were encouraged to suggest any potential changes to the two applications to improve their usability. Participants feedback was gathered and contributed to a thematic analysis.

By undertaking these usability test sessions, we measured the engagement users had with each application and evaluated the overall design of each application from a user experience perspective. We can also conclude how "fit for purpose" both applications are for the training of air traffic service staff and gained feedback from participants on how to improve the usability of each application. We can then reflect upon the usability tests to conclude what elements make for an engaging digital training application for the air traffic industry.

RESULTS

Quantitative Analysis

Usability assessments were used to quantify participant's reactions to the usability of each application based on SUS methodology. Over seven of the nine usability sessions, usability scores were averaged ($n = 39$, see **Figure 10**). LI scored an average usability score of 80 and The ACPO Starter Pack scored an average usability score of 74 (see **Figure 11** for SUS score variability depicted using boxplots). On the SUS scale, this is defined as "Good" with a letter grade of B (Usability.gov, n.d.). Two usability testing sessions were disregarded due to a logistical error.

A thematic analysis (see **Appendix** in Supplementary Material) was undertaken based on the qualitative feedback received via the open-ended questions on the usability questionnaires and the discussions between researcher and participants. A thematic analysis has been undertaken for all nine of the usability testing sessions. The common thematic trends seen through all usability testing sessions were identified and are discussed in the following sections.

Qualitative Analysis

Location Indicators: Themed Feedback

Unclear aim

Participants when first presented with LI were unsure of the applications aim. This misunderstanding of the purpose of the application directly impacted the way in which users initially attempted to interact with the application. Many participants were confused with what they needed to do when using the application, with one participant thinking that they needed to stipulate if the code and the location matched on a true or false basis. "Is this for true or false?" was a common piece of feedback throughout usability sessions. One participant commented that "I was thinking I had to type something in" and many participants wrongly assumed the LI application as a "testing" tool rather than a tool for "learning" and "assessment." This meant that participants felt that that "people will cheat themselves" as they only had to tell the game they remembered a code or not. This extended to participants feeling they needed more explanation, one participant is quoted in saying "generally good application, needs more explanation at the start though." With participants regularly asking "What do I need to do?" when first presented with LI.

Simplicity

Simplicity was a core theme throughout the feedback received. Participants commented on the colour choices of red and green for remembering and not remembering as a good visual design decision. A total of seven participants remarked that simplicity was LI's key value. Participants particularly liked the simplicity of the cards which made it easier for them to focus on the task at hand.

Fit for purpose and well-designed learning tool

Participants were favourable toward LI, with participants speaking positively of the design of the application. Participants stated that the LI app was an accessible, simplistic tool and participants expressed that "the app would work well for the intended purpose" and the application was "quick to pick up when you have some spare time." Many participants commented they had gained additional knowledge of ICAO codes even after using the application briefly, with one participant stating that "With no previous knowledge I was quickly able to recognise some of the names as they appeared more frequently."

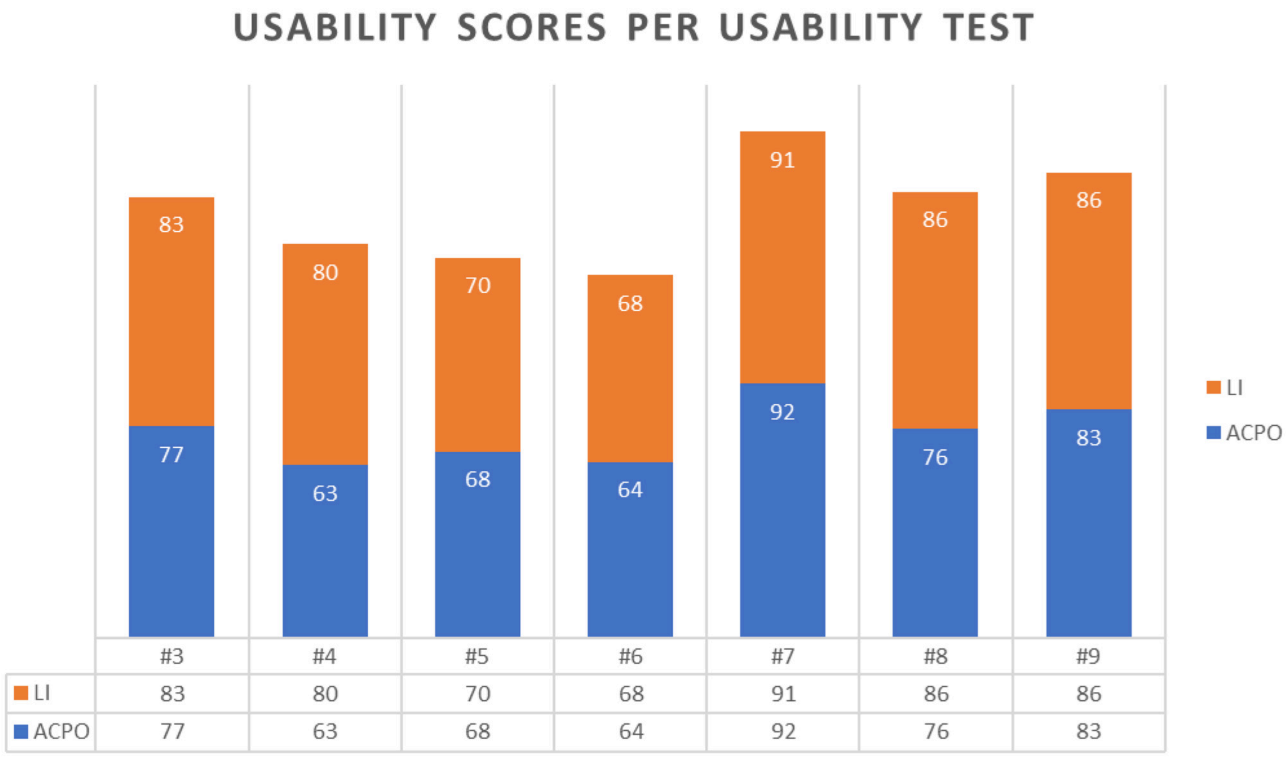


FIGURE 10 | Usability scores per usability test.

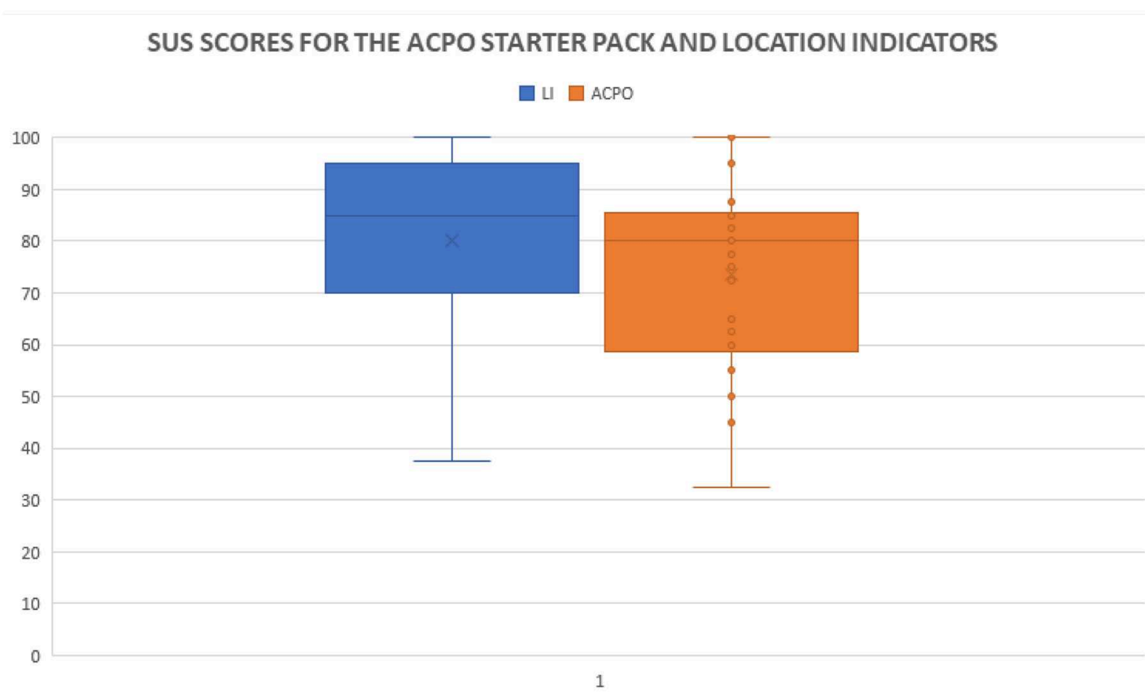


FIGURE 11 | Average usability scores for each application.

Participants felt that LI was fit for purpose at teaching trainees and could be used in a real-world capacity. Participants felt that the application was very easy to use and that would be suitable for trainees of a younger age. Some participants expressed some mixed response toward LI as an effective learning tool. This was due to participants not feeling LI was optimal for reinforcing their knowledge, due to the random nature of the codes appearing.

Inappropriate codes being displayed

Participants from usability sessions that contained experienced air traffic staff had many participants comment on the codes implemented within LI. There were multiple suggestions from participants that LI should contain more commonly known codes. Many ICAO codes that were implemented within the application were deemed too specific and were only required knowledge when trainees undertake sector-specific training. Participants felt that these low scores could potentially discourage further use of LI. Participants expressed that there should be a stronger focus on prioritising major airports as well as codes for sectors that the trainees would be managing in the future.

Participants expressed negativity toward the codes being repeated throughout a game session and specifically that too many uncommon ICAO codes were repeated with one participant stating that there was "too much repetition". Many participants highlighted the typos seen within LI. Participants felt that "map information would be beneficial" with other participants echoing the use of maps to enhance learning. Participants felt that this would help guide the learning process and make the information less abstract in nature. Many participants expressed that they didn't like "The fact that all location indicators are given at once" with this being one participant's least favourite element of the game. One participant commented that "Not knowing how large the list of Location Indicators could be, I felt discouraged."

The ACPO Starter Pack: Themed Feedback

Presentation of information

A large amount of the feedback was focused on the fact that participants felt that the application was busy, with one participant saying that "It (ACPO Starter Pack) felt rushed, too much use of graphics and text." Participants felt that the application contained too many images, gifs, and videos per page, with some pages having too much on-screen movement. This made it hard for participants to concentrate on the learning content as "There is so much information condensed into a short experience." This made pages too cluttered leading to participants feeling unfocused, unable to digest information and in one case one participant almost felt motion sick. Participants found it hard to focus on certain pages and would even just skip pages without reading through the contents. Participants did this as they were not able to focus on certain pages or felt uncomfortable doing so. The combination of horizontal and vertical animation was a pain point for a few participants as they felt it was very distracting and uncomfortable. Participants felt that various images which were in LI were unnecessary and gave the user no educational benefit.

Improvement over existing training

Participants generally held a positive feeling toward the two applications through the usability sessions. Participants indicated that they particularly felt that the ACPO Starter Pack was easy to understand and had a good layout with a clean design. The level of engagement was commented on with one participant saying that the ACPO Starter Pack was "Dynamic and subsequently more engaging." Participants felt that the application would be very useful, this was particularly highlighted by participants experienced in air traffic controller training. Participants regularly stated that the ACPO Starter Pack was a significant improvement over current training materials. Participants with experience in undertaking ATC training consistently commented that the ACPO Starter Pack on iPad was a "huge upgrade" to the existing training equivalent.

Participants commented on the use of visuals as one of the main highlights of the ACPO Starter Pack. The simplification of complex concepts via the use of simple language as well as visual media was commented on various times by participants. A number of participants who had spent considerable amount of time with application felt that wasn't much of a difference between the paper-based training and the digital stating that they "Felt like the paper learning just stuck onto an app." These comments are possibly due to the core content of the existing ACPO Starter Pack being identical to the content of the mobile application.

Interaction with application

Participants frequently commented on the interactivity they had with the ACPO Starter Pack, focusing on the sound and sub-pages within the application. The comments focused on the inconsistency within the application regarding what can be interacted within the ACPO Starter Pack and what cannot be interacted with. Participants are quoted as saying "some maps could not be clicked on" and "Some links aren't clear you can touch to see other stuff." Participants also commented on the fact that some buttons aren't working and other elements within the application were not interactive when they should have been and there was a lack of interactivity with "Some elements should interact, such as the phonetic alphabet." This was commented upon from various participants in the usability sessions consisting of users who had used the two applications for a longer period of time. One participant commented that they perceived the lack of interactivity "made it difficult to retain info." One participant echoed previous feedback by commenting that "It could be more interactive to make it feel more as an actual app rather than a textbook." The interactivity itself being the reason why comparisons were drawn between textbook and application.

Sound was commented as jarring within the application due to audio cues being used inconsistently within the application. Audio was also considered to be unnecessary and did not require to be implemented within the application as it was not perceived to enhance learning in any way.

Many participants commented on the issues they had with navigating throughout the application. Participants stated that the navigation of content could be improved, as the forward arrow button offered an overly linear way of moving through content (only allowing forwards or backwards navigation). Participants said that navigation would be improved if a contents page could be accessed to directly access certain sections. Some participants disagreed and felt that the ACPO Starter Pack was "very easy to navigate."

Many participants suggested the use of a play button to play and pause certain animations with one participant stating that "The gifs [videos] were a bit much—I'd have liked them to play once and then need to be tapped to play again." One participant agreed about the animations being distracting suggesting "the animations could be paused/ played."

The ACPO Starter Pack vs. Location Indicators

The final usability results suggest that there is no clear preference from users regarding the two apps. The usability scores when averaged are near identical at 80 and 74 with the qualitative feedback from participants having a positive reaction to both LI and The ACPO Starter Pack. Clear trends were easily identifiable through the qualitative feedback and the suggested changes participants would like to see are apparent. The authors believe that further research would be beneficial to explore the use of UX methodology vs. the use of gamification principles. This further work would benefit the future design of training applications and other e-learning content.

DISCUSSION

The ACPO Starter Pack had an average usability score of 74 whilst LI had an average usability score of 80. This is in line with other studies that prove well-designed mobile applications are effective learning tools (Sánchez et al., 2008; Su et al., 2009; Sánchez and Espinoza, 2011; Tsuei et al., 2013; Ahmad and Azahari, 2015; Armstrong and Wilkinson, 2016). The SUS scores seen in **Table 1** ranged from 74.25 to 90.6 for an average of 79.4.

Participants found the lack of progression an issue, as they had no method of assessing their own learning. Participants suggested the use of summaries, evaluation elements such as quizzes to evaluate their knowledge of the content as well as reinforce their own learning. Other studies have utilised gamification mechanisms such as quizzes, leader boards, badges, achievements and avatars (Gordillo et al., 2013; Su and Cheng, 2014; Ngan et al., 2016; Pechenkina et al., 2017) and have produced satisfactory results with respect to usability. The Zeigarnik effect states that unfinished tasks are far better remembered that finished tasks (Zeigarnik, 1938) whilst Nunes and Drèze (2006) explains "a phenomenon whereby people provided with artificial advancement toward a goal exhibit greater persistence toward reaching the goal." Both Nunes and Drèze (2006) and Zeigarnik speak about matters regarding the motivation of individuals and that when there exists a semblance of progression, there is an increase in engagement. A progression tracker is an example of the Zeigarnik effect where unfinished tasks are displayed to users in order to motivate. Whilst functioning as an e-book may provide some value as a learning tool, implementing evaluation elements and breaking the section up, may produce an engaging flow and ensure trainees are learning the contents of the ACPO starter pack.

Participants initially found LI's interactivity unintuitive. Participants stated that swiping should be reversed due to their exposure to mobile dating apps (David and Cambre, 2016). This is an example of an "affordance" which Norman (1988) outlines as "the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used." In layman's terms the design of objects influences the perceived interaction (Norman, 1988). Referred to as "Swiping ambiguity" by Budiu and Nielsen (2010), issues with the swiping interaction is common for of iPad apps and websites, where the swipe interaction is both inconsistent and confusing (Budiu and Nielsen, 2010). Miller et al. (2017) conducted a usability study into an iPad health app for the vulnerable population. They found that participants had little to no problems interacting with the application. They explained this was due to closely considering their users and the fact they possessed "low computer literacy." They purposefully created a simple interface, using large buttons for navigation which the user tapped (Miller et al., 2017). Despite LI being designed with simplicity in mind, Miller et al. (2017) presents an even simpler interaction solution which caused little usability issues with users. A study into gestured interaction design saw that commonly people use tapping or swiping. The same study saw 70% of participants using tapping gestures and 45% using swiping gestures (Leitão, 2012). Whilst swiping is natural to most users, the act of tapping is even more natural to users with the act of swiping having significant ambiguity.

Participants were also unsure of what they were signifying when interacting with LI and what the application wanted them to indicate. In LI, users are presented with ICAO codes, after they reveal the answer the user then needs to stipulate if they either "remember" or "don't remember" the code. Several participants were not initially aware of how the application worked, with many feeling that this was not naturally obvious to users. Participants initially thought they had to stipulate to the game if it was true or false that the ICAO code designated the displayed location. Participants expressed that quizzing participants by giving them multiple locations to choose from, was a more preferred method of learning. Studies that focus on the effectiveness of digital flashcard mobile applications present different methods of how to interact with flashcards digitally. Previous studies have simply used digital buttons to reveal the answer and to signify if the user remembered or did not remember the answer. A previous study used large buttons that simply say "check," "correct," and "incorrect" (Edge et al., 2012). Whilst theoretically, not being suited to a universal audience, another study uses buttons with ticks and crosses inside them and which are coloured red and green for a more universal approach (Edge et al., 2011). Both studies reported study participants having no problems interacting with the applications (Edge et al., 2011, 2012). A study which used the application Study Blue also reported no problems with the commercial application presenting flashcards which users tapped on to reveal the answer. Users of Study Blue swipe the card left or right to signify if the user knows the answer or not (Burgess and Murray, 2014). This is also seen in the popular self-learning application Quizlet, which uses the exact same interaction technique as Study Blue. Quizlet itself has proven to be an effective learning tool (Barr, 2016). It's clear that the design of LI's core interaction is overly confusing, with clear alternative interaction methods existing that present more usable interaction techniques. Participants repeatedly felt that they were cheating themselves due to LI relying on the user's honesty. Due to there being nothing stopping users from marking codes as "remembered", participants felt that they could easily cheat. This feeling of cheating may be due to the fact that participants had no motivation to actively learn the codes as part of the usability sessions. Future work will look at longer term testing which will assess participant's motivation to learn. There is a strong case based on qualitative data that an alternative method of evaluation is needed for LI.

Reviewing the qualitative feedback from participants of other usability studies issues pertaining to the presentation of information and navigation were common, with the presentation of text—its colour, size and amount of text being commented upon. This was also seen in the feedback the ACPO Starter Pack received, with the presentation of text being commented upon. One participant is quoted as saying "The text content needs a huge improvement to be easier to understand." Reviewing the findings of this study and other relevant studies show that how text is presented is a big factor to consider in terms of the usability of mobile applications. Participants from the other studies reacted negatively due to the lack of navigation buttons for "faster learning of the application" (Hashim et al., 2011) and being annoyed as they couldn't access certain content. Users of the ACPO Starter Pack regularly commented that they weren't able to access the content of the sub-pages. This was due to the interaction needed to access the sub-pages, being unclear. Presenting information is very important and doing so in a clear and obvious way is important for mobile devices due to their small screen size. The way in which users navigate mobile learning applications, should be closely considered.

The qualitative data highlights that the core mechanic for LI is not obvious or effective. The core mechanic in LI needs to change, as one participant commented, having multiple choices of locations to choose from to progress through LI is an option. Several participants believed the game was asking if the code and location matching up was true or false, this is also a solution. Mobile phones now commonly use on-screen keyboards, the application could easily require users to type in a code or location when either a location or code is presented with them. Expanding on this, the game could also ask users to provide the remaining letter of an unfinished location or code. The usage of maps and other imagery could also be explored with the real-world locations making the learning of ICAO codes less abstract, this has proven to lead to positive outcomes (Zbick et al., 2015).

The following guidelines are recommended for the design of mobile learning applications for ATCs, based on the findings of the study.

ATCs and other workers in the aeronautical industry have jobs that require extensive training in order to undertake work which is safety critical. ATCs have diverse background with anyone with five GSCE's able to become a trainee ATC (BBC, 2018a,b). Due to trainees coming into the aeronautical industry being of different ages and demographics, they also have varying degrees of experience using technology. Designing any sort of technology for trainee ATCs, needs to accommodate individuals who possess little experience at using technology in their day to day lives. To accommodate this, the user experience of mobile applications produced for trainee ATCs needs to be considered. Interfaces need to be designed with icons, colours, and other elements targeting a wide audience. Interactions are required to be as simple and user friendly as possible. Simplicity is something that should be targeted through any application design, from the qualitative feedback received, simplicity was highlighted consistently as one of LI's most admired elements. Feedback from the qualitative analysis also provided evidence that the applications produced are not suited to the individuals learning style, with some testing participants wanting more interactivity when using the ACPO Starter Pack and other individuals wanting a more tailored experience for LI. With the diverse background of ATC trainees and individuals wanting differing self-learning experiences, creating applications that can be flexible and contain different learning modalities is key. Due to the critical role that air traffic service industry personnel undertake, self-learning tools should also be kept up to date with the latest information. Most importantly information should always be accurate. Using inaccurate information in any learning material, would deem it to be unusable and if used, could pose a real danger. Lastly, one of the most common pieces of feedback received focused on the unclear purpose of LI. The expected learning outcomes and core learning is never disclosed to the user. Not only was the aim of the application unclear to participants, how they were meant to interact with the application was also often confusing to participants. Because of this we recommended that any future learning applications, present the purpose of the given application and how to interact with the application, in either an information menu or through a tutorial or onboarding process.

CONCLUSIONS AND FURTHER WORK

Our work has focused on producing two mobile applications for the aeronautical industry. These mobile applications serve as an alternative to existing training undertaken in the aeronautical industry. By undertaking usability testing sessions with various participants of differing demographics and experience with ATC training, we have assessed the design and usability of the two mobile applications produced. We have seen that approval rating is high from the quantitative usability results whilst qualitative results suggest that there are clear improvements to be considered for future versions of both applications.

In the methodology section of this paper, we defined three questions.

- What elements of the applications contribute to higher or lower usability scores?
- What changes can be done to both applications to improve their usability?

- Are both applications perceived to be fit for purpose to train the aeronautical workforce?

Both LI and ACPO achieved a usability score which based on the literature can be considered a good and excellent rating. The feedback gathered from the usability sessions provides a clear pathway for further work to improve the usability of both mobile applications.

By analysing the qualitative data taken from participants, there are clear elements that contribute to lower and higher usability scores for both apps. Simplicity was a common comment with five out of the nine usability sessions consisting of multiple participants mentioning simplicity as their favourite element of the LI app. Another element that participants reacted well to was the gamification implemented in LI via the feedback given after every play session.

In terms of the elements that contribute to lower usability scores of LI, the core interaction of swiping was a contentious theme of feedback. Participants regularly mentioned that they had problems interacting with LI with participants commenting that swiping felt "clunky" and "fiddly [sic]." LI also had persistent comments on its purpose, with participants unsure what the aim of LI was. Often participants commented that "people will cheat themselves" For the ACPO Starter Pack, participants reacted well to the use of visuals, with one participant commenting that "The use of animation is a definite advantage over a regular textbook/worksheet." Many comments from participants focused on how "digestible" the content of the ACPO Starter Pack was with one commenter saying that the ACPO Starter Pack "Made some aspects easier to understand." Participants consistently had issues when navigating the ACPO Starter Pack as it was not clear that certain sub-pages could be accessible. One participant commented that "An icon to represent clickable content would be beneficial." Participants also had problems with the amount of content viewable at any one time, with participants commenting that they felt the ACPO Starter Pack felt "busy."

This paper contributes to the research area of digital training and digital transformation. This research has focused on the aeronautical industry, proposing, and testing the use of the digital apps to improve the current air traffic controller training. In the study we present two mobile learning applications which demonstrate that presenting mobile applications, designed through a user-centered design approach can be received well by a wide audience. This audience includes experienced ATC staff, ATC trainees and individuals with no experience with ATC or ACPO training. We've considered a wide audience and shown that self-learning in the air traffic services industry can theoretically be innovated through mobile learning applications. This study is beneficial for the air traffic services industry as it considers and evaluates the use of mobile applications for the air traffic services industries. The study implicates that the use of mobile technology can be beneficial for ATC trainees. ATCs encapsulate a wide audience of individuals who have no particular demographic, this study presents varying demographics having a positive response to two applications. This is especially noteworthy due to the wide demographic of individuals who could possibly apply to be an ATC, with only basic qualifications needed to become an ATC (BBC, 2018a,b). As far as we are aware this paper is unique in that it presents a usability study of digital apps for the aeronautical industry.

Future work is already being undertaken by the authors, in which they are exploring the effectiveness of both applications as learning tools. A study is underway to compare the learning effectiveness of the existing training in comparison to LI and the ACPO Starter Pack. This research builds directly on the results of this paper. One group is tasked with learning via the existing paper-based training in current use, the other group is utilizing the two applications to learn the identical learning content. At the end of the study each group will complete the same assessment. After the assessments are completed, the average learning score of both groups will be compared. The aim is to evaluate how effective the approaches are for learning retention, but this could only be undertaken after a successful design could demonstrate high user engagement in digital tools.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by this study and was conducted under the auspices of the Abertay University School of Design and Informatics Research Ethics Committee, having passed the ethics process and acquired approval. The University of Abertay Dundee also approved of the consent form given to all participants. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PS was the Research Associate on the project. ID was the project lead and the primary academic supervisor and contributed to the overall project development. RF was the knowledge base support and contributed to the project development. ID, RF, and KS-B assisted in academic matters including the structure and writing of the paper.

FUNDING

This research was funded by a Knowledge Transfer Partnership (KTP). The KTP was part-funded by Innovate UK on behalf of the UK government.

ACKNOWLEDGMENTS

The author would like to thank all the participants of the study as well as Steve Graham, Andrea Kirkhope, Robert Becker, Simon Messer, and Melanie Hayes. As a Knowledge Transfer Project Andrea Kirkhope acted as primary company supervisor and contributed with assistance related to the assisting aeronautical

company. This included organising usability tests with their colleagues at the aeronautical company they were employed at. Robert Becker conducted two usability sessions on the behalf of the author. Steve Graham was the company sponsor. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00022/full#supplementary-material

REFERENCES

Abras, C., Maloney-Krichmar, D., and Preece, J., (2004). "User-centered design," in *Berkshire Encyclopedia of Human-computer Interaction, Vol. 1*, ed W. S. Bainbridge (Great Barrington: Berkshire Publishing Group), 763–768.

Ahmad, W. F. W., and Azahari, I. N. N. (2015). *Visual Approach of a Mobile Application for Autistic Children: Little Routine*. Los Angeles, CA: HCI, 436–442.

Arachchilage, N. A. G., Love, S., and Maple, C. (2013). Can a mobile game teach computer users to thwart phishing attacks?. *Int. J. Infonomics 6,* 720–730. doi: 10.20533/iji.1742.4712.2013.0083

Armstrong, P., and Wilkinson, B. (2016). "Preliminary usability testing of ClaMApp: a classroom management app for tablets," in *OzCHI '16: Proceedings of the 28th Australian Conference on Computer-Human Interaction* (Launceston), 654–656.

Barr, B. W. B. (2016). Checking the effectiveness of Quizlet® as a tool for vocabulary learning. *The Center for ELF J.* 1, 36–48. doi: 10.15045/ELF_0020104

BBC (2018a). *Call for Students as Air Traffic Service Demand Rises.* Available online at: https://www.bbc.co.uk/news/uk-wales-45187872 (accessed December 6, 2019).

BBC (2018b). *Hundreds More Air Traffic Controllers Needed*. Available online at: https://www.bbc.co.uk/news/uk-scotland-scotland-business-45057103 (accessed November 25, 2019).

Benedikz, R. (2019). *ICAO CODES.* Available online at: https://www.key.aero/article/icao-codes (accessed December 16, 2019).

Brooke, J. (1996). "SUS - a quick and dirty usability scale," in *Usability Evaluation in Industry* eds P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland (London: Taylor and Francis).

Budiu, R., and Nielsen, J. (2010). *Usability of iPad Apps and Websites.* Fremont, CA: Nielsen Norman Group.

Buley, L. (2013). *The User Experience Team of One: A Research and Design Survival Guide*. New York, NY: Rosenfeld Media.

Burgess, S. R. and Murray, A. B. (2014). Use of traditional and smartphone app flashcards in an introductory psychology class. *J. Instr. Pedagogies* 13, 1–7. Available online at: https://eric.ed.gov/?id=EJ1060082 (accessed June 24, 2020).

Burke, B. (2014). *Gamify: How Gamification Motivates People to Do Extraordinary Things*. New York, NY: Bibliomotion Inc.

Cordeil, M., Dwyer, T., and Hurter, C. (2016). *Immersive Solutions for Future Air.* Niagara Falls, NY: ACM.

Darejeh, A., and Salim, S. S. (2016). Gamification solutions to enhance software user engagement—a systematic review. *Int. J. Hum. Comput. Inter.* 32, 613–642. doi: 10.1080/10447318.2016.1183330

David, G., and Cambre, C. (2016). Screened intimacies: tinder and the swipe logic. *Soc. Media Soc.* 2, 1–11. doi: 10.1177/2056305116641976

de Jong, C., Klemp, K., Mattie, E., and Goodwin, D. (2017). *Ten Principles for Good Design: Dieter Rams: The Jorrit Maan Collection.* Munich: Prestel.

Deloitte (2017). *UK Public Are 'Glued To Smartphones' as Device Adoption Reaches New Heights*. Available online at: https://www2.deloitte.com/uk/en/pages/press-releases/articles/uk-public-glued-to-smartphones.html (accessed January 30, 2020).

Dowell, J., and Long, J. (1998). Target Paper: Conception of the cognitive. *Ergonomics 41,*126–139. doi: 10.1080/0014013981 87125

Edge, D., Searle, E., Chiu, K., Zhao, J., and Landay, J. A. (2011). "MicroMandarin: mobile language learning in context," in *CHI 2011 Conference on Human Factors in Information Systems* (Vancouver, BC: ACM).

Edge, D., Whitney, M., Landay, J., and Fitchett, S. (2012). "MemReflex: adaptive flashcards for mobile microlearning," in *MobileHCI 2012 Conference on Human Computer Interaction with Mobile Devices and Services* (San Francisco, CA: MobileHCI), 431–440.

Febretti, A., and Garzotto, F. (2009). "Usability, playability, and long-term engagement in computer games," in *Conference: Proceedings of the 27th International Conference on Human Factors in Computing Systems*, *CHI 2009, Extended Abstracts Volume* (Boston, MA).

Federal Aviation Administration (2017). *Retirement.* Available online at: https://www.faa.gov/jobs/employment_information/benefits/csrs/ (accessed *December* 6, 2019).

Fowler, M., and Highsmith, J. (2001). *Manifesto for Agile Software Development.* Available online at: https://agilemanifesto.org/ (accessed April 1, 2020).

Gordillo, A., Gallego, D., Barra, E., and Quemada, J. (2013). "The city as a learning gamified platform.," in *2013 IEEE Frontiers in Education Conference (FIE)* (Oklahoma City, OK), 372–378, doi: 10.1109/FIE.2013.6684850

Haissig, C. (2013). "Air traffic management modernization: promise and challenges," in *Encyclopedia of Systems and Control*, eds J. Baillieul and T. Samad (London: Springer), 1–7.

Hart, J., Sutcliffe, A., and De Angeli, A. (2012). "Using affect to evaluate user engagement," in *CHI EA: Conference on Human Factors in Computing Systems* (New York, NY), 1811–1834.

Hashim, A. S., Ahmad, W. F. W., and Ahmad, R. (2011). "Mobile learning course content application as a revision tool: the effectiveness and usability," in *International Conference on Pattern Analysis and Intelligence Robotics* (Putrajaya), 184–187.

Heinonen, R., Luoto, R., Lindfors, P., and Nygård, C.-H. (2012). Usability and feasibility of mobile phone diaries in an experimental physical exercise study. *Telemed. e-Health* 18, 115–119. doi: 10.1089/tmj.2011.0087

Humphreys, C. (2017). *Engaging a New Generation of Air Traffic Controllers.* Towards Maturity.

Kultima, A. (2015). *Defining Game Jam*. Pacific Grove: FDG.

Lain, D., and Aston, J. (2004). *Literature Review of Evidence on e-Learning.* Brighton: Institute for Employment Studies.

Lane, K. (2018). *Air Traffic Controllers Overcome Complexities, Challenges*. Available online at: https://news.erau.edu/headlines/air-traffic-controllers-overcome-complexities-challenges (accessed April 20, 2020).

Law, L.-C. E., Roto, V., Hassenzahl, M., Vermeeren, A. P. O. S., and Kort, J. (2009). *Understanding, Scoping and Defining User Experience*: *A Survey Approach.* Boston, MA: ACM.

Leitão, R. (2012). *Creating Mobile Gesture-based Interaction Design Patterns for Older Adults: a study of tap and swipe gestures with Portuguese seniors* (Master thesis).

Little, A., Cook, B., Hughes, H., and Wilkerson, K. (2017). *Utilizing Guided Simulation in Conjunction with Digital Learning.* Daytona Beach, FL: Embry-Riddle Aeronautical University.

Little, J. R., Pavliscsak, H. H., Cooper, M., Goldstein, L., Tong, J., and Fonda, S. J. (2017). Usability of a Mobile Application for Patients Rehabilitating in their Community. *J. Mobile Tech. Med.* 6, 14–24. doi: 10.7309/jmtm.6.3.4

Lovell, S. (2011). *As Little Design as Possible.* London: Phaïdon.

ManpowerGroup (2016). *Millennial Careers: 2020*. Vision. ManpowerGroup.

Matt, C., Hess, T., and Benlian, A. (2015). Digital transformation strategies. *Bus. Inform. Syst. Eng.* 57, 339–343. doi: 10.1007/s12599-015-0401-5

McCarthy, J., and Wright, P. (2004). *Technology as Experience.* Cambridge: MIT Press.

McKendrick, J. (2015). *Why Digital Transformation is So Difficult, and 8 Ways to Make It Happen*. Available online at: https://www.forbes.com/sites/joemckendrick/2015/12/29/eight-ways-to-make-digital-transformation-stick/#26fdc32c756a (accessed Feburary 11, 2020).

Metz, S. (2011). Motivating the Millennials. *Sci. Teach.* 78:6. Available online at: https://www.questia.com/read/1G1-267811175/motivating-the-millennials (accessed 24 June, 2020).

Miller, D. P. Jr., Weaver, K. E., Case, L. D., Babcock, D., Lawler, D., Denizard-Thompson, N., et al. (2017). *Usability of a novel mobile health iPad app by vulnerable populations. JMIR Mhealth Uhealth.* 5:e43. doi: 10.2196/mhealth.7268

Myers, K., and Sadaghiani, K. (2010). Millennials in the workplace: a communication perspective. *J. Bus. Psychol. 25,* 225–238. doi: 10.1007/s10869-010-9172-7

National Research Council (1997). *Flight to the Future: Human Factors in Air Traffic Control.* Washington, DC: National Academy Press.

NATS(a) (n.d.). *Careers*. Available online at: https://www.nats.aero/careers/trainee-air-traffic-controllers/#games (accessed March 26, 2020).

NATS(b) (n.d.). *Introduction to Airspace*. Available online at: https://www.nats.aero/ae-home/introduction-to-airspace/ (accessed January 31, 2020).

NATS(c) (n.d.). *Training to Be an Air Traffic Controller.* Available online at: https://www.nats.aero/careers/trainee-air-traffic-controllers/ (accessed February 7, 2020).

Ngan, H. Y., Lifanova, A., Jarke, J., and Broer, J. (2016). "Refugees welcome: supporting informal language learning and integration with a gamified mobile application," in *European Conference on Technology Enhanced Learning* (Lyon) 521–524.

Norman, D. (1988). *The Design of Everyday of Things, 2013 Edn*. Cambridge: Basic Books.

Noyes, D. (2018). *Could Air Traffic Controller Shortage Have Impact on Safety?*. Available online at: https://abc7news.com/travel/could-air-traffic-controller-shortage-have-impact-on-safety/4468996/ (accessed December 6, 2019).

Nunes, J., and Drèze, X. (2006). The endowed progress effect: how artificial. *J. Cons. Res.* 32, 504–512. doi: 10.1086/500480

Oblinger, D. (2003). *Boomers, Gen-Xers, and Millennials: Understanding the "New Students"*, Vol. 500. Educause, 36–47. Available online at: https://www.naspa.org/images/uploads/main/OblingerD_2003_Boomers_gen-Xers_and_millennials_Understanding_the_new_students.pdf (accessed June 24, 2020).

Oprescu, F., Jones, C., and Katsikitis, M. (2014). *I PLAY AT WORK—ten principles for transforming work processes through gamification*. *Front. Psychol.* 5:14. doi: 10.3389/fpsyg.2014.00014

Pechenkina, E., Laurence, D., Oates, G, Eldridge, D., and Hunter, D. (2017). Using a gamified mobile app to increase student engagement, retention and academic achievement. *Int. J. Educ. Technol. Higher Educ.* 14:31. doi: 10.1186/s41239-017-0069-7

Perry, J., Luedtke, J., Little, A., Cook, B., and Hughes, H. (2017). *Incorporating Digital Learning Tools in Conjunction with Air Traffic.* Daytona Beach, FL: Embry-Riddle Aeronautical University.

Plaza García, I., Sánchez, C. M., Espílez, Á. S., García-Magariño, I., Guillén, G. A., and García-Campayo, J. (2017) Development and initial evaluation of a mobile application to help with mindfulness training and practice. *Int. J. Med. Inform.* 105, 59–67. doi: 10.1016/j.ijmedinf.2017.05.018

Pollard, E., and Hillage, J. (2011). *Exploring eLearning.* Brighton: The Institute for Employment Studies.

Preece, J., Sharp, H., and Rogers, Y. (2002). *Interaction Design: Beyond Human-Computer Interaction. 1st Edn.* Chichester: John Wiley and Sons.

Sánchez, J., and Espinoza, M. (2011). "mGuides, design and usability of a mobile system to assist learning in critical situations," in *Conference: Universal Access in Human-Computer Interaction. Context Diversity - 6th International Conference, UAHCI 2011, Held as Part of HCI International 2011* (Orlando, FL), 415–424.

Sánchez, J., Flores, H., and Sáenz, M. (2008). "Mobile science learning for the blind," in *Extended Abstracts Proceedings of the 2008 Conference on Human Factors in Computing Systems, CHI 2008* (Florence), 3201–3206.

Savill, R. (2009). *Air Traffic Control Service NATs Found Guilty of Age Discrimination*. Available online at: https://www.telegraph.co.uk/news/uknews/4862480/Air-traffic-control-service-NATs-found-guilty-of-age-discrimination.html (accessed December 6, 2019).

Schwaber, K., and Beedle, K. (2002). *Agile Software Development with Scrum. Vol. 1.* Upper Saddle River, NJ: Prentice Hall.

Schwertner, K. (2017). Digital transformation of business. *Trakia J. Sci.* 15, 388–393. doi: 10.15547/tjs.2017.s.01.065

Seixas, L. D. R., Gomes, A. S., and de Melo Filho, I. J. (2016). Effectiveness of gamification in the engagement of students. *Comput. Hum. Behav.* 58, 48–63. doi: 10.1016/j.chb.2015.11.021

Shackel, B., and Richardson, S. (1991). *Human Factors for Informatics Usability.* Cambridge, MA: Cambridge University Press.

Sousa, M. J., Cruz, R., and Martins, J. M. (2017). "Digital learning methodologies and tools–a literature review," in *Edulearn17 Proceedings* (Barcelona). doi: 10.21125/edulearn.2017.2158

Strother, J. (2002). An assessment of the effectiveness of e-learning incorporate training programs. *Int. Rev. Res. Open Dist. Learn* 3, 1–13. doi: 10.19173/irrodl.v3i1.83

Su, C. -H., and Cheng, C. -H. (2014). A mobile gamification learning system for improving the learning motivation and achievements. *J. Comput. Assis. Learn.* 31, 268–286. doi: 10.1111/jcal.12088

Su, K.-W., Lee, K.-Y., Huang, P.-H., and Chen, I.-T. (2009). Developing a usable mobile flight case learning system in air traffic control miscommunications. *Int. Conf. Hum. Comput. Inter* 5613, 770–777. doi: 10.1007/978-3-642-02583-9_83

Swink, S. (2009). *Game Feel: A Game Designer's Guide to Virtual Sensation*. New York, NY: CRC press.

The Economist (2016). *Antiquated Air Traffic Control Systems Are Becoming a Serious Threat to Safety*. Available online at: https://www.economist.com/science-and-technology/2016/08/10/antiquated-air-traffic-control-systems-are-becoming-a-serious-threat-to-safety (accessed April 20, 2020).

Thom, J., Millen, D. R., and Dimicco, J. M. (2012). "*Removing gamification from an enterprise* SNS," in *CSCW '12: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 1067–1070. doi: 10.1145/2145204.2145362

Tortorella, R., and Kinshuk, A. (2017). A mobile context-aware medical training system for the reduction of pathogen transmission. *Smart Learn Environ.* 4, 1–13. doi: 10.1186/s40561-017-0043-9

Tsuei, M., Chou, H.-Y., and Chen, B.-S. (2013). "Measuring usability of the mobile mathematics curriculum-based measurement application with children," in *International Conference of Design, User Experience, and Usability* (Las Vegas, NV), 304–310.

Unity (n.d.). *Build Once, Deploy Anywhere, Captivate Everyone.* Available online at: https://unity3d.com/unity/features/multiplatform (accessed February 12, 2020).

Usability. gov (n.d.). *System Usability Scale (SUS).* Available online at: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html (accessed January 21, 2020).

Wang, A. I., Øfsdahl, T., and Mørch-Storstein, O. K. (2008). "An evaluation of a mobile game concept for lectures," in *21st Conference on Software Engineering Education and Training* (Charleston), 197–204.

Winslow, B. D., Ruble, M., and Huber, Z. (2018). *Mobile, game-based training for myoelectric prosthesis control. Front. Bioeng. Biotechnol.* 6:94. doi: 10.3389/fbioe.2018.00094

Wobbrock, J. O., Morris, M. R., and Wilson, A. D., (2009). "User-defined gestures for surface computing," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1083–1092.

Zbick, J., Nake, I., Milrad, M., and Jansen, M. (2015). "A web-based framework to design and deploy mobile learning activities: evaluating its usability, learnability and acceptance," in *2015 IEEE 15th International Conference on Advanced Learning Technologies* (Hualien), 88–92.

Zeigarnik, B. (1938). "On finished and unfinished tasks," in *A Source book of Gestalt Psychology,* ed W. D. Ellis (New York, NY: Harcourt, Brace, & World), 300–314.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fcomp.2020.00028

Check for updates

A Technical and Conceptual Framework for Serious Role-Playing Games in the Area of Social Skill Training

Julia Othlinghaus-Wulhorst and H. Ulrich Hoppe*

COLLIDE Group, Department of Computer Science and Applied Cognitive Science, Faculty of Engineering, University of Duisburg-Essen, Duisburg, Germany

OPEN ACCESS

Edited by:
Carlos Vaz De Carvalho,
Polytechnic Institute of Porto, Portugal

Reviewed by:
Patricia Martin-Rodilla,
University of A Coruña, Spain
António Fernando Coelho,
University of Porto, Portugal
Ramon Fabregat Gesa,
University of Girona, Spain

***Correspondence:**
Julia Othlinghaus-Wulhorst
othlinghaus@collide.info

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *28 December 2019*
Accepted: *24 June 2020*
Published: *31 July 2020*

Citation:
Othlinghaus-Wulhorst J and Hoppe HU (2020) A Technical and Conceptual Framework for Serious Role-Playing Games in the Area of Social Skill Training. Front. Comput. Sci. 2:28. doi: 10.3389/fcomp.2020.00028

Virtual role-playing games can provide an authentic experience of situated learning and allow for trying out different problem-solving and communication strategies without consequences in the real world. This is of particular interest and benefit for the training of social skills. This article presents a conceptual and technical framework for serious role-playing games for the training of specific social skills in virtual 2D learning environments involving chatbots in dialog-centric settings. It summarizes different use cases and evaluation results from prior studies. From the design perspective, several distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) separate phases of immersion and reflection to facilitate a change of perspective that is considered conducive for learning, (3) the learning process is emphasized by means of adaptive feedback based on individual analyses. We propose a system architecture that is based on three components: (1) AI-controlled chatbots that adapt to the player's behavior, (2) a multi-agent blackboard system as the backbone in order to keep components independent and optimize performance due to parallel processing, and (3) intelligent support for an automated evaluation of the player's performance and feedback generation. The training scenarios presented and discussed in this article include workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management. First evaluation studies indicate that the scenarios may be well-suited for real training situations. Due to its flexible architecture, our framework and approach can easily be tailored to different settings and use cases and thus serve as a basis for future research focusing on the adaptation to other contexts and systems. On the basis of these developments, we elaborate important design dimensions, reflect and discuss general issues and major challenges, summarize and contrast different approaches and strategies, and identify opportunities for serious role-playing games in the area of social skills training.

Keywords: virtual role play, intelligent support, serious games, social skills, chatbots

INTRODUCTION

In recent years, serious games have been established as an efficient medium in education and professional training (Michael and Chen, 2006; Marr, 2010). The serious gaming approach attempts to use the appeal of digital games not only for entertainment purposes but also to convey "serious" content and to train practice-oriented skills (Ritterfeld et al., 2009). The combination of the serious gaming approach with role play scenarios is particularly promising. Role play enables learners to explore new situations and train how to act and react in these situations (Martens et al., 2008). Virtual role-playing games provide mobile, safe, and continuable environments, whereas traditional role plays can be time-consuming, costly, and difficult to administer (Totty, 2005). In addition, they lack repeatability. One general problem in the evaluation of role play experiences for educational purposes is the effort involved in analyzing and reflecting on the actual role play following the enactment. Traditional scenarios typically rely on video recording and, if applicable, note-taking. However, virtual learning environments enable structured recordings with integrated indexing, navigation instruments, search functions, and cross-references between different media and data sources. In addition, computer-supported analyses can help to evaluate and track the learners' performance. This is an important aspect, since without feedback and post-role-play reflection, the transfer to real word situations cannot be ensured (Lim et al., 2009). An additional important advantage of serious role-playing games in contrast to other virtual learning activities and environments is the motivational component, which may lead to intense and passionate involvement of learners (Susi et al., 2007).

Based on a series of different instances of role-playing games for the training of specific social skills, this article presents the underlying conceptual and technical framework that facilitated the implementation of the different applications. This framework is characterized by using scripted chatbots as training cases in a dialogic setting. A multi-agent architecture supports both the actual dialogic processing as well as the evaluation of the dialogs and the generation of adaptive feedback. Conceptual and technical aspects of this framework are described in chapters Framework: Conceptual Approach and Framework: Technical Approach, following up on a discussion of related work in this area (chapter Related Work). Chapter Case Studies assembles several case studies conducted with different instances of virtual role-playing environments based on the framework, reporting on experience and evaluation results. Chapter Dimensions of the Design of Serious Role-Playing Games for the Training of Social Skills combines this specific experience with general issues in the design of serious role-playing games to devise a set design dimensions in the sense of important aspects to be considered in the design, description and comparison of serious role-playing games.

RELATED WORK

Serious Role-Playing Games for the Training of Social Skills

Serious games can be defined as "any form of interactive computer-based game software of one or multiple players to be used on any platform and hat has been developed with the intention to be more than entertainment" (Ritterfeld et al., 2009) and with an explicit focus on education. Games of this category are supposed to convey specific knowledge or train certain skills by using the attractiveness of entertainment games (Susi et al., 2007). Serious games can generally cover many different subject areas, but their application is mainly found in healthcare, education, and training, including military or employee training in companies (Marr, 2010). Serious games are widely accepted as an important and efficient medium with respect to education, training, and behavioral change (Michael and Chen, 2006). They are recognized to have several benefits: Serious games facilitate learning experiences while not having negative or harmful impacts (Ritterfeld et al., 2009). Games in general not only have a positive effect on the development of the player but can also be conducive to many different skills. Among others, Mitchell and Savill-Smith suggest that such target competences can be related to cognitive, social, analytical, and strategic aspects (Mitchell and Savill-Smith, 2005). Squire and Jenkins also made a comparable assessment (Squire and Jenkins, 2003). Further advantages include the reduction of costs and time associated with the use of serious games. They make it possible to recreate situations or working conditions that would otherwise not be possible in the real world (Corti, 2006; Susi et al., 2007). Serious games intend to facilitate deep and sustained learning (Gee, 2007) and prove to be more effective than traditional pedagogy and other educational technologies (Prensky, 2000; Ritterfeld et al., 2009).

Michael and Chen differentiate between games that educate and games that train (Michael and Chen, 2006). Games that educate should convey knowledge, facts and processes in a playful way, thereby contributing to education, while games that train are intended to improve the learners' skills in virtual environments or simulations. Our work is focused on the second category, more specifically on serious games for the training of social skills based on role play. Social skills can be seen as a sub-category of soft skills. The term *soft skills* refers to a broad concept that describes a set of personal attributes or traits expressing how persons know and manage both themselves and their relationship with other people (Dell'Aquila et al., 2017). While no universal definition of the term "soft skills" is available, Dell'Aquila et al. combine several different approaches to the following definition (Dell'Aquila et al., 2017): "Soft skills are not domain or practice specific; experientially based; both self and people orientated; goal-related behaviors; inextricably complementary to hard technical knowledge and skills enabling completion of activities and accomplishment of results; and crucial for effective leadership performance." Social skills refer to soft skills related to interaction with other people. It describes "the ability to interact with others in a given social context in specific ways that are societally acceptable or valued and at the same time personally beneficial, mutually beneficial, or beneficial primarily to others" (Combs and Slaby, 1977) and includes, e.g., communication, cooperation, assertion, responsibility, empathy, engagement, and self-control (Gresham and Elliott, 2008). Role play is a great instrument to train interaction with other people. Assuming roles provides the opportunity to train to act and re-act in new situations. It facilitates the creation of knowledge and meaning through concrete experiences (Lim et al., 2009). Also, the observation of role play can lead to conclusions about own behavior (Martens et al., 2008). The integration of role play in a serious gaming context seems to be particularly promising, as this combination (a) incorporates a highly motivational character and (b) creates opportunities for exploration and experimentation in a protective environment without any consequences in the real world. In addition, virtual role plays may be much more effective than conventional approaches in settings where the social component is a crucial factor (Lim et al., 2009).

Several serious role-playing games for the training of social skills are available. They can be assigned to three main categories of relevant social skills: (1) leadership skills, (2) communication skills, and (3) conflict management. Examples for serious role-playing games for training leadership skills are *Virtual Leader* (Knode and Knode, 2011), *TeamUp* (Bezuijen, 2012), and *Learn to Lead* (Di Ferdinando et al., 2011). *Virtual Leader* is a simulation game in which students practice leadership styles and approaches within a 3D environment using avatars and intelligent agents in order to create a preferably realistic environment (Knode and Knode, 2011). Players participate in virtual business meetings with animated characters and are required to make a series of decisions in five scenarios with increasing complexity. *TeamUp* is a collaborative game for the training of teamwork and leadership skills, developed at the TU Delft (Bezuijen, 2012). In this game, four players need to work together to overcome several challenges, each designed to cover a specific element of effective teamwork. In *Learn to Lead*, the players have to lead a simulated team of employees (e.g., workers in a bank, a post-office, or a local government office) that is competing against other teams (Di Ferdinando et al., 2011). In this game, the players have two main objectives: First, they need to ensure that the company is running efficiently and productively. Second, they need to ensure that their teams develop in the desired manner. The Productive Leadership Game is a simulation game that is supposed to foster leadership competencies to improve team-based and organizational productivity (Kesti et al., 2017). A recapitulatory overview of serious role-playing games for training leadership skills can be seen in **Table 1**.

There are various examples for serious role-playing games aiming at the training of communication skills: *ENACT* (Marocco et al., 2015) is an online game for the standardized psychometric assessment and training of negotiation skills based on Rahim's model of conflict handling styles (Rahim and Bonoma, 1979). In this game, players assume different characters to negotiate with computer-controlled virtual 3D agents in various scenarios representing everyday life situations. They can always choose one of four possible pre-defined sentences to communicate with the agents. In *DREAD-ED*, players become part of a crisis management team that is dealing with an emergency situation (Haferkamp et al., 2011). The game is organized into a series of timed rounds, separated by phases in which a tutor can provide feedback to the players. Bosse et al. developed a game targeted at police academy students that focuses on decision-making aspects in critical situations like the so-called "door scene" in which a police officer has been informed about an incoming emergency call and is supposed to find out if it is indeed a case of domestic violence or not (Bosse and Gerritsen, 2016). The players interact with virtual characters in a realistic 3D environment by using a relatively simple interaction paradigm based on multiple choice and dialog trees. In the game *deLearyous*, players assume the role of a manager who just announced that the parking facilities of the company are no longer free and needs to deal with the reaction of an employee (Vaassen and Wauters, 2012) by using unconstrained written natural language input. The design of the virtual character representing the employee is based on a framework for interpersonal communication called Leary's Rose (Leary, 1957). *JUST-TALK* is a serious game to train law enforcement personnel for encounters with persons showing symptoms of serious mental illness (Hubal et al., 2003). The players interact with these computer-controlled characters using spoken natural language. They are supposed to look for indications of particular forms of mental illness so that they can adapt their approach in an appropriate way und thus defuse the situation. In *POINTER*, a game developed for interview training targeted at police officers, the players assume the role of a police officer interacting with a subject in the context of a police interview (Linssen et al., 2014). The subject here is a virtual agent who is not cooperating during the interview. The players' task is to interact with the subject in a way that makes it cooperate in order to gather information from them. *ELECT BiLAT* is a simulation game in which soldiers practice bilateral engagements within a cultural context (Lane and Hays, 2008). The recruits are supposed to conduct meetings and negotiations with local leaders. *Maritime City* is a game targeted at social workers. It aims at training the ability to read emotional states of persons and improving communication skills in verbal and non-verbal forms (Flynn et al., 2011). In this game, players are asked to investigate a disturbance at a house where a woman is living with her two children and need to investigate a range of approaches for each part of the scenario. *TARDIS* is a scenario-based serious game simulation platform that supports social training and coaching in the context of job interviews (Gebhard et al., 2018). It is specifically intended to be used by young people and job-inclusion associations to explore, practice, and improve their skills in a diverse range of possible interview situations by interacting with virtual agents acting as recruiters. *Communicate!* is a serious role-playing game designed to support practicing interpersonal communication between health care professionals such as doctors, pharmacists, or psychologists and a patient or client (Jeuring et al., 2015). In the scenarios included in the game, the players find themselves in a consultation with a virtual character during which they can choose between

TABLE 1 | Serious games for the training of leadership skills (overview).

Game	Author	Use of AI	Mode	Learning objective	Underlying framework/model/theory
Virtual Leader	Knode and Knode (2011)	Yes	Singleplayer	Leadership styles	–
TeamUp	Bezuijen (2012)	No	Multiplayer	Teamwork, leadership skills	–
Learn to Lead	Di Ferdinando et al. (2011)	No	Singleplayer	Leadership skills	Full-range theory
Productive Leadership Game	Kesti et al. (2017)	No	Singleplayer	Leadership competencies	Human capital production function

various options. They receive immediate feedback through the utterances and emotions of the conversational partner. The game *SALVE* (Augello et al., 2016) is using AI-controlled chatbots participating in medical consultations and is based on the Social Practice Theory (Schatzki, 1996). In contrast, Even et al. developed a serious game primarily targeting schizophrenia patients to support rehabilitation programs for social skills (Even et al., 2016). This approach is combining role play with problem-solving exercises on which remediation therapies rely. A recapitulatory overview of serious role-playing games for the training of communication skills can be seen in **Table 2**.

Conflict management is an important social skill that has been the subject of serious role-playing games in the past. *Choices and Voices,* for example, is an interactive simulation game for preventing violent extremism. In it, players explore and discuss issues and influences leading to tension and disruption in communities (Memarzia and Star, 2011). In this game, players face several moral dilemmas in which their decisions determine the outcome of the game (for themselves, their family, and their friends). This is supposed to show the significant consequences real life decisions can have. The storytelling game *Façade* asks players to resolve a conflict between a married couple. Through communication with the conflicted parties, they are to investigate the causes of their issues and provide counseling (Mateas and Stern, 2003). The emphasis here is on believable characters, natural language conversation, and a dynamic storyline. In *Office Brawl* the player assumes the role of a mediator, who is moderating a conflict between two parties in a workplace-oriented setting, using AI-controlled virtual characters (Glock et al., 2011). As a project manager in the game, the player needs to handle an argument between two members of a team. *FearNOT!* is a virtual drama for anti-bullying education targeted at children (Aylett et al., 2005). In this game, the bullying behavior of one of the characters is leading to dramatic episodes. The victim is seeking advice of the player who can interact with this character by using free text input. It is supposed to allow children to explore what happens in bullying situations in which they take responsibility for what happens to a victim without feeling victimized themselves. The game *LOITER* lets prospective police officers enact street interventions with loitering juveniles (Linssen et al., 2014) and aims to improve their social awareness. Here, players can experiment with different ways of interacting with the juveniles. *Self City* is a serious game developed for emotionally impaired adolescents, which is supposed to help them develop skills such as process-oriented thinking and conflict resolution (Van Dijk et al., 2008). In this game, players can walk around online in a virtual city. On their way to the cinema, they experience challenging social situations and learn how to deal with them. Players are accompanied by a daemon that provides advice in conflict situations and suggests alternative actions. The *Junior Detective Computer Game* has been developed as part of a multi-component social skills intervention for children with Asperger syndrome (Beaumont and Sofronoff, 2008). Here, players take the role of a trainee at the Detective Academy and are taught how to recognize complex emotions in computer-animated and human characters. They need to complete several missions, such as dealing with bullying, playing with others, and trying out new things. A recapitulatory overview of serious role-playing games for training conflict management can be seen in **Table 3**.

Frameworks for the Design of Serious Games

There is a number of existing models and frameworks for the general design of serious games, which describe fundamental components of such systems and support formal approaches to game design. A very general approach is the so-called *MDA (Mechanics—Dynamics—Aesthetics) framework* (Hunicke et al., 2004). It proposes three different perspectives for understanding and designing games: *Mechanics* refer to the actual implementation of the game. They describe its particular components (actions, behaviors and control mechanisms) at the level of data representation and algorithms. *Dynamics* relate to the overarching design goals and run-time behavior of the mechanics acting on player inputs and each other's output over time. *Aesthetics* refers to the resulting game experience. They describe the desirable emotional responses evoked in players, when interacting with the game system. Although the MDA framework is widely accepted and practically employed, it has weaknesses and limitations (Walk et al., 2017): It focuses too much on game mechanics, neglecting many design aspects of games, including an over-arching narrative. Therefore, it is not really suitable for all types of games, including particularly gamified content or any type of experience-oriented design.

Another approach toward serious game design is the *Four-Dimensional Framework* suggested by De Freitas and Oliver (2006). It postulates four main dimensions of learning processes to be considered in the design process of serious games: the *context* in which learning takes place (e.g., classroom-based or outdoors, access to equipment, technical support), the *learner specification* (e.g., learner profile, pathways, learning background), the *mode of representation* (e.g., level of fidelity,

TABLE 2 | Serious games for the training of communication skills (overview).

Game	Author	Use of AI	Mode	Learning objective	Underlying framework/model/theory
ENACT	Marocco et al. (2015)	No	Singleplayer	Negotiation skills	Model of conflict handling styles
DREAD-ED	Haferkamp et al. (2011)	No	Multiplayer	Disaster Communication	Theories of crisis and emergency risk management
The "Door Scene"	Bosse and Gerritsen (2016)	No	Singleplayer	Communication skills (police domain)	Education program of police academy students
deLearyous	Vaassen and Wauters (2012)	Yes	Singleplayer	Communication skills (workplace)	Interpersonal circumplex (Leary's Rose)
JUST-TALK	Hubal et al. (2003)	Yes	Singleplayer	Communication skills (law enforcement)	–
POINTER	Linssen et al. (2014)	Yes	Singleplayer	Communication skills (police domain)	Cognitive model for social interaction
ELECT BiLAT	Lane and Hays (2008)	Yes	Singleplayer	Cultural social conventions (military domain)	–
Maritime City	Flynn et al. (2011)	Yes	Singleplayer	Communication skills (social work domain)	–
TARDIS	Gebhard et al. (2018)	No	Singleplayer	Communication skills (job interview)	–
Communicate!	Jeuring et al. (2015)	No	Singleplayer	Communication skills (health care domain)	–
SALVE	Augello et al. (2016)	Yes	Singleplayer	Communication skills (healthcare domain)	Social practice model
Serious game for schizophrenia patients	Even et al. (2016)	No	Singleplayer	Communication skills (emotion recognition)	Social skills programs for schizophrenia

TABLE 3 | Serious games for the training of conflict management (overview).

Game	Author	Use of AI	Mode	Learning objective	Underlying framework/model/theory
Choices and Voices	Memarzia and Star (2011)	No	Singleplayer	Prevention of violent extremism	National curriculum
Façade	Mateas and Stern (2003)	Yes	Singleplayer	Conflict resolution	–
Office Brawl	Glock et al. (2011)	Yes	Singleplayer	Mediation	–
LOITER	Linssen et al. (2014)	Yes	Singleplayer	Street interventions	Cognitive model for social interaction, Virtual Storyteller (VST)
Self City	Van Dijk et al. (2008)	No	Singleplayer	Process-oriented thinking, conflict resolution	Dialogical self-theory
Junior Detective Computer Game	Beaumont and Sofronoff (2008)	No	Singleplayer	Bullying, conflict resolution	Social skills programs for individuals with Asperger syndrome

interactivity, and immersion used in the game), and *pedagogic considerations* (e.g., learning models, approaches for learning support). Like the MDA framework, this framework is a high-level model, meaning that it specifies a limited number of generic concepts that can or should be taken into consideration when designing or evaluating serious games, but only on a very general level with no concrete design or evaluation guidelines (Mayer et al., 2014).

This also applies to the *RETAIN (Relevance Embedding Translation Adaptation Immersion & Naturalization) model* by Gunter et al. (2006). This model was developed to support game development and to assess whether a serious game is appropriate for educational purposes, how well the academic or pedagogical content is immersed and embedded in the game's narrative and how knowledge transfer is promoted. *Relevance* means that the information students learn in the game should be relevant to the game world as well as to the players' targeted objectives. *Embedding* should be done in a way that learning objectives and fantasy are tightly coupled. *Transfer* refers to how well players can recognize and apply newly learned information outside the game environment. *Adaptation* means that players apply their learned knowledge to create new scenarios that apply literacy skills in a new domain. *Immersion* should be facilitated by the game environment and the ability to create customizable social presence. *Naturalization* means that players should be encouraged to gradually use their own skills to gain the knowledge necessary for success in other problems and subject areas (Kenny and Gunter, 2011).

The *Triadic Game Evaluation* (TGE) (Harteveld, 2011) approach stresses three different perspectives for the design and evaluation of serious games: reality, meaning and play. The *reality* component determines the game subject, variables and definitions. It could be represented by players from the real world or a representation of the real world inside the game. Evaluation criteria in regards to this component include fidelity, realism, and validity. The *meaning* component of the framework considers how a meaningful effect beyond the game experience can be achieved and incorporates aspects such as communication, learning, rhetoric, and opinions. Evaluation criteria include reflection, transfer, and relevance. The *play* component refers to the fact that games are primarily highly interactive and engaging tools that immerse players into a fictitious situation, and is related to game elements like actors, rules, resources, challenges, and competition. Evaluation criteria for this component are engagement, fun, and immersion. The TGE framework claims that games need to be designed equally along these three components (Kortmann and Harteveld, 2009). In contrast to the aforementioned models, this framework comes with a concrete agile development model that describes different software engineering phases and decision moments in the creation process. However, specific design and implementation guidelines are not included.

In summary, the various promising approaches to training social skills by means of role-playing games are still defined on a very general level. Our aim is to provide a comprehensive conceptual and technical framework for the concrete design and implementation of serious role-playing games for the training of social skills in dialog-centric settings with virtual characters through which we would support more efficient and effective design and implementation of such game environments.

FRAMEWORK: CONCEPTUAL APPROACH

From the design perspective, several distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) separate phases of immersion (role-playing) and reflection to facilitate a change of perspective that is considered conducive for learning, (3) the learning process is emphasized by means of adaptive feedback based on individual analyses.

Chatbots in Virtual Role-Playing Environments

Chatbots are computer programs (conversational agents) that communicate with users in natural language. Their purpose is to simulate a human conversation via text or voice interactions. Originally, chatbots were developed for entertainment purposes. However, especially in today's world, in which the possibilities of computer use are becoming more and more diverse, the use of chatbots can be extended to many other areas. Chatbots are found in daily life now, such as personal assistants (like Google Assistant, Amazon Alexa, or Apple's Siri), search engines, customer service and support, and healthcare coaching (Winkler and Söllner, 2018). They can be used in a variety of domains including business, e-commerce, entertainment, medicine, and others (Kerly et al., 2006; Shawar and Atwell, 2007).

Chatbots can also be used successfully for learning. Past studies even show that chatbots present feasible means to improve learners' results (Kerly et al., 2006). They have been used for a variety of purposes including medical education and therapy, language learning, as well as receiving feedback and strengthen motivation and self-efficacy (Winkler and Söllner, 2018). Chatbots have also been used in serious role-playing games, as shown in the examples in chapter Related Work. The use of chatbots in serious role-playing games has several advantages. First, having a chatbot interact with the player instead of a human ensures a certain level of standardization that could never be achieved in a setting with human actors. Second, scenarios including a chatbot are repeatable, independent of time and place, and no additional resources are needed. An important part of chatbots is the creation of dialogs. A chatbot can only be as good as its knowledge base used for answer generation (Abdul-Kader and Woods, 2015). The problem of the "classic" chatbots is that they do not allow to store the course of the conversation and have no real understanding of the answers. However, a realistic and responsive behavior of chatbots is important to increase the players' engagement and contribute to the immersive nature of role plays. To achieve this, our approach proposes several technical workarounds that will be explained in detail in chapter Multi-Agent Architecture.

Immersion and Reflection

The educational impact of serious role-playing games highly draws on the "willing suspension of disbelief" by the players who commit to the role they are supposed to play (Lim et al., 2009). Thus, this kind of system intends to create a certain degree if immersion. Janet H. Murray defines the term immersion as follows: "A stirring narrative in any medium can be experienced as a virtual reality because our brains are programmed to tune into stories with an intensity that can obliterate the world around us... The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus" (Murray, 2017). When players identify themselves with the character they are assuming in the game and are immersed, their motivation to proceed and succeed in the game increases (Annetta, 2010). This intrinsic way of motivating learners is something conventional instruction modes do not have (Yee, 2006). Players become immersed in a game because they find it satisfying, and through this intrinsic motivation, they get more engaged in the learning task (Annetta, 2010).

In terms of experience-based, authentic learning, it seems reasonable to carry out the enactment in an immersive situation. However, there is reason to believe that the immersion tends to impede the critical self-reflection that is important for the learning process (Malzahn et al., 2010). Reflection is a successful tool to improve the learning process (Jonassen et al., 1993), and it is needed to ensure the transfer to real-life situations (Lim et al., 2009). During the reflection process, people recapture, rethink, and evaluate their experiences to develop new understandings and appreciations (Boud et al., 1985). It is to be expected that the amount of reactive attention required for immersion impedes the players' ability to distance themselves from the role, which in turn interferes with self-reflection. Thus, the requirement of role distance in phases of reflection suggests that the mode should be changed to help the learner step out of his role and adopt a different perspective. Based on this assumption, we decided to separate the actual role-playing game from the reflection session in our framework.

Adaptive Feedback

As stated above, an important challenge for serious role-playing games is shaping the narrative experience and the pedagogical outcomes that generally depend on post-role-play reflection and feedback (Lim et al., 2009). Feedback on the performance of the player(s) during the role-playing session is necessary to ensure the transfer to real-life settings. It is supposed to help learners to improve their performance by providing information about the correctness of their actions (Shute, 2008). Johnson et al. identified four feedback characteristics: (1) the *type* of feedback (e.g., outcome-based or process-based feedback), (2) the *timing* of feedback after an action (i.e., immediate or delayed feedback), (3) the *modality* in which the feedback is presented (e.g., spoken or text-based feedback), and (4) *adaptation* to learner characteristics (e.g., in regards to prior knowledge or spatial ability) (Johnson et al., 2017).

Our framework relies on adaptive feedback based on an automated, individual performance analysis. We differentiate between three types of feedback: The first one is implicit feedback during the role-playing session through the reactions of the chatbot (*ingame feedback*). These reactions can be non-verbal (e.g., facial expressions) or verbal. Real-life situations are simulated through both types of reactions to the players' actions. The second one is a general summary of the analysis results (*aftergame feedback*). Players should receive an overall feedback on their performance during the role play that summarizes the most important aspects (positive and negative). The third type is direct and specific feedback on single incidents during the role play that can be provided through prompts in a replay of the conversation. A replay offers several advantages: The whole conversation can be shown again step by step and augmented with individual feedback at certain points, commenting on specific actions of the player. Also, it provides the possibility to navigate between the different phases of a conversation, pause the replay, or jump to the next feedback marker. As a result, it is much more flexible and searchable than, e.g., a video of a conventional role play.

FRAMEWORK: TECHNICAL APPROACH

In our approach, the technical implementation of such systems entails three main challenges: (1) dialog modeling of the chatbot, (2) implementing a multi-agent system as the backbone in order to keep components independent and optimize performance, and (3) performance analysis and feedback generation. The following section will present our approach toward each of these aspects in detail.

Dialog Modeling

In our framework, the *Artificial Intelligence Markup Language* (AIML) is used for the implementation of the chatbots' conversational logic. It is a common XML-based solution for passive AI-controlled chatbots, which comes with an easy syntax and a small number of control structures (Wallace, 2004). AIML relies on a simple pattern matching. It consists of categories, each containing a pattern and a template. If the user input is matching a pattern, the template defines the answer or action to be given. Recursion and wildcards allow for many different inputs matching one single pattern, while the ability to store a context and the use of variables and conditions allow a complex and sophisticated chatbot design.

Although AIML has a long history and is a common solution for chatbots used in educational contexts, it has certain limitations. One problem is the passive nature of AIML. An AIML chatbot only reacts to an input it receives, it cannot take the initiative. This behavior can be bypassed by using external triggers to make the bot become active when required in certain situations. Another problem is that an AIML chatbot (as is true for all artificial natural language processing) cannot truly grasp the sense of what has been said. The AIML chatbot only checks the user input against predefined patterns; if there is no match, it can at most output some default statements (which need to be predefined as well). To solve this problem, our framework proposes the use of sentence openers in dialog-centric role play scenarios. This means that players always have to select a sentence opener from a predefined set and supplement it with free text input to compose a message.

This approach has several advantages: First, a sentence opener already defines the general gist of a message (e.g., affirmation, rejection, proposal, inquiry). As a result, it is at least possible to provide a default answer that is tailored to the selected sentence opener even if the free text input following the opener does not match a predefined pattern. Furthermore, if each phase of the chat conversation has unique sentence openers, the chatbot always has some kind of context information. Second, the use of sentence openers reduces the complexity of the dialog scripts dramatically because the possible starting points of all input sentences are already known. Third, sentence openers provide support to the players and help them phrase their messages. In addition, sentence openers improve the overall atmosphere of the simulated conversation and make it seem more realistic and natural. Last, sentence openers (in contrast to fully predefined text messages) still allow for free text input that can be analyzed in detail and influence the course of the game.

Multi-Agent Architecture

Our technical framework is based on a uniform multi-agent system architecture with a blackboard as the communication and integration mechanism. The blackboard is realized through a so-called *tuple space*. The components (agents) in this system are loosely coupled, i.e., they do not communicate with each other directly but only via entries on a central tuple space server (Gelernter, 1985). These entries have a simple tuple structure that contains primitive data types (integers, characters, booleans) and strings. According to the original concept of Gelernter, there are only a few generic operations (read, write, take, wait-to-take, etc.) to interact with such a blackboard. In contrast to a pure database solution, however, there are active trigger mechanisms such as notifications. The *SQLSpaces* developed in the COLLIDE group itself serve as a specific implementation basis in our framework (Weinbrenner, 2012). While the server itself is implemented in Java, the system framework of SQLSpaces provides clients for the agent programming in various programming languages. SQLSpaces also facilitates the logging of relevant data of each gaming session, which can later be used for analysis and comparison.

The overall system consists of a user interface and various agents, each of which is responsible for one task in either dialog analysis, feedback creation, or game control. The user interface in the three implemented training scenarios described in this article have been implemented as a web application using HTML, CSS, and JavaScript (2D frontend). Previous implementations were based on OpenSimulator3 (3D frontend), but since there were no specific advantages of 3D environments over 2D environments, we decided to go ahead with a 2D approach (Malzahn et al., 2010). As described above, the client (user interface) and all agents are writing and reading tuples from the tuple space server without communicating with each other directly, which results in a loosely coupled and adaptive system. That means, agents can easily be adapted, added, or removed depending on the actual application scenario.

The agents can be divided into three groups, depending on their functionality. *Pervasive agents* are overarching agents, which are crucial in connecting the individual game components. The *register agent,* for example, is managing the log-in of the player (or players in a collaborative scenario). When a new client is logging in, the register agent receives a request via the tuple space (callback) and starts a new gaming session. The *silence agent* reacts if a player has been inactive for a certain amount of time, in which case the agent is triggered and sends an internal message to which the chat bot responds. After the fourth internal message from the silence agent, the conversation ends. *Pre-processing agents* are used to pre-process the player's input before the answer to it is generated in order to provide the best possible answer. This pre-processing is mainly used to overcome the limited capabilities of AIML: Analyzing certain aspects separately helps to prioritize specific behaviors, i.e., make sure that the chatbot is reacting adequately to rude or aggressive behavior. In addition, this procedure reduces the structure of the AIML scripts and supports the feedback creation. Each of the implemented scenarios uses different pre-processing agents depending on the context. All pre-processing agents analyze the player's input regarding one specific aspect. **Figure 1** shows the basic architecture.

Performance Analysis and Feedback Generation

Both performance analysis and feedback generation always depend on the context and the learning objectives of the serious role-playing game. As described above, our architecture is using analysis agents, each of which is responsible for the evaluation of one specific aspect of the player's communication behavior. They are divided into pre-processing agents and regular agents.

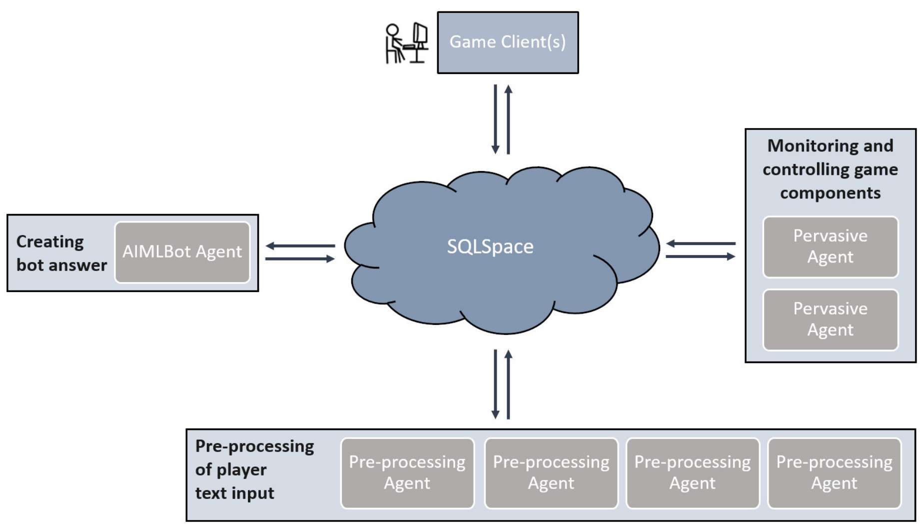


FIGURE 1 | Basic system architecture.

Pre-processing is necessary for generating a suitable chatbot response. For example, if a player acts aggressive or rude, the chatbot should react to this behavior regardless of any other information the player's message contains. The results of the pre-processing are collected, and if an immediate reaction to a specific behavior is required, the text input is modified. If, for instance, a swearword has been detected in a player's message, the complete input string is replaced by a specific trigger ("swearword"), causing the bot to react appropriately. The same applies to other behaviors. In case the pre-processing agents do not find anything that needs an immediate reaction of the chatbot, the bot receives the original text input. Simultaneously, all other analysis agents evaluate the message and add their feedback to it in the form of feedback tags (e.g., #praise#, #interruption#, or #criticize#). These feedback tags mark any situations in which the player is supposed to receive feedback during the replay that is taking place in the reflection phase following the role play session. The tags are filtered out during the chat session; the players do not get to see them during the game, but they play an important role in the feedback generation.

CASE STUDIES

Based on the framework described above, the research group COLLIDE at the University of Duisburg-Essen has conducted various case studies with different instances of virtual role-play environments. The training scenarios include workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management.

Case Study: Conflict Management

ColCoMa (Collaborative Conflict Management) is a collaborative serious game for the training of conflict management strategies in an organizational context within a role-playing scenario, developed at the COLLIDE group in 2012. It involves two players in a conversation with an AI-controlled chatbot acting as a mediator in a 2D virtual environment. The following description of the approach and game design is based on the work of Emmerich et al. (2012).

Approach

In ColCoMa, two players have a conversation about a fictitious conflict, moderated by an AIML chatbot in the role of a mediator. The main goal of the players is to resolve the conflict by showing constructive and appropriate behavior during the conversation. Each player is assigned a predefined role in this fictitious scenario: As a member of the computer support hotline of a big software company, Mr. Meier is conscientiously taking much time for his customers. Mrs. Schmidt is his supervisor. She is dissatisfied with Mr. Meier's way of working. She notices that he takes too much time for the customers and therefore does not work efficiently in her eyes. Mr. Meier does not agree with her, and the situation escalates after a negative appraisal of Mr. Meier's performance. In order to support immediate understanding of the situation and empathy with the assigned role, the scenario is kept as simple and comprehensible as possible and focuses on the main conflict as well as the person's feelings.

Game Design

The players are introduced to the game and the scenario through a cartoon-like picture story that is told from their respective role's perspective and is supposed to result in conflicting points of view. The conversation itself takes place in a chat window where graphical representations of the mediator and the other player's character are shown to create the association of sitting opposite each other. The dialog partners can communicate via simple text messages. Facial animations can be evoked via common emoticons. The interface also includes a notepad with hints as well as a help section that offers additional information on the game controls and the fictitious scenario if needed. **Figure 2** is showing the basic user interface.

The conversation is divided into five phases according to Proksch (2010): (1) framing phase, (2) topic collection, (3) working on the conflict, (4) looking for a solution, and (5) contract. The framing phase represents the starting point of the mediation talk and is important for establishing certain rules for the conflicted parties and their behavior toward each other. The actual conflict is not yet the focus. Instead, the participants state their personal hopes and mediation goals and reflect on their own point of view as well as the opponent's position. In the second phase, both parties are supposed to name relevant topics they would like to put on the agenda during the mediation talk, like performance review, working conditions, the participants' perspective in the company as well as their behavior toward each other. The mediator chatbot recognizes the topics based on a list of keywords and phrases. In order to be able to advance in the game, the two players need to name three topics; otherwise the mediator terminates the conversation due to a lack of contribution. If only two topics are volunteered, the mediator will suggest a third one. The mediation talk itself takes place in the third phase. The main task during this phase is to discuss the selected topics in detail. Both players are given the opportunity to explain why a topic is important to them, what changes they would like to see in regard to the specific topic, and what they themselves can contribute to realize these changes. They are also given the opportunity to comment on whether the other party's perception is correct and to rectify their position if this is not the case. The aim of the fourth phase is to find solutions for the different topics that are acceptable for both parties. Finally, in phase five, they are supposed to agree to adhere to the solutions they came up with and enter into a contract.

The mediation talk is followed by a reflection phase in which both players receive feedback on their performance in order to help them reflect on their behavior. At the start of this phase, players get the opportunity to directly exchange feedback with each other in a free chat without the mediator. After this free chat, each of them receives an overall feedback on the own performance during the mediation talk. Finally, the players take part in a replay session of the whole chat conversation, but this time augmented with individual feedback commenting on especially positive and negative contributions of the players. A change of the graphical interface during the replay is supposed to reinforce role distance, which is assumed to be conducive for learning (see chapter Immersion and Reflection).

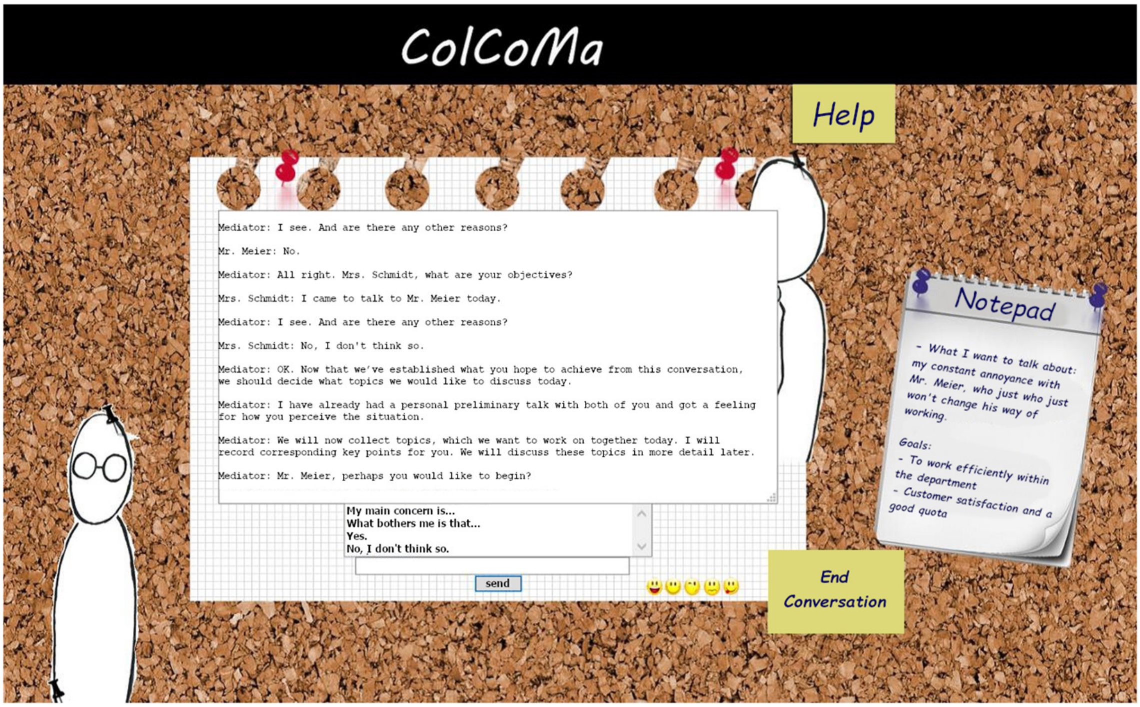


FIGURE 2 | ColCoMa chat interface.

The performance analysis and assessment is based on general rules that conflicting parties have to adhere to during a mediation talk, such as not being aggressive or rude, not being reproachful, and not impairing the opponent's autonomy (Stauss and Seidel, 2010). Instead, the participants are supposed to have an open and constructive attitude, name topics and issues in a concrete way, and help the other party understand their perspective. The evaluation of the players' performance during the mediation talk is done by several analysis agents, each responsible for one specific behavioral aspect, e.g., rudeness (by comparing the players' input to a list of swearwords and defamations), aggression (e.g., by checking for multiple exclamation marks or use of all-caps spelling), emotion-showing (e.g., use of emoticons), or the use of I- and you-statements (by counting the amount of words referring to the speaker and those referring to the dialog partner). Some of the analysis results are used just for the overall feedback that is provided to players after the conversation.

Evaluation Results

In 2018, an eye-tracking study has been conducted in collaboration with the Dortmund University of Applied Sciences and Arts (Othlinghaus-Wulhorst et al., 2018). The results of this study will be summarized and discussed in this section.

Apart from getting feedback on the prototype, the main goal of the study was to investigate the question if there is a correspondence between gaze synchronicity of the two players and the quality of collaboration. Twenty subjects (average 22.8, SD = 2.84, 5 females, 15 males) participated in the study and have been tested in dyads, using two desktop-based eye-trackers to track the players' gaze during the experiment. To investigate the research question, three main hypotheses have been examined: The first hypotheses postulated "a positive relation between the convergence of visual foci of attention (gaze synchronicity) and the successful completion of the game (achievement score)" (Othlinghaus-Wulhorst et al., 2018). In this study, *gaze synchronicity* has been defined as the extent to what the two players have been looking at the same areas of interest in the same time interval during the course of a gaming session. The so-called *achievement score* has been used to measure the success in the game and reflects the players' performance during the mediation talk based on three criteria: (1) automated feedback generated by the system, which summarizes the players' behavior during the game, (2) the successful completion of the topic collection phase (which has been considered a major milestone in the game), and (3) the successful completion of the game, which is achieved when both players sign a contract, which includes the agreements and rules they worked out together with the mediator. Referring to the hypothesis, a highly significant correlation between the gaze synchronicity and the achievement score has been found on the aggregate level (taking overall eye-tracking convergence as a global parameter).

In the second hypothesis it is assumed that "there is a positive relation between the convergence of visual foci of attention (gaze synchronicity) and the quality of collaboration in the chat." (Othlinghaus-Wulhorst et al., 2018). In order to define the *quality of collaboration*, a rating scheme has been developed, which includes five dimensions: (1) argumentation (players discuss or bring forward justifying arguments), (2) agreement/disagreement (players endorse or dissent from one another), (3) collaborative orientation (players refer to each other, ask questions, provide feedback or refer to topics brought up by the other player), (4) solution orientation (players try to find or propose a solution), and (5) shared awareness/reinforcing shared history (players share common knowledge or explain their situation). Based on this scheme, all chat messages have been analyzed and checked against the five dimensions and assigned a total quality score. Finally, all matches of a gaming session have been added up to a percentage indicating the overall quality of the collaboration for a pair of players. Relating to the hypothesis, a high correlation between the gaze synchronicity and the collaboration quality has been found, especially for the dimension's agreement/disagreement, solution orientation and shared awareness.

The third hypothesis proposes "a dynamic (time-related) congruence between similar eye movements (synchronicity) and the quality of collaboration in the chat" (Othlinghaus-Wulhorst et al., 2018), meaning that there is not only a gaze synchronicity on the aggregate level, but also synchronicity between convergent eye-tracking and chat interaction during the course of the game. This hypothesis could not be verified. It is assumed, that the specific nature of the chat might be a reason for this, as three persons are involved (the two players and the mediator chatbot) and thus the two human actors do not really communicate directly, but only to the mediator. They answer his questions and do not really have the chance to communicate with each other directly, which is resulting in a predefined structure of the chat conversation and rather long time interval between the utterances of the two players.

Case Study: Patient-Centered Medical Interview

In 2013, a training scenario for medical interviews has been developed at the COLLIDE group. It is supposed to give medical students the opportunity to train doctor-patient conversations autonomously and systematically in the form of role plays with simulated patients. The following description of the approach and game design is based on the work of Behler et al. (2013):

Approach

This scenario for the training of doctor-patient communication has the basic goal to train the communication strategies between doctor and patient and is tailored to the target group of medical students. Here, the player takes the role of a locum doctor for family medicine whose goal is to uncover all the symptoms of a patient in a given time. To achieve this, they have to use methods of the GOG (*Gesundheitsortientierte Gesprächsführung*, engl. "health-oriented negotiation") (Schwantes and Kampmann, 2007), in order to create a pleasant conversation atmosphere. To successfully master the game, it is necessary to behave in accordance with this concept and to bring in the guidelines in the course of the conversation. Another important learning objective is to build trust and empathize with the patient, as these aspects play a central role in the doctor-patient conversation (Kruse, 2000). Medical diagnosis is not a learning objective in this scenario, so the game can be used independently of progress in medical studies.

Game Design

At the beginning, the player enters the waiting room, where several patients are already sitting and waiting for their call. The patients represent different scenarios, which differ in the content and level of difficulty. The level of difficulty is determined by the number of symptoms to be identified and the willingness to talk about his or her condition. In the waiting room, the trophies and high scores already achieved by the current player are also displayed. By choosing a patient, the player starts the scenario and enters the doctor's office, where the actual interview with the patient takes place.

In the office, the player communicates with the patient via text input. The player chooses a suitable sentence opener and completes the sentence with free text. The sentence openers are related to GOG phases. In addition to verbal interaction, the player can also use items from the doctor's bag (information leaflets, stethoscope, pills, and syringe) and conduct non-verbal actions like nodding, smiling or touching the patient, which are also important in real interpersonal communication (Ziebarth et al., 2014). The items provide a playful added value. The player has to find out when which item is reasonable to use and receives bonus points for this, but only in combination with appropriate topics—otherwise, points are deducted. **Figure 3** is showing the basic user interface of this scenario.

In order to win the game, i.e., to achieve the highest possible score, the player has to collect points for recognized symptoms as well as points for trust-building and empathic contributions and actions. The accumulated sum of trust and empathy points in the game represents the conversation atmosphere and serves as a threshold value that defines how quickly a symptom is revealed by the patient. The patient reveals symptoms when the player addresses a scenario relevant topic and has reached the corresponding threshold value.

The main conflict is between the limited time available to the player to find the symptoms and the patient who only reveals them under certain conditions. This situation resembles a doctor's real conflict between time pressure and the desire to help patients comprehensively. Each scenario of this game contains a side mission to increase replayability. While the main task includes finding relevant symptoms, a secondary task could be, e.g., to point out the benefits of assisted living to an elderly patient to ensure long-term care. Side missions give more depth to the game as they refer to the social situation of the patient and thus lead to more immersion (McMahon and Ojeda, 2008). Players receive bonus points and trophies for solving side missions. As an additional incentive system, the total number of points is entered to a leaderboard, which all players can see. According to Festinger's theory of social comparison (Festinger, 1954), this motivates players to improve their own abilities, which are represented by the points.

The gaming session is followed by a reflection phase. First, the players are presented with their individual score in the fields trust, empathy and symptoms. Afterwards, they receive detailed feedback in the form of an augmented replay, in

FIGURE 3 | The doctor's office (basic game interface).

which the analysis results are presented. In this analysis, the player interaction is, e.g., checked for the use of paraphrases, emotions expressed to the patient, showing choices to the patient, addressing him or her by name and the use of all phases of the GOG. Although the phases do not have to be passed linearly, goal guidance and explanations for example are particularly relevant toward the end of the conversation. Pauses, nods and facial expressions are evaluated as well. In addition, behaviors are taken into account that do not directly lead to an improvement of the score but influence the course of the conversation. For example, the patient reacts verbally to excessive talking of the player and a lack of balance between the doctor and the patient as the subjects of the player's statements. This leads to a loss of time, which increases the central conflict of the game. As in a real situation, the player receives his feedback directly from the patient and can react to it in the process of the conversation.

Evaluation Results

The prototype has been evaluated in two studies, both performed by the COLLIDE group in cooperation with the Department of Family Medicine of the Charité in Berlin (Ziebarth et al., 2014). The results of these studies will be summarized and discussed hereinafter:

The focus of the studies was the examination of usability and playability, as well as immersion and reflection. The following key questions have been deducted from the global objectives: (1) Does the flow of the game feel natural? (playability), (2) Is the player able to manage the game well? (usability), (3) Is the game immersive? (immersion), (4) Is the reflection phase at the end of the game perceived as helpful? (reflection), (5) Which functions are used?, (6) What are the difficulties in using them?, (7) How is the game perceived by the target group?

Playability has been evaluated using self-created items relating to the clarity of the goal, structural problems regarding game flow (i.e., the use of sentence openers to create chat messages), "functional" playability (i.e., the extend of feeling understood by the patient), and the complexity of the game (Ziebarth et al., 2014). *Usability* has been assessed based on the following categories of ISO 9241-110: conformity with user expectations, suitability for learning, self-descriptiveness, and error tolerance. The items of the questionnaire were phrased based on the German inventories Isonorm[1] and IsoMetrics[2]. The aspect of *immersion* was measured based on the approach developed by Jennett et al. for measuring immersion in digital games (Jennett et al., 2008). The items selected for the studies address the subjective enjoyment of the game's representation, fun factor, immersion, and emotional involvement (Ziebarth et al., 2014). For the assessment of the *reflection* support, participants were asked what they thought the game is aiming to train, if they viewed the annotated replay, and if they thought about what they could have done differently (Ziebarth et al., 2014).

The first experiment was an observational study with 7 medical students (average 21, SD = 2.582, 6 females, 1 male). Although the results indicate that the idea and approach of the game were assessed quite positively, the observations showed slight problems with the general usage of the game. While the interaction principles have been generally well-understood, a few participants reported problems with expressing themselves using the predefined sentence openers. Also, the free text supplementing the sentence opener was often not understand by the chatbot, because some topics have not been considered in the

[1]https://abeto-online.de/ep/index,id,3314.html
[2]http://www.isometrics.uni-osnabrueck.de/

design. Apart from these limitations, the students liked the game as an alternative for the training of medical interviews before performing them with human patients. The second experiment was an online study with 21 medical students (average 23.05, SD = 4.295, 15 females, 6 males). The online questionnaires (n = 21) as well as the questionnaires completed by the participants on paper during the observation (n = 7) were included in the evaluation of the questionnaire in the subject areas playability, usability, immersion and reflection. The results mostly support the findings gained in the observation study. While most of the participants liked the user interface [M = 4.18 (of 5), SD = 0.819] and understood the goal of the game [M = 3.68 (of 5), SD = 0.905], its suitability for learning [M = 3.82 (of 5), SD = 0.782] and the self-descriptiveness [M = 3.33 (of 5), SD = 0.603] were considered good (above average), and the imaginative immersion [M = 3.29 (of 5), SD = 0.076] as well as the emotional involvement (M = 2.93, SD = 1.086) showed only average values. In addition to the questionnaires, a total of 36 conversation transcripts were evaluated in order to uncover possible weak points in the text recognition module of the system, which was used to fix and further improve the AIML scripts.

Case Study: Customer Complaint Management

The case study CuCoMaG (*Customer Complaint Management Group reflection*) is a serious role-playing game for the training of customer complaint handling based on theories of consumer psychology and complaint management, originally developed in 2016 at the COLLIDE group in the context of a student master project (Doberstein et al., 2016; Othlinghaus and Hoppe, 2016). It has been re-designed and evaluated in 2019 by Othlinghaus-Wulhorst et al. (2019). The following description of the approach and game design is based on these works.

Approach

In this game, the player assumes the role of a customer service employee in the fictitious company *LittleOnes*, a producer and seller of personalized clothing for children via an online shop. The player is confronted with a chatbot in the role of a complaining customer, who is reporting a certain problem. The player is communicating with the customer through a simple chat environment. Like in the other scenarios, the player has to select a sentence opener and supplement it with free text in order to formulate a chat message. The chat setting is ideal in this use case, as it simulates everyday work situations for people working in the customer support sector. **Figure 4** is showing the general user interface of this scenario. This game has one distinctive feature that sets it apart from the previous ones presented: It offers explicit support for group reflection (Othlinghaus and Hoppe, 2016). Group reflection enables a collective exchange and thus collaborative learning (Schuster, 2010). The group reflection session, which is supposed to take place subsequent to the role play session, is designed to be guided by a trainer. This trainer is given a special group reflection tool that he can use to arrange an interactive after-action review process (Othlinghaus and Hoppe, 2016). The tool allows him to show and discuss important sequences from chat conversations of several players, review specific actions and aspects of their communication behavior. The data provided by the tool can be used to give feedback to the players and initiate group discussions to help them reflecting their actions and improving their performance.

Game Design

The game includes three different scenarios. The scenarios differ according to the type of customer used, especially in terms of conversation style (Rahim and Bonoma, 1979) and the problem situation of the customer, and thus in the level of difficulty. The first scenario serves as a base level and tutorial. The customer's problem can be solved quite easily by the player, since the conversation is reduced to the conversation phases in which only information content has to be collected and no pure "soft skill" phases have to be passed through. The conversation therefore only includes the greeting phase, the problem-solution phase and the conclusive phase. The customer in this scenario can be classified as an *integrating* customer according to the model of Rahim and Bonoma (Rahim and Bonoma, 1979), who differentiated five different styles of handling interpersonal conflicts, and is therefore open to reach a solution acceptable for both parties and is showing problem-solving behavior. According to a study conducted by Cho et al. (2002), the customer's problem is the third most common cause of non-public online customer complaints: delivery problems. The aim of this scenario is to help the player becoming acquainted with the user interface and let him walk through the basic milestones of the complaint conversation.

In the second scenario, the level of difficulty is increased. The customer is emotionally aroused because of his problem and must be calmed down. According to the classification of Rahim and Bonoma (1979), this customer is considered a *compromising* customer. The customer's problem is the most common problem within non-public online complaints (Cho et al., 2002): he has (among other things) problems with the customer service. The individualization on the delivered product is wrong and in a previous attempt to complain, the customer did not achieve a satisfactory result because there was a misunderstanding between the customer and the other member of the support staff. This makes the customer also a *follow-up complainant*, as it is the second time that he has contacted the customer service about the same problem (Stauss and Seidel, 2010). The result of the scenario is that after retrieving the database, the player learns that an error in production caused the incorrect individualization. Possible solutions to the problem in this scenario are replacement with the correct product or a refund. The goal of the scenario is to successfully pass through all five phases of a complaint process.

The third scenario is the one with the highest level of difficulty. Unlike the previous scenarios, it is less about processing the information milestones than about showing patience and applying soft skills. This customer can be classified as a *dominating* customer (Rahim and Bonoma, 1979) and a grouser (Stauss and Seidel, 2010). He is only focused on his own needs and shows little or no understanding for the other side. He tries to force a solution that is optimal for him and is looking for a continuation of the conflict. He has problems with the

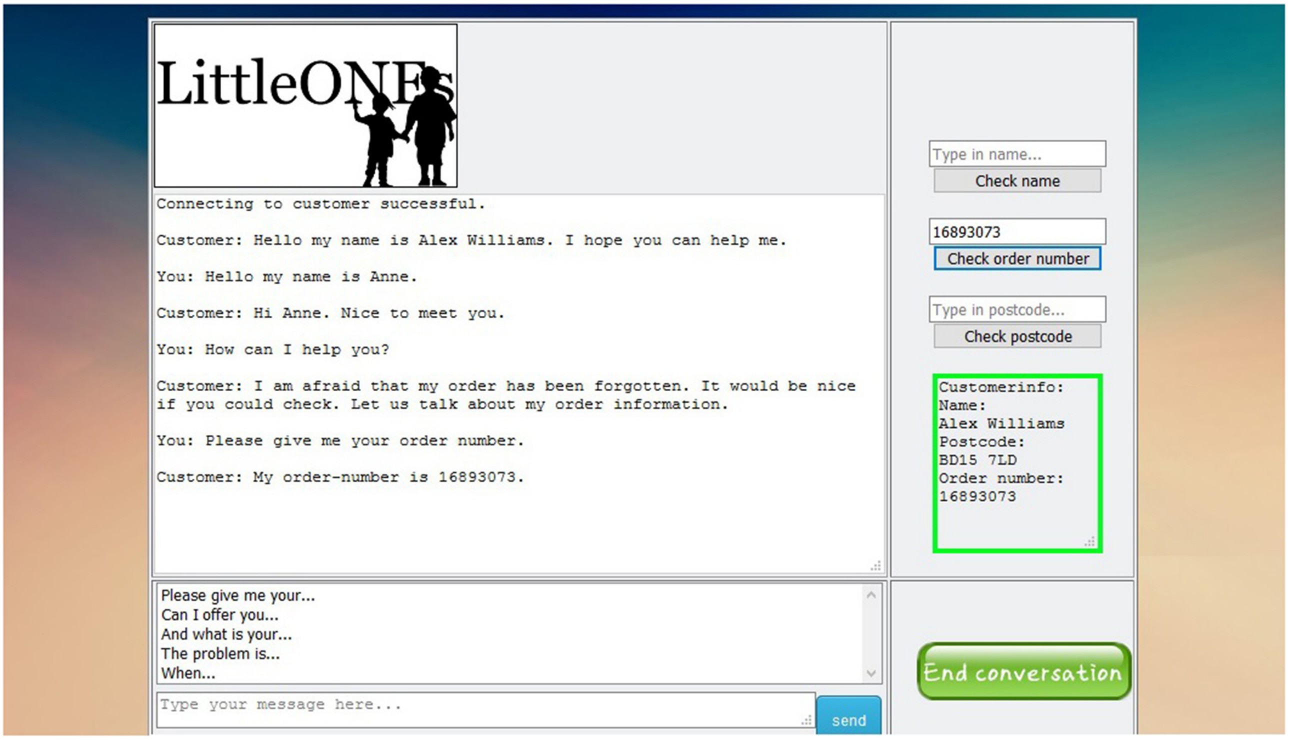


FIGURE 4 | Basic Interface of CuCoMaG.

business rules and conditions, which is the second most common problem with non-public online complaints according to Cho et al. (2002). The costumer is not reasonable and reacts abusive. When checking the database, the player learns that the customer is regularly complaining. The player's best result may be to not respond to the customer's provocations and finally end the conversation, since in this scenario, the player is not able to step into the next phase of the complaint process. This is called *active farewell* (Stauss and Seidel, 2010). The goal of this scenario is to deal with extreme situations. In this scenario, the player has to prove his ability to deal with provocations and difficult customers.

Evaluation Results

The playability, game experience, as well as the perception of the serious game itself and the interaction with the chatbot in particular have been evaluated in a mixed method study combining qualitative and quantitative methods (Othlinghaus-Wulhorst et al., 2019). The results of this study will be summarized and discussed in this section.

To investigate, whether the scenarios are perceived as realistic, if the developed scenarios' chatbots behave as intended, and whether their style of conversation is influencing how the players experience the chatbots, three hypotheses have been formulated (Othlinghaus-Wulhorst et al., 2019)—mainly relying on the game experience and perception of the chatbots (subjective measures) and evaluation of the chat transcripts (objective measures):

1. *"Participants who play the second scenario ("compromising") achieve different results in the game experience questionnaire (GEQ) dimensions tension, negative affect, and challenge than participants who play the third scenario ("dominating")."*
2. *"Participants who play the second scenario ("compromising") achieve different results in the Holtgraves questionnaire dimensions comfortable, thoughtful, polite, responsive, and engaging than participants who play the third scenario ("dominating")."*
3. *"Participants with prior experience/knowledge in complaint management achieve better results than participants without prior experience/knowledge."*

20 subjects (average 26.05, SD = 7.99, 15 females, 5 males) participated in the study. of the subjects indicated that they had prior experience in customer complaint management. All participants of the study played two scenarios, either the first ("integrating customer") and the second ("compromising customer"), or the first and the third ("dominating customer"). The distribution was randomized. Subsequent to the gaming session, the participants were asked to answer several post-experiment questionnaires to collect their experiences and perceptions during the game: (1) The *Game Experience Questionnaire* (GEQ) (Ijsselsteijn et al., 2013), (2) a questionnaire for the evaluation of educational role-playing games (Dell'Aquila et al., 2017), and (3) a questionnaire for measuring the human-like qualities of the chatbots developed by Holtgraves et al. (2007). In addition to the questionnaires and a subsequent

qualitative interview, the chat transcripts have been evaluated in regard to the answer quality of the chatbot. Based on human coding, every answer of the chatbot during the gaming session was assigned to one of three categories: *constructive*, *comprehensive*, and *nonsensical*. In order to estimate which of the predefined sentence openers have been used frequently, rarely or not at all, the frequency of uses for each one was counted.

The first hypothesis could only be partially confirmed. There were only significant differences in the dimension *negative affect*, but not in the dimensions *tension* and *challenge*. The lack of significant results could be possibly caused by methodical conditions. First, the number of participants was rather small. Second, the participants were asked to evaluate the perception of both played scenarios combined. The second hypothesis could be partially confirmed as well. Players who played the second scenario indeed showed significant differences in the dimensions *thoughtful*, *polite*, *responsive*, and *engaging*, but not in the dimension *comfortable*. As predicted, there were no significant differences in the dimensions *human* and *skilled*. This suggests that there is a difference in the style of conversation but not in the quality of the chatbots' implementation. In order to be able to examine hypothesis 3, we needed to define "success in the game." It has been determined by (a) a relative score calculated by the system and (b) the total number of inputs, as it was assumed that fast completion is an indicator for effective complaint management. Unfortunately, this hypothesis could not be tested due to the small sample size ($n = 4$).

The analysis of the chat transcripts revealed that *constructive* chatbot responses were the ones occurring most often, followed by *comprehensible* responses. The number of *nonsensical* responses has been quite low for all scenarios, which underlines the quality of the chatbot scripts. Reponses that were categorized as comprehensible, but not constructive, were default outputs, which were implemented for every sentence opener in case the free text part of the chat message was not understood by the chatbot. This way, the chatbot was able to show that he still understands the general gist or intention of the message. Sentence openers used to obtain information from the customers (e.g., "Tell me…", "Please describe…") were used most frequently, as well as the sentence opener "I am sorry…", which is not surprising, since apologies are almost always suitable in the given situation and clearly associated with polite behavior. In general, the results of the study showed that the idea and approach of the game were rated positively, but the evaluation also revealed problems, e.g., with the use of the sentence openers. It could be validated that the chatbots' style of discussion is influencing the players' perception of them, which emphasizes the successful design of the dialog scripts.

DIMENSIONS OF THE DESIGN OF SERIOUS ROLE-PLAYING GAMES FOR THE TRAINING OF SOCIAL SKILLS

The previous chapter has assembled examples instances of serious role-playing environments and ensuing empirical studies. In this chapter, we combine and inter-relate this experience with general issues in the design of serious role-playing games to devise and propose a set of design dimensions that constitute important aspects for the conceptualization, description, and comparison of serious role-playing games.

Learning Context

Many practical considerations have to be taken into account when designing serious role-playing games for the training of social skills. The probably most important is (as it is the case for every educational game) to have a clear educational purpose for using them (Whitton and Hollins, 2008). Digital games have a great motivational potential, but this potential needs to be utilized to convey the pedagogical goals and learning objectives. The goal of the game should be aligned to the learning outcomes as much as possible, otherwise learners may learn something, but it may not be what was intended. Learning objectives and intents need to be translated into concrete mechanical elements of gameplay by mapping learning mechanics and game mechanics onto each other. In games that pursue the goal to impart and train certain skills, learners should be given the opportunity to put these skills into practice in order to facilitate skills acquisition and provide a context in which these skills are useful (Naido et al., 2000).

Furthermore, the setting of the game needs to be appropriate for the learning context (Whitton and Hollins, 2008). As many studies mentioned in chapter Related Work show, role-playing games are an ideal instrument for the assessment and training of soft skills. However, the chosen scenario and storyline need to be appropriate in the given thematic context and should be described adequately for the players, so they are able to develop immediate understanding and empathy with the role they are assigned. The storyline may be fictitious, but the concepts used in it should be real to ensure that a transfer to real-word settings is possible (Pivec, 2009). Also, the desired learning outcome will not be achieved unless the correct game situation is chosen for the selected topic (Salen and Zimmerman, 2004). Another important point is that the educational design must be based on an underlying corpus of background theories. This includes general psychological and pedagogical concepts and guidelines for the design of serious games, as well as theoretical foundations of the learning material itself.

Technical Architecture and Set-Up

From the technical perspective, major issues regarding the implementation of serious role-playing games are flexibility, reusability, and extensibility/adaptability. Thanks to the use of a multi-agent blackboard architecture, our framework for scenario-based game development can easily be adapted and tailored to different settings and use cases, while the web-based gaming environment ensures easy access and platform independence. To adapt the framework to a new scenario, the following elements are needed: (1) a new GUI including sentence openers to be provided in menu selection, (2) new AIML scripts, and (3) modified or additional agents in the backend. The actual effort of course depends on the expertise of the developer.

The use of chatbots in serious role-playing games entails some major challenges. There are different approaches and technologies for natural language processing, each coming with specific advantages and disadvantages. In our approach, we use AIML as technological basis for the dialog modeling of the chatbots, but of course, there are many more approaches (e.g., data-driven technologies).

Dialog Models and Degrees of Freedom in Communication

There is a range of possibilities for introducing dialogs with virtual characters (not necessarily chatbots) in digital role-playing games. Brusk and Björk summarize different dialog models in games (Brusk and Björk, 2009): In some games, dialogs are the only way of interacting with the game, meaning that the dialog *is* the gameplay. In other games, dialogs are integrated as separate modes. Either they are taking place concurrently to other actions, or solely with no other activity occurring at the same time. In our scenarios, dialog is indeed the main gameplay element. There may be side tasks, but the focus is on the communication behavior of the players and their interaction with other characters (chatbots and/or other players, depending on the setting of the game).

In dialog-centric role play settings, one major design decision is related to the degrees of freedom in communication the players have. There is a range of communication models from fully predefined single choice inputs to free text composition. The choice is mainly depending on the setting, the narrative structure of the game, and the technical implementation. Using single choice inputs within underlying conversation trees are rather easy to implement, but provide the least freedom in communication. The players always have to select a predefined answer from a given set, and have no possibility to express themselves. Also, the game plot and the structure of the dialog is predefined. There may be decision nodes in the communication tree allowing for different lines of action, but the freedom of choice is very limited.

In our approach, we decided to integrate chatbots as dialog partners for the players. As illustrated in chapter Multi-Agent Architecture, it is very hard for natural language processing artificial intelligence to really grasp the sense of what has been said and a sophisticated chatbot design and implementation is a complex task. Thus, free text input poses a big challenge for developers. The use of sentence openers appears to be a compromise between these two ends of the spectrum. On the one hand, it limits the possible inputs, which reduces the complexity of the AIML scripts immensely and helps the chatbot to understand the general gist of a text input. On the other hand, it still offers the players the possibility to formulate their own inputs and express themselves more freely.

Feedback and Scaffolding Elements and Mechanisms

As we have shown in chapter Adaptive Feedback, feedback is crucial for ensuring the success of any serious role-playing game. It allows the learners to reflect on what happened during the role play and to analyze the consequences of their actions. In our approach, we differentiate between *ingame* and *aftergame* feedback. Ingame feedback refers to implicit feedback during the role-playing session. We realize this kind of feedback mainly through the reactions of the chatbots. Other feedback mechanisms are conceivable, but they should not corrupt or break the immersion during the role play situation. The balance between keeping the realism and immersion on the one hand and providing information on the status of the conversation as well as the players' performance and progress on the other hand is proposing a major design challenge for this kind of games. In our approach, aftergame feedback is an important point for enabling reflection processes. We consider a combination of an overall summary presented after the role play session and some kind of augmented replay of the dialog particularly helpful and promising.

Another important challenge for research and development in the area of serious role-playing games is to establish intelligent mechanisms for support and guidance (scaffolding). Learners should be provided with appropriate support in order to enable them to master the challenges of the game and achieve the learning goals. Ideally, a serious game should also adapt to the learners' level of knowledge, skills, as well as progress and current performance, as adaptation and personalization are considered key factors for education (Bellotti et al., 2010). Kickmeier-Rust and Albert suggest the introduction of micro-adaptive interventions (Kickmeier-Rust and Albert, 2010). This approach allows for interventions, support, guidance or feedback in a meaningful and personalized way, embedded in the game flow. These adaptive educational mechanisms are supposed to support the learner by hinting or providing appropriate feedback in certain situations, e.g., when misconceptions occur or when the progress is unsatisfactory (Kickmeier-Rust and Albert, 2010). The idea is to provide help to the learners by intelligently monitoring and interpreting their behavior in a non-invasive manner, which we consider a very promising approach. At this point of time, scaffolding, adaptation and personalization are incorporated in our framework only to a limited extend, thus augmenting these dimensions in our approach proposes a significant challenge for future research.

Relation Between Immersion and Reflection

As described in chapter Immersion and Reflection, one major advantage of games is their motivational and immersive potential. Immersion holds the potential to motivate learners and make them get more engaged in learning task and this potential needs to be used to full capacity in the role play situation. Getting immersed in a game requires some degree of (perceived) realism, because if learners do not perceive a scenario as realistic, they are likely to regard the game experience as irrelevant to their understanding of the real world (Sutcliffe, 2002). Thus, realism is an important characteristic of any successful serious role-playing game design. Ribbens and Malliet identified seven factors of perceived game realism: (1) simulation realism, (2) freedom of choice, (3) character involvement, (4) perceptual pervasiveness, (5) authenticity regarding subject matter, (6) authenticity regarding characters, and (7) social realism (Ribbens and Malliet, 2010).

A properly designed serious role-playing game also needs to provide support for reflection, allowing the learners to re-think and reflect on their actions. There are approaches claiming that it could be beneficial to have reflection taking place within the game itself without letting the learner step out of the game world by offering reflection activities within the game (Yusoff et al., 2009). However, as we have argued in chapter Immersion and Reflection, there is reason to assume that immersion tends to hinder the critical self-reflection, and based on this assumption, we decided to separate the actual role play phase from the reflection phase in our framework, allowing the learners to step out of the game world and their role and take over a distant perspective during the reflection phase. Following Malzahn et al. (2010), we claim that reflection needs role distance, which is not compatible with a high degree of immersion (although this is desirable during the actual role play). Accordingly, phases of enactment (role play) should be separated from reflection. Reflection phases should enable to take a third-person perspective on the prior experience, which requires an accessible/readable representation of this experience. During this phase, immersion is explicitly undesirable in order to help learners to view their own actions from the perspective of an external observer.

Collaboration Support

An increasing popularity of multi-user virtual environments and games is causing a growing interest in the use of collaborative technologies for learning scenarios and recent research is indicating the positive effects of collaborative learning (Whitton and Hollins, 2008). Collaborative learning in the context of games describes a learning situation in which more than one learner participates in a learning (game) activity pursuing a common goal (Romero et al., 2012). In collaborative scenarios, learners work together on a common goal, they share and construct a certain level of knowledge, expertise and understanding (Romero et al., 2012). Major pedagogical benefits of bringing collaborative elements in gaming environments are (among others) providing multiple perspectives, creating self-awareness of the learning process, and thus making learning authentic and relevant (Whitton and Hollins, 2008). Serious games can provide a context for solving tasks and learning together with others. Integrating collaborative elements in a serious game may increase the players' motivation and foster the development of cognitive skills (Romero et al., 2012). In addition, collaborative virtual environments allow for a detailed recording of all collaborative interactions and thus may help to get a better understanding of those (Dillenbourg, 1999). Dillenbourg claims that it should be the aim of research to determine under which conditions collaborative learning is efficient (Dillenbourg et al., 1996).

He identifies three main criteria for rich and successful collaborative learning interactions (Dillenbourg, 1999): interactivity, synchronicity and negotiability. *Interactivity* is an integral part of any collaborative situation. It is not the frequency of interactions that defines the degree of interactivity, but the extent to which the interactions influence the other persons' cognitive processes. *Synchronicity* means that persons involved in a collaborative situation wait for messages from others and process them immediately. *Negotiability* relates to the structure of collaborative dialog being more complex than a hierarchical situation. That means one person will not impose her view only based on her authority, but will (to a certain extent) argue for her standpoint, justify, negotiate, and try to convince.

We have provided an example of a collaborative scenario in chapter Case Study: Conflict Management., in which two human players are involved in a mediation talk moderated by an AI-controlled mediator. However, not in all scenarios it is desirable and reasonable to include collaborative elements. It always depends on the context, the scenario and the learning objectives. If a task can be solved by one player, there is no need for collaboration. Thus, the tasks incorporated in the game should be only solvable if players act together and there should be a common goal (Wendel et al., 2013).

CONCLUSION

In this article, we presented a technical and conceptual framework for serious role-playing games for the training of specific social skills in virtual learning environments involving chatbots in dialog-centric settings. From the design perspective, three distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) phases of immersion (role-playing) and reflection are separated to facilitate a change of perspective that is considered conducive for learning, and (3) the learning process is emphasized by means of adaptive feedback based on individual analyses. The technical conception is based on three main components: (1) AI-controlled chatbots that adapt to the player's behavior, (2) a multi-agent blackboard system as the backbone in order to keep components independent and to optimize performance due to parallel processing; and (3) intelligent support for an automated evaluation of the player's performance and feedback generation.

Different use cases based on this framework have been presented, including scenarios for the training of workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management. First evaluation studies indicate that this approach is assessed positively, the scenarios are perceived as useful and realistic and may qualify for real training situations. Due to the flexible architecture, our framework can easily be tailored to different settings and use cases and thus serve as a basis for future research focusing on the adaptation to other contexts and systems.

Our framework facilitates the building of serious virtual role playing games in that it allows for tailoring and adapting a given component architecture with very limited effort, comprising the provision of a specific GUI with sentence openers, a new set of AIML scripts (chatbots), and (possibly) a modification/extension of the backend agents. The framework provides all the basic mechanisms such as the inter-operability between GUI, chatbots, and agents through a tuple space. The basic architecture is available as a kind of modifiable prototype. Other than model-driven development (Schmidt, 2006), our approach does not use meta-level descriptions in combination with generators in the overall systems engineering process. Only the AIML-based specification of chatbot behavior could be conceived as a meta-level element. However, this is limited to one of the components and only imported and exploited in our application framework. Although our system architecture and basic mechanisms are predefined, these premises do not preclude the quality of the ensuing application instances. These depend very much on the specification of chatbot scripts as well as on the GUI design. Accordingly, our evaluations have relied on standard instruments to measure game experience and usability as the main human-oriented factors.

Based on our experience, we formulated a set of general dimensions and challenges in the design of serious role-playing games for the training of social skills. In summary, we identified six major aspects: The *learning context* builds the basis of each serious game and relates to its theoretical foundation and the desired learning outcomes. The *technical architecture and set-up* refer to technologies and tools that are used for the technical implementation of such games and the underlying system architecture. *Dialog models and degrees of freedom in communication* relates to the question of how the communication with the non-playing dialog partner(s) is carried out, structured, and controlled, from predefined answers to sentence openers to free text input. *Feedback and scaffolding elements and mechanisms* are essential for the transfer of learning to application in the real world and can be integrated in many different ways. The *relation between immersion and reflection* refers to the question whether phases of immersion and reflection overlap or occur separate from each other. *Collaboration support* relates to the number of (human) players involved in the game and the question whether it enables collaborative learning.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

JO-W has been involved in a guiding role (main responsibility) in the design, implementation, and evaluation of the serious role-playing games ColCoMa and CuCoMaG described in chapter Case Studies. JO-W and HH are (co-)authors of the relevant publications connected to these developments. JO-W wrote the main manuscript with text input from HH. All authors have reviewed and approved the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank several students and employees of the University of Duisburg-Essen who were involved in the design, development, evaluation, and publication of the mentioned use cases, namely Katharina Emmerich, Katja Neuwald, and Anna Jedich (ColCoMa), Philipp Behler, Ingo Börsting, Heike Choi, Evelyn Fricke, Stefan Liszio, Christian Klöpfel, and Anna Kizina (training scenario for medical interviews), as well as Dorian Doberstein, Nadja Agreiter, Marco Bäumer, Menglu Cui, Shaghayegh Abdollahzadegan, Diba Heidari, Nan Jiang, Markus Mentzel, Huangpan Zhang, Hao Zheng, Viet Hung Dinh, and Anne Mainz (CuCoMaG).

REFERENCES

Abdul-Kader, S. A., and Woods, J. (2015). Survey on chatbot design techniques in speech conversation systems. *Int. J. Adv. Comput. Sci. Appl.* 6, 72–80. doi: 10.14569/IJACSA.2015.060712

Annetta, L. A. (2010). The "I's" have it: a framework for serious educational game design. *Rev. Gen. Psychol.* 14, 105–113. doi: 10.1037/a0018985

Augello, A., Gentile, M., and Dignum, F. (2016). "Social agents for learning in virtual environments," in *International Conference of Games and Learning Alliance* (Cham: Springer), 133–143.

Aylett, R. S., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). "FearNOT! - an experiment in emergent narrative," in *International Workshop on Intelligent Virtual Agents* (Berlin; Heidelberg: Springer), 305–316.

Beaumont, R., and Sofronoff, K. (2008). A multi-component social skills intervention for children with asperger syndrome: the junior detective training program. *J. Child Psychol. Psychatry* 49, 743–753. doi: 10.1111/j.1469-7610.2008.01920.x

Behler, P., Börsting, I., Choi, H., Fricke, E., Liszio, S., Klöpfel, C., et al. (2013). "Eigentlich geht es mir gut - Entwicklung eines Serious Games zur patientenzentrierten Gesprächsführung," in *DeLFI 2013: Die 11. E-Learning Fachtagung Informatik*, eds A. Breiter and C. Rensing (Bonn: Gesellschaft für Informatik e.V.), 11–23.

Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2010). Supporting authors in the development of task-based learning in serious virtual worlds. *Br. J. Educ. Technol.* 41, 86–107. doi: 10.1111/j.1467-8535.2009.01039.x

Bezuijen, A. (2012). *Teamplay: the further development of TeamUP, a teamwork focused serious game* [Master's thesis]. Delf University of Technology, Delft, Netherlands.

Bosse, T., and Gerritsen, C. (2016). "Towards serious gaming for communication training - a pilot study with police academy students," in *International Conference on Intelligent Technologies for Interactive Entertainment* (Cham: Springer), 13–22.

Boud, D., Keogh, R., and Walker, D. (1985). "Promoting reflection in learning: a model," in *Reflection: Turning Experience Into Learning*, eds D. Boud, R. Keogh, and D. Walker (London: Kogan Page), 18–40.

Brusk, J., and Björk, S. (2009). "Gameplay design patterns for game dialogues," in *Proceedings of the 2009 Digital Games Research Association Conference* (London).

Cho, Y., Im, I., Hiltz, R., and Fjermestad, J. (2002). "An analysis of online customer complaints: implications for web complaint management," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (Big Island, HI: IEEE), 2308–2317. doi: 10.1109/HICSS.2002.994162

Combs, M. L., and Slaby, D. A. (1977). Social-skills training with children. *Adv. Clin. Child Psychol.* 1, 161–201. doi: 10.1007/978-1-4613-9799-1_5

Corti, K. (2006). Games-based Learning; a serious business application. *Informe PixelLearn*. 34, 1–20. Available online at: http://www.pixelearning.com/docs/seriousgamesbusinessapplications.pdf

De Freitas, S., and Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Comput. Educ.* 46, 249–264. doi: 10.1016/j.compedu.2005.11.007

Dell'Aquila, E., Marocco, D., Ponticorvo, M., Di Ferdinando, A., Schembri, M., and Miglino, O. (2017). *Educational Games for Soft-Skills Training in Digital Environments*. Cham: Springer. doi: 10.1007/978-3-319-06311-9

Di Ferdinando, A., Schembri, M., Linchan, C., Linehan, C., and Miglino, O. (2011). *Learn to Lead - A Web Based Game to Teach Leadership Theories in Vocational Courses*. *Games and Creativity in Education and Training*. Naples: Fridericiana Editrice Universitaria.

Dillenbourg, P. (1999). What Do You Mean by Collaborative Learning? *Collaborative Learning: Cognitive and Computational Approaches*. Oxford: Elsevier, 1–19.

Dillenbourg, P., Baker, M., Blaye, A., and O'Malley, C. (1996). The Evolution of Research on Collaborative Learning. Learning in Humans and Machine: Towards an Interdisciplinary Learning Science. Oxford: Elsevier, 189–211.

Doberstein, D., Agreiter, N., Bäumer, M., Cui, M., Abdollahzadegan, S., Heidari, D., et al. (2016). "CuCoMaG - group reflection support in role-playing environments," *in DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik*, ed U. Lucke et al. (Bonn: Gesellschaft für Informatik e.V.), 327–329.

Emmerich, K., Neuwald, K., Othlinghaus, J., Ziebarth, S., and Hoppe, H. U. (2012). "Training conflict management in a virtual environment," *International Conference on Collaboration and Technology* (Berlin; Heidelberg: Springer), 17–32.

Even, C., Bosser, A. G., Ferreira, J., Buche, C., Stéphan, F., Cavazza, M., et al. (2016). "Supporting social skill rehabilitation with virtual storytelling," in *Proceedings of the 29th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2016)* (Key Largo, FL), 329–334.

Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140. doi: 10.1177/001872675400700202

Flynn, R., McKinnon, L., Bacon, E., and Webb, J. (2011). "Maritime city: using games technology to train social workers - some initial results," in *International Conference on Entertainment Computing*. (Berlin; Heidelberg: Springer), 415–418.

Gebhard, P., Schneeberger, T., André, E., Baur, T., Damian, I., Mehlmann, G., et al. (2018). Serious games for training social skills in job interviews. *IEEE Trans. Games* 11, 340–351.

Gee, J. P. (2007). *Good Video Games and Good Learning: Collected Essays on Video Games, Learning, and Literacy*. New York, NY: Lang.

Gelernter, D. (1985). Generative communication in Linda. *ACM Transac. Prog. Lang. Syst.* 7, 80–112. doi: 10.1145/2363.2433

Glock, F., Junker, A., Kraus, M., Lehrian, C., Schäfer, A., Hoffmann, S., et al. (2011). ""Office Brawl" - A Coversational Storytelling Game and its Creation Process," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE 2011)* (New York: ACM).

Gresham, F., and Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) Rating Scales*. Bloomington, IN: Pearson Assessment.

Gunter, G. A., Kenny, R. F., and Vick, E. H. (2006). A case for a formal design paradigm for serious games. *J Int. Dig. Media Arts Assoc*, 3, 93–105. Available online at: https://www.academia.edu/download/49480493/Gunter_20Kenny_20Vick_20paper.pdf

Haferkamp, N., Kraemer, N., Linehan, C., and Schembri, M. (2011). Training desaster communication by means of serious games in virtual environments. *Entertain. Comput.* 2, 81–88. doi: 10.1016/j.entcom.2010.12.009

Harteveld, C. (2011). *Triadic Game Design*. London: Springer.

Holtgraves, T. M., Ross, S. J., Weywadt, C. R., and Lin, H. T. (2007). Perceiving artificial social agents. *Comput. Hum. Behav.* 23, 2163–2174. doi: 10.1016/j.chb.2006.02.017

Hubal, R. C., Frank, G. A., and Guinn, C. I. (2003). "Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training," in *Proceedings of IUI* (Miami, FL), 85–92. doi: 10.1145/604045.604062

Hunicke, R., Leblanc, M., and Zubek, R. (2004). "MDA: a formal approach to game design and game research," in *Proceedings of the Challenges in Games AI Workshop, 19th National Conference of Artificial Intelligence* (San Jose, CA: AAAI Press), 1-5.

Ijsselsteijn, W. A., De Kort, Y. A., and Poels, K. (2013). *The Game Experience Questionnaire*. Eindhoven: Technische Universiteit Eindhoven.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* 66, 641–661. doi: 10.1016/j.ijhcs.2008.04.004

Jeuring, J., Grosfeld, F., Heeren, B., Hulsbergen, M., Ijntema, R., Jonker, V., et al. (2015). *Communicate! - A Serious Game for Communication Skills. EC-TEL 2015*. Cham: Springer, 513–517.

Johnson, C. I., Bailey, S., and Van Buskirk, W. L. (2017). "Designing effective feedback messages in serious games and simulations: a research review," in *Instructional Techniques to Facilitate Learning and Motivation of Serious Games*, eds P. Wouters and H. van Oostendor (Cham: Springer), 119–140.

Jonassen, D., Mayes, T., and McAleese, R. (1993). "A manifesto for a constructivist approach to uses of technology in higher education," in *Designing Environments for Constructive Learning*, eds T. M. Duffy, J. Lowyck, and D. H. Jonassen (Berlin; Heidelberg: Springer), 231–247.

Kenny, R., and Gunter, G. (2011). Factors affecting adoption of video games in the classroom. *J. Interact. Learn. Res.* 22, 259–276.

Kerly, A., Hall, P., and Bull, S. (2006). "Bringing chatbots into education: Towards natural language negotiation of open learner models," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (London: Springer), 179–192.

Kesti,M. O., Leinonen, J., and Kesti, T. (2017). "The productive leadership game: fromtheory to games-based learning," in *Public Sector Entrepreneurship and the Integration of Innovative Business Models*, eds M. Lewandowski and B. Kozuch (Hershey: IGI Global), 238-260. doi: 10.4018/978-1-5225-2215-7.ch010

Kickmeier-Rust, M. D., and Albert, D. (2010). Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. *J. Comput. Assist. Learn.* 26, 95–105. doi: 10.1111/j.1365-2729.2009.00332.x

Knode, S., and Knode, J.-D. (2011). "Using a simulation program to teach leadership," in *Proceedings of the 2011 ASCUE Summer Conference* (North Myrtle Beach, SC), 86–92.

Kortmann, R., and Harteveld, C. (2009). "Agile game development: lessons learned from software engineering," in *Proceedings of the 40th Annual Conference of the International Simulation and Gaming Association* (Singapore).

Kruse, J. (2000). *Diagnostische Falleinschätzung bei Patienten mit Psychischen und psychosomatischen Beschwerden und Störungen in hausärztlichen Praxen* [Professorial dissertation]. Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany.

Lane, H. C., and Hays, M. J. (2008). "Getting down to business: teaching cross-cultural social interaction skills in a serious game," in *Workshop on Culturally Aware Tutoring Systems (CATS)* (Montreal, QC), 35–46.

Leary, T. (1957). *Interpersonal Diagnosis of Personality: Functional Theory and Methodology for Personality Evaluation*. Oxford: Ronald Press.

Lim, M. Y., Aylett, R., Enz, S., Kriegel, M., Vannini, N., Hall, L., et al. (2009). Towards intelligent computer assisted educational role-play. *Edutainment* 2009, 208–219. doi: 10.1007/978-3-642-03364-3_27

Linssen, J., de Groot, T., Theune, M., and Bruijnes, M. (2014). "Beyond simulations: serious games for training interpersonal skills in law enforcement," in *Proceedings of European Social Simulation Association*, 604–607. Available online at: http://ddd.uab.cat/record/125597

Malzahn, N., Buhmes, H., Ziebarth, S., and Hoppe, H. U. (2010). "Supporting reflection in an immersive 3D learning environment based on role-play," in *European Conference on Technology Enhanced Learning* (Berlin; Heidelberg: Springer), 542–547.

Marocco, D., Pacella, D., Dell'Aquila, E., and Di Ferdinando, A. (2015). "Grounding serious game design on scientific findings: the case of ENACT on soft skills training and assessment," in *Design for Teaching and Learning in A Networked World (9307)*, eds G. Conole, C. Klobučar, J. Rensing, J. Konert, and É. Lavoué (Switzerland: Springer), 441–446.

Marr, A. C. (2010). *Serious Games für die Informations- und Wissensvermittlung. Biblotheken auf Neuen Wegen*. Wiesbaden: Dinges & Frick.

Martens, A., Diener, H., Malo, S., Pan, Z., Cheok, A., Müller, W., et al. (2008). Games-based learning with computers - learning, simulations, and games. *Transac. Edutain* 1, 172–190. doi: 10.1007/978-3-540-69744-2_15

Mateas, M., and Stern, A. (2003). "Façade: an experiment in building a fully-realized interactive drama," in *Proceedings of the Game Developers Conference (GDC '03)* (San Jose,CA).

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., et al. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *Br. J. Educ. Technol.* 45, 502–527. doi: 10.1111/bjet.12067

McMahon, M., and Ojeda, C. (2008). "A model of immersion to guide the design of serious games," in *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education* (Las Vegas), 1833–1842.

Memarzia, M., and Star, K. (2011). "Choices and voices - a serious game for prevent violent extremism," in *Intelligence Management: Knowledge Driven Frameworks for Combating Terrorism and Orgamized Crime*, eds S. Yates and B. Akhgar (Heidelberg: Springer), 130–142.

Michael, D., and Chen, S. (2006). *Serious Games: Games that Educate, Train, and Inform*. Boston, MA: Thomson Course Technology.

Mitchell, A., and Savill-Smith, C. (2005). *The Use of Computer and Video Games for Learning. A Review of the Literature*. London: LSDA.

Murray, J. H. (2017). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge; London: MIT press.

Naido, S., Ip, A., and Linser, R. (2000). Dynamic goal-based role-play simulation on the web: a case study. *J. Educ. Technol. Soc.* 3, 190–202. Available online at: https://www.jstor.org/stable/10.2307/jeductechsoci.3.3.190

Othlinghaus, J., and Hoppe, H. U. (2016). "Supporting group reflection in a virtual role-playing environment," in *International Conference on Intelligent Technologies for Interactive Entertainment* (Cham: Springer), 292–298.

Othlinghaus-Wulhorst, J., Jedich, A., Hoppe, H. U., and Harrer, A. (2018). "Using eye-tracking to analyze collaboration in a virtual role play environment," in *International Conference of Collaboration and Technology* (Cham: Springer), 185–197.

Othlinghaus-Wulhorst, J., Mainz, A., and Hoppe, H. U. (2019). "Training customer complaint management in a virtual role-playing game: a user study," in *European Conference on Technology Enhanced Learning* (Cham: Springer), 436–449.

Pivec, P. (2009). *Game-Based Learning or Game-Based Teaching*? British Educational Communications and Technology Agency (BECTA), Report No. 1509.

Prensky, M. (2000). *Digital Game-Based Learning*. New York, NY: McGraw-Hill.

Proksch, S. (2010). *Konfliktmanagement in Unternehmen. Mediation als Instrument für Konflikt- und Kooperationsmanagement am Arbeitsplatz.* Heidelberg: Springer.

Rahim, A., and Bonoma, T. V. (1979). Managing organizational conflict: A model for diagnosis and intervention. *Psychol. Rep.* 44, 1323–1244. doi: 10.2466/pr0.1979.44.3c.1323

Ribbens, W., and Malliet, S. (2010). Perceived digital game realism: a quantitative exploration of its structure. *Pres. Teleoperat. Virtual Environ*. 19, 585–600. doi: 10.1162/pres_a_00024

Ritterfeld, U., Cody, M., and Vorderer, P. (2009). *Serious Games: Mechanics and Effects.* New York, NY: Routledge.

Romero, M., Usart, M., Ott, M., Earp, J., and de Freitas, S. (2012). "Learning through playing for or against each other? Promoting collaborative learning in digital game based learning," in *Proceedings of the 20th European Conference on Information Systems (ECIS) 2012* (Barcelona). Available online at: http://aisel.aisnet.org/ecis2012/93

Salen, L., and Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge; London: The MIT Press.

Schatzki, T. R. (1996). *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge: Cambridge University Press.

Schmidt, D. C. (2006). Model-driven engineering. *IEEE Comput.* 39:25. doi: 10.1109/MC.2006.58

Schuster, R. J. (2010). Gruppenreflexion als Kommunikationsinstrument. Management und Wirtschaft. *Schriftenreihe zur Wirtschaftlichen Forschung*, 13, 7–23. Available online at: http://www.fhvie.ac.at/Forschung/Publikationen/Schriftenreihe

Schwantes, U., and Kampmann, M. (2007). Gesundheitsorientierte Gesprächsführung. *Der Mensch* 38. Available online at: https://www.dachverband-salutogenese.de/cms/der-mensch/heft-38-12007/

Shawar, B. A., and Atwell, E. (2007). Chatbots: are they really useful? *Ldv Forum* 22, 29–49.

Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795

Squire, K., and Jenkins, H. (2003). Harnessing the power of games in education. *Insight* 3, 5–33.

Stauss, B., and Seidel, W. (2010). "Complaint management," in *Introduction to Service Engineering*, eds G. Salvendy and W. Karwowski (Hoboken, NJ: John Wiley & Sons), 414–432.

Susi, T., Johannesson, M., and Backlund, P. (2007). *Serious Games: An Overview. Technical Report HS-IKI-TR-07-001*. Skövde: School of Humanities and Informatics, University of Skövde.

Sutcliffe, M. (2002). "Simulations, games and role-play," in *Handbook for Economics Lecturers*, ed P. Davies (Bristol: The Higher Education Academy, The Economics Network), 1–26.

Totty, M. (2005). Better training through gaming. *Wall Street J. Eastern Ed.* 245:R6.

Vaassen, F., and Wauters, J. (2012). "deLearyous: training interpersonal communication skills using unconstrained text input," in *Proceedings of ECGBL* (Cork), 505–513.

Van Dijk, D., Dick, R., Hunneman, R., and Wildlevuur, S. (2008). Self City: "Training social skills in a game," in *Proceedings of Second European Conferences on Game-based Learning* (Academic Publishing Limited) (Barcelona), 481–488.

Walk, W., Görlich, D., and Barrett, M. (2017). "Design, dynamics, experience (DDE): an advancement of the MDA framework for game design," in *Game Dynamics* eds O. Korn and N. Lee (Cham: Springer), 27–45.

Wallace, R. (2004). *The Elements if AIML Style*. San Francisco, CA: ALICE AI Foundation.

Weinbrenner, S. (2012). *SQLSpaces - A Platform for Flexible Language-Heterogeneous Multi-Agent Systems* [Doctoral dissertation]. Verlag Dr. Hut, Munich, Germany.

Wendel, V., Gutjahr, M., Göbel, S., and Steinmetz, R. (2013). Designing collaborative multiplayer serious games. *Educ. Inform. Technol.* 18, 287–308. doi: 10.1007/s10639-012-9244-6

Whitton, N., and Hollins, P. (2008). Collaborative virtual gaming worlds in higher education. *ALT-J* 16, 221–229. doi: 10.3402/rlt.v16i3.10900

Winkler, R., and Söllner, W. (2018). *Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis.* Chicago, IL: Academy of Management Annual Meeting (AOM).

Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Pres. Teleoperat. Virtual Environ.* 15, 309–329. doi: 10.1162/pres.15.3.309

Yusoff, A., Crowder, R., Gilbert, L., and Wills, G. (2009). "A conceptual framework for serious games," in *2009 Ninth IEEE International Conference on Advanced Learning Technologies* (Riga: IEEE), 21–23. doi: 10.1109/ICALT.2009.19

Ziebarth, S., Kizina, A., Hoppe, H. U., and Dini, L. (2014). "A serious game for training patient-centered medical interviews," in *14th International Conference on Advanced Learning Technologies* (Athens: IEEE), 213–217. doi: 10.1109/ICALT.2014.69

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 07 August 2020
doi: 10.3389/fcomp.2020.00023

Check for updates

A Virtual Tour of a Hardly Accessible Archaeological Site: The Effect of Immersive Virtual Reality on User Experience, Learning and Attitude Change

Christos Kyrlitsias[1,2†], Maria Christofi[1,2†], Despina Michael-Grigoriou[1,2], Domna Banakou[3] and Andri Ioannou[2,4]*

[1] GET Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus, [2] CYENS Centre of Excellence, Nicosia, Cyprus, [3] Event Lab, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Spain, [4] CIL Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus

OPEN ACCESS

Edited by:
Elvira Popescu, University of Craiova, Romania

Reviewed by:
Alin Moldoveanu, Politehnica University of Bucharest, Romania
Marco Temperini, Sapienza University of Rome, Italy

***Correspondence:**
Despina Michael-Grigoriou
despina.grigoriou@cut.ac.cy

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *30 January 2020*
Accepted: *09 June 2020*
Published: *07 August 2020*

Citation:
Kyrlitsias C, Christofi M, Michael-Grigoriou D, Banakou D and Ioannou A (2020) A Virtual Tour of a Hardly Accessible Archaeological Site: The Effect of Immersive Virtual Reality on User Experience, Learning and Attitude Change. Front. Comput. Sci. 2:23. doi: 10.3389/fcomp.2020.00023

Some archaeological sites are not easily accessible by visitors due to mobility or geographical restrictions. Digital technology can make such sites virtually accessible and provide educational information at the same time. Toward this goal, we created a digital reconstruction of the archaeological site of Choirokoitia. Given that a 3D digital reconstruction can be used along with different technologies, we designed and developed an interactive application, where users can navigate and get information about the site, for two different systems: Virtual Reality (VR) systems and desktop computers. A feasibility study was conducted where we compared aspects of the two systems so as to allow the suggestion of the proper technology to utilize according to a user's aims. The results showed higher levels of presence and more positive experience by the participants who used the VR system compared to those who used the desktop version. On the other hand, greater learning gains were demonstrated in participants who used the desktop version compared to those who used the VR version. No differences were shown between the two groups regarding the participants' change of attitudes toward the archaeology of Cyprus.

Keywords: virtual reality, cultural heritage, immersion, learning, user experience, attitude change

INTRODUCTION

People around the world visit archaeological and historical sites mainly for pleasure or to learn about the place as such and its history, or even the combination of both, in which case it is called educational tourism. Important archaeological sites around the world are not always easily accessed by everyone due to the remote location of the site itself or mobility difficulties of possible visitors. With the current advances in technology, such places can be accessed remotely through interactive multimedia applications. With ubiquitous computing evolving and devices, such as VR headsets allowing more immersive experiences compared to traditional desktop computing, virtual heritage is gaining increasing interest. According to Roussou (2002), a well-designed virtual heritage application aiming to benefit public education should visualize and provide "access" to sites and places that no longer exist, are geographically remote, or are unreachable.

The archaeological site of Choirokoitia (pronounced "Khirokitia"), a Neolithic settlement in Cyprus, is the site of interest in this work. The site was selected carefully in order to meet all the criteria we would like to address. These include the following:

(i) It is a site that is not easily accessible by everyone: the physical terrain of Choirokoitia's settlement is rough and non-uniform, with a lot of changes in elevations, and it is located on a natural steep-sloped hill, making access to people with mobility difficulties (e.g. wheelchairs) impossible, according to the official website [The Deputy Ministry of Tourism (2019)].
(ii) It is a site that attracts educational interest: the country's Ministry of Education and Culture organizes educational visits for primary school students since learning material on this specific settlement is included in the history curriculum in public schools. However, the visits are possible only for students in nearby cities. The latter adds an extra dimension to the limited access to the settlement as listed above in (i), in this case, due to geographical restrictions.
(iii) It is one of the most visited tourist attractions in Cyprus: Choirokoitia is an archaeological site with ruins dated back to 7000BC and is ranked among the top oldest ruins in the world. It was also listed by UNESCO as a United Nations Educational, Scientific and Cultural Organization World Heritage Site (1998), giving additional motivation for tourists to visit it.

The aim of this work is two-fold:

(i) Enable virtual access to the archaeological site of Choirokoitia that can be beneficial for people with disabilities and for those separated from it by geographical and political boundaries.
(ii) Suggest the most appropriate technology to use according to the user's aims, which may be related to learning or to user experience.

The first goal was addressed with the development of two versions of the application, one for VR systems and the other for personal computers. The second goal has been addressed by performing a feasibility study comparing the two different versions and assessing them on a number of different aspects, namely, participants' sense of presence, level of user experience, change in attitudes toward the archaeology of Cyprus, and learning performance.

Although several studies have been conducted on Virtual Heritage applications the topic needs to be further addressed since there have been no clear conclusions. A previous study (Christofi et al., 2018) conducted specifically for Choirokoitia was limited to assessing a VR application integrating its virtual reconstruction, and it lacked a comparison between different types of technologies. Also, some other limitations of the study, such as the small number of participants and the motion sickness experienced by a substantial percentage of them, made the results less robust.

In another study, Tost and Economou (2009) investigated the suitability of immersive VR for learning about archaeology and the past in cultural heritage settings at the Hellenic Cosmos (the exhibition center of the Foundation of the Hellenic World in Athens) but again lacked in comparison with other technologies as to its effectiveness. The results of their study demonstrated that each exhibit supported a different kind of learning, and the exhibits were considered suitable for obtaining a global idea of spatial details. In a study by Michael et al. (2010), different museum exhibits, one traditional and five interactive Information and Communication Technologies exhibits, were compared with the main emphasis being on the assessment of user experience (UX). Their results showed that the Information and Communication Technologies exhibits were rated higher than the traditional exhibit. In another comparative study, Wrzesien and Raya (2010) compared the E-Junior application, a serious virtual world (SVW) for teaching children natural science and ecology, with a traditional type of class. With regard to learning effectiveness, the results did not present statistically significant differences between the two groups. However, students from the virtual group reported enjoying the class more, being more engaged, and having greater intentions to participate than students from the traditional group. Similarly, a study by Zaharias et al. (2013) investigated the user experience (UX) and learning effectiveness of the "Walls of Nicosia," a 3D multi-touch table, as compared to a traditional approach where a group of students took a guided tour throughout the museum and learned about the walls of Nicosia through printed maps exhibited at the museum. Results showed no statistically significant differences in the learning performance, but the virtual group reported user experience at significantly higher levels. Finally, Loizides et al. (2014), presented two museums using an immersive head-mounted display (HMD)-based technology and a stereoscopic Powerwall and evaluated the users' overall experience. The results of the user evaluation revealed that both ways of presenting the museum received an equal usability score among the users.

In the next section, we describe the two versions (VR and Desktop) that were developed of an interactive application representing the Choirokoitia settlement and the experimental study that was conducted to evaluate their feasibility. The analysis of the collected data and results are then described, while a discussion elaborating on findings, describing the present study's limitations, and outlining potential future work follows.

MATERIALS AND METHODS

Materials

We used a three-dimensional (3D) digital reconstruction of the archaeological site of Choirokoitia (**Figures 1**, **2**), which was created based on real photographs taken on the site and highly accurate modeling of the geometry. We also recorded a set of audio clips with historical information about Choirokoitia. Using the 3D reconstruction and the recorded audio, we developed a new version of an existing Interactive VR Application described in Christofi et al. (2018). Additionally, we developed an Interactive Desktop Application in such a way so as to provide exactly the same information with a VR Application in order to compare the medium. The two applications, the VR Application and the Desktop Application, that have been developed and evaluated through the feasibility study are described below.

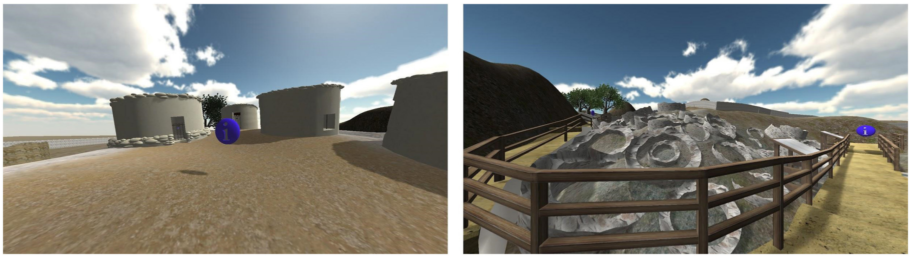


FIGURE 1 | Views from the 3D reconstruction of the archaeological site of Choirokoitia where information points are shown. The virtual reconstruction visualizes the archaeological site as it stands today, including the real reconstructed houses located near the settlement **(left)** and the ancient ruins **(right)** of the settlement.

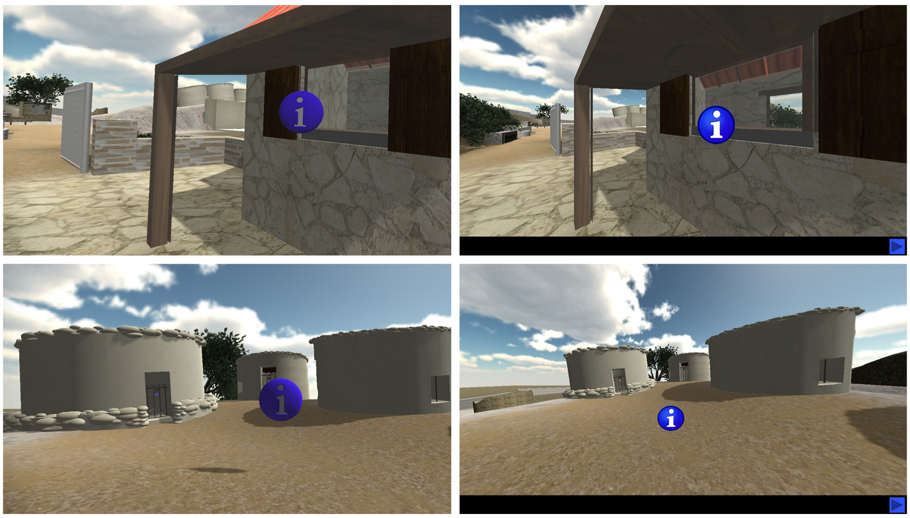


FIGURE 2 | Views from the 3D reconstruction of the archaeological site of Choirokoitia, as shown in the VR Application (top **left** and bottom **left**) and the Desktop Application (top **right** and bottom **right**).

VR Application

The VR Application used in this study is an improved version of the previously developed VR Choirokoitia application. The new version integrates a recent technique developed to minimize the dizziness that was occurring to a great extent to the participants, as is stated in the relevant publication (Fernandes and Feiner, 2016). This technique requires a dynamic reduction of the field of view (FoV) depending on the user's movement and rotation in the virtual environment. Specifically, the FoV was gradually decreased, using a vignetting effect (**Figure 3**) as the user's navigation (movement and rotation) speed increased, and it increased as the user's navigation speed decreased. The change in the FoV was slow and gradual, so as not to distract the user. Note that the dynamic reduction of the FoV only takes into account the navigation using the video game controller, that causes the dizziness, and not the movement and rotation applied using the HMD's head tracking. To further improve the previous version of the VR Application, we also minimized the intervention with the participants during the time that they were immersed in the VR Application. In this new version, we gave them all of the instructions through the application and achieved their familiarization with the application's functionality by having a training session at the very beginning of the experience. The training phase took place in a virtual space outside the virtual archaeological site, in a neutral scene, to avoid receiving information about Choirokoitia before the main application began.

In the main VR Application, participants were free to navigate through the 3D reconstruction of Choirokoitia using a video game controller. The application required that the participants put on a VR HMD (**Figure 4**, left), which offered a first-person stereoscopic view of the environment and the ability to physically change the looking direction with head rotations. While navigating, participants listened, through the HMD headsets' headphones, to eight audio recordings, in total, containing information related to the archaeological site. The audio clips lasted ∼30 s each. The recordings were triggered by collision with 8 information points, visualized by floating 3D icons (**Figures 1**, **2**, top left; **Figure 2**, bottom left) that were located at various parts of the virtual archaeological site. The information presented in each of those recordings was related to the specific area the participant was exploring at that time.

FIGURE 3 | Participants using the VR Application **(left)** and Desktop Application **(right)**.

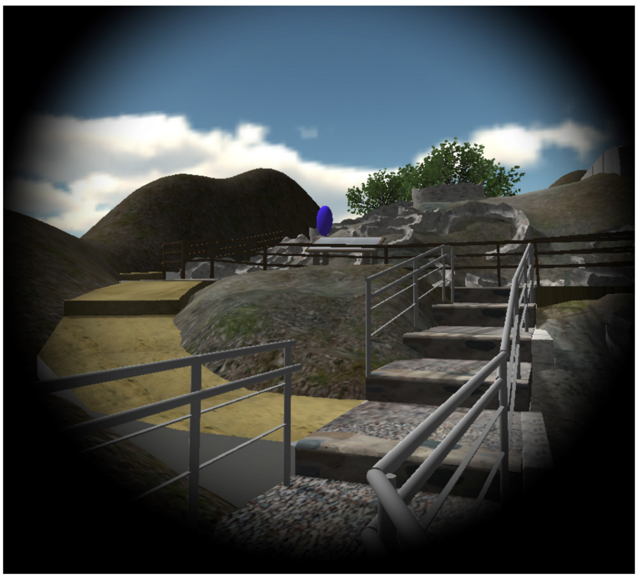


FIGURE 4 | Example of dynamic reduction of the field of view in the VR Application.

Desktop Application

In order to test the impact of VR, we designed a non-immersive Desktop Application as a control. It provides the user with the same visual and audio information but without the immersive features of the VR technology (3D stereo projection, free navigation, free head movements based on head tracking). The Desktop Application was designed in such a way that it would be easy for a person with basic computer skills to develop, in contrast to the VR Application, which requires advanced skills to develop, such as programming skills. The Desktop Application was developed using 2D rendered images based on the same 3D reconstruction of the site as was used in the VR Application instead of real photographs of the archaeological site, so as to not provide different visual content. The 2D rendered images were created in regard to specific points of view selected in such a way so as to have the rendering of areas that included all eight points of interest, where information points existed, as well as areas along the route between any two consecutive information points (**Figure 2**). Similar to the VR Application, in order to minimize the intervention with the user while conducting the study, we provided all the instructions through digital material at the beginning of the Desktop Application.

Interactivity was achieved with the mouse button (**Figure 2**, top right; **Figure 2**, bottom right). The participants were able to move to the next rendering of a virtual area by clicking on a button that was placed on the bottom right corner of the screen. Information points were triggered by clicking on corresponding small icons similar to those in the VR group. The participants in the Desktop group received identical audio recorded information as those in the VR group while they were looking at the rendering from the corresponding viewpoint.

Technical Setup

The 3D reconstruction of Choirokoitia was created in Autodesk Maya 2015 and was textured using as reference the original photographs taken from a visit to the archaeological site. The reconstruction was done with high precision and accuracy based on photos and maps and included not only the archaeological buildings but also the surrounding area as it stands today (**Figure 1**). The stereo audio recordings were edited in Adobe Audition CS6. The VR Application was developed with Unity 2017.1.2 software, and the virtual environment was displayed through an Oculus Rift CV HMD. This has two 1,080 × 1,200 pixel OLEDs per eye at a 90-Hz display refresh rate, coupled with a positional tracker and built-in headphones. A video game controller was used for navigation inside the virtual space. The interactive Desktop Application was created using Microsoft PowerPoint. We used screenshots of renders of 3D reconstructions within the Unity software in order to achieve a visual resemblance to the VR Application. The application was displayed to participants through a 1,920 × 1,080 pixel 15.6-inch computer screen. A computer mouse was used for interaction with the application.

TABLE 1 | Experimental design and distribution of participants by condition.

	Condition	
	VR	Desktop
N (Males/Females)	20 (7/13)	20 (10/10)
Mean ± S.D Age	24.75 ± 5.48	28.55 ± 8.9
Median Code VR Experience (IQR)	2 (1)	2 (1)
Median Code Archaeological Site Visits (IQR)	3 (1)	3 (1)

Experimental Design

The study had a between-groups design. Participants were randomly assigned to one of two experimental groups: the VR group or the Desktop group. The VR Application was used by the participants assigned to the VR group, and the Interactive Desktop Application was used by the participants in the Desktop group.

Ethics Statement

All participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individuals for the publication of any potentially identifiable images or data included in this article.

Participants

Forty adults ($N = 40$), male and female participants, aged 20–54 years (mean ± SD age 26.65 ± 7.53), participated in the study. After signing an informed consent form, participants were randomly allocated to either the VR or the Desktop group.

Participants had no or little prior experience with VR technology. **Table 1** presents relevant descriptive data for each condition, the total number of participants, mean ages, median and IQR values for experience in VR (1 = 0, 2 ="<1," 3 ="1–2, … 5 =">5"), and frequency of previous visits to archaeological sites. Codes refer to a 1–5 Likert scale, where 1 indicates least agreement and 5 most agreement with the statement.

Measurements

Participants were assessed on their sense of presence, level of user experience, attitudes toward the archaeology of Cyprus, and learning performance through questionnaires.

Presence

The sense of presence was measured with a three-item questionnaire (**Table 2a**) based on the presence questionnaire developed by Slater et al. (1994). Presence can be characterized as "the illusion of 'being there' in the environment depicted by the VR displays" (Slater and Sanchez-Vives, 2016). Presence was rated on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). This questionnaire was given to participants right after their experience.

User Experience

In order to assess the user experience, we designed a four-item questionnaire (**Table 2b**) evaluating the overall experience, the image and audio quality, and whether participants would be willing to try a similar experience in the future. Participants rated their experience on a Likert scale ranging from 1 (negative) to 5 (positive). This questionnaire was also given to the participants right after their experience.

TABLE 2 | The questionnaires given to participants.

Variable	Question
(a) Presence	
There	How much did you feel as if you were present at the archaeological site? (1. Not at all. 5. Very much)
Reality	Were there moments during your experience that the virtual world became reality for you and you almost forgot about the real world where the study took place? (1. Never. 5. All the time)
Place	Did you have a stronger feeling that you were in the real world of the laboratory or at the virtual Choirokoitia? (1. Laboratory. 5. Choirokoitia)
(b) User experience	
Image	How do you rate the image quality? (1. Very bad. 5. Excellent)
Audio	How do you rate the audio quality? (1. Very bad. 5. Excellent)
Pleasure	How pleasant did you find the experience? (1.Very unpleasant. 5.Very pleasant)
Repeat	Would you like to try a similar experience in the future? (1.Not at all. 5.Very much)
(c) Attitudes	
VisitChoirokoitia	Do you wish to visit the Choirokoitia site in the future? (1.Not at all. 5.Very much)
VisitSites	Do you wish to visit any other archaeological sites in the future? (1.Not at all. 5.Very much)
LearnChoirokoitia	Do you wish to learn more about the site of Choirokoitia? (1.Not at all. 5.Very much)
LearnArchaeology	Do you wish to know more about the archaeology of Cyprus? (1.Not at all. 5.Very much)

Attitudes

A four-item questionnaire (**Table 2c**) on a Likert scale ranging from 1 (negative) to 5 (positive) was designed to measure participants' attitudes toward the archaeology of Cyprus. This is interpreted as the intention to acquire further knowledge about Choirokoitia or to visit this or a different archaeological site in the future. This questionnaire was given to the participants before their exposure (preAttitudes) and again right after their exposure (postAttitudes). The variable of interest was the *Change in Attitude* (dAttitudes = postAttitudes – preAttitudes), where positive values indicate an increase in participants' intention to learn more about the archaeology of Cyprus or visit archaeological sites and negative values a decrease.

Learning Performance

Participants' learning about the archaeological site was assessed with a 10-statement multiple-choice questionnaire that evaluated participants' knowledge of the archaeological settlement of Choirokoitia (e.g., "What was the shape of the houses in the archaeological site?," "What was the average life expectancy of the inhabitants of the archaeological site?," etc.). These questions were based on the information that participants had been given (audio clips) while using the applications. Each correct answer received 1 point, while wrong answers received 0 points. The knowledge test was administered two times, one prior to participants' virtual exposure (preScore) and the second immediately after their virtual exposure (postScore). The response variable of interest *Learning Performance* was the difference between the two (dScore = postScore – preScore), which shows the degree of improvement (positive values) or decline (negative values) in score after exposure.

Procedure

Upon arriving at the laboratory, and after completing the consent form, participants were randomly assigned to the VR or the Desktop group and were asked to complete the Knowledge and Attitudes questionnaires alongside some demographic questions. Once this process was completed, participants in the VR group were asked to put on the VR HMD (**Figure 4**, left), and participants in the Desktop group were asked to sit in front of the computer screen (**Figure 4**, right). Then, both groups were instructed to follow the respective tutorial instructions in order to become familiar with the use of the corresponding system and application. Once the tutorial was completed (with an approximate duration of 5 min), the main experimental session began, which lasted ∼8 min for both groups. After completing their navigation within virtual Choirokoitia, the participants of both groups again completed the Knowledge and Attitudes questionnaires and also filled in the User Experience and Presence questionnaires.

Statistical Analysis

For the questionnaire data on Presence, User Experience, and Attitudes (**Table 2**), factor analysis was carried out to reduce the number of questionnaire variables. This method also has the advantage of transforming ordinal variables to continuous ones. Corresponding factor scores were used, and the interpretation of each factor was identified. For factor analysis on ordinal variables, such as in our case, Polychoric PCA analysis can be used as a test (this treats ordinal variables as if they were derived from cut-offs sampled from a normally distributed variable) (Olsson, 1979) and scores derived from those, so this approach was employed. Box plots of all raw questionnaire scores are provided in the **Supplementary Material**. Subsequent analysis of the derived variables was done using independent samples *t*-test. All results were obtained with Stata 13 software.

RESULTS

Factor Analysis of the Questionnaires

A single factor was retained in the case of Presence (**Table 2a**), and the factor loadings on the scoring variable (*Presence)* are shown in **Table 3**. In the interpretation of the factor loadings, these capture the amount of overall variance in the observed variables. The scoring coefficients are the coefficients of the equations describing the factor scores in terms of a linear combination of the original variables. In **Table 3**, F1 is explained

TABLE 3 | Factor analysis for presence resulting in a single factor *F1*, and the scoring coefficients for the factor score *Presence*.

Variable	Factor loadings	Scoring coefficients
	F1	Presence
There	0.937	0.356
Reality	0.926	0.352
Place	0.946	0.360

TABLE 4 | Factor analysis for user experience resulting in a single factor F1, and the scoring coefficients for the factor score *UserExperience*.

Variable	Factor loadings	Scoring coefficients
	F1	User experience
Pleasure	0.854	0.299
Image	0.924	0.324
Audio	0.772	0.270
Repeat	0.824	0.288

TABLE 5 | Factor analysis for participants' attitudes before the experience resulting in a single factor F1, and the scoring coefficients for the factor score preAttitudes.

Variable	Factor loadings	Scoring coefficients
	F1	preAttitudes
VisitSites	0.844	0.272
VisitChoirokoitia	0.927	0.299
LearnArchaeology	0.940	0.303
LearnChoirokoitia	0.802	0.259

by *There, Reality*, and *Place* (**Table 2a**), which is reflected in the corresponding scoring coefficients for *Presence*. The latter interprets this factor as *"the illusion of being at the archaeological site of Choirokoitia instead of in the physical laboratory."*

The User Experience questions (**Table 2b**) resulted in a single factor, and the factor loadings on the scoring variable *UserExperience* are shown in **Table 4**. This factor is explained by *Pleasure, Image, Audio*, and *Repeat,* which is interpreted as "*the level of pleasantness of the experience.*"

The attitudes questionnaire was administered before and after the participants' experience. Factor analysis on the questions (**Table 2c**) resulted in a single factor, and the factor loadings on the scoring variables *preAttitudes* and *postAttitudes*, respectively, are shown in **Tables 5**, **6**. The factors are explained by *VisitSites*, *VisitChoirokoitia, LearnArchaeology*, and *LearnChoirokoitia* (**Table 2c**) and are interpreted as participants' "*interest in acquiring new historical knowledge in the future.*"

Presence Analysis

Figure 5 shows the bar chart of the means and standard errors of the derived factor analysis variable *Presence.* Participants in the VR condition felt a stronger sense of "being" at the archaeological

TABLE 6 | Factor analysis for participants' attitudes after the experience resulting in a single factor F1, and the scoring coefficients for the factor score postAttitudes.

Variable	Factor loadings	Scoring coefficients
	F1	postAttitudes
VisitSites	0.835	0.263
VisitChoirokoitia	0.931	0.293
LearnArchaeology	0.940	0.296
LearnChoirokoitia	0.856	0.269

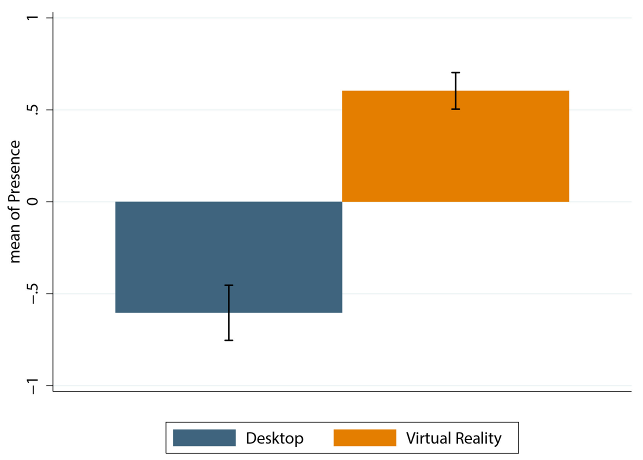


FIGURE 5 | Bar chart showing means and standard errors for the factor score *Presence* by condition.

site of Choirokoitia (0.59 ± 0.098) compared to participants in the Desktop condition (−0.59 ± 0.148). An independent samples t-test showed that the above differences are significant [$t_{(38)} = -4.712, p = 0.000$].

User Experience Analysis

Figure 6 shows the bar chart of the means and standard errors of the derived factor score *UserExperience*. Participants in the VR condition rated the overall experience of using the application more positively (0.394 ± 0.172) compared to participants in the Desktop condition (−0.394 ± 0.238). An independent samples t-test showed that the above differences are significant [$t_{(38)} = -2.6809, p= 0.0108$].

Attitudes Analysis

We investigated whether participants' attitudes toward archaeology were affected by the use of the VR and Desktop Applications. The variable of interest, *dAttitudes* = *postAttitudes* – *preAttitudes*, reflects the degree of change in attitudes regarding acquiring new knowledge in the future (positive values). **Figure 7** shows the bar chart of the means and standard errors of *dAttitudes*, derived as the difference from the factor scores *preAttitudes* and *postAttitudes*. Although it can be seen that there is a slight positive change for the VR condition (0.013 ± 0.102) and a negative change for the Desktop condition

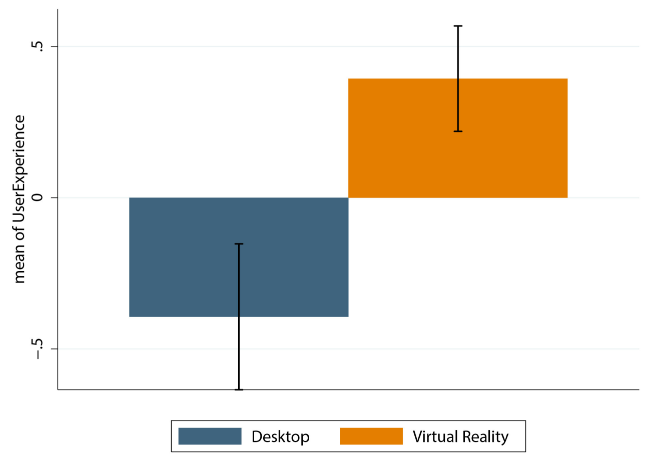


FIGURE 6 | Bar chart showing means and standard errors for the factor score *UserExperience* by condition.

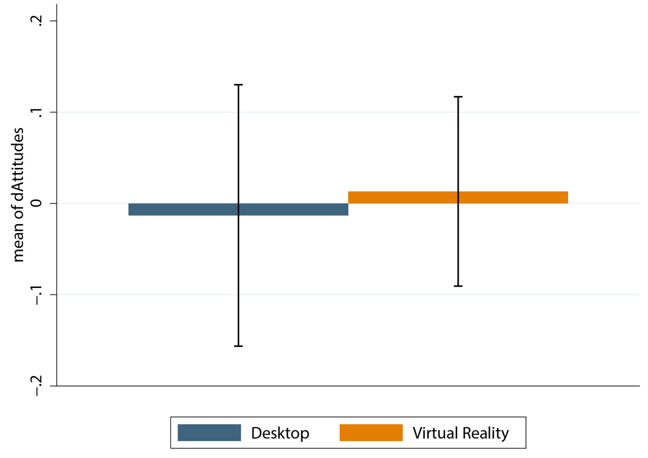


FIGURE 7 | Bar chart showing mean and standard error of dAttitudes by condition.

(−0.013 ± 0.141), the differences between the two groups are not significant [$t_{(38)} = -0.151, p = 0.88$].

Learning Performance Analysis

The response variable of interest here is *dScore* = *postScore* – *preScore*, showing the degree of improvement (positive values) or decline (negative values) in score. The score is defined as the total number of correct responses to the knowledge questionnaire before and after the exposure. **Figure 8** shows that the mean change in *dScore* was greater in the Desktop than in the VR condition. The means and standard errors are 2.5 ± 0.336 and 1.35 ± 0.386, respectively, with Cohen's d = 0.71, which is a medium to large effect size. An independent samples t-test showed a significant difference between the means $t_{(38)} = 2.247$, $p = 0.031$. The residual errors are compatible with normality (Shapiro-Wilk $P = 0.35$).

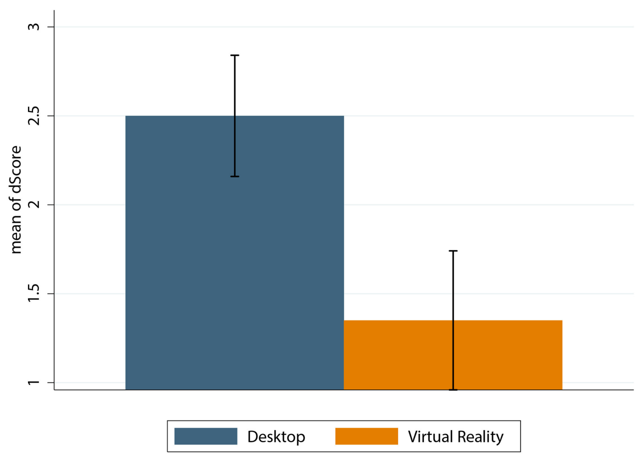


FIGURE 8 | Bar chart showing mean and standard error of dScore by condition.

Correlations

Finally, we performed a Pearson correlation analysis in order to explore relationships between the data collected. The correlations for the Desktop and the VR condition are summarized in **Supplementary Tables 1**, **2**, respectively.

The sense of Presence strongly correlated with User Experience in both the Desktop [$r_{(18)} = 0.83$, $p < 0.001$] and VR [$r_{(18)} = 0.836$, $p < 0.001$] conditions. This result means that participants who reported a higher sense of *Presence* in the virtual environment, also reported a higher level of *UserExperience* (how positively they rated the overall experience of using the application).

Presence is correlated with *preScore* in the Desktop condition [$r_{(18)} = 0.61$, $p < 0.004$]. Also, in the Desktop condition, *Presence* correlated with *preAttitudes* [$r_{(18)} = 0.561$, $p = 0.01$] and *postAttitudes* [$r_{(18)} = 0.619$, $p = 0.004$]. *UserExperience* and *preScore* were correlated in the Desktop condition [$r_{(18)} = 0.606$, $p = 0.005$]. Finally, *UserExperience* correlated with *preAttitudes* [$r_{(18)} = 0.491$, $p = 0.028$] and *postAttitudes* [$r(18) = 0.569$, $p = 0.009$] in the VR condition.

DISCUSSION

The aims of this study were (i) to enable virtual access to the archaeological site of Choirokoitia and (ii) to examine the feasibility of the use of the immersive interactive VR application developed in comparison to an application that can be easily developed by a person with basic computer skills. For this reason, we virtually reconstructed the archaeological site of Choirokoitia and developed and compared two applications, an immersive VR Application and a Desktop Application. This comparison was focused on four main aspects, namely, participants' sense of presence, user experience, attitudes toward archaeology, and learning performance regarding historical facts of Choirokoitia, in order for the results to suggest the most appropriate technology to utilize according to a user's aims, which may be related to learning or to user experience.

First, we found that the VR Application has a critical advantage in terms of the sense of presence delivered to participants, which refers to the sense of "being there" in the virtual world (here, the virtual Choirokoitia) (Slater and Sanchez-Vives, 2016). The VR Application succeeded in giving participants the feeling of virtually visiting the archaeological site of Choirokoitia, more so than the Desktop Application, and it was also found that the overall user experience was significantly better in terms of image and audio quality, overall pleasantness, and willingness to try a similar application in the future. This is in line with previous research and proposed models for complex conceptual learning (Slater and Wilbur, 1997; Salzman et al., 1999; Lee et al., 2010; McMahan, 2013). The latter suggests that immersive environments create a strong sense of presence due to the very powerful emotional impact they can have, which then leads to higher engagement in the experience, motivation, and cognitive processing of the material.

However, our results demonstrated no advantage of the VR Application regarding participants' attitudes toward archaeology. The use of the VR Application did not seem to impact participants' motivation to acquire new knowledge in the future more than the Desktop Application. Furthermore, the VR Application was found to be less effective in acquiring and memorizing new information about the archaeological site. Interestingly, participants in the Desktop group gained more knowledge than the participants in the VR group. Previous studies have shown that low-immersion simulations, such as computer games and other desktop applications, can result in better cognitive performance and attitudes toward learning (Bonde et al., 2014; Clark et al., 2016; Makransky et al., 2016; Thisgaard and Makransky, 2017). However, research evidence as to whether high-immersive virtual reality applications lead to increased motivational outcomes and learning is still controversial. In the study described in Passig et al. (2016), the authors showed that teaching in immersive VR environments contributed to students' cognitive modifiability more than traditional learning experiences. Similar results were reported by Alhalabi (2016) when using a VR system to enhance students' education in engineering, and by Webster (2016) when aiming to improve learning on basic corrosion prevention and control in military personnel. In contrast, other studies have yielded negative results when comparing learning in immersive VR environments and desktop applications. In one example, Moreno and Mayer (2002) investigated how desktop VR and immersive VR, the second using an HMD and navigation techniques, compared with multimedia learning material. It was found that the two media did not affect students' performance differently, with equal improvement results for both. Similarly, Richards and Taylor (2015) found that biology students' knowledge did not improve after exposing them to virtual simulations with 3D models more than with simulations using two-dimensional (2D) models. The authors concluded that this could have occurred due to the additional cognitive load imposed by the 3D models. Likewise, in previous research comparing an immersive VR human anatomy application with traditional slide presentations, it was found that although both methods increased participants' performance, this was higher for participants in the slide-presentation group (Michael-Grigoriou et al., 2017). Our present results are in line with the aforementioned studies, showcasing that indeed, immersive VR technologies can have an overwhelming impact on participants, leading to diminished attention and a lower increase in their learning performance.

According to the cognitive load theory (CLT), if one engages in excessive amounts of extraneous processing (i.e., cognitive processing that does not support the goal, caused for instance by distractions), then there is not an adequate capacity for critical processing and thus meaningful learning outcomes (Sweller, 1994). Additionally, Thisgaard and Makransky (2017) suggests that highly immersive environments might not necessarily result in higher learning and transfer outcomes due to their highly hedonic or utilitarian nature. According to this theory, fun or pleasurable experiences within immersive VR environments can lead users to disregard their instrumental value, and instead, concentrate on the entertainment value such systems offer.

Based on the above and in line with previous studies, we speculate that something similar happened in the present study, with participants in the VR group spending more time exploring and navigating through the new (to them) immersive environment, thus leading to poor attention to other stimuli and consequently, worse performance in the knowledge test. This has been previously supported in the context of immersive virtual gaming apps, where the authors explain that navigation through virtual rooms can draw attention away from the main task and negatively influence players' learning (Freina and Canessa, 2015). Further, and according to Van der Heijden (2004), we could argue that participants focused more on enjoying the environment rather than on learning the material presented to them. We do not know, however, how performance would have been affected if the participants had had more experience in using VR environments. In our study, all participants had little or no prior knowledge of VR systems, and therefore, this point cannot be addressed here, but we leave it as an open question for further research. Also, regarding motivation, it should be noted that participants in both experimental groups reported positive attitudes toward archaeology even before their virtual exposure (see **Supplementary Figure 3**), which could possibly explain why there were no differences in attitudes after the exposure. Further, attitudes and motivation were based on self-reported questionnaires, which are oftentimes subject to social desirability bias (Yu et al., 2018). We propose that future studies should look at alternative and more indirect ways of measuring motivation, such as, for example, follow-up forum access and participation related to activities on the given subject (either on the same day or in the long-term).

A limitation of our study is that there are differences not only in the technology used in the two groups (VR vs. Desktop) but also in the features of the applications as such. For example, the VR Application offers the user the possibility to freely navigate in a 3D environment, while the Desktop version provides the visual information in 2D renders. While this simplification on the Desktop Application was intensional, as it allows its development by people with basic computer skills, it does not allow us to safely attribute the differences we detected between the VR group and Desktop group either to the type of the technology used (immersive or not) or to the application as such. A future study where exactly the same application is used, visualized through a VR system or a Desktop computer, will allow a conclusion on this aspect.

This VR Application of Choirokoitia attempted to provide the best possible representation of the archaeological site, giving the user an experience that simulates a physical visit. The reconstruction used in this paper represents the current state of the archaeological site. Achieving this goal can make the site (at least virtually) accessible, overcoming mobility or geographical constraints. But it only exploits a part of the possibilities of VR technologies. VR can overcome the boundaries of reality, time, or space. For instance, users could experience "being" in the Choirokoitia of the Neolithic period, when the settlement was inhabited. Such an application would give the user visual information on how the settlement looked and the way of life and occupations of the inhabitants. This paper has examined whether VR can enhance and promote archaeological content in comparison to a non-immersive medium. Although VR seems to have an advantage in terms of Presence and User Experience, regarding Learning Performance, new ways to convey information to the users should be explored. According to Tost and Economou (2009) and Roussou (2002), in order to be suitable for learning, cultural heritage virtual worlds should not just be visually represented in a photorealistic manner; they need to be complete, interactive and present the virtual world in a meaningful and engaging way. In our application, the information that the knowledge test was based on was conveyed using audio. It is possible, that the participants in the VR group, overwhelmed by the technology, paid less attention to the audio information, and this might explain their lower increase in learning performance. Therefore, in future studies, a Neolithic-period Choirokoitia application in which the information would be rendered mainly visually could overcome the limitation of knowledge transfer.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CK and MC made a substantial contribution to the design of the study, the development of the VR Application, the data acquisition, the analysis and interpretation of data, and the drafting of the article. DM-G contributed substantially to the conception and design of the study, the interpretation of the data, and the drafting and critical revision of the manuscript and supervised and coordinated all of the steps of the study. DB contributed substantially to the analysis and interpretation of the data for the work and to the drafting and critical revision of the manuscript. AI contributed to the interpretation of the data for the work and revising the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

ACKNOWLEDGMENTS

The authors would like to acknowledge Zoe Anastasiadou, Maria Michaelidou, Ioanna Papamichael, and Katerina Pieri, students of the course MGA444: Virtual Reality of the Department of Multimedia and Graphic Arts of the Cyprus University of Technology for sharing with us the original version of the Choirokoitia VR Application, which was developed under the supervision of the course's instructor, DM-G. We would also like to thank Christos Hadjipanayi and Maria Matsangidou, research associates with the Research Centre on Interactive Media Smart Systems and Emerging Technologies for proofreading and partially reviewing some sections of the article, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00023/full#supplementary-material

REFERENCES

Alhalabi, W. (2016). Virtual reality systems enhance students' achievements in engineering education. *Behav. Inform. Technol.* 35, 919–925. doi: 10.1080/0144929X.2016.1212931

Bonde, M. T., Makransky, G., Wandall, J., Larsen, M. V., Morsing, M., Jarmer, H., et al. (2014). Improving biotech education through gamified laboratory simulations. *Nat. Biotechnol.* 32, 694–697. doi: 10.1038/nbt.2955

Christofi, M., Kyrlitsias, C., Michael-Grigoriou, D., Anastasiadou, Z., Michaelidou, M., Papamichael, I., et al. (2018). "A tour in the archaeological site of Choirokoitia using virtual reality: a learning performance and interest generation assessment," in *Advances in Digital Cultural Heritage* (Cham: Springer), 208–217.

Clark, D. B., Tanner-Smith, E. E., and Killingsworth, S. S. (2016). Digital games, design, and learning: a systematic review and meta-analysis. *Rev. Educ. Res.* 86, 79–122. doi: 10.3102/0034654315582065

Fernandes, A. S., and Feiner, S. K. (2016). "Combating VR sickness through subtle dynamic field-of-view modification," in *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (Greenville, SC: IEEE), 201–210. doi: 10.1109/3DUI.2016.7460053

Freina, L., and Canessa, A. (2015). "Immersive vs desktop virtual reality in game based learning," in *European Conference on Games Based Learning*, Steinkjer, Norway (Reading: Academic Conferences International Limited), 195.

Lee, E. A. L., Wong, K. W., and Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Comput. Educ.* 55, 1424–1442. doi: 10.1016/j.compedu.2010.06.006

Loizides, F., El Kater, A., Terlikas, C., Lanitis, A., and Michael, D. (2014). "Presenting cypriot cultural heritage in virtual reality: a user evaluation," in *Euro-Mediterranean Conference* (Cham: Springer), 572–579.

Makransky, G., Thisgaard, M. W., and Gadegaard, H. (2016). Virtual simulations as preparation for lab exercises: assessing learning of key laboratory skills in microbiology and improvement of essential non-cognitive skills. *PLoS ONE* 11:e0155895. doi: 10.1371/journal.pone.0155895

McMahan, A. (2013). "Immersion, engagement, and presence: a method for analyzing 3-D video games." in *The Video Game Theory Reader* (New York, NY: Routledge), 89–108.

Michael, D., Pelekanos, N., Chrysanthou, I., Zaharias, P., Hadjigavriel, L. L., and Chrysanthou, Y. (2010). "Comparative study of interactive systems in a museum," in *Euro-Mediterranean Conference* (Berlin; Heidelberg: Springer), 250–261.

Michael-Grigoriou, D., Yiannakou, P., and Christofi, M. (2017). "Intuitive interaction for exploring human anatomy in a VR setup," in *2017 23rd International Conference on Virtual System and Multimedia (VSMM)* (Dublin: IEEE), 1–4.

Moreno, R., and Mayer, R. E. (2002). Learning science in virtual reality multimedia environments: role of methods and media. *J. Educ. Psychol.* 94:598. doi: 10.1037/0022-0663.94.3.598

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika* 44, 443–460. doi: 10.1007/BF02296207

Passig, D., Tzuriel, D., and Eshel-Kedmi, G. (2016). Improving children's cognitive modifiability by dynamic assessment in 3D Immersive Virtual Reality environments. *Comput. Educ.* 95, 296–308. doi: 10.1016/j.compedu.2016.01.009

Richards, D., and Taylor, M. (2015). A Comparison of learning gains when using a 2D simulation tool versus a 3D virtual world: an experiment to find the right representation involving the Marginal Value Theorem. *Comput. Educ.* 86, 157–171. doi: 10.1016/j.compedu.2015.03.009

Roussou, M. (2002). Virtual heritage: from the research lab to the broad public. *Bar Int. Ser.* 1075, 93–100.

Salzman, M. C., Dede, C., Loftin, R. B., and Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Pres. Teleoperat. Virtual Environ.* 8, 293–316. doi: 10.1162/105474699566242

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI.* 3:74. doi: 10.3389/frobt.2016.00074

Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Pres. Teleoperat. Virtual Environ.* 3, 130–144. doi: 10.1162/pres.1994.3.2.130

Slater, M., and Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Pres. Teleoperat. Virtual Environ.* 6, 603–616. doi: 10.1162/pres.1997.6.6.603

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learn. Instruct.* 4, 295–312. doi: 10.1016/0959-4752(94)90003-5

The Deputy Ministry of Tourism. (2019). *Choirokoitia Archaeological Site (Neolithic Settlement)*. Available online at: www.visitcyprus.com/index.php/en/discovercyprus/rural/sites-monuments/item/757-choirokoitia-archaeological-site-neolithic-settlement

Thisgaard, M., and Makransky, G. (2017). Virtual learning simulations in high school: effects on cognitive and non-cognitive outcomes and implications on the development of STEM academic and career choice. *Front. Psychol.* 8:805. doi: 10.3389/fpsyg.2017.00805

Tost, L. P., and Economou, M. (2009). Worth a thousand words? The usefulness of immersive virtual reality for learning in cultural heritage settings. *Int. J. Architect. Comput.* 7, 157–176. doi: 10.1260/147807709788549367

United Nations Educational, Scientific and Cultural Organization World Heritage Site (1998). *Choirokoitia*. Available online at: whc.unesco.org/en/list/848

Van der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quarterly* 28, 695–704. doi: 10.2307/25148660

Webster, R. (2016). Declarative knowledge acquisition in immersive virtual learning environments. *Interact. Learn. Environ.* 24, 1319–1333. doi: 10.1080/10494820.2014.994533

Wrzesien, M., and Raya, M. A. (2010). Learning in serious virtual worlds: evaluation of learning effectiveness and appeal to students in the E-Junior project. *Comput. Educ.* 55, 178–187. doi: 10.1016/j.compedu.2010.01.003

Yu, P. L., Shen, C., and Smith, G. S. (2018). *Relevance and Application of Heritage in Contemporary Society* (New York, NY: Routledge). doi: 10.4324/9780203702277

Zaharias, P., Michael, D., and Chrysanthou, Y. (2013). Learning through multi-touch interfaces in museum exhibits: an empirical investigation. *J. Educ. Technol. Soc.* 16, 374–384. Available online at: www.jstor.org/stable/jeductechsoci.16.3.374

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

BRIEF RESEARCH REPORT
published: 25 August 2020
doi: 10.3389/fcomp.2020.00034

The Use of a Serious Game to Assess Inhibition Mechanisms in Children

Maura Crepaldi [1], Vera Colombo [2], Stefano Mottura [2], Davide Baldassini [2], Marco Sacco [2], Alice Cancer [3] and Alessandro Antonietti [3]*

[1] Department of Human Science, University of Bergamo, Bergamo, Italy, [2] STIIMA (Institute of Intelligent Industrial Systems and Technologies for Advanced Manufacturing), CNR, Milan, Italy, [3] Department of Psychology, Catholic University of the Sacred Heart, Milan, Italy

The design and implementation of a serious game (SG) concerning inhibition skills in children are presented. The SG consists of a set of activities, each eliciting the tendency to respond in an immediate and inappropriate (wrong) way. The SG is based on the Dual Pathway model of attention-deficit/hyperactivity disorder (ADHD) proposed by Sonuga-Barke and on the Unity/Diversity model of executive functions proposed by Miyake. In the SG, children must block impulsive tendencies, reflect upon the situation, inhibit irrelevant thoughts, and find the non-immediate solution. A study was carried out by testing the SG on typically developing primary school children (30 children, 16 boys; age, $M = 9.30$ years, $SD = 0.87$) to verify that it measures the same variables addressed by tests usually employed to assess attention ability in children and to diagnose ADHD. Three standardized tasks belonging to the Italian Battery for ADHD were administered, as well as an *ad hoc* questionnaire devised to check the acceptability, usability, and comprehensibility of the SG. Positive correlations between impulsiveness as measured by standard tests and impulsiveness scores in the SG emerged. These findings support the notion that skills associated with the control of impulsivity are involved in the SG. Furthermore, self-report ratings in the questionnaire showed that the SG is easy to be understood, is engaging, and elicits positive reactions in children.

Keywords: ADHD, hyperactivity, attention, impulsiveness, serious game, children, dual pathway model

OPEN ACCESS

Edited by:
Carlos Vaz De Carvalho, Polytechnic Institute of Porto, Portugal

Reviewed by:
Camino Fidalgo, University of Zaragoza, Campus de Teruel, Spain
Walter Gerbino, University of Trieste, Italy
Antonio Rizzo, University of Siena, Italy

***Correspondence:**
Maura Crepaldi
maura.crepaldi@unibg.it

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: 27 December 2019
Accepted: 10 July 2020
Published: 25 August 2020

Citation:
Crepaldi M, Colombo V, Mottura S, Baldassini D, Sacco M, Cancer A and Antonietti A (2020) The Use of a Serious Game to Assess Inhibition Mechanisms in Children. Front. Comput. Sci. 2:34. doi: 10.3389/fcomp.2020.00034

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a neurodevelopmental disorder included in the Diagnostic and Statistical Manual of Mental Disorders (DSM-V; American Psychiatric Association, 2013). ADHD is characterized by inattention and hyperactivity. Such symptoms involve the core of executive functions (EFs) (Miyake et al., 2000; Diamond, 2013), namely:

- *Inhibition* is defined as the deliberate overriding of a dominant or assertive response (Miyake and Friedman, 2012). Inhibition is a multicomponential construct itself and comprehends different abilities such as managing impulses and interferences (Nigg et al., 2004; Diamond, 2013).
- *Impulsivity*, which occurs in rapid actions that are taken without reflection, resulting from a desire for immediate reward or an inability to delay gratification (Martinez et al., 2016). Impulsivity is not negative at all, but it becomes disabling when it interferes with decision making and doing actions (Kahneman and Tversky, 1974). Impulsivity is often linked with a multitude of behaviors or responses that are poorly conceived, premature, or inappropriate and that frequently result in unwanted or deleterious outcomes (Daruna and Barnes, 1993).

- *Cognitive flexibility* or *shifting*, namely, the capacity to change perspective over a task. To do so, people need to deactivate previous perspectives and to retrieve or devise a new one. This involves the ability to adjust and change demands or priorities, as well as to take advantage of unexpected opportunities (Diamond, 2013).
- *Working memory* (WM) refers to the ability to hold in mind and manipulate useful information available for a short time while completing a task when the information is no longer present (Baddeley and Hitch, 1994; Smith and Jonides, 1999). Diamond (2013) theorized a bidirectional relation between inhibition and WM, by describing them as separate dimensions yet strongly related.

The features of ADHD can be described also from a pathophysiological point of view (Cortese and Castellanos, 2012; Sonuga-Barke and Taylor, 2015). The currently dominant neurocognitive model of ADHD stresses the role of dysfunctions underpinned by disturbances in the frontodorsal striatal circuit and associated dopaminergic branches (Sonuga-Barke, 2005). In contrast, motivationally based accounts focus on altered reward processes and implicate frontoventral striatal reward circuits and those mesolimbic branches that terminate in the ventral striatum, especially the nucleus accumbens (Sonuga-Barke, 2003, 2005).

ADHD is meant as a motivational style characterized by attempts to escape or avoid delay of rewards, which arises from fundamental disturbances in reward centers (Sonuga-Barke, 2003). While traditionally regarded as competing, these models have been presented as complementary accounts of two psychopathophysiological subtypes of ADHD with different developmental pathways, underpinned by different corticostriatal circuits and modulated by different branches of the dopamine system (Sonuga-Barke, 2005). In the Dual Pathway model of ADHD, the first pathway (the cognitive one) concerns EFs, including WM and inhibition, while the motivational path is linked to delay sensitivity and aversion (i.e., the tendency to choose a smaller immediate reward rather than waiting for a larger delayed reward) (Sjöwall et al., 2013). In this perspective, ADHD is considered as depending on hypersensitivity to reward-related delay and characterized by an abnormal sensitivity to reinforcement (reward, punishment, and response cost) (Sonuga-Barke, 1996).

The cognitive path of the Dual Pathway model shows some overlaps with the Unity/Diversity model of EFs proposed by Miyake et al. (2000), and Miyake and Friedman (2012), which focuses on three aspects of EFs: updating WM, shifting, and inhibition. The model by Miyake states that the key requirement for response inhibition is the ability to be captured by common EFs, whereas stopping itself may be relatively automatic (Chatham et al., 2012). Hence, inhibition deficits should merit special attention in the assessment and treatment of ADHD.

Considering the neurocognitive aspect of ADHD is fundamental in order to plan functional and useful long-term effect training (Chacko et al., 2018).

Technological Tools Addressed to ADHD

Technology represents a potentially useful instrument to help children with ADHD to manage and monitor behavior (Simons et al., 2016; Powell et al., 2019). The literature presents some applications and games supporting parents in managing and monitoring children (e.g., by suggesting daily routines, such as getting dressed and going to bed); other apps, instead, aim at enabling children to monitor their own behavior (Powell et al., 2017). With respect to gaming solutions, Kulman (2012) claimed that playing a computer game involves EFs and metacognitive skills. In fact, the problem-solving skills needed to accomplish a task within the game imply to identify the relevant strategies to be applied (Antonietti et al., 2000), an ability that is poorly developed in people with ADHD. Furthermore, since children with ADHD get bored fast, it is important to constantly catch their attention. Serious games (SGs) can meet these requirements because the player has to stay focused all the time. Moreover, all the senses of the players are stimulated, and this makes the activities more interesting (Wronska et al., 2015). Finally, children with ADHD prefer immediate reward, and they need to maintain the experience at a "minimum waiting level" (Sonuga-Barke, 1996). All these aspects can be found in technological solutions such as SGs, which, therefore, represent valuable tools to engage children with ADHD.

Several studies demonstrate the usefulness of SGs for enhancing some positive attitudes (Beale et al., 2007) in children with ADHD (Rassin et al., 2004) by increasing problem-solving strategies (Coyle et al., 2007) and by modifying abnormal behaviors (Walshe et al., 2003). An example is the Harvest Challenge BCI Videogame (Muñoz et al., 2015), which consists of three mini-games that work with the use of neurofeedback. Plan-It Commander (Bul et al., 2015) is an online computer game structured in two parts: the mission game, formed of three isolated mini-games with embedded learning goals, and a closed social community interacting through predefined messages. Braingame Brian is an SG aimed at improving WM in children with ADHD aged 8–12 years (Van der Oord et al., 2014). PlayMancer (Fernández-Aranda et al., 2012) has been proven able to change underlying attitudinal, behavioral, and emotional processes of patients with impulse-related disorders, which are similar to ADHD. The Virtual Classroom (Rizzo et al., 2002; Coleman et al., 2019) was developed for the study, assessment, and rehabilitation of cognitive processes in patients with different forms of central nervous system dysfunction. It offers an assessment in a real-world dynamic simulation with distractors that mimic situations typically occurring in a classroom (Parsons et al., 2019).

In addition to SGs, recent studies applied Augmented Reality to rehabilitation of children with ADHD. As an example, the project Beyond the tReatment of the Attention deficit hyperactiVity disOrder (BRAVO) aims at implementing an immersive therapeutic game context to improve the relationships between young patients and therapies. The solution, based on an ICT system, combines the use of SGs, wearable equipment, and Virtual and Augmented Reality devices (Barba et al., 2019).

SGs as Assessment Tools for Children With ADHD

Even though the assessment was one of the aims of the SGs mentioned before, they were mainly designed as training tools. Therefore, data supporting their use as reliable tests for assessment are lacking. Furthermore, most of them are not grounded on a specific model of the mechanisms underlying ADHD. Technology, which should be reliably helpful, must meet some quality standards; however, SGs are often developed rapidly and are based on little evidence (Powell et al., 2017). Hence, the need for an SG based on updated and well-reputed theories and properly tested as an assessment tool is still unmet. The goal of the SG described in this paper is to address such a need.

Being inhibition involved in the cognitive pathway of ADHD, according to the development of the Unity/Diversity model, the two main goals of the SG are to assess the control of impulsivity and to prevent non-adaptive behaviors and irrelevant thoughts through inhibition. The purpose is to provide a technological tool able to assess to what extent children with ADHD handle impulsivity by implementing functional inhibition strategies.

Symptoms of ADHD are basically assessed with clinical interviews, behavioral observation, standardized tests or questionnaires, and information gathered from parents, teachers, and clinicians (Mühlberger et al., 2020). Recent studies show that SGs concerning EFs are useful for exploring the differences between children with ADHD and controls, as well as among ADHD subtypes (Areces et al., 2019). In this paper, we describe the structure of an SG aimed to assess inhibition mechanisms in children. We verified whether the abilities the SG involves correspond to those measured by some of the standard tests usually employed to diagnose lack of inhibition in ADHD, in order to be sure that the SG in the future could be used to assess ADHD.

The SG is addressed to children aged 8–11 because this is the period when children begin to master efficient self-management strategies and are able to generalize them in different everyday-life contexts (Welsh et al., 2006; Prencipe et al., 2011).

Antonyms: A SG for Children With ADHD

The SG, Antonyms, was designed as a single storyline that includes different scenarios. In this way, the child has an overview of the activities he/she will have to perform, which should help him/her to carry out every single adventure. Many existing SGs are made up of mini-games. This approach allows the child to focus on one aspect at a time. Antonyms consist of a series of activities, each eliciting the tendency to respond in an immediate, impulsive way (see the flow chart in **Appendix 1**). In the SG, children must block such a tendency, reflect upon the situation, and find the non-intuitive solution.

The SG can run on a tablet, a smartphone, or a touchscreen laptop. The player interacts with the environments by tapping on the screen or by dragging and dropping objects. If no touch screen is available, he/she can interact through the keyboard using the "Space bar" and Return key. Antonyms has been developed in Unity3D, a game development platform, based on C# programming language. A more detailed description of the SG architecture is reported by Crepaldi et al. (2017). Scores computed in each part of the game can be exported in an Excel database, and all the results are saved immediately for each player.

The SG is characterized by the presence of a narrative frame that involves players in all parts of the story. The player personifies a superhero (called Atansyon), who is asked to save a realm on the opposite side of the Earth (Antonyms). Atansyon will face different steps to free the planet from enemies. The tasks that require to inhibit responses are contextualized. As a consequence, the activation of inhibition mechanisms has a specific meaning, which is not arbitrary (as it happens in many SGs).

The complete version of the SG consists of three different scenarios. In the present study, only two scenarios are taken into account: *Training School* and *Central Building*. At the beginning of the game, the player can choose the main settings of the game: screen resolution, graphic quality, and monitor. After the main settings, the player is required to enter his/her name, surname, age, and gender to create a new user profile. For each scenario of the game, there is a user menu with the options: "Start," "Settings," "Instructions," "Audio," "Back," and "Quit" (**Figure 1**).

Each mini-game has different levels of difficulty so that the activities become more challenging. The transition from a level to the next one is constrained to the number of errors made by children. It is worth mentioning the multimodal, detailed, and immediate feedback provided by the SG in the form of both visual and auditory messages, which have been proven to be highly beneficial to people with ADHD (Fabio and Antonietti, 2012). Moreover, it is possible to examine the player's behavior during the SG by saving the performance in the form of different types of errors (e.g., errors in waiting, wrong answer, etc.) and time spent.

The activities of the SG were inspired by classical neuropsychological tasks addressing several subcomponents of inhibition and attention, such as (1) the *Flanker task* (Eriksen and Eriksen, 1974), in which the target is flanked by non-target stimuli, which correspond either to the same direction response as the target, to the opposite response, or to neither (respectively: congruent, incongruent, and neutral flankers); (2) the *Stroop test* (Stroop, 1935; Dalrymple-Alford, 1972), assessing the ability to inhibit an automatic, incorrect response and to produce a non-automatic, correct response; and (3) the *Stop signal task* (Lappin and Eriksen, 1966; Logan, 1994) in which the subject must respond to an arrow stimulus by selecting one of two options, depending on the direction in which the arrow points. If an audio tone is present, the subject must withhold the production of that response.

The mechanisms underlying these tasks were adapted to create a new assessment tool with a more ecological background, addressing inhibitory control abilities. In a different investigation, Antonyms was administered to a small group of children with ADHD (Crepaldi et al., 2020) who showed significantly lower performances than a matched group of typically developing children, thus suggesting the SG sensitivity.

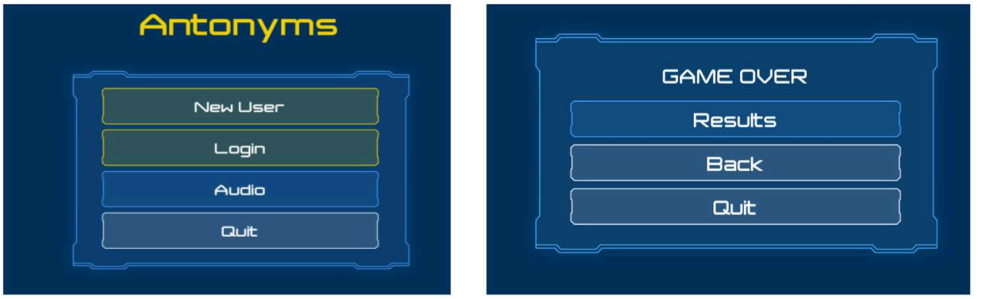


FIGURE 1 | Settings in antonyms.

METHOD

SG Concurrent Validity: A Pilot Study

Teachers and children, belonging to schools different than those involved in the study, have been involved in the first part of the project, when they were shown examples of the tasks included in the SG and could play a prototype (pilot application) of the SG. A series of features of the SG were changed, according to the feedbacks received about the pilot application. This phase of the project, even though not based on a systematic assessment of the usability and ergonomics of the game but rather on qualitative feedbacks, was nevertheless useful to check the acceptability of the SG. The new version of the game better matched the initial idea to create an SG that is realistic and able to engage children.

Design of the SG: Scenarios and Activities

Two activities from the Antonyms SG were tested in the present study:

- *Training School (TS)* (**Figure 2**) is a same-different task addressing visual selective attention and inhibitory control. In the center of the screen, a target stimulus (an object useful to the mission) appears, and the player must compare it with the object that appears on a shelf: if they are the same, the player takes the object from the shelf and puts it in the backpack; otherwise, he/she has to put it in the trash in the shortest possible time. The activities are divided into two different levels: a warm-up (level 1) and the actual game (level 2). The objects from the shelf are dragged and dropped using touch; they should be the same as the given target, but slight differences can hamper the success. A total of 30 items appear in each level of the task, 15 of which are "different" trials, and 15 are "same" trials. The order of appearance of the objects and their position on the shelf are random with the only constraint that the two correct objects from two consecutive groups of stimuli cannot appear in the same position. Errors are classified as Trash Errors (putting the correct object in the trash) and Bag Errors (putting a wrong object in the bag), by referring to the signal detection theory (Swets, 1964), which categorizes, in same-different tasks, incorrect responses in either false alarms or miss responses. In the Training School task, Trash Errors are considered false alarms, since the player incorrectly detected a difference between identical stimuli, whereas Bag Errors are considered miss responses, since the player failed to detect the difference between stimuli. The correct responses and errors in the second level of the game (since the first level was a warm-up task, the actual performance was recorded during the most challenging level) were used to compute a sensitivity score as a measure of discriminability (Macmillan and Creelman, 2004).
- *Central Building (CB)* (**Figure 3**) addresses sustained attention and the ability to hold automatic responses. The player walks through the corridors to arrive at the central room. Movements are driven by colored lights appearing on the floor that the player has to select either touching directly the screen or by pressing the Space bar on the keyboard. He/she must walk only when the path is highlighted with a proper light (green light) according to a randomized sequence whose timing is referred to the Conners (1994) Continuous Performance Test, the most used instrument to measure ADHD impairment in impulsivity and inhibition (Mühlberger et al., 2020), which requires respondents to select a response to a particular stimulus that appears at regular intervals (Fang and Dai Han, 2019). The interstimulus intervals are 1, 2, and 4 s with display time of 250 ms. When a different light (blue light) appears or there is no light on the screen, the player must stop. Scores are assigned to the child based on the speed of the response (touch the screen/press the spacebar): five points for an answer in <1 s and 10 points for an answer from 1 to 3 s. Two paths have been created differing in direction and predictability. Both paths can be completed in about 5 min in case of no errors. Starting from these two paths, we designed two levels of difficulty for each route, and, as in the previous scenarios, the transition from one level to another depends on the number of errors. Errors are classified as follows: Anticipation Errors = errors while waiting for the appearance of the light (the player clicks but no light has appeared yet); Position Errors = failure to select the position; impulsivity errors = in level 2, the player selects a blue light; Omission Errors = non-selection. In the cases of Position or Omission Errors, the player retreats by an amount of distance equal to a quarter of the displacement expected to reach the light just missed, for a maximum of four times (i.e., the player has moved backward to the previous position). The instructions of the SG translated in English are available in **Appendix 2**.

FIGURE 2 | Training school.

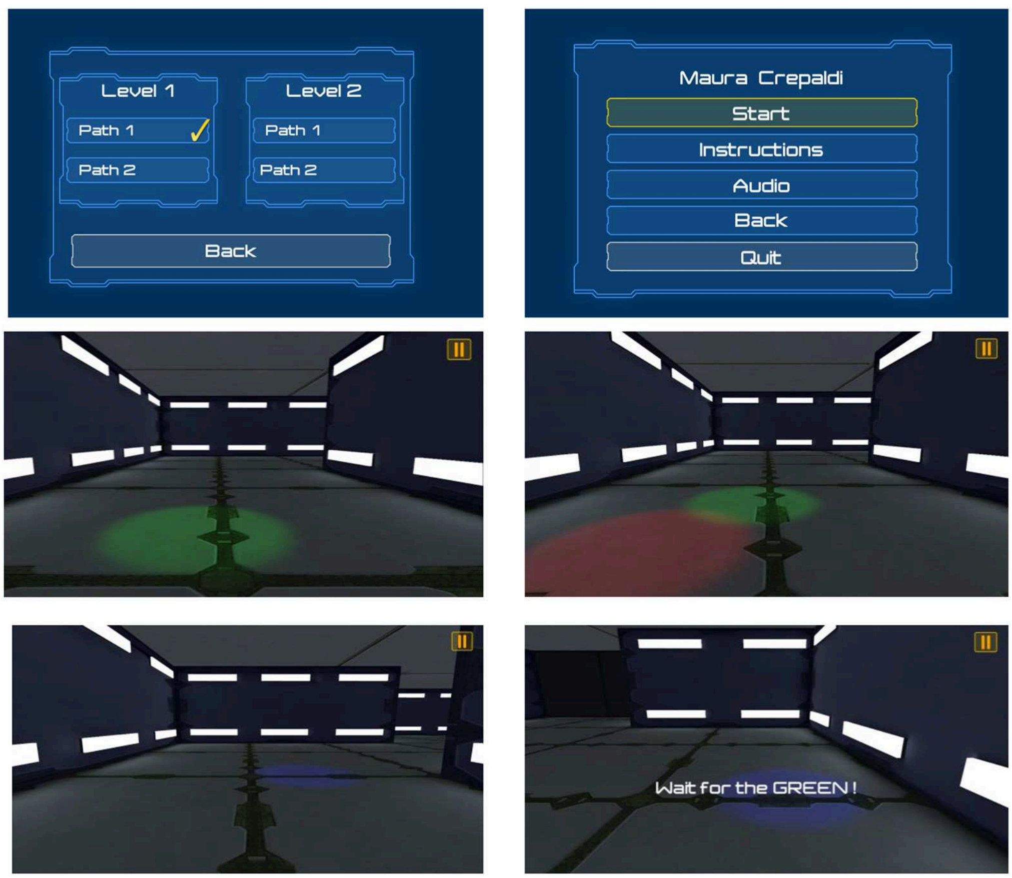


FIGURE 3 | Central building.

Participants

The SG was tested in two primary schools near Milan (Italy), which have had the access to the Service of Learning and Educational Psychology of the Catholic University of the Sacred Heart of Milan in previous circumstances when they took part in research projects. We obtained the permission, through written informed consent, from parents to involve 70 8–11 years old typically developing children, excluding *a priori* those who had a diagnosis of cognitive impairment, ADHD, or other neurodevelopmental disorders. They were administered the Scala per il Disturbo di Attenzione e Iperattività (Scale for the Attention Deficit and Hyperactivity) (SDAI) (Marzocchi and Cornoldi, 2000) included in the Italian Battery for ADHD (BIA: Marzocchi et al., 2010). SDAI is the most commonly used screening tool for ADHD in Italy. Its standard use is based on the frequency with which each of the behaviors reported in the questionnaire is manifested. The 18-item scale is completed by the teacher, who must evaluate, for each of the behaviors listed, the frequency with which it appears. It is composed of two subscales: attention (odd items) and hyperactivity/impulsivity (even items). Responses are on a Likert scale (0–3) for each item (0 if the child never shows that behavior; 1 if the child sometimes presents that behavior; 2 if the child presents it often; 3 if the child presents it very often). According to Marzocchi and Cornoldi (2000), the cutoff score to suspect the presence of ADHD is >14 in one of the two subscales. Since we were interested in impulsivity, only the score of the relevant subscale was taken into account. In the sample, no child obtained a score of 14 or higher in the hyperactivity/impulsivity subscale. From the initial sample of 70 children, 30 children−15 from the lowest range (0–6) and 15 from the highest range (7–13)—were randomly extracted by researchers for the next step so to have a reasonable (in term of time needed for each subject) number of participants who could studied individually and a heterogeneous sample (but still constituted by typically developing children). Children were distributed evenly in each grade (third, fourth, and fifth). Gender was equally distributed in all grades [$\chi^2_{(2)} = 0.268; p = 0.875$] and in the two SDAI level groups [$\chi^2_{(1)} = 0.536; p = 0.464$].

Assessment Instruments

The standard tests chosen from the Italian Battery for ADHD (BIA, Marzocchi et al., 2010) were the following:

- *Ranette* (small frogs), which is a mirror of the "Walk/Don't walk" task. The test involves sustained attention (the test lasts about 10 min), selective attention (children must detect the target sound), and motor inhibition (children must stop the impulsive response to go ahead). In this task, the errors represent a failed inhibition of the answer to "GO" to the right time (difficulty in controlling inhibitory, typical of children with ADHD). The total score consists of the number of correct responses in the 20 items presented. In the analyses we carried out, we considered the complementary score (i.e., the number of errors) as a measure of sustained attention, to facilitate the interpretation of the relations between different measures.
- *Number Stroop task*, which measures interference control and the ability to stop an automatic, but incorrect, response. The task consists of two parts: a baseline task, in which individuals are asked to count how many asterisks are present in each box, and a Stroop task, in which the request is to count how many elements (numbers) are present in each cell. Two kinds of errors may occur: identity errors (when the written number is read instead of the number of stimuli presented; interference of stimuli occurs) and counting errors (when there are errors in counting the stimuli). For the sake of the analyses, the number of interference errors was considered as a measure of inhibition.
- *MF20,* an adaptation of the Matching Familiar Figure Test (MFFT), which consists of 20 items. The task measures visual selective attention and inhibitory control of impulsive responses. Each item is made up of two pages: on the first page, a target figure is shown; on the second page, six figures are shown; they are similar to the target but only one is just the same as the target. The task is to choose the figure that is identical to the target. Then, the following variables are computed: Time of the First answer and Number of Errors. We decided to include in the analyses the number of errors score only, namely, the most discriminant measure, as reported by previous literature on the use of the MF task for measuring impulsivity (Homatidis and Konstantareas, 1981; Brown and Wynne, 1984; Milich and Kramer, 1984; Marzocchi and Cornoldi, 1998).
- *Ad hoc questionnaire.* Children were asked how they felt playing the game (Likert scale); if it was enjoyable or boring (Likert scale); if there was something that they would add in the scenarios (open questions); if the story was simple or difficult (Likert scale); and to summarize the storyboard (open question). The questionnaire is reported in **Appendix 3**.

Procedure

The protocol included three individual sessions for each child. Children carried out the tasks at school in a quiet room in the morning with the presence of a psychologist. All children played all the levels but in a slightly different time, depending on errors and execution speed. The overall time, however, varied in a not very broad range, for a total of 45 min for each section. In the first session, each child completed the Ranette test, the Stroop test, the first level of the Training School (in Antonyms), and the first path of the Central Building (in Antonyms). In the second session, MF20 was proposed together with the second level of the Training School and the remaining paths of the Central Building. In the last session, the *ad hoc* questionnaire was proposed. Each child performed the same tests in the same order and the same activities of the SG. Neither reinforcements nor credits were given to the participants.

Statistical Methods

Statistical analyses were performed with Jamovi R-based software, version 1.0.8.0 (2019), The Jamovi project (2020), IBM Corp (2016), and SPSS, version 24 (2016). Given the count nature of error scores for both the standardized tests and the SG activities, assumptions of normality (all Shapiro–Wilk ps < 0.01, with the only exception of the MF20 error score), homogeneity of error variance, and linearity of the associations were not met. Therefore, non-parametric analyses have been performed.

As for the Training School task, the proportion of hits (i.e., items correctly put in the trash) and false alarm errors (i.e., items incorrectly put in the trash) was used to compute a sensitivity score as a measure of discriminability {$d' = [z(Hit) - z(FA)]$} (Macmillan and Creelman, 2004). A criterion score, representing the bias toward the same (i.e., items put in the bag) response type ($C > 0$) or different (i.e., items put in the trash) response type ($C < 0$), has been computed as well {$C = -[z(Hit) + z(FA)]/2$} and considered in the analyses.

First, descriptive analyses have been carried out, and gender, school-grade, and SDAI's level differences for all measures have been tested. The *p*-value was adjusted to compensate for multiple comparisons, using the Bonferroni correction (alpha = 0.05/9 = 0.005). Afterwards, we computed correlations (Spearman's rho) between the measures of the BIA (Ranette, Number Stroop, and MF20). The rationale of these analyses was to check that these tests, usually employed for diagnostic aims, were adequate to the purpose of the present study: Coherent patterns of correlation, in the expected directions, between them would support their relevance. Then, correlations between the BIA tests and the performance in Antonyms (scores in Training School and Central Building) were computed to verify, according to the main goal of the paper, if the SG can be considered a valid instrument to measure impulsivity tendencies, and therefore testing its concurrent validity. When school-grade differences have been found, partial Spearman's correlations controlling by school grade have been computed. Finally, responses to the *ad hoc* questionnaire were analyzed to draw conclusions about the subjective experience with the SG.

RESULTS

Neither SG performance scores nor the other standardized tasks from the BIA battery (i.e., Ranette and Number Stroop) differed between gender groups (all ps > 0.46) (**Table 1**).

Similar performances were measured in different school grades (i.e., third, fourth, and fifth) in all tasks (all ps > 0.01, **Table 2**). The SDAI level (medium-low vs. medium-high) was not found to discriminate children's performance in either the SG tasks, or the standardized tests, except for the Ranette subtest (**Table 3**), although in the normal range, a lower SDAI level was found to be associated with a lower level of sustained attention, as measured by the Ranette task.

Within BIA tests, we found that scores in the Ranette test (number of errors in the auditory sustained attention test) were positively correlated to the interference errors in the Number Stroop test (the score indicates the errors that children make in pronouncing the stimuli; rho = 0.37; $p < 0.05$, one-tailed). High Ranette scores (right actions) were associated to less impulsive behaviors and more attention in the tasks. There was a negative correlation between errors in MF (when children choose the wrong figure) and time of the performance (more time spent in the performance is linked with a greater accuracy of the answer and fewer errors; rho = 0.62; $p < 0.001$, one-tailed). This means that the more the child was impulsive, the less time he/she took to complete the task and the more mistakes he/she did. These patterns of correlations were expected and are reported just to support the validity of the assessment tools we used in the kind of population the sample belongs to.

Concerning the relationships between the standardized tests and performances in the SG activities, errors in MF20 performance correlated negatively with the sensitivity score in the Training School (**Table 4**). This association suggests that a lower performance in the SG task (lower sensitivity scores) is indicative of a failed inhibition of the automatic response (higher number of errors in MF20) and, therefore, emphasizes that the two tasks measure the same construct involving the similar basic principle. Conversely, the bias score in the Training School task was not correlated with any standardized measure. Scores in the Ranette test (number of errors in the sustained attention task) were positively correlated to anticipation errors in the Central Building scenario (measure of the difficulty for children to stop their irrelevant actions) (**Table 4**). Errors of interference in the Stroop Test (when children say the number written in the box and not the correct number of stimuli inside the box) correlated with errors in anticipation and omission in the Central Building (failure in touching the light) but not with sensitivity in Training School. This may be because the performance request in the Training School is more ecological and helps children in considering details and differences; the Stroop test, instead, is more specific, and there is more conflict between the two codes considered. To sum up, there was evidence of correlations between standard tests, like the tasks from BIA and performances in the SG.

Concerning the *ad hoc* questionnaire, instructions were considered easy, and the only difficulties that children underlined were linked to some words [three third-grade children and one fourth-grade child marked as difficult words: *perfido* (perfidious), *impulsivo* (impulsive)]. Twenty-eight children reported that the story is easy, and only two claimed that the story is difficult to follow. In general, all players understood the game and remembered that Antonyms is the planet of "contrary thoughts and actions," that is, it is a "slowly" planet and that there were enemies who captured the inhabitants. All participants accomplished the different tasks without the help of an adult.

The main difficulties described in the Training School were that some objects differed because of small details. This may depend on the fact that the background color of the shelf is sometimes misleading with respect to the color of the objects that appear. The main difficulties reported in Central Building were that the player must wait for different time lapses before the light appears, and he/she must remember not to click on the blue light. In the second level, when the lights were smaller, it was difficult for children to properly select the light. We suggest that such difficulty can be explained by the higher level of required sustained attention necessary to detect a small visual cue, therefore addressing one of the main ADHD-related symptoms, and not by any issues in the perception of a small visual target.

Given the question "Was Antonyms enjoyable or boring?," 29 children answered "enjoyable" and only one fifth-grade child

TABLE 1 | Gender differences in SG activities (TS, Training School; CB, Central Building) and in the standardized tests (BIA battery).

		Male ($N = 16$) M (SD)	Female ($N = 14$) M (SD)	Comparison U(1); *p*
SG activities	TS sensitivity	2.441 (0.72)	2.587 (0.868)	87.0; 0.305
	TS bias	−0.084 (0.313)	−0.051 (0.202)	103.0; 0.720
	CB omission errors	3.875 (2.918)	4.643 (4.36)	111.5; 1.00
	CB impulsivity errors	0.125 (0.5)	0.286 (1.07)	110.5; 0.923
	CB anticipation errors	1.125 (1.455)	1 (1.88)	99; 0.544
	CB position errors	3 (2.53)	2.857 (2.91)	106; 0.814
Standardized test (BIA)	Ranette errors	2.19 (1.87)	2.14 (2.14)	104.5; 0.767
	Number stroop errors	2.06 (1.84)	1.5 (1.4)	94.5; 0.467
	MF20 errors	4.69 (2.41)	2.71 (2.09)	59.5; 0.029

Differences were tested using the Mann-Whitney test. Alpha = 0.005.

TABLE 2 | School grade differences in SG activities (TS, Training School; CB, Central Building) and in the standardized tests (BIA battery).

		3rd grade ($N = 10$) M (SD)	4th grade ($N = 10$) M (SD)	5th grade ($N = 10$) M (SD)	Comparison H(2); *p*
SG activities	TS sensitivity	2.25 (0.90)	2.83 (0.51)	2.45 (0.84)	2.97; 0.230
	TS bias	−0.10 (0.33)	0.02 (0.27)	−0.12 (0.18)	1.95; 0.380
	CB omission errors	6.4 (4.12)	4.3 (3.23)	2 (2)	6.083; 0.048
	CB impulsivity errors	0.0 (0)	0.4 (1.26)	0.2 (0.632)	1.038; 0.595
	CB anticipation errors	2 (1.89)	0.6 (1.35)	0.6 (1.35)	8.691; 0.013
	CB position errors	4 (2.83)	2.6 (2.84)	2.2 (2.2)	2.45; 0.294
Standardized test (BIA)	Ranette errors	3.0 (2.05)	2.0 (2.05)	1.5 (1.65)	3.614; 0.164
	Number stroop errors	2.5 (2.22)	1.4 (1.17)	1.5 (1.27)	1.754; 0.416
	MF20 errors	3.8 (2.2)	3.3 (1.89)	4.2 (3.22)	0.594; 0.743

Differences were tested using the Kruskal–Wallis test. Alpha = 0.005.

found it "boring." Twenty-five participants liked the game, and only five liked quite the game; none answered that he/she did not like it.

When requested to provide suggestions to improve the SG, three children would add the avatar of Antonyms to modify the hero and make him/her similar to themselves. According to the responses given in the questionnaire, we can conclude that the general aims of the SG were understood by children. They reported to have fun in playing the games and appeared to be fully immersed in the task. All these elements support the notion that Antonyms reached a satisfactory level of acceptability.

DISCUSSION AND CONCLUSIONS

Chacko et al. (2018) pointed out that, although the main interventions now available for children with ADHD consist of a combination of pharmacological treatments and behavioral training [which usually have short-term effects: Daley et al. (2014), DuPaul et al. (2012), and Evans et al. (2014)], it is important to focus attention also on neurocognitive aspects since they, if properly stimulated, have long-term effects. Precisely for this reason, it is important to have engaging tools aimed at strengthening cognitive functions and having a neuropsychological basis, such as the SG presented here.

The optimal intervention should be based on psychosocial skills and take into consideration the neurocognitive mechanisms and processes involved in a specific functional ability together with targeted training to improve and enhance the mechanisms and processes on which these skills are based (Chacko et al., 2018).

The SG was designed, referring to the Dual Pathway model by Sonuga Barke because the basis of the game is the cognitive path that concerns EF aspects involved in ADHD. In particular, inhibition (in the SG, the most immediate action is the wrong one), shifting (cognitive flexibility is requested in order to remember changes in the rules of the game), and WM (remembering the storyboard is needed during the activities) are considered. Considering the two main objectives of the proposed SG (assessing impulsive control and preventing non-adaptive behaviors and irrelevant thoughts through inhibition), the performance and data obtained seem to show a positive result in this direction.

Results suggest that the activities embedded in the SG appear to be associated with the performance in standard tests usually employed to assess inattention, impulsivity, and hyperactivity levels in children. Few significant correlations emerged, but it is interesting to observe that these concerned only those aspects of the SG that closely match the specific skills measured by those tests. No significant differences emerged in the SG scores between

TABLE 3 | SDAI level differences in SG activities (TS: Training School; CB: Central Building) and in the standardized tests (BIA battery).

		Medium-low SDAI (*N* = 15) M (SD)	Medium-high SDAI (*N* = 15) M (SD)	Comparison U(1); *p*
SG activities	TS sensitivity	2.54 (0.76)	2.48 (0.83)	103.5; 0.723
	TS bias	−0.03 (0.27)	−0.10 (0.26)	91.5; 0.388
	CB omission errors	2.93 (3.19)	2.93 (2.12)	106.5; 0.814
	CB impulsivity errors	1.07 (1.67)	1.07 (1.67)	110.0; 0.923
	CB anticipation errors	0 (0)	0.4 (1.12)	97.5; 0.164
	CB position errors	5 (3.59)	3.47 (3.60)	84.5; 0.250
Standardized test (BIA)	Ranette errors	1 (1)	3.33 (2.02)	**34.5; 0.001**
	Number stroop errors	1.47 (1.46)	2.13 (1.81)	91.5; 0.382
	MF20 errors	3.27 (2.52)	4.27 (2.34)	91.5; 0.382

Differences were tested using the Mann-Whitney test. Alpha = 0.005 (statistical significance is marked in bold).

TABLE 4 | Spearman's correlation matrix between SG activities (TS, Training School; CB, Central Building) and standardized tests (BIA battery).

	Ranette errors	Number stroop errors	MF20 errors
TS sensitivity	−0.206	0.023	**−0.451****
TS bias	0.015	0.052	−0.200
CB omission errors	0.243	**0.358***	−0.022
CB impulsivity errors	0.299	0.155	0.076
CB anticipation errors	**0.366***	**0.431****	0.009
CB position errors	0.065	0.097	−0.024

*Alpha = 0.05. *p < 0.05, **p < 0.01, one-tailed. Statistical significance is marked in bold.*

SDAI levels (i.e., medium-low vs. medium-high), which reflects the trend obtained in standardized tests. These results are not surprising, since the sample was composed only by typically developing children. In the study by Crepaldi et al. (2020) on ADHD children, Antonyms scores were found to be significantly different between a group of children with a diagnosis of ADHD and a paired subgroup of typically developing children (which was extracted from the larger group recruited for the present study). Since in Antonyms there are no significant differences between different levels of the normal SDAI score range, but instead differences emerged when compared to a clinical sample (SDAI score in the clinical range, >14) (Crepaldi et al., 2020), this leads us to suppose that Antonyms could be a useful tool to discriminate between ADHD and healthy children and not between different levels within a non-clinical range.

When designing the SG, we chose to propose activities with a reference to everyday life. The generalizability of tasks to everyday life remains a limit of many existing computer activities, which we tried to overcome in Antonyms. For example, the path in the Central building may be compared to the attention that children must pay when they are walking on a street: in this situation, children must be careful where they put their feet, stop when there is a danger, and proceed only when it is allowed. The activities in the Training School can mirror preparing the backpack for school, which is an activity in which children with ADHD often struggle.

The SG has been ultimately designed to assess impulsivity and inhibition in children with ADHD and ADHD-related characteristics. However, since the abilities that are lacking in ADHD are not all/nothing variables, a continuum in performance can be assumed. Hence, the SG can be employed also to check if inhibitory skills are poorly expressed in typically developing children. Another advantage is that scoring is automatized and, therefore, more accurate and quicker. Furthermore, only a laptop is needed to administer the SG, whereas traditional tests require a series of different materials (sheets, cards, audio-recorder, pencils), which are continuously presented and removed from the child, thus by distracting him/her.

In the future, some critical ergonomic points stressed by the participants might be improved, and the ergonomic aspects of the SG might be tested in a systematic way. We also plan to implement more mini-games with other activities to test the efficacy of Antonyms in enhancing the ability to keep attention in daily life situations and to propose the game to a large sample of children diagnosed with ADHD.

Furthermore, it will be advisable to verify the influence and role of the greater engagement of the SG compared to the pencil paper tasks, but it seems that, in comparison to standardized test, the SG is more motivating for children. If, as it seems, this is verified, it could prevent evaluators to struggle to attract and keep attention of the patient, a problem which is rather common in the assessment of ADHD.

We expect that children who will be engaged in this SG could improve attention and learn strategies to manage impulsivity so to inhibit irrelevant thoughts and thus an enlarged version of Antonyms could be proposed as a rehabilitation tool.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethical committee of the Catholic University of Milan. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MC: project, game design, tasks administration to children, data analysis, and writing. AA: project, game design supervision, data analysis, and contacts with schools. MS, SM, VC, and DB: game design and implementation and supervision. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.00034/full#supplementary-material

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders: DSM 5, 5th Edn*. Arlington, VA: American Psychiatric Publishing. doi: 10.1176/appi.books.9780890425596

Antonietti, A., Ignazi, S., and Perego, P. (2000). Metacognitive knowledge about problem-solving methods. *Br. J. Educ. Psychol.* 70, 1–16. doi: 10.1348/000709900157921

Areces, D., García, T., Cueli, M., and Rodríguez, C. (2019). Is a virtual reality test able to predict current and retrospective ADHD symptoms in adulthood and adolescence? *Brain Sci.* 9:274. doi: 10.3390/brainsci9100274

Baddeley, A. D., and Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology* 8, 485–493. doi: 10.1037/0894-4105.8.4.485

Barba, M. C., Covino, A., De Luca, V., De Paolis, L. T., D'Errico, G., Di Bitonto, P., et al. (2019). "BRAVO: a gaming environment for the treatment of ADHD," in *International Conference on Augmented Reality, Virtual Reality and Computer Graphics* (Cham: Springer), 394–407. doi: 10.1007/978-3-030-25965-5_30

Beale, I. L., Kato, P. M., Marin-Bowling, V. M., Guthrie, N., and Cole, S. W., (2007). Improvement in cancer-related knowledge following use of a psychoeducational video game for adolescents and young adults with cancer. *J. Adolesc. Health*. 41, 263–270. doi: 10.1016/j.jadohealth.2007.04.006

Brown, R. T., and Wynne, M. E. (1984). Attentional characteristics and teachers ratings in hyperactive, reading disabled, and normal boys. *J. Clin. Child Adolesc. Psychol.* 13, 38–43. doi: 10.1207/s15374424jccp1301_6

Bul, K. C. M., Franken, I. H. A., Van der Oord, S., Kato, P. M., Danckaerts, M., Vreeke, L. J., et al. (2015). Development and user satisfaction of "plan-it commander," a serious game for children with ADHD. *Games Health J*. 4, 502–512. doi: 10.1089/g4h.2015.0021

Chacko, A., Bedard, A. C., Marks, D., Gopalan, G., Feirsen, N., Uderman, J., et al. (2018). Sequenced neurocognitive and behavioral parent training for the treatment of ADHD in school-age children. *Child Neuropsychol.* 24, 427–450. doi: 10.1080/09297049.2017.1282450

Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M., T., et al. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. *PLoS ONE* 7:e31546. doi: 10.1371/journal.pone.0031546

Coleman, B., Marion, S., Rizzo, A., Turnbull, J., and Nolty, A. (2019). Virtual reality assessment of classroom-related attention: an ecologically relevant approach to evaluating the effectiveness of working memory training. *Front. Psychol.* 10:1851. doi: 10.3389/fpsyg.2019.01851

Conners, C. K. (1994). *The Conners Continuous Performance Test*. Toronto, ON: Multi-Health Systems.

Cortese, S., and Castellanos, F. X. (2012). Neuroimaging of attention-deficit/hyperactivity disorder: current neuroscience-informed perspectives for clinicians. *Curr. Psychiatry Rep*. 14, 568–578. doi: 10.1007/s11920-012-0310-y

Coyle, D., Doherty, G., Matthews, M., and Sharry, J. (2007). Computers in talk-based mental health interventions. *Interact. Comput*. 19, 545–562. doi: 10.1016/j.intcom.2007.02.001

Crepaldi, M., Colombo, V., Baldassini, D., Mottura, S., and Antonietti, A. (2017). "Supporting rehabilitation of ADHD children with serious games and enhancement of inhibition mechanisms," in *International Conference on Virtual Reality and Augmented Reality* (Cham: Springer), 167–181. doi: 10.1007/978-3-319-72323-5_11

Crepaldi, M., Colombo, V., Mottura, S., Baldassini, D., Sacco, M., Cancer, A., et al. (2020). Antonyms: a computer game to improve inhibitory control of impulsivity in children with attention deficit/hyperactivity disorder (ADHD). *Information* 11:230. doi: 10.3390/info11040230

Daley, D., Van der Oord, S., Ferrin, M., Danckaerts, M., Doepfner, M., Cortese, S., et al. (2014). Behavioral interventions in attention-deficit/hyperactivity disorder: a meta-analysis of randomized controlled trials across multiple outcome domains. *J. Am. Acad. Child Adolesc. Psychiatry* 53, 835–847. doi: 10.1016/j.jaac.2014.05.013

Dalrymple-Alford, E. C. (1972). Associative facilitation and interference in the stroop color-word task. *Percept. Psychophys.* 11, 274–276. doi: 10.3758/BF03210377

Daruna, J. H., and Barnes, P. A. (1993). "A neurodevelopmental view of impulsivity," in *The Impulsive Client: Theory, Research and Treatment*, eds W. G. McCown, J. L. Johnson, and M. B. Shure (Washington, DC: American Psychological Association), 23–37. doi: 10.1037/10500-002

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol*. 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

DuPaul, G. J., Eckert, T. L., and Vilardo, B. (2012). The effects of school-based interventions for attention deficit hyperactivity disorder: a meta-analysis 1996–2010. *School Psychol. Rev*. 41, 387–412.

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Percept. Psychophys.* 16, 143–149. doi: 10.3758/BF03203267

Evans, S. W., Owens, J. S., and Bunford, N. (2014). Evidence-based psychosocial treatments for children and adolescents with attention-deficit/hyperactivity disorder. *J. Clin. Child Adolesc. Psychol.* 43, 527–551. doi: 10.1080/15374416.2013.850700

Fabio, R. A., and Antonietti, A. (2012). Effects of hypermedia instruction on declarative, conditional and procedural learning in ADHD students. *Res. Dev. Disabil*. 33, 2028–2039. doi: 10.1016/j.ridd.2012.04.018

Fang, Y., and Dai Han, H. L. (2019). A virtual reality application for assessment for attention deficit hyperactivity disorder in school-aged children. *Neuropsychiatr. Dis. Treat*. 15, 1517–1523. doi: 10.2147/NDT.S206742

Fernández-Aranda, F., Jiménez-Murcia, S., Santamaría, J. J., Gunnard, K., Soto, A., Kalapanidas, E., et al. (2012). Video games as a complementary therapy tool in mental disorders: playMancer, a European multicentre study. *J. Ment. Health* 21, 364–374. doi: 10.3109/09638237.2012.664302

Homatidis, S., and Konstantareas, M. M. (1981). Assessment of hyperactivity: isolating measures of high discriminant ability. *J. Consult. Clin. Psychol*. 49:533. doi: 10.1037/0022-006X.49.4.533

IBM Corp (2016). *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.

Kahneman, D., and Tversky, A. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Kulman, R. (2012). "Generalization of game-based learning for children with ADHD," in *Proceedings of The Annual American Psychological Association Convention* (Orlando, FL), 2–5. doi: 10.1037/e636962012-001

Lappin, J. S., and Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *J. Exp. Psychol*. 72, 805–811. doi: 10.1037/h0021266

Logan, G. D. (1994). "On the ability to inhibit thought and action: a users' guide to the stop signal paradigm," in *Inhibitory Processes in Attention, Memory, and Language*, eds D. Dagenbach and T. H. Carr (New York, NY: Academic Press), 189–239.

Macmillan, N. A., and Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Washington, DC: Psychology Press. doi: 10.4324/9781410611147

Martinez, L., Prada, E., Satler, C., Tavare, M. C. H., and Tomaz, C. (2016). Executive dysfunctions: the role in attention deficit hyperactivity and post-traumatic stress neuropsychiatric disorders. *Front. Psychol.* 7:1230. doi: 10.3389/fpsyg.2016.01230

Marzocchi, G. M., and Cornoldi, C. (1998). Disturbi di impulsività e ricerca visiva in bambini con deficit di attenzione/iperattività. *Psichiatr. Dell'infanz. Dell'adolesc.* 65, 173–186.

Marzocchi, G. M., and Cornoldi, C. (2000). Una scala di facile uso per la rilevazione dei comportamenti problematici dei bambini con deficit di attenzione e iperattività. *Psicol. Clin. Dello Sviluppo* 4, 43–64. doi: 10.1449/584

Marzocchi, G. M., Re, A. M., and Cornoldi, C. (2010). *BIA-Batteria Italiana Per l'ADHD. Batteria Per La Valutazione Dei Bambini Con Deficit Di Attenzione/Iperattività*. Trento: Edizioni Centro Studi Erickson.

Milich, R., and Kramer, J. (1984). Reflections on impulsivity: an empirical investigation of impulsivity as a construct. *Adv. Learn. Behav. Disabil.* 3, 57–94.

Miyake, A., and Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Curr. Dir. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., and Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Adv. Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Mühlberger, A., Jekel, K., Probst, T., Schecklmann, M., Conzelmann, A., Andreatta, M., et al. (2020). The influence of methylphenidate on hyperactivity and attention deficits in children with ADHD: a virtual classroom test. *J. Atten. Disord*. 24, 277–289. doi: 10.1177/1087054716647480

Muñoz, J. E., Lopez, D. S., Lopez, J. F., and Lopez, A. (2015). "Design and creation of a BCI videogame to train sustained attention in children with ADHD," in *2015 10th Computing Colombian Conference (10CCC)* (Bogota: IEEE), 194–199. doi: 10.1109/ColumbianCC.2015.7333431

Nigg, J. T., Goldsmith, H. H., and Sechek, J. (2004). Temperament and attention deficit hyperactivity disorder: the development of a multiple pathway model. *J. Clin. Child Adolesc. Psychol.* 33, 42–53. doi: 10.1207/S15374424JCC3301_5

Parsons, T. D., Duffield, T., and Asbee, J. (2019). A comparison of virtual reality classroom continuous performance tests to traditional continuous performance tests in delineating ADHD: a meta-analysis. *Neuropsychol. Rev*. 29, 338–356. doi: 10.1007/s11065-019-09407-6

Powell, L., Parker, J., and Harpin, V. (2017). ADHD: is there an app for that? A suitability assessment of apps for the parents of children and young people with ADHD. *JMIR MHealth UHealth* 5:e149. doi: 10.2196/mhealth.7941

Powell, L., Parker, J., Harpin, V., and Mawson, S. (2019). Guideline development for technological interventions for children and young people to self-manage attention deficit hyperactivity disorder: realist evaluation. *J. Med. Internet Res*. 21:e12831. doi: 10.2196/12831

Prencipe, A., Kesek, A., Cohen, J., Lamm, C., Lewis, M. D., and Zelazo, P. D. (2011). Development of hot and cool executive function during the transition to adolescence. *J. Exp. Child Psychol.* 108, 621–637. doi: 10.1016/j.jecp.2010.09.008

Rassin, M., Gutman, Y., and Silner, D. (2004). Developing a computer game to prepare children for surgery. *AORN J.* 80, 1095–1102. doi: 10.1016/S0001-2092(06)60689-3

Rizzo, A. A., Bowerly, T., Buckwalter, J. G., Schultheis, M. T., Matheis, R., Shahabi, C., et al. (2002). "Virtual environment for the assessment of attention and memory processes: the virtual classroom and office," in *Proceeding of the 4th International Conference on Disability, Virtual Reality and Associated Technology* (Hungary: University of Reading, Vresprem), 3–11.

Rodríguez, C., García, T., and Areces, D. (2017). New and future challenges concerning the use of virtual reality tools for assessing ADHD. *Curr. Dev. Disord. Rep.* 4, 8–10. doi: 10.1007/s40474-017-0103-4

Simons, L., Valentine, A. Z., Falconer, C. J., Groom, M., Daley, D., and Craven, M. P. (2016). Developing mHealth remote monitoring technology for attention deficit hyperactivity disorder: a qualitative study eliciting user priorities and needs. *JMIR MHealth UHealth* 23:e31. doi: 10.2196/mhealth.5009

Sjöwall, D., Roth, L., Lindqvist, S., and Thorell, L. B. (2013). Multiple deficits in ADHD: executive dysfunction, delay aversion, reaction time variability, and emotional deficits. *J. Child Psychol. Psychiatry* 54, 619–627. doi: 10.1111/jcpp.12006

Smith, E. E., and Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science* 283, 1657–1661. doi: 10.1126/science.283.5408.1657

Sonuga-Barke, E. J. (1996). When "impulsiveness is delay aversion": a reply to Schweitzer and Sulzer-Azaroff (1995). *J. Child Psychol. Psychiatry* 37, 1023–1025. doi: 10.1111/j.1469-7610.1996.tb01499.x

Sonuga-Barke, E. J. (2003). The dual pathway model of AD/HD: an elaboration of neurodevelopmental characteristics. *Neurosci. Biobehav. Rev.* 27, 593–604. doi: 10.1016/j.neubiorev.2003.08.005

Sonuga-Barke, E. J. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biol. Psychiatry* 57, 1231–1238. doi: 10.1016/j.biopsych.2004.09.008

Sonuga-Barke, E. J., and Taylor, E. (2015). "Disorders of attention activity," in *Rutter's Child and Adolescent Psychiatry, 6th Edn*, ed A. Thapar (Oxford: Blackwell), 738–756. doi: 10.1002/9781118381953.ch55

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Swets, J. A. (1964). *Signal Detection and Recognition by Human Observers*. New York, NY: Wiley. doi: 10.1037/e444572004-001

The jamovi project (2020). *jamovi (Version 1.2) [Computer Software]*. Retrievedfrom: https://www.jamovi.org (accessed June 30, 2020).

Van der Oord, S., Ponsioen, A. J. G. B., Geurts, H. M., Ten Brink, E. L., and Prins, P. J. M. (2014). A pilot study of the efficacy of a computerized executive functioning remediation training with game elements for children with ADHD in an outpatient setting: outcome on parent- and teacher-rated executive functioning and ADHD behavior. *J. Atten. Disord*. 18, 699–712. doi: 10.1177/1087054712453167

Walshe, D. G., Lewis, E. J., Kim, S. I., O'Sullivan, K., and Wiederhold, B. K. (2003). Exploring the use of computer games and virtual reality in exposure therapy for fear of driving following a motor vehicle accident. *CyberPsychol. Behav.* 6, 329–334. doi: 10.1089/109493103322011641

Welsh, M. C., Friedman, S. L., and Spieker, S. J. (2006). "Executive functions in developing children: current conceptualizations and questions for the future," in: *Blackwell Handbook of Early Childhood Development*, eds K. McCartney and D. Phillips (Blackwell Publishing), 167–187. doi: 10.1002/9780470757703.ch9

Wronska, N., Garcia-Zapirain, B., and Mendez-Zorrilla, A. (2015). An iPad-based tool for improving the skills of children with attention deficit disorder. *Int. J. Environ. Res. Public Health* 12, 6261–6280. doi: 10.3390/ijerph120606261

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

REVIEW
published: 31 August 2020
doi: 10.3389/fcomp.2020.00030

Serious Pervasive Games

António Coelho [1,2], Rui Rodrigues [1,2], Rui Nóbrega [2,3], João Jacob [1], Leonel Morgado [2,4], Pedro Cardoso [1,2,5], Maria van Zeller [1], Liliana Santos [1] and A. Augusto Sousa [1,2]*

[1] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, [2] INESC TEC, Porto, Portugal, [3] NOVA LINCS, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Lisbon, Portugal, [4] Universidade Aberta, CIAC, & LE@D, Coimbra, Portugal, [5] Faculdade de Belas Artes, Universidade do Porto, Porto, Portugal

Serious Pervasive Games extend the *magic circle* (Huizinga, 1938) to the players' context and surrounding environment. The blend of both physical and fictive game worlds provides a push in player engagement and promotes situated learning approaches. Space and time, as well as social context, acquire a more meaningful impact on the gameplay. From pervasive learning towards science communication with location-based games, this article presents research and case studies that exemplify their benefits and related problems. Pervasive learning can be defined as "learning at the speed of need through formal, informal and social learning modalities" (Pontefract, 2013). The first case study—the BEACONING project—aims to contextualize the teaching and learning process, connecting it with problem-based game mechanics within STEM. The main goal of this project is to provide the missing connection between STEM subjects and real-world interactions and applications. The pedagogical foundation is supported on problem-based learning (PBL), in which active learning is in the center, and learners have to work with different tools and resources in order to solve problems (quests). Teachers create, facilitate, and assess pervasive and gamified learning activities (missions). Furthermore, these quests are gamified in order to provide non-linear game plots. In a second case study, we demonstrate and evaluate how natural heritage can benefit from pervasive games. This study is based on a set of location-based games for an existing natural park, which have been developed in order to provide enhanced experiences, as well as additional information about some species that are more difficult to observe or that are seasonal. Throughout the research and development of these projects, we have encountered and identified several problems, of different nature, present in pervasive games.

Keywords: serious pervasive games, pervasive games, serious games, location-based games, pervasive learning, science communication

OPEN ACCESS

Edited by:
Jože Rugelj, University of Ljubljana, Slovenia

Reviewed by:
Cesar Collazos, University of Cauca, Colombia
Camino Fidalgo, University of Zaragoza, Spain

***Correspondence:**
António Coelho
acoelho@fe.up.pt

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: *28 January 2020*
Accepted: *03 July 2020*
Published: *31 August 2020*

Citation:
Coelho A, Rodrigues R, Nóbrega R, Jacob J, Morgado L, Cardoso P, van Zeller M, Santos L and Sousa AA (2020) Serious Pervasive Games. Front. Comput. Sci. 2:30. doi: 10.3389/fcomp.2020.00030

INTRODUCTION

Games can be considered a cultural media, and source of formative experiences, in the broad definition of "Ludification of Culture" (Raessens, 2006).

Serious games take advantage of the potential of games to abstract and simplify reality and engage players in activities other than entertainment (Sawyer and Rejeski, 2002). However, games are also played as a closed formal system, called the magic circle (Huizinga, 1938). This concept was later adapted to digital games by Salen and Zimmerman (2004). This promotes disengaging from reality, suspension of disbelief, which amplifies the potential to generate learning and behavior change. Although this can be good for many serious games applications, by focusing on the essential aspects of the intended experience, it may fail in other cases. This is the case when the focus of the player should be in the surrounding objects, as in a museum visit, for example. In this case, the game should not disengage the visitor from the primary focus on the museum experience.

Pervasive games extend this magic circle to the real world, exploring the context of the player (Montola, 2005; Montola et al., 2009). These games combine features of ubiquitous gaming and live role-playing games (Falk and Davenport, 2004). This type of pervasiveness significantly differs from the area it is inspired from, pervasive computing (Nieuwdorp, 2007). In the case of pervasive games, context can be related to the players' profile, location, time of play, or other related information such as the weather or calendar events. Considering a serious game for learning, for example, the learning process occurs in distinct times and places, either in school or at home, but also with friends when and where is more convenient. Serious games for learning should take into consideration this situation to promote the best learning activities for all these distinct contexts. In class, the teacher can introduce competitive or collaborative dynamics to encourage individual or group activities; at home, different actors and dynamics can be introduced, involving friends and family, personal backgrounds, and space. The intersection of Serious Games and Pervasive Games defines the concept of Serious Pervasive Games. Arango-López et al. (2017) define its concept as "A pervasive game delivers to the player an enriched experience of game through an evolvement of the dynamics of the game, expanding the space of the game according to the context where it is played. This allows breaking the boundaries of the game world, making reality part of it and that the elements in that reality have an influence during the game."

This article presents two case studies that take advantage of serious pervasive games to promote learning "anytime and anywhere" (Section Pervasive Learning), with a concrete case of communicating science in natural parks (Section Pervasive Games in Science Communication for Natural Parks). Section Issues in Serious Pervasive Games provides a discussion of the issues that must be taken into consideration when designing serious pervasive games, and the last section presents some conclusions and future challenges.

PERVASIVE LEARNING

Pervasive learning has its roots in mobile learning (m-learning), which refers to learning facilitated by mobile devices. The primary aim of m-learning is to provide users with a learning environment that is not restricted to a specific location or time (Laine and Joy, 2009).

Pervasive learning pushes m-learning a little further, using pervasive computing for increased immersion in the learning process. Pervasive learning is an immersive experience that mediates between the learner's mental (e.g., needs, preferences, prior knowledge), physical (e.g., objects, other learners), and virtual (e.g., content accessible with mobile devices, artifacts) context. The intersection of these contexts is referred to as a pervasive learning environment (Syvänen et al., 2005).

Related Work

Felipe et al. (2018) accounts for early efforts to develop a methodology for creating pervasive serious games, extending a methodology called Process Model for Development of Serious Games, by leveraging the development case of a serious games for learning of computer networking. Arango-López et al. (2019) focus on the narrative and geolocation applied to close spaces for an enriched experience to the player. The related work presents a very good review on frameworks for developing pervasive games and, as a main contribution, describes the creation of a platform that integrates tools for creating stories and editing narrative content in the game experiences. Several other studies have also focused on specific Software Engineering techniques for pervasive games (Valente and Feijó, 2014; Viana et al., 2014; Valente et al., 2017).

Santos and Burguès (2017) investigated authoring of Chronica Mobilis, a serious pervasive game for raising awareness and discussion of social, political, and historical uses in contemporary cities, via Chronica Mobilis, a situated experience about gentrification, by placing players amid conflict. They detail design aspects to balance narrative and ludology, with a method that combines scripting, design, and implementation, integrating production and dissemination roles as co-creative.

Arango-López et al. (2018) present a systematic review of the literature on the results of pervasive learning games, identifying different research projects that improve the learning process. A recent survey (Lima et al., 2020) of problems and contributions in the field of m-learning demonstrated that research effort is mostly focused on creating activities and monitoring them, and somewhat on exploring new technologies for assessing their possibilities. It also established that there is a lack of research on structural and logistics aspects, and on assessment aspects of m-learning, which are essential components of large scale, regular deployment of learning activities in formal education. Efforts to collect data from serious pervasive games while they are ongoing, as part of educational information systems such as those developed for BEACONING (detailed in the next section), are thus critical for widespread adoption of these games by educational systems.

An example of a game that was developed for pervasive learning is the EduPark app. This game-like application (app) is an interactive, interdisciplinary quiz with AR, based on Geocaching principles, intended to be played in a specific urban park (Carvalho and Guimarães, 2018). The EduPARK project followed a framework proposed by Parker (2011), since it allows developing enhanced learning environments comprising various cycles of refinement of the prototype (Carvalho and Guimarães, 2018). The game Savannah allows learning about lions' behavior through collaboration between players, thus exploring a more educational component (Benford et al., 2004).

Finde Vielfalt Simulation is a simulation game for school classes from grade 7. As the players get pulled into a story, they have to solve tasks to discover biodiversity, in real nature location places, while dealing with the dilemma between economic interest and nature conservation. Schaal et al. (2018) showed that playing Finde Vielfalt Simulation had a positive impact on biodiversity-related knowledge and that the players felt closer to nature after playing it.

Games have another very relevant characteristic, which is the fact that information on gameplay (game metrics/analytics) can be used also as indicators for the efficiency end efficacy of learning. Taborda et al. (2019) present a systematic literature review on the effectiveness and fun metrics in a pervasive game experience. The main conclusions show that only a few metrics were identified, so there is a need for further study in this area in order to improve the user experience. This topic is further explored in Martinez-Ortiz et al. (2019) as part of the BEACONING platform, in the next section.

The BEACONING Project

The BEACONING project (Bourazeri et al., 2017) was designed to address the needs of pervasive learning, as reflected in the full name of the project: "Breaking Educational Barriers with Contextualized, Pervasive and Gameful Learning." It was focused on "anytime, anywhere" learning by exploiting pervasive, context-aware, and gamified techniques and technologies, framed under the Problem-Based Learning approach. This project was concluded in April 2019 and achieved the following results:

- Integrated BEACONING ecosystem: a cloud-based modular platform that supports the delivery of integrated learning. This platform provides a set of core services and tools that integrate the whole community in a gamified workflow. Game designers create games to foster the learning activities, organized by educators, that make them available to teachers to deploy to the students.
- Gamified Lesson Path (GLP): the core of the BEACONING Project and experience. The GLP [Bourazeri et al.] embeds learning in the daily life of the students, while providing deep engagement through their Game Plots overlaid over the pedagogical content. The game plot is the game narrative that drives the student in the learning activity. The game plot can include adventures and/or location-based narratives. On top of this game plot, the teachers integrate the gamified learning activities (missions) and the specific challenges (problems) to solve (quests), including the analytics that provide feedback to the teacher about the students' performance (**Figure 1**).
- BEACONING Platform GUI: graphical user interfaces (GUIs) used in the BEACONING Platform form the front-end user experience (UX) for important stakeholders such as the Teacher and the Students. Teachers can select and edit the GLP that they find more useful for their classes and deploy them at the most appropriate time. Through a dashboard, the teachers can monitor the progression of the students' activities and their performance (**Figure 2**), being able to adapt the GLP for each student according to their needs. The students have their dashboard that allows them to be aware of the missing activities and get feedback on their progress and performance.
- Authoring tool: The authoring tool (Cardoso et al., 2020) is also a core component as it allows the creation and editing of the GLP, integrating the pedagogical contents with the game plots. So, its architecture integrates most of the BEACONING components, and the GUI is intuitive and based on a visual paradigm (**Figure 3**) in order for a teacher to understand how the game plot will drive the educational activities. The tool sets up the mini-games (quests) that drive the educational content (missions) and assigns them to the events of the game plot. Finally, it submits the GLP to the core services, in order to make it available for the teachers to deploy it to the students.
- Game plot editor: Game plots can be created by game companies using this tool, making new game plots available for instructional design of GLP. Although a more specific tool, this game plot editor is meant for anyone with knowledge on game design to implement the game mechanics and level design without any knowledge of programming (**Figure 4**).
- Learning Analytics: The BEACONING ecosystem provides educational activities as GLP instances. These activities generate a comprehensive set of analytics that are available for teachers to monitor students' performance (Martinez-Ortiz et al., 2019). The analytics are defined on the authoring tool at the time of the learning design. This enables the teacher to get analytics already integrated in the GLP as they deploy it.

An Example of a Beaconing Pervasive Learning Activity

A set of large-scale pilots of the BEACONING ecosystem was carried out throughout Europe on February 10, 2018. In Portugal, the authors created an activity at the University of Porto's Botanical Garden. This activity was focused on exploring scientific content in botany, along with a story that embraced the literary work of Sophia de Mello Breyner Andresen, a celebrated Portuguese writer whose grandparents lived in this house at the beginning of the 20th century (thus a place which played a role in her work). With this in mind, the authors explored events and the lore in "O Rapaz de Bronze" (The Bronze Boy) in a way for the visitors (the players) to be able to travel through the Garden and explore its diversity. Guided through this story, this activity was composed of seven sequential challenges, each located in a particular Point of Interest (POI), in which the former challenges triggered clues for the players to find the next (**Figure 5**). Within a predefined radius, each POI presented players with particular challenges, consisting of different mini-games, from puzzles to quizzes, action games, riddles, and friendly conundrums (**Figure 6**), based on a visit we had first undertook with Garden's staff.

The overall activity valued the presence of the players *in loco*, meaning that many of the challenges (**Figure 7**) were focused on inspecting players' actual physical surroundings, sometimes requiring them to take their sensorial perception into account to progress. The goal was to make visitors engaged with the Garden itself, exposing them to academic content that could be witnessed or experienced there. Ergo, blending the content of the game with the content of the Garden transformed its space into the space of the game itself, and it was by merging these two spaces that this activity was successful in promoting experiential learning.

At the end, we surveyed 19 random players. Despite some issues, we found that about 79% of the interviewees found the experience rewarding at a scientific level, around 83% found

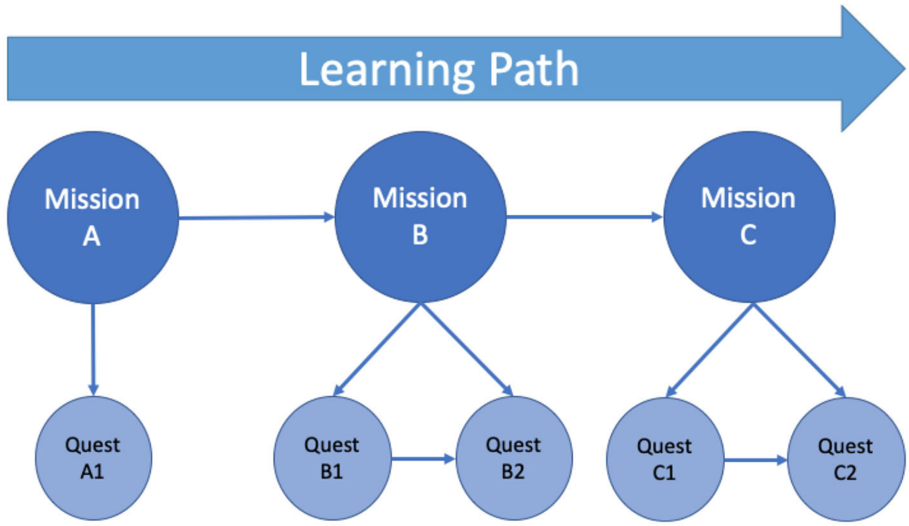


FIGURE 1 | Learning path leveling up—a meta-game narrative/story contextualizes the missions and quests (Arnab et al., 2016).

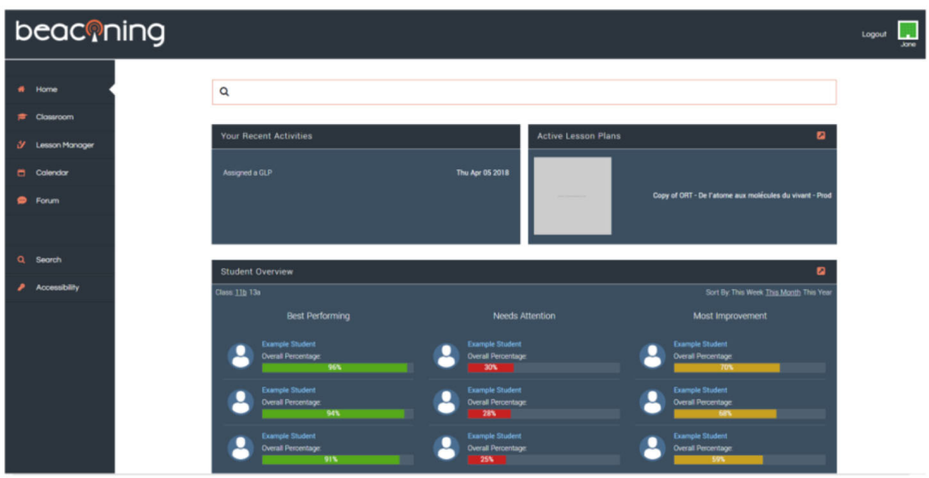


FIGURE 2 | GUI: Teacher dashboard[1].

that the games made the experience more fun, and about 95% understood that the mini-games helped them learn the content.

PERVASIVE GAMES IN SCIENCE COMMUNICATION FOR NATURAL PARKS

Pervasive serious games allow amplifying and enriching the experience of visiting nature parks since a layer of digital information augments our perception of the real world. This augmentation is quite relevant as animals are generally evasive, and wildlife has biorhythms that are not always suitable for the visitors. As an example, some animals only appear at night, and some flowers can only be seen at specific times of the year (e.g., Spring). Furthermore, the intricacy of ecosystems makes it difficult to communicate, and pervasive games can improve this by providing specific gameplay, at the location of the visitor.

Following this challenge, we have developed a set of prototypes to analyze how serious pervasive games can improve the visitors' experience and the effectiveness of the science communication.

As a first study, we have developed the game of the jay (Santos et al., 2017) that is unlocked in an oak tree forest inside the park. It is a serious game to learn about the vital role the Eurasian jay has on oak tree forest reforestation. The player has to make the jay collect and plant acorns, and collapse clouds to make it rain, in order to put out wildfires and make the sprouts grow (**Figure 8**). To win, the player must reforest the scenery until the ecosystem can reach the maximum number of oak trees.

[1]Beaconing deliverable D4.5 BEACONING Platform GUI, 2018. Available online at: https://beaconing.eu/wp-content/uploads/deliverables/D4.5.pdf.

FIGURE 3 | The authoring tool GUI[2].

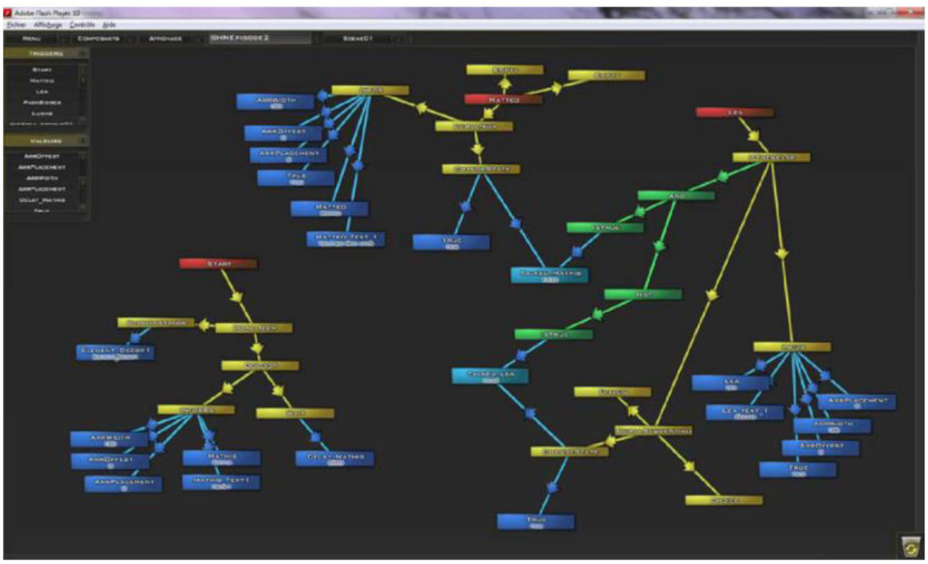


FIGURE 4 | The Game plot editor[3].

This game took advantage of simple mechanics and abstraction to deliver relevant information on location, about the particular ecosystem the visitor is exploring. A preliminary study performed on location with 56 subjects concluded that the game was effective in communicating the educational content (Santos et al., 2017). From these findings, a decision has been taken to improve the pervasiveness on the next prototype.

The Virtual Animal Detector (Santos et al., 2016) is a "treasure hunt" and collection game about the animals that live in freedom in a park. Some of them are difficult to see in the wild since they are elusive or can only be seen in schedules not compatible with visits (e.g., at night). With this game, visitors can explore the park and learn more about these species. In the game, a radar detects the approach to a predefined set of coordinates where the animal is found more often, sending a notification to the user. By touching the radar, it displays the information on that species. The player has to find and register information on each species. The user is then able to collect the "sticker" of that species. By gamifying the visit through the

[2]Beaconing deliverable D4.4 Prototype of the Authoring and Procedural System Framework Component, 2018. Available online at: https://beaconing.eu/wp-content/uploads/deliverables/D4.4.pdf.

[3]Beaconing deliverable D3.5 Game Design Document, 2016. Available online at: https://beaconing.eu/wp-content/uploads/deliverables/D3.5.pdf.

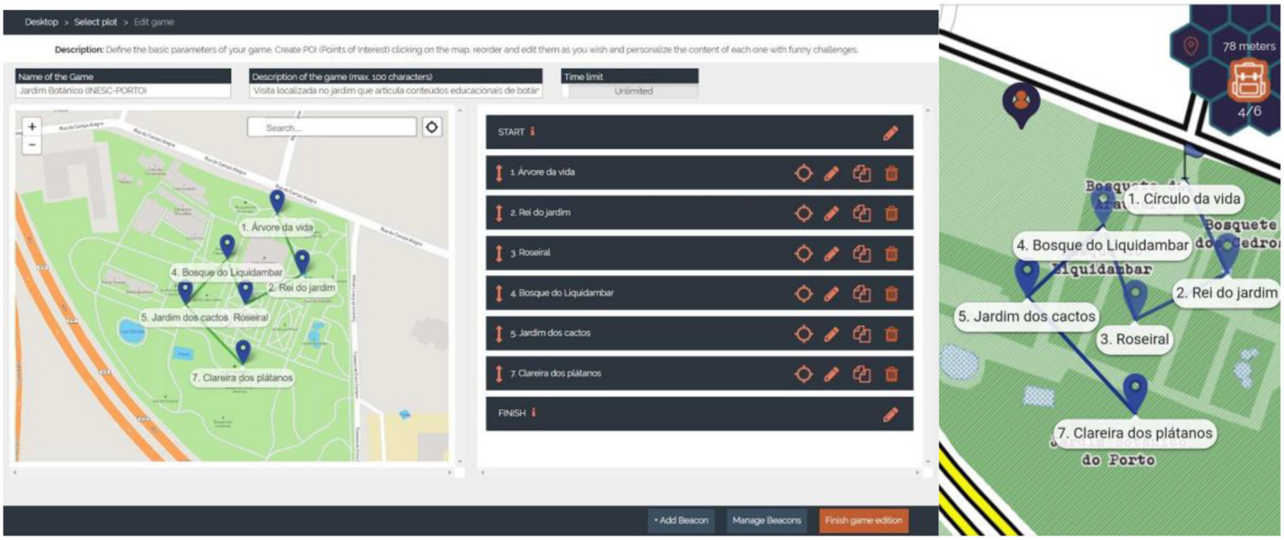


FIGURE 5 | Screenshot of the map displaying the entire path of the activity and the respective Points of Interest (authoring tool on the left; mobile app on the right).

FIGURE 6 | Screenshot and real-world aspects of playing the first challenge, which asks players to find the answer to the riddle in the circle at the entrance of the garden that is a map of the species on Earth. Translation of the picture: "Circle of the Species," "Search the origin of this story. Find our name in the circle of life".

dynamics of a sticker card collection (Coelho and Costa, 2017), the visit becomes more engaging, and a narrative is unveiled to the user as the collectibles are all captured. Curiosity to explore the location is improved by using an engaging GUI (**Figure 9**) such as the radar metaphor. This approach allows less disengagement from the natural environment by moving the *magic circle* toward the surrounding environment with a more pervasive application. Nevertheless, some issues about the usability of the mobile application have arisen. Usability is a highly relevant issue for pervasive games as the *magic circle* expands.

Other mechanics were introduced in this second prototype—the Footprints Memory game (Santos et al., 2017)—where the player learns to recognize some animal species by their footprints. Associations must be made between cards with images of animals and cards with footprint illustrations of those species. The goal is

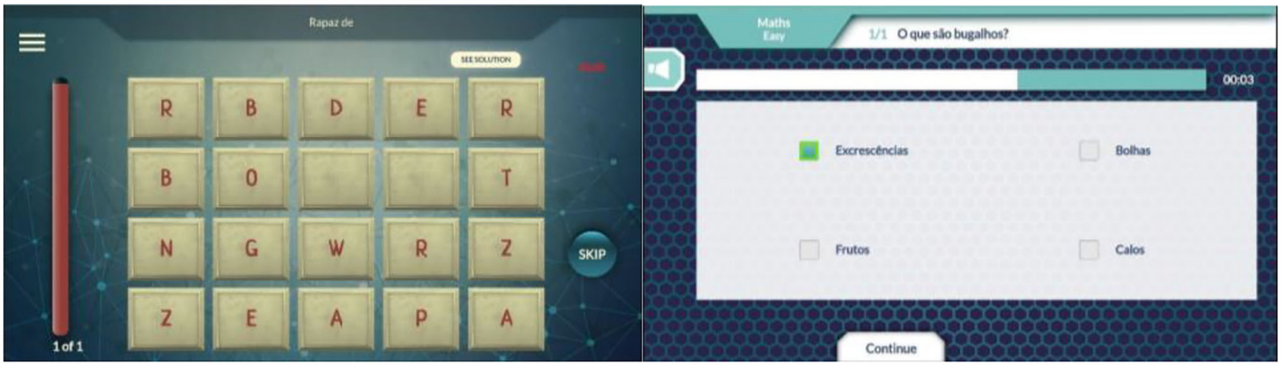


FIGURE 7 | Screenshots of two mini-games encountered in POI and actual ongoing activity within the garden space.

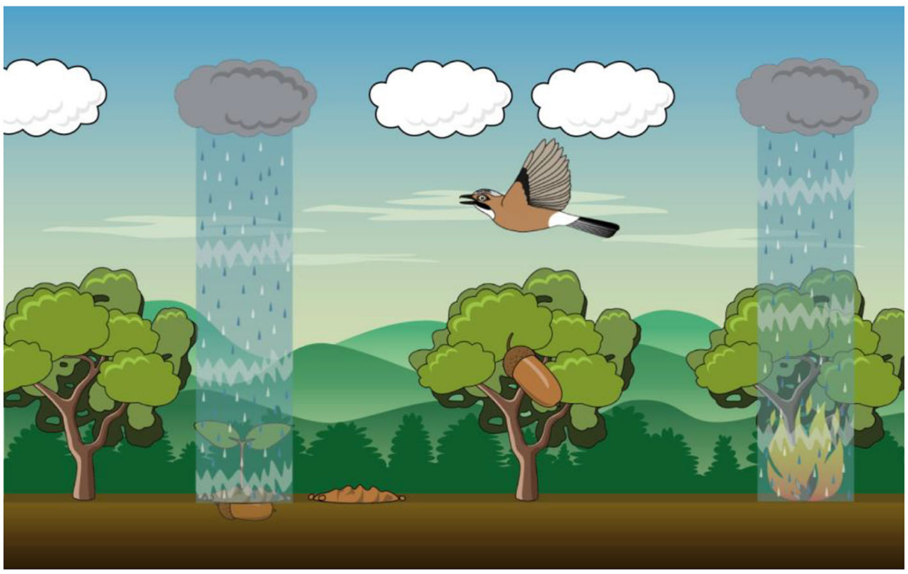


FIGURE 8 | Game of the Jay (Santos et al., 2017).

to be able to associate each footprint to their species as quickly as possible by revealing them two at a time (**Figure 10**).

Focusing on the need to increase awareness for behavior change, the Invasive Plants game (Santos et al., 2018) is a serious game to inform and raise awareness about invasive plant species in Portugal (**Figure 11**). It makes the visitor aware of the proper way of removing invasive species so that they will not spread throughout the local ecosystem, disrupting it, and avoiding damaging the ecosystem with those measures. The main goal of the game is to remove invasive plants, which consume all the oxygen from the lake's water and, due to their fast reproductive rate, can prevent sunlight from reaching the underwater flora. A user-centered design approach was used to develop the prototype, maximizing the usability and the fun factor, while keeping the educational content effective. This was performed through two iterations, including respectively 33 users in the first iteration and 30 users in the second. As a result of this process, the perception that the message of the game was clear increased from 75 to 83%, and the effectiveness of the science communication component increased from 76 to 80%. About the fun factor, the percentage of users that found the game fun increased from 48 to 80%. Overall, there is a need to introduce a user-centered approach in the development of this type of game to ensure the effectiveness of the science communication and fun in the game. The use of user-centered design is very important to achieve the goals on pervasive games. The main reason is that it values the needs and the critical opinions of the users to design the best opportunities to challenge them without disrupting the user experience of the contact with the surrounding environment.

FIGURE 9 | The Virtual Animal Detector (Santos et al., 2016).

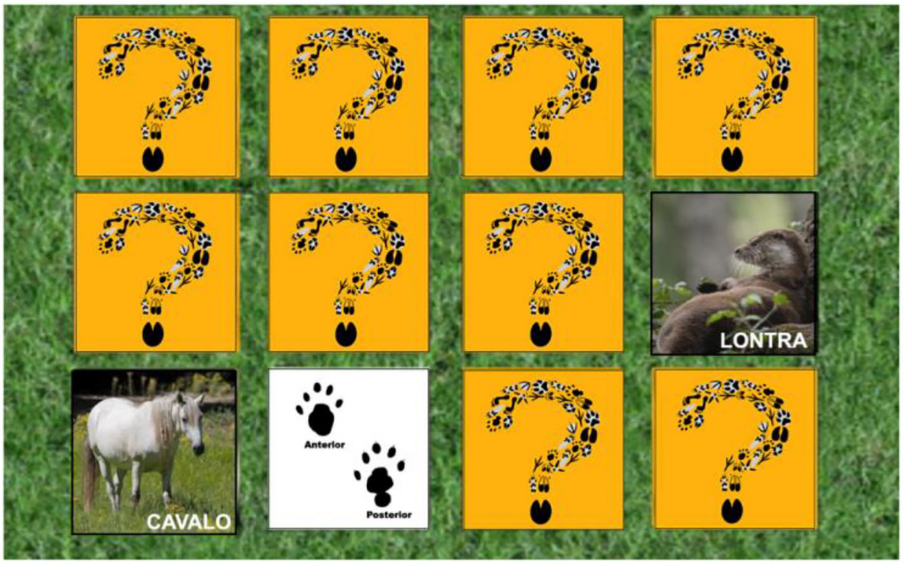


FIGURE 10 | Footprints Memory game (Santos et al., 2017).

To further extend the pervasiveness of the games, Augmented Reality was introduced in a new game—Birds of Prey (Santos et al., 2020). The purpose of this mini-game is to engage the visitors to spot *augmented* birds of prey flying above them and guess the species' name based on the birds' silhouette. When a bird is selected for the first time, it will be immediately unlocked

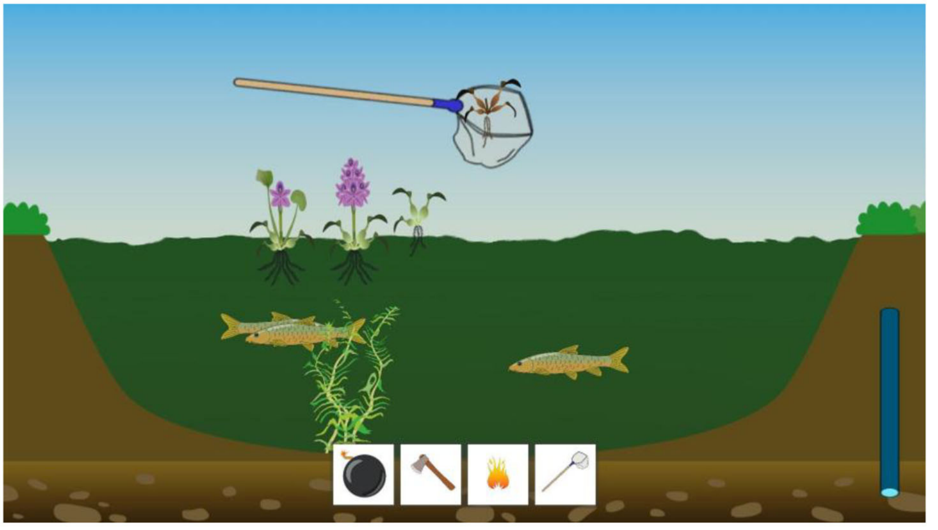


FIGURE 11 | The Invasive Plants game (Santos et al., 2018).

FIGURE 12 | The Birds of Prey game—binocular view (Santos et al., 2020).

and added to the collectibles board. The player then can access the Collectibles and view more information about the newly discovered bird. The game has an option to use *binoculars* that, when activated, will zoom in the game (as real binoculars), as can be observed in **Figure 12**. This in-game feature makes it easier to observe the differences between the birds' silhouettes.

A total of 34 volunteers participated in the evaluation of this prototype, where a simple questionnaire was introduced, with a 1–5 Likert scale. A median of 5 was obtained for fun and visuals and a median of 4 for the usability of the game. As expected, the introduction of Augmented Reality in pervasive games increases engagement and usability.

The Game of the Jay (Santos et al., 2017), the Virtual Animal Detector (Santos et al., 2016), the Footprints Memory (Santos et al., 2017), Invasive Plants (Santos et al., 2018), and Birds of Prey (Santos et al., 2020) are some examples of game prototypes for a nature park. Each explores different mechanics and pervasiveness on science communication and environmental awareness. The goal with these games is to amplify and enrich the experience of visiting natural parks, augmenting our perception of the natural world so we can better understand it and protect it. Nevertheless, these examples of Serious Pervasive Games can be pushed to other areas and explored for other educational contexts or behavior change.

ISSUES IN SERIOUS PERVASIVE GAMES

Serious games development has been studied, and some guidelines have already been proposed for learning and training effectiveness (Catalano et al., 2014). Nevertheless, for serious pervasive games, several additional issues must be taken into consideration, such as the previous work on location-based games (Jacob and Coelho, 2011). These issues can affect all magic circle expansions (Spatial, Temporal, and Social) (Nevelsteen, 2015). For instance, the player may be at a location unsuitable for the game (lack of geo-referenced content in the game), or the time of play may be inappropriate (bad weather, incorrect or inadequate time of day to play the game) or the player may lack companions or adversaries to play the game with (no other players physically present). While some of these issues are minor and could be addressed through adaptivity mechanisms or procedural content generation, other issues may pose health or safety concerns (Jacob et al., 2017). Specifically, techniques such as Action Prediction (McGee and Abraham, 2010), Player Modeling (Yannakakis et al., 2013), Adaptive Pacing (Thompson, 2014), Dynamic Game Difficulty Adjustment (Hawkins et al., 2012), and Learning Classifier Systems (Shafi and Abbass, 2017) have been applied successfully to games and could also be adapted to pervasive games to help mitigate some of these issues.

From the learning design perspective to the game design, the following subsections describe the main issues encountered.

Learning Design Issues

An issue impacting the use and dissemination of serious pervasive games is their associated workload from a teacher's perspective (Marklund and Taylor, 2016). To employ such a game within the constraints of time and curriculum, teachers need to realize how it will contribute to the learning goals planned for students, and monitor how the game is progressing. It is necessary to plan ahead several deployment tasks, such as instructions for students, instruments for keeping track of progress and collecting assessment elements. After deployment, it is necessary to manage the pervasive gaming activity while it takes place, ensuring that students get adequate and timely support and feedback, that there is adequate provision for teacher regulation, self-regulation, and co-regulation of learning among peers. Once the activity is completed, it is necessary to have elements to provide adequate assessment: otherwise, teachers may resort to use pervasive games only for awareness of motivational purposes, withdrawing them from assessment or worse, conducting assessment based only on right/wrong answer or item completion, thus bypassing all the context-aware meaningful capabilities of this medium. BEACONING's approach of combining game progress and activities with learning goals in a gamified lesson path, via a triadic assessment model (Baptista et al., 2016) is a significant contribution toward this issue (Cardoso et al., 2020). However, we believe that further advances are required in extracting meaning from pervasive activities. Teachers should be able to identify students' actions as part of foreseen (or unforeseen) choreographies of events, rather than mere marks of progress or completion. This can support more contextualized live feedback, by identifying that a student is pursuing a known deadlock or a best practice, and the same rich information about actions can then be employed for much richer assessment and post-activity feedback. There are tantalizing prospects of achieving this, by pursuing the concept of virtual choreographies as units of analysis, recording, and replaying of actions, as put into practice in proof-of-concept systems for energy consumption gamification (Cassola et al., 2017) and platform-independent virtual animations replay (Lacet et al., 2020). This is surely an issue that is necessary to be tackled for large-scale adoption of serious pervasive games.

For the design of this type of applications, some points must be considered:

1. In games-based learning, a positive feedback is required on different activities, creating a learning environment motivated by success. So, even failure must contain pertinent information that leads to learning. The success message should reinforce what has been learned.
2. Activities should be directed to *in situ* observation and should not involve prior knowledge of the subject. This allows one to broaden the age range of the applications, making them feasible for different groups.
3. The contents should have a simple and straightforward language using short sentences. Each activity should be limited to an idea that is in tune with the game mechanics.
4. All the information and navigation must be compatible with accessibility tools of the mobile devices (auditive speech, augmented lettering, contrast text, shortcut keys, and haptic information)
5. The objects found at the POIs must inspire, enthrall, and enlighten the player. At its core is the ability to inspire curiosity and stimulate learning.
6. Creative activities should allow the use of various tools and should have the possibility of social sharing.

Serious games must consider the different player profiles (active, passive)—explorer, collector, creative, and collaborative. They must bear in mind different game mechanics and the emotional involvement that is typical of this game typology.

Game and Level Design Issues

Since a location-based game is played in the spatial context of the player, extending the *magic circle* from the mobile device to the surrounding environment of the user, several relevant issues must be taken into consideration. The player must be kept safe during gameplay by avoiding dangerous places (e.g., road traffic or cliffs) or reducing the challenges' intensity in these places. This can be done by using indirect control, providing restrictions to the movement of the player, by introducing obstacles or enemies. Another consideration is that level design should be adapted to each of the players' locations or only allow for the player to play the game at the designated location. Also, there is always an unpredictability issue, considering that the environment may have changed due to construction works, events, or weather conditions. When navigation is required, intuitive navigational aids, such as maps or navigation cues, should be provided. Finally, user-centered design methodologies should be used to assure that the game is fun and, at the same time, keeps the effectiveness.

Hardware Limitations

Pervasive games are played through mobile devices. In location-based games, there is a need to position the player in the functional space of the game. In outdoor locations, it is simpler to use a Global Navigation Satellite System (GNSS) compatible device, although its low accuracy needs to be taken into consideration when designing the game. If the locations of the challenges are close together, that may lead to the deployment of the wrong challenge. Therefore, a radius with a dimension appropriate to the location error margin should be considered for checking the location, based on the hardware and the interference of the surrounding environment (e.g., buildings). If higher accuracy is required, devices such as Bluetooth beacons or RFID tags can be used. This is also the case of indoor environments, where GNSS technologies do not operate.

Another critical issue is also the Internet coverage for using client/server application to support the use of the players' context in pervasive games. If WiFi is available, the user may need to login to a specific network. Nevertheless, since the game is played in specific places, the game developers should ensure full coverage of the game area. If this WiFi network is not available, then the user should use the cellular data network from the player's data plan. This can also bring constraints as the user may incur additional costs, or the player's operator may not provide full coverage for that area.

Players may not be willing to use their personal smartphones to play the game. Many reasons may lead to this, be the battery drain, inadequate data plan, or the memory shortage to install the app, or being afraid to install an additional app. For this latter issue, the app can be developed in HTML 5, and be executed on a web browser.

Information Availability and Suitability

Pervasive games are based on contextual information, which may not be available for several reasons. Not all parts of the globe provide access to specific information, such as geographic information (e.g., maps) or related information (e.g., weather). Furthermore, profile information may not be available.

Player's Fitness and Pace

In location-based games, the player's movement around the real world is used as an input, often to move the avatar around the game world. Not all players can walk around the same amount of distance or overcome specific obstacles that involve climbing, for example, potentially leading to situations where both the player's and bystanders' safety is at risk. From children to senior adults, there should be a way to customize games to fit the players' fitness profile or provide adaptive gameplay (Jacob et al., 2017). In intense gameplays, there should also be considered the need to make a pause or adapt the distance or pace of the game to the weather conditions, events, or disruptions that may occur.

Player's Data Protection

When playing pervasive games, there are several data that may be private to the user, such as the location and date of the games. When considering multiplayer or social games this fact could be relevant. Furthermore, pervasive apps should comply with the legislation of the place of play. Regarding the storage of this sensitive information, it is preferable to store it remotely in a server as this option provides a safer place for the player to store personal data, particularly considering how easy it is to lose a mobile phone or someone unauthorized accessing its contained information.

CONCLUSION AND FUTURE CHALLENGES

Pervasive games have a potential for engaging the user with contextual challenges, establishing a connection to the surrounding environment. Serious Pervasive Games lie at the intersection of pervasive games with serious games and raise questions on how to design for experiential learning, which by its nature implies a design that favors players' attention to both the physical domain as for the digital realm.

Serious pervasive games have a high potential for Education, particularly for problem-based learning methods. Pervasiveness can be adjusted to many conditions: from spatial contextualized games to augmented reality games, they provide distinct levels of engagement. Nevertheless, several issues should be considered when designing a pervasive game.

One of the important issues of our research is related to co-creation and the monitoring that must be given to users/creators of this type of games. In this sense, a research methodology centered on content creators is required, which will allow co-creation more focused on learning content.

The technical support of a game designer in the initial conception can be essential for the acquisition of the technology and mechanics involved in mini-games. In this sense, we will conduct workshops with focus groups to identify the monitoring needs that this type of user needs.

The creation processes, difficulties, and concerns of these creators will be identified, with the objective of creating a system based on procedures. In the future, it may be important to improve in any system that involves the creation of pervasive games and augmented reality experiences.

Another important issue concerns user personalization; that is, depending on the player's profile, the system can adapt to the mechanics that underlie it, which may prove essential in terms of user satisfaction and adequacy to the profile and needs, being particularly important in terms of accessibility.

Summarizing, the future challenges of serious pervasive games include authoring tools that facilitate the creation of pervasive content for learning or other application areas, game engines suitable for its development, and robust frameworks that overcome the issues reported.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This work was financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-030740 - PTDC/CCI-COM/30740/2017. Part of this work has also been supported by the European Union's Horizon 2020 - The EU Framework Programme for Research and Innovation 2014–2020, under grant agreement No. 687676. We would also acknowledge the research grant from the Operation NORTE-08-5369-FSE-000049 supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Social Fund (ESF), and also the FCT-Austin grant Ref. PD/BD/142893/2018.

REFERENCES

Arango-López, J., Cerón Valdivieso, C. C., Collazos, C. A., Gutiérrez Vela, F. L., and Moreira, F. (2019). CREANDO: Tool for creating pervasive games to increase the learning motivation in higher education students. *Telematics Informatics* 38, 62–73. doi: 10.1016/j.tele.2018.08.005

Arango-López, J., Collazos, C. A., Velas, F. L. G., and Moreira, F. (2018). Using pervasive games as learning tools in educational contexts: a systematic review. *Intern. J. Learn. Technol.* 13, 93–114. doi: 10.1504/IJLT.2018.092094

Arango-López, J., Gallardo, J., Gutiérrez, F. L., Cerezo, E., Amengual, E., and Valera, R. (2017). Pervasive games: Giving a meaning based on the player experience. *ACM Intern. Conf. Proceed. Ser.* Part F1311. doi: 10.1145/3123818.3123832

Arnab, S., Tombs, G., Duncan, M., Smith, M., and Star, K. (2016). "Towards the blending of digital and physical learning contexts with a gamified and pervasive approach," in *Games and Learning Alliance. GALA 2015. Lecture Notes in Computer Science*, Vol. 9599, eds A. De Gloria and R. Veltkamp (Cham: Springer). doi: 10.1007/978-3-319-40216-1_50

Baptista, R., Coelho, A., and Vaz de Carvalho, C. (2016). "Relation between game genres and competences for in-game certification," in *Serious Games, Interaction, and Simulation. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Vol. 161, eds C. Vaz de Carvalho, P. Escudeiro, and A. Coelho (Cham: Springer). doi: 10.1007/978-3-319-29060-7_6

Benford, S., Rowland, D., Hull, R., Reid, J., Morrison, J., Facer, K., et al. (2004). "Savannah: designing a location-based game simulating lion behavior," in *International Conference on Advances in Computer Entertainment Technology*. Retrieved from http://eprints.lincoln.ac.uk/659/

Bourazeri, A., Arnab, S., Heidmann, O., Coelho, A., and Morini, L. (2017). Taxonomy of a gamified lesson path for STEM education: the beaconing approach. *Proc. 11th Euro. Conf. Games Based Learn. ECGBL.* 1, 29–37.

Cardoso, P., Morgado, L., Coelho, A. (2020). *Authoring Game-Based Learning Activities that are Manageable by Teachers.* ERCIM News. 120, 17–18.

Carvalho, A. A. A., and Guimarães, D. (2018). "The EduPARK game-like app with Augmented Reality for mobile learning in an urban park," in *Atas do 4.o Encontro sobre Jogos e Mobile Learning*.

Cassola, F., Iria, J., Paredes, H., Morgado, L., Coelho, A., and Soares, F. (2017). *Using Choreographies to Support the Gamification Process on the Development of an Application to Reduce Electricity Costs. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Springer Verlag), 256–257.

Catalano, C. E., Luccini, A. M., and Mortara, M. (2014). Guidelines for an effective design of serious games. *Intern. J. Seri. Games* 1. doi: 10.17083/ijsg.v1i1.8

Coelho, A., and Costa, L. M. (2017). "The integration of augmented reality and the concept of sticker album collection for informal learning in museums," in *Immersive Learning Research Network. iLRN 2017. Communications in Computer and Information Science*, *Vol. 725,* ed D. Beck et al. (Cham: Springer). doi: 10.1007/978-3-319-60633-0_9

Falk, J., and Davenport, G. (2004). *Live Role-Playing Games: Implications for Ubiquitous Computer Game Interfaces*.

Felipe, E., Massa, S., and Rico, C. A. (2018). "Pervasive Serious Game for Development Skills in Computer Networking," in *2018 IEEE Biennial Congress of Argentina (ARGENCON)* (IEEE), 1–7.

Hawkins, G., Nesbitt, K., and Brown, S. (2012). Dynamic difficulty balancing for cautious players and risk takers. *Intern. J. Comp. Games Technol.* 2012, 1–10. doi: 10.1155/2012/625476

Huizinga, J. (1938). *Homo Ludens: Proeve Ener Bepaling Van Het Spelelement Der Cultuur. Groningen*, Wolters-Noordhoff cop. (original Dutch edition - A Study of the Play-Element in Culture).

Jacob, J., and Coelho, A. (2011). Issues in the development of location-based games. *Intern. J. Comp. Games Technol.* 2011:7. doi: 10.1155/2011/495437

Jacob, J., Lopes, A., Nóbrega, R., Rodrigues, R., and Coelho, A. (2017). "Player adaptivity and safety in location-based games," in *Advances in Computer Entertainment Technology. ACE 2017. Lecture Notes in Computer Science, Vol. 10714,* eds A. Cheok, M., Inami and T. Romão (Cham: Springer), 219–238. doi: 10.1007/978-3-319-76270-8_16

Lacet, D., Penicheiro, F., Morgado, L., and Coelho, A. (2020). "Preserving story choreographies across multiple platforms: an approach to platform-independent reuse of characters' behaviors for games, simulations, animations and interactive videos," in *ARTECH 2019 Proceedings* (ACM). doi: 10.1145/3359852.3359916

Laine, T. H., and Joy, M. (2009). Survey on context-aware pervasive learning environments. Intern. *J. Interactive Mobile Technol.* 3, 70–76.

Lima, C., Schlemmer, E., and Morgado, L. (2020). "Managing learning in environments where students move: a panorama of problems and contributions," in *Atas do 5° Encontro sobre Jogos e Mobile Learning, Centro de Estudos Interdisciplinares do Século 20* (Universidade de Coimbra, Portugal), 248–258.

Marklund, B. B., and Taylor, A. S. A. (2016). Educational games in practice: the challenges involved in conducting a game-based curriculum. *Electronic J. e-Learn.* 14, 122–135.

Martinez-Ortiz, I., Perez-Colado, I., Rotaru, D. C., Freire, M., and Fernandez-Manjon, B. (2019). *From Heterogeneous Activities to Unified Analytics Dashboards*. (Zenodo). doi: 10.1109/EDUCON.2019.8725222

McGee, K., and Abraham, A. (2010). "Real-time team-mate AI in games: a definition, survey, and critique," in *Proceedings of the Fifth International Conference on the Foundations of Digital Games (FDG'10). Association for Computing Machinery* (New York, NY), 124–131. doi: 10.1145/1822348.1822365

Montola, M. (2005). Exploring the edge of the magic circle: defining pervasive games. *Proc. DAC* 1966, 16–19.

Montola, M., Stenros, J., and Waern (2009): *A. Pervasive Games: Theory and Design. Experiences on the Boundary Between Life and Play*. Amsterdam: Morgan Kaufmann. doi: 10.1201/9780080889795

Nevelsteen, K. J. L. (2015). "A survey of characteristic engine features for technology-sustained pervasive games," in *Springer Briefs in Computer Science* (Cham: Springer).

Nieuwdorp, E. (2007). The pervasive discourse: an analysis. *Comp. Entertain.* 5:13. doi: 10.1145/1279540.1279553

Parker, J. (2011). "A design-based research approach for creating effective online higher education courses," in *26th Annual Research Forum: Educational Possibilities* (Fremantle, WA: Western Australian Institute for Educational Research Inc; University of Notre Dame).

Pontefract, D. (2013). *Flat Army: Creating a Connected and Engaged Organization*. Boise, ID: Wiley.

Raessens, J. (2006): Playful identities, or the ludification of culture. *Games Cult.* 1, 52–57. doi: 10.1177/1555412005281779

Salen, K., and Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge; London: The MIT Press.

Santos, L., Coelho, A., and Nóbrega, R. (2018). "Game design for environmental awareness - the case of 'Invasive Plants'," in *10th Conference on Videogame Sciences and Arts - Videojogos* 2018.

Santos, L., Pereira, D., Beça, P., and Coelho, A. (2017). Aplicação móvel para divulgação do património natural no turismo. *Turismo Desenvolv.* 27/28, 1461–1473.

Santos, L., Pereira, D., Nóbrega, R., Beça, P., and Coelho, A. (2016). Jogos sérios digitais para a promoção do património natural – o caso de um jogo de exploração de um parque na busca dos seus animais [poster + demo] {BEST DEMO}. 33:2991.

Santos, L., Silva, L., Nóbrega, R., Almeida, R., and Coelho, A. (2020). "An interactive application framework for natural parks using serious location-based games with augmented reality," in *Proceedings of GRAPP 2020 - The International Conference on Computer Graphics Theory and Applications*, (to be published). doi: 10.5220/0008947602470254

Santos, V., and Burguès, R. P. (2017). "Authoring a serious pervasive game for reflecting upon urban spaces," in *Interactivity, Game Creation, Design, Learning, and Innovation,* (Cham: Springer), 74–84. doi: 10.1007/978-3-319-76908-0_8

Sawyer, B., and Rejeski, D. (2002). *Serious Games: Improving Public Policy Through Game-based Learning and Simulation*. Washington, DC: Woodrow Wilson International Center for Scholars.

Schaal, S., Otto, S., Schaal, S., and Lude, A. (2018). Game-related enjoyment or personal pre-requisites–which is the crucial factor when using geogames to encourage adolescents to value local biodiversity. *Int. J. Sci. Educ.* 8, 213–226. doi: 10.1080/21548455.2018.1441571

Shafi, K., and Abbass, H. (2017). A survey of learning classifier systems in games. *IEEE Comput. Intellig. Magazine*, 12, 42–55. doi: 10.1109/MCI.2016.2627670

Syvänen, A., Beale, R., Sharples, M., Ahonen, M., and Lonsdale, M. (2005). "Supporting pervasive learning environments: adaptability and context awareness in mobile learning," in *Proceedings of the 2005 IEEE International Workshop on Wireless and Mobile Technologies in Education* (Tokushima).

Taborda, J. P., Arango-López, J., Collazos, C. A., Vela, F. L. G., and Moreira, F. (2019). Effectiveness and fun metrics in a pervasive game experience: a systematic literature review. *Adv. Intellig. Syst. Comp.* 932, 184–194. doi: 10.1007/978-3-030-16187-3_18

Thompson, T. (2014). *In the Directors Chair: Left 4 Dead. AI and Games.* Available online at: http://t2thompson.com/2014/12/01/in-the-directors-chair-left-4-dead/

Valente, L., and Feijó, B. (2014). "Extending Use Cases to Support Activity Design in Pervasive Mobile Games," in *Brazilian Symposium on Games and Digital Entertainment, SBGAMES, 2014-December(December)*, 193–201. doi: 10.1109/SBGAMES.2014.11

Valente, L., Feijó, B., and Leite, J. C. S. do P. (2017). Mapping quality requirements for pervasive mobile games. *Requirem. Eng.* 22, 137–165. doi: 10.1007/s00766-015-0238-y

Viana, J. R. M., Viana, N. P., Trinta, F. A. M., and De Carvalho, W. V. (2014). "A systematic review on software engineering in pervasive games development," in *Brazilian Symposium on Games and Digital Entertainment, SBGAMES, 2014-Decem(December)*, 51–60. doi: 10.1109/SBGAMES.2014.16

Yannakakis, G. N., Spronck, P., Loiacono, D., and Andre, E. (2013). "Player modeling," in *Artificial and Computational Intelligence in Games*, eds S. M. Lucas, M. Mateas, M. Preuss, P. Spronck, and J. Togelius (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik), 45–59. Available online at: https://www.um.edu.mt/library/oar/handle/123456789/29725

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 08 September 2020
doi: 10.3389/fpsyg.2020.02254

Social Conformity in Immersive Virtual Environments: The Impact of Agents' Gaze Behavior

Christos Kyrlitsias[1,2], Despina Michael-Grigoriou[1,2], Domna Banakou[1,3] and Maria Christofi[1,2]*

[1] GET Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Limassol, Cyprus, [2] Research Centre on Interactive Media, Smart Systems and Emerging Technologies – RISE, Nicosia, Cyprus, [3] Event Lab, Department of Clinical Psychology and Psychobiology, University of Barcelona, Barcelona, Spain

Immersive virtual reality (IVR) can induce an experience of "social presence" which can, in turn, increase social influence. Non-verbal behavior such as eye contact is an important component of human communication and, therefore, an important factor in creating social presence. This paper presents an experimental study that elaborates social influence through conformity with a group of virtual agents within an immersive virtual environment (IVE). Specifically, it investigates the impact of the agents' gaze behavior on social presence and influence. An experiment based on the Asch (1951) paradigm using two levels of agents' gaze behavior (Eye Contact condition vs. No-Eye Contact condition) was conducted. The results showed that participants conformed with the agents as they gave significantly more incorrect responses to the trials that the agents also gave an incorrect response, compared to those trials that the agents gave correct answers. However, no impact of the agents' gaze behavior on conformity was observed, even if the participants in the Eye Contact condition reported a higher sense of social presence. In addition, self-reported measures showed a number of social effects that occurred only in the eye contact condition, indicating that the agents' gaze behavior has an impact on participants' experience.

Keywords: virtual reality, agents, behavioral realism, social influence, conformity, social presence, eye contact

OPEN ACCESS

Edited by:
*Jože Rugelj,
University of Ljubljana, Slovenia*

Reviewed by:
*Pradipta Biswas,
Indian Institute of Science (IISc), India
Pradeep Raj Krishnappa Babu,
Duke University, United States*

****Correspondence:***
*Despina Michael-Grigoriou
despina.grigoriou@cut.ac.cy*

Specialty section:
*This article was submitted to
Human-Media Interaction,
a section of the journal
Frontiers in Psychology*

Received: *30 January 2020*
Accepted: *11 August 2020*
Published: *08 September 2020*

Citation:
*Kyrlitsias C, Michael-Grigoriou D,
Banakou D and Christofi M (2020)
Social Conformity in Immersive Virtual
Environments: The Impact of Agents'
Gaze Behavior.
Front. Psychol. 11:2254.
doi: 10.3389/fpsyg.2020.02254*

INTRODUCTION

In our daily lives, our decisions are greatly influenced by others. Our attitudes, our beliefs, and our behavior are influenced in a way that meets the demands of our social environment. This act of matching attitudes, beliefs, and behaviors to group norms, known as conformity, is one of the most powerful forms of social influence. This effect was initially studied by Jenness (1932), who asked the participants to estimate the number of beans in a bottle, individually. Then the participants were divided into small groups and were asked to discuss the task and to provide a common estimate, and finally, they were provided with the opportunity to revise their initial individual estimates. The results showed that the majority of the participants changed their initial estimation toward that of the group. Sherif (1935) conducted a series of experiments using the autokinetic effect, the illusion of movement in the absence of a reference point (spot of light in a dark room). When the participants were asked to individually estimate how far the light moved, their answers varied considerably. Yet, when they were asked to do the same in groups of three, stating their estimates out loud, Sherif found that their estimates converged. The most famous experimental approach of conformity, however, was the one carried out by Asch (1951, 1955, 1956). Asch conducted an experiment to investigate the extent to which social pressure from the majority can influence an individual and make him/her conform. Participants were placed in a room along with seven confederates and were asked to answer some simple line-length comparison tests. The confederates' responses had been agreed in advance. The participant was led to believe that the other six attendees were also real participants and not part of the experiment's scenario. The results demonstrated that the participants were affected by the pressure of the majority of others. Approximately one third of all estimates in the critical group were distorted in the direction of the majority.

One of the most common uses of Immersive Virtual Reality (IVR) technologies is the simulation of real or hypothetical scenarios (Slater and Sanchez-Vives, 2016) in entertainment, education (Stavroulia et al., 2018), training (Bombari et al., 2015), and research (Blascovich et al., 2002; Bombari et al., 2015), just to name a few. The use of Virtual Reality (VR) applications is expanding dramatically as IVR technologies are becoming more affordable and have the ability to provide a controlled, realistic, safe, and accessible experience to the user. Many of these scenarios include social interactions with agents, computer-generated representations of humans whose behaviors are determined by the computer (Blascovich et al., 2002; Von der Pütten et al., 2010) and who can play various social roles such as that of an audience (Nazligul et al., 2017) and teachers and students (Stavroulia et al., 2018). The social interaction with the agents plays a crucial role in the effectiveness of these applications, which requires that the agents behave as real people, so that the user reacts to them realistically (Bombari et al., 2015).

Recent studies have explored whether conformity can also be caused by virtual humans. The results of a study (Kraemer, 2013) replicated the Asch (1951) experiment within Second Life, a non-immersive virtual world that enables users to create virtual representations of themselves and interact with other users. The confederates in that study were avatars controlled by other users, and the participants were aware of that. The results showed that the participants were more likely to make the same choices as the confederates, compared to participants tested alone. A different study (Midden et al., 2015) compared the conformity with a group of humans, computers, and virtual agents displayed on computer screens. The results showed that conformity can be exerted by artificial majorities as participants conformed to the virtual agents' group and the computer's group only in a high-ambiguity task. The impact of the ambiguity of the task on conformity with non-humans was also addressed by Weger et al. (2015), who demonstrated that conformity was greater in more ambiguous tasks. Similar were the findings of a study reported by Hertz and Wiese (2016) that investigated social conformity with agents using three levels of human-likeness (computer, robot, and human). The results showed greater conformity in the high-ambiguity condition. Conversely, no effect of human-likeness on conformity was observed.

These studies, however, are limited to physical agents such as robots, and conversational agents on computer screens without the use of IVR technologies. IVR technologies have several advantages, which make them a very useful tool for reproducing social scenarios (Blascovich et al., 2002). One of the advantages of using IVR as a tool for psychological experiments is its ability to offer a high level of experimental control. This enables the researcher to conduct experiments which would be otherwise very difficult and inefficient.

An IVR system, thanks to its capability to provide a multisensory and interactive representation of a virtual environment, can induce to the user the illusion of being in this environment. This sense of "being there" is called presence (or place illusion) (Slater, 2009). Furthermore, IVR can induce an experience of "social presence" (or "co-presence"), the feeling of sensing another entity being present, which can in turn increase social influence (Oh et al., 2018). In this paper, we define social presence as the "sense of being together," including the feeling that the "other" is a sentient human being (Oh et al., 2018). For example, a recent study (Bailey et al., 2019) showed that children in an IVR condition demonstrated social compliance to a greater extent than children in a non-immersive condition, suggesting that IVR may elicit differential cognitive and social responses compared to less immersive technologies.

Nonetheless, the exploration of social conformity within immersive virtual environments (IVEs) is still limited. The results of a study by Bailenson et al. (2008a,b) showed that the participants conformed to the virtual classmates who exhibited either positive (attentive and focused their gazes on the teacher) or negative (distracted and did not pay attention to the teacher) learning behaviors. Specifically, the participants' learning abilities were affected by the virtual co-learners' behavior in a virtual classroom.

A recent study (Kyrlitsias and Michael, 2016; Kyrlitsias and Michael-Grigoriou, 2018) replicated the Asch experiment in order to investigate conformity with virtual humans in IVEs using two experiments. The results of the first experiment showed that participants' response time was affected by the virtual agents' answers, indicating some level of social pressure, but the participants' judgments were not affected. In a follow-up experiment, a similar procedure was used in order to investigate the effect of behavioral realism (gaze behavior) and agency (the extent to which the user believes that a virtual human is controlled by a real human rather than the computer) on conformity. The results showed that participants conformed to some extent to the virtual confederates, but no effects of agency and gaze behavior on conformity were observed. However, the level of conformity was very low, as only 4 of the 52 participants (7.69%) conformed to some extent. Nonetheless, this minimal conformity rate did not allow any clear conclusions to be drawn regarding the effect of gaze behavior and agency on conformity. **Supplementary Table 1** summarizes the IVR experiments using the Asch paradigm, including the present study.

The primary aim of the current study was to investigate social conformity with a group of virtual agents within an IVE. To test that, we followed a procedure similar to the original Asch (1951) experiment, using the line-length comparison task. We designed an IVR version of Asch's experiment, with four virtual agents as confederates. Based on our previous findings, and in order to achieve a higher conformity rate than in our previous study (Kyrlitsias and Michael-Grigoriou, 2018), we increased the task difficulty and also reduced the sense of anonymity of the participants. In order to increase the difficulty of the task, we limited the trial card projection duration, which is a common way of increasing the task difficulty in such experiments (Hertz and Wiese, 2016; Midden et al., 2015). Based on pilot tests, we balanced the task difficulty (projection duration and line-length differences), so that it was challenging enough, but the participants could still figure out the correct answer. In order to reduce the sense of anonymity, participants were asked to introduce themselves to the agents before the procedure began. Our prediction was that participants would conform to the virtual humans' judgments by giving more incorrect answers to trials where the confederates gave a wrong response, than to the trials where the confederates answered correctly.

Additionally, by manipulating the agents' non-verbal behavior, and specifically the gaze behavior, we examined whether this factor has an impact on the level of conformity. Non-verbal behavior (such as eye contact, interpersonal distance, facial expressions, and gestures) is an important component of human communication (Bente et al., 2007) and, therefore, an important factor for the creation of social presence (Oh et al., 2018). In this study, we created two levels of gaze behavior for the agents, the Eye Contact (EC) condition and the No Eye Contact (NEC) condition. In the NEC condition, the agents had no gaze behavior, and therefore they did not make any eye contact with the participant or the other agents (**Figure 1**, left). In the EC condition, during the answering phases of the procedure, the agents turned their gaze toward the one who was responding at that moment, whether that was the participant or another agent (**Figure 1**, right). We presume that by adding an extra social cue, such as eye contact, a stronger sense of social presence will be induced to the participants, and, as social presence has an impact on social influence (Oh et al., 2018), we predicted that conformity in the EC condition would be greater than in the NEC condition.

This manipulation of the agent confederates' gaze behavior is an example of the advantage of the use of IVR technologies for this kind of experiments. By using human confederates, it would be almost impossible to control their gaze behavior between the trials and the experimental sessions. Here the recruitment of virtual agents allowed us to have total control over the experimental protocol and study the impact of the confederates' gaze behavior on participants' conformity.

In addition, we measured the participants' self-esteem (Rosenberg, 1965), a personality characteristic that is related to conformity (Gergen and Bauer, 1967), in order to understand how a person's individual characteristics may affect social behavior with agents in IVEs. We also addressed the participants' subjective evaluations regarding their experience in the virtual environment. The participants were asked to assess their sense of presence in the virtual environment and evaluate the agents' behavioral realism. Moreover, we asked the participants to state how confident they felt about their answers and whether they were influenced by the responses of the virtual humans. Finally, using the head tracking provided by the VR headset, we recorded the participants' gaze direction during their experience. Specifically, we recorded the duration that each participant looked toward the agents and the duration of the mutual gaze between the participant and the agents. This was done to examine the impact of social pressure and the agents' behavior on the participants' non-verbal behavior.

MATERIALS AND METHODS

Experimental Design

Overall, 41 volunteers over 18 years of age participated in the study. One of them did not complete the experiment, while two participants were excluded as they were aware of the original Asch experiment on which the study was based, and which could have biased their responses. Therefore, in total, data collected from 38 participants, of whom 26 were female, were used in this study. This was a between-group design, and participants were randomly assigned to one of the two conditions, the EC group or the NEC group (**Table 1**). All participants signed a consent form which was a prerequisite for participation in the study.

Ethics Statement

All participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individuals for the publication of any potentially

FIGURE 1 | The virtual environment and the virtual agents as seen from the participant's perspective in the NEC **(left)** and EC conditions **(right)** at a time in which the participant is required to provide an estimate.

TABLE 1 | Distribution of participants over conditions and summarized pre-VR questionnaire measures.

	Conditions	
	NEC	EC
N (Males)	18 (6)	20 (6)
Mean ± S.E. Age	27.88 ± 1.845	24.4 ± 3.872
Median VR Experience (IQR)	2 (3)	3 (4)
Median 3D Experience (IQR)	2 (3)	2 (2)
Mean ± S.E. Self-Esteem	33.06 ± 0.979	29.05 ± 1.05

identifiable images or data included in this article or the **Supplementary Material**.

Technical Setup

The experiment was performed using a PC equipped with an NVidia GeForce GTX 1060 graphics card. The setup included the Oculus Rift (oculus.com/rift) head-mounted display (HMD) with 2160 × 1200 resolution (1080 × 1200 per eye), 110° field of view, and 90 Hz refresh rate for 3D immersive viewing, head rotational, and positional tracking, and providing spatialized audio. The application was created using the Unity (version 2018.2.1) game engine and the environment using Autodesk Maya and Adobe Photoshop. The virtual characters were designed and rigged using Autodesk Character Generator. For the lip synchronization feature, the SALSA plug-in for Unity was used.

Virtual Agents

Four animated virtual agents were created for the experiment, two male and two female. Two body animation clips were created for each agent, an "idle" and an "answering" animation. The "idle" animation included breathing movements and was repeated for most of the time. The "answering" animation clip included some movement of the body and the hands and was playing each time the agents gave their answers. The above animation clips were slightly different for each agent. Also, to improve the realism, the agents performed blinking and lip movement animations using blend-shapes. The lip-movement animation was synchronized with the audio to simulate speaking. The audio clips used for the agents' voice were pre-recorded.

An inverse kinematic technique directed by a scrip was used for the agents' head movement and gaze manipulation. When the trials were projected, in both conditions the agents turned their heads toward the board. During the answering phase, in the EC condition all the agents turned their heads toward the one answering, including the participant (**Figure 1**, right), performing eye contact. The participants' head position was tracked dynamically using the HMD's positional tracking. During the answering phase in the NEC condition, the agents were looking straight ahead (**Figure 1**, left). An amount of randomness was applied to the delay and the speed of the agents' head movement in order to make it look more natural and less robotic.

Procedure

Upon their arrival at the laboratory, the participants received written information about the study and filled in the consent form. Then, they were asked to complete a pre-VR questionnaire that included demographic information as well as the Rosenberg self-esteem test.

After they were fitted with the virtual reality HMD and the necessary calibration was done, the virtual room with the 4 virtual agents (**Figure 1**) was presented and a familiarization period of 30 s began. Next, the instructions phase began, where pre-recorded instructions explaining the task and the process of the study were played back to the participant. During this phase, which lasted 2 min, the agents and the participant were asked to verbally introduce themselves by stating their first names, their age, and their occupation. This was done so that the participants could better understand the order and the way in which they would give their responses during the different trials, and to reduce the sense of the participants' anonymity. Thirty seconds after the instruction phase was completed, the trial session began. Each trial was presented on the virtual boards for 5 s and then the agents and the participant gave their judgments in sequence. The participant was placed in the last (fifth) position and, therefore, gave his/her judgment after listening to the other four agents' judgments. This procedure was repeated for all the 18 trials. Examples of different trial sessions for both experimental conditions are presented in **Supplementary Video 1**. More details about the trials are presented in the Trials section and **Table 2**. During this session, the participants' answers to each trial (Trial Error) and the participants' Response Time to each trial were automatically recorded by the software. Also, using the head tracking provided by the VR HMD, the Eye Contact Time, and the Mutual Gaze Time (in the EC group only) were recorded. More information about the recorded data can be found in the Measures section.

After the trials session, participants were asked to complete a post-VR questionnaire regarding their experience (**Table 3**). Finally, the participants were verbally asked whether they were familiar with the original Asch's (1951) conformity experiment and they were debriefed.

Trials

Overall, there were 18 line-length comparison trials. Each trial was presented for 5 s and had only one correct answer. The agents gave their answers in all trials unanimously.

Six of the trials (trials 1, 2, 5, 10, 11, and 14) were "non-critical," and the agents gave the correct answer to all of them. The non-critical trials were used as training trials and were not considered in the analysis. The use of non-critical trials is a technique used in this kind of experiments (e.g., Asch, 1956; Hertz and Wiese, 2016), and their purpose was to avoid causing any confusion to the participants regarding the length comparison task and generate some trust toward the agents. This is the reason that the opening trials are non-critical.

The remaining 12 trials (3, 4, 6, 7, 8, 9, 12, 13, 15, 16, 17, and 18) were the "critical" trials. The agents gave the correct answer to the 6 critical trials (3, 6, 8, 13, 16, and 18) and a wrong answer

TABLE 2 | Trials with the correct answers and the answers given by agents.

Trial Number	1*	2*	3	4	5*	6	7	8	9	10*	11*	12	13	14*	15	16	17	18
Correct Answer	A	B	A	C	A	C	C	B	A	A	B	A	C	A	C	C	B	A
Agents' Answer	A	B	A	A	A	C	A	B	B	A	B	C	C	A	B	C	C	A

Non-critical trials.

TABLE 3 | The questions of the post-VR questionnaire.

Measures	Variable name	Question
a. Presence	there	How do you assess your sense of presence in the virtual room where the trials were carried out?
	real	To what extent, during your experience, the virtual world has become the "reality" for you?
	lab*	During your experience, which sensation was stronger, the feeling that you were in the virtual room, or the feeling that you were in the laboratory where the study was being carried out?
b. Perceived Behavioral Realism	behave	The other participants in the study behaved like real people.
	move	The other participants were moving like real people.
	talk	The other participants spoke as real people.
	feel	I had the feeling that the other participants were real people.
c. Social Presence	sameRoom	I had the feeling that the other participants were with me in the same room.
	otherPerceived	I had the feeling that the other participants were aware of my presence.
	otherListen	I had the feeling that the other participants were listening to my answers.
	alone*	I had the feeling that I was alone in the room.
d. Response Confidence	correctAnswers	The answers I gave to the study were correct.
	difficult*	The tests were difficult.
	doubts*	I have doubts about the correctness of the answers I gave to the examination.
	confidentAnswers	I felt confident about my answers.
e. Self-Reported Conformity	myOpinion*	The answers I gave to the study were mainly based on my own opinion.
	otherOpinion	The answers given by the other participants in the study affected my own.

Reverse interpretation.

to the other 6 (4, 7, 9, 12, 15, and 17). The correct answers to each trial as well as the answers given by the agents are summarized in **Table 2**.

The first 9 trials were identical with the other 9 in the same order (trial 1 was the same as trial 10, trial 2 as trial 11, and so on). In this way, each participant was asked to respond to each critical trial twice, once after the agents gave the correct answer and once after they unanimously gave a wrong answer. This was done in order to balance the task difficulty between the critical trials that the agents responded correctly to, and the critical trials where they gave a wrong answer. Participants were not aware of this manipulation.

Measures

Pre-VR Questionnaire

Using a questionnaire that was given to the participants before their exposure to the virtual world, we recorded various data on demographic characteristics such as gender, age, intimacy with 3D environments and virtual reality, and self-esteem. These are summarized in **Table 1**.

Participants' self-esteem was measured using the Greek version (Galanou et al., 2014) of the Rosenberg Self-Esteem Scale (Rosenberg, 1965). It includes a total of ten questions on a 1–4 consensus scale, and the score can range between 10 and 40. Higher scores are interpreted as higher self-esteem.

Post-VR Questionnaire

After their virtual exposure, participants were asked to complete a questionnaire (post-VR questionnaire) on their experience in the virtual world. All questions, which were evaluated on a 1–7 Likert scale, are presented in **Table 3**. The sense of Presence (Slater, 2009), the subjective sense of being in the virtual world, was recorded using three questions (**Table 3**, a) based on the Slater–Usoh–Steed (Slater et al., 1994) questionnaire. Four additional questions (**Table 3**, b) rated the realism of the agents' behavior. Social presence was measured using four questions (**Table 3**, c) based on a questionnaire by Bailenson et al. (2003). With the use of 4 questions (**Table 3**, d), the participants stated the degree of confidence they felt about the answers they gave to the study, whereas two questions (**Table 3**, e) addressed whether they were influenced by the agents' responses.

Trial Error

The responses given by the participants in each trial were recorded. Using these responses, a Conformity Error scale and a Non-Conformity Error scale were created. The Conformity Error represents the number of incorrect answers given by the participants in the trials that the agents gave the wrong answer. The Non-Conformity Error represents the wrong answers given in the critical trials where the agents gave correct answers. Additionally, a Conformity Index (CI) was constructed. The CI is a scale that describes the conformity magnitude in the agents' responses. This scale was calculated from the difference of the Conformity Error and Non-Conformity Error (CI = Conformity Error – Non-Conformity Error) and describes the level of the participant's conformity.

Response Time and Participants' Gaze Behavior

Participants' response time in each trial was recorded. Response time was the time distance between the moment the participants were called to respond and the moment they gave their answer in each trial.

The total duration that the participants were looking at the agents (we refer to this as Look-At Duration) was recorded. Due to the fact that the participants wore the VR headset that is not equipped with an eye tracker, a separate eye tracker could not be used. Thus, this measurement relied on the direction of the participant's head, using the head tracking feature of the VR headset. Finally, the duration that the participants were looking at the agents when it was their turn to respond was recorded. At that time, in the Eye-Contact condition the agents also looked at the participants, which we refer to as Mutual Gaze Duration.

RESULTS

All results were obtained by analyzing data using the IBM SPSS Software v.24. The dataset generated for this study is provided in the **Supplementary Data Sheet 1**.

Pre-VR Questionnaire

In the NEC condition, the mean value of self-esteem was 33.06 while in the EC condition it was 29.05. The mean and the standard error of self-esteem score for each condition are shown in **Figure 2**. The mean value for both experimental conditions was 30.89, which is considered moderate self-esteem.

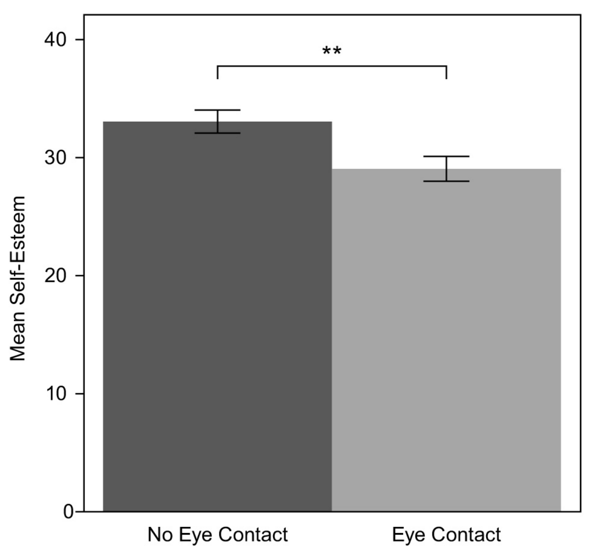


FIGURE 2 | Means and standard errors of self-esteem in both experimental conditions.

An unexpected statistically significant difference was observed between the two conditions. A Mann–Whitney test showed that Self-Esteem was higher among participants in the NEC (M = 33.06, SD = 4.038) than those in the EC group (M = 29.5, SD = 4.696); U = 85.0, p = 0.009. This difference is taken into account in further analysis.

Post-VR Questionnaire

In order to reduce the number of variables from the post-VR questionnaire (**Table 3**), a principal component analysis (PCA) was performed. A single factor emerged from each set of variables and the factor loadings in the scoring variables Presence, Perceived Behavioral Realism, Social Presence, Responses Confidence, and Self-Reported Conformity are shown in **Table 4**. The scoring coefficients are the coefficients of the equations describing the factor scores in terms of the linear combination of the original variables. The factor that emerged from the questions about presence (**Table 3**, a) is interpreted as "the feeling of 'being' in the virtual room." The factor that emerged from the questions on agents' behavioral realism (**Table 3**, b) is interpreted as "the extent to which the agents behaved like real people." The factor that resulted from the social presence questions (**Table 3**, c) is interpreted as "the sense of being together with the agents." The factor that emerged

TABLE 4 | Factor loadings and corresponding scoring coefficients for the factors resulted from principal component analysis.

Variable	Factor loadings	Scoring coefficients
a. Presence		
	F1	Presence
there	0.858	0.467
real	0.830	0.452
lab	0.641	0.349
b. Behavioral Realism		
	F1	Perceived Behavioral Realism
behave	0.895	0.320
move	0.821	0.293
talk	0.773	0.276
feel	0.852	0.305
c. Presence		
	F1	Social Presence
sameRoom	0.830	0.309
otherPerceived	0.866	0.322
otherListen	0.893	0.333
Alone	0.670	0.249
d. Responses Confidence		
	F1	Responses Confidence
correctAnswers	0.724	0.389
difficult	0.666	0.358
doubts	0.725	0.390
confidentAnswers	0.605	0.326
e. Self-Reported Conformity		
	F1	Self-Reported Conformity
myOpinion	0.924	0.541
otherOpinion	0.924	0.541

from the questions regarding participants' responses confidence (**Table 3**, d) is interpreted as "the participants' confidence in their responses." Questions about self-reported conformity (**Table 3**, e) resulted in a single factor interpreted as "the statement that they were influenced by the agent's answers."

There was a statistically significant difference in Social Presence between the two experimental conditions. Specifically, participants in the EC condition reported higher sense of Social Presence (0.337 ± 0.187) than those in the NEC condition (−0.374 ± 0.249). An independent sample t-test showed that the above difference is significant; t(36) = −2.311, $p = 0.027$. An independent sample t-test showed a statistically significant difference in Response Confidence between the two conditions; $t(36) = 2.485$, $p = 0.018$. Participants in the EC condition (−0.358 ± 0.238) reported lower confidence about their responses to those in the NEC condition (0.398 ± 0.181). The means and standard errors of the derived variables are shown in **Figure 3**.

Conformity

Initially, it was examined whether the participants' responses were influenced by the agents' responses in the critical trials. In order to do that, we compared the participants' wrong answers in the trials where the agents replied correctly (Non-Conformity Error) with the participants' wrong answers in the trials where the agents also replied wrongly (Conformity Error). A Wilcoxon signed-rank test showed a statistically significant difference between Conformity Error and Non-Conformity Error in both NEC condition; $Z = -2.113$, $p = 0.035$, and EC condition; $Z = -3.001$, $p = 0.003$. This result suggests that participants in both conditions conformed with the agents' judgments, as they made wrong estimates more often in the trials where the agents gave a wrong answer, than in the trials where agents gave a correct answer. The means and standard errors of Conformity Error and Non-Conformity Error in the two experimental conditions are shown in **Figure 4**.

For the purposes of the analysis, we built the CI scale (described in Section "Trial Error") to describe the level of the participants' conformity. The median of the CI was 1, while the mean was 1.47. In the NEC condition, the mean was 1 while in the EC condition the mean was 1.5. An independent sample t-test indicated that this difference was not statistically significant; $p(36) = -0.225$, $p = 0.823$. This result suggests that the agents' gaze behavior did not affect the conformity level. The mean and the standard error of CI for the two conditions are shown in **Figure 5**.

Response Time and Participants' Gaze Behavior

The response time of the participants' answers in each of the critical trials was recorded, and the Mean Response Time was calculated. The mean response time of participants in the NEC group was 1.6 s, while in the EC Group it was 1.75 s. No significant difference in response time was observed between the two experimental groups; $t(26) = -1.285$, $p = 0.210$.

For the Look-At Duration, in the NEC condition the mean was 205.32 s while in the EC condition the mean was 192.8 s. This difference between the two conditions was not statistically significant ($t(36) = 0.232$, $p = 0.818$). However, this comparison between the two conditions could be influenced by the baseline difference in Self-Esteem (reported in Section "Pre-VR Questionnaire"), as the Look-At Duration was found correlated

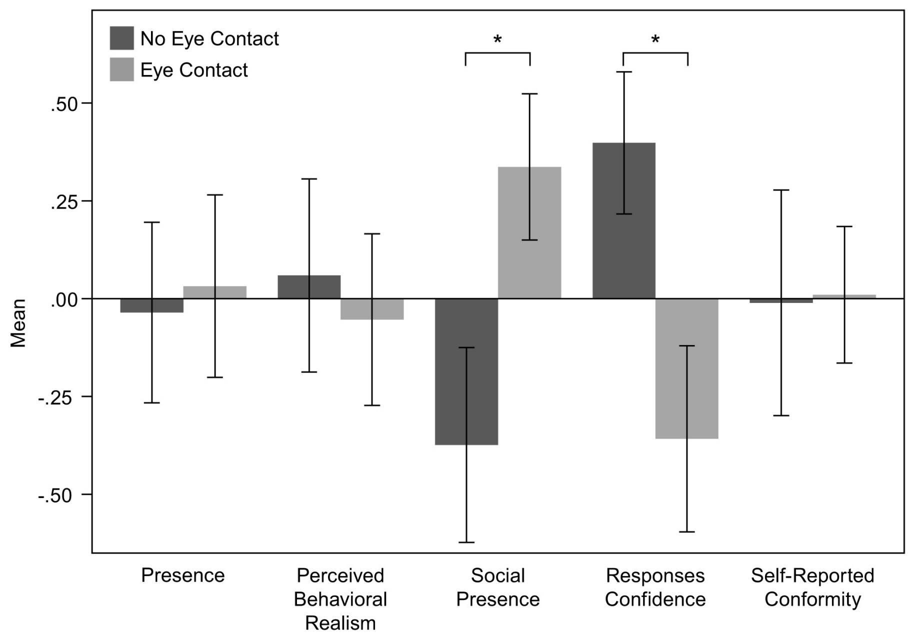


FIGURE 3 | Means and standard errors of the variables resulted from factor analyses.

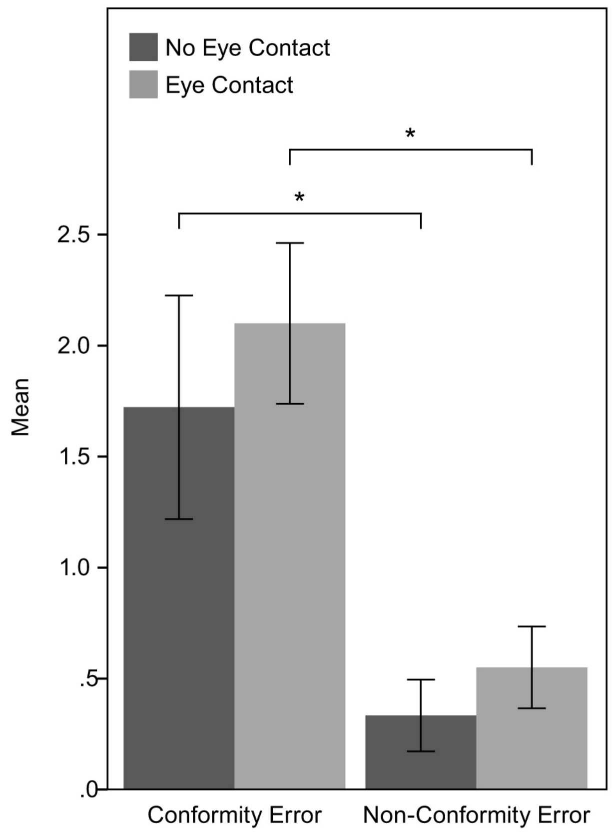


FIGURE 4 | Means and standard errors of Conformity Error and Non-Conformity Error in both experimental conditions.

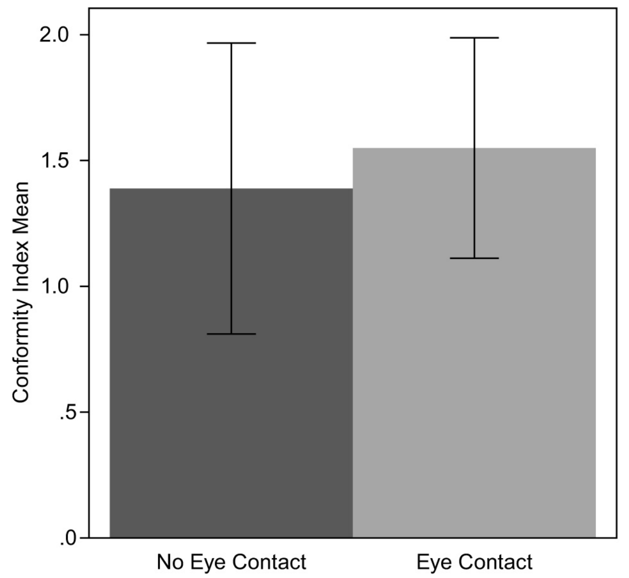


FIGURE 5 | Means and standard errors of CI in both experimental conditions.

(reported in Section "Correlations") with the participants' Self-Esteem in the EC condition. For Mutual Gaze in the EC condition, the mean was 8.34 s. At the corresponding periods of the process, in the NEC condition, the participants looked at the agents altogether for an average of 8.26 s.

Correlations

Participants' Self-Esteem did not seem to be associated with the level of participant conformity. A Pearson product–moment correlation coefficient does not reveal any correlation between Self-Esteem and CI in any of the two experimental conditions. This result is important as it indicates that the difference in the baseline level of Self-Esteem that emerged between the two experimental conditions does not impact the results. Self-Esteem was only correlated with Look-At and Mutual Gaze duration in EC condition.

Finally, we looked at the correlations between the dependent variables in both experimental conditions and some interesting results have emerged. In both experimental groups, a correlation between conformity (CI) and Self-Reported conformity was found (NEC: $r = 0.801$, $n = 18$, $p < 0.001$; EC: $r = 0.575$, $n = 20$, $p = 0.008$), indicating that participants' conformity was conscious. Another interesting result was that, in the EC condition, participants who stated that they were more confident about their responses responded more rapidly to the trials ($r = -0.700$, $n = 15$, $p = 0.004$). This correlation was not presented in the NEC condition. The correlation values and significance levels for the dependent variables in NEC and EC conditions are summarized in **Tables 5**, **6**, respectively.

DISCUSSION

The first goal of this study was to investigate whether social conformity occurs with a group of virtual agents within an IVE. Our prediction was confirmed as participants' judgments were significantly influenced by those of the agents. The participants gave significantly more incorrect responses to the trials where the agents gave a correct response, than the trials where the agents gave the correct response. This result has shown that within IVEs, conformity can be caused by the false judgments of a unanimous majority, even if the majority consists of artificial agents. In addition, the correlation between conformity and self-reported conformity, in both experimental conditions, indicates that the participants were consciously affected by the agents.

This finding is in line with the results of a previous study (Kyrlitsias and Michael-Grigoriou, 2018) where a similar result occurred. However, in the present study, the level of conformity is evidently higher, as only 7.69% of participants in Kyrlitsias and Michael-Grigoriou (2018) conformed with the agents, a percentage fairly small, in contrast to 63.16% of this study. We speculate that the increased level of conformity can be attributed to several differences between the two studies, which include the increased task difficulty, the sense of anonymity, and likely the VR equipment itself. With respect to the task difficulty, the literature has shown that the ambiguity of the task is a critical factor affecting the degree of conformity (Coleman et al., 1958). Specifically, participants tend to yield more easily to social

TABLE 5 | Correlations between depended variables in the NEC condition.

	1	2	3	4	5	6	7	8	9
1. Conformity Index	-								
2. Self-Esteem	−0.132	-							
3. Social Presence	−0.247	−0.228	-						
4. Presence	−0.386	−0.241	0.433	-					
5. Self-Reported Conformity	0.801**	−0.246	−0.322	−0.308	-				
6. Perceived Behavioral Realism	−0.272	0.261	0.560*	−0.026	−0.362	-			
7. Response Confidence	−0.356	0.133	0.346	−0.076	−0.316	0.327	-		
8. Mean Response Time	−0.516	−0.426	0.007	0.403	−0.111	−0.524	−0.055	-	
9. Look-At Duration	0.072	−0.092	0.344	−0.044	−0.138	0.321	0.200	−0.288	-
10. Mutual Gaze Duration	0.020	−0.033	0.410	−0.049	−0.155	0.369	0.219	−0.283	0.918**

***Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).*

TABLE 6 | Correlations between dependent variables in the EC condition.

	1	2	3	4	5	6	7	8	9
1. Conformity Index									
2. Self-Esteem	0.306								
3. Social Presence	−0.120	−0.375							
4. Presence	0.014	−0.118	0.498*						
5. Self-Reported Conformity	0.575**	0.026	−0.160	0.117					
6. Perceived Behavioral Realism	0.156	−0.357	0.660**	0.503*	0.123				
7. Responses Confidence	−0.123	0.050	0.327	0.094	−0.256	0.293			
8. Mean Response Time	0.020	−0.282	−0.068	0.334	0.200	−0.182	−0.700**		
9. Look-At Duration	0.300	0.484*	0.007	0.189	0.232	−0.154	0.203	0.074	
10. Mutual Gaze Duration	0.281	0.467*	0.134	0.184	0.148	−0.075	0.277	−0.048	0.921**

***Correlation is significant at the 0.01 level (2-tailed). *Correlation is significant at the 0.05 level (2-tailed).*

pressure in a more difficult or ambiguous task than in an easier task. The difficulty of the task is also associated to the type of influence. In easy and obvious tasks, social conformity is attributed to normative influence (Deutsch and Gerard, 1955), as individuals change their judgment in order to match the group, but they keep their opinions private. On the other hand, with a difficult or unclear task, conformity can also be attributed to informational (Deutsch and Gerard, 1955) influence, as individuals change their judgment in order to be correct. In this study, we increased the difficulty of the task by projecting the stimulus for a limited duration (5 s). Another factor that may have affected the level of conformity is anonymity. Past studies have shown that the conformity rate is noticeably reduced when the responses are private (Gavish and Gerdes, 1998; Lea et al., 2001; Tsikerdekis, 2013). In VR, users are usually represented by virtual characters different to one's own self-representation, which may give them the perception of some kind of anonymity. In this study, the participants were deliberately asked to verbally introduce themselves by stating their first names, their age, and their occupation, in order to decrease any sense of anonymity. Further studies need to investigate the impact of the user's sense of anonymity on conformity in IVR. Finally, another factor that may have increased conformity is the level of immersion. In this study, we used an enhanced HMD (higher display resolution, better head tracking, etc.) than in our previous study (Kyrlitsias and Michael-Grigoriou, 2018). Past studies (e.g., Bailey et al., 2019) suggest that the level of immersion may affect social responses. Again, regarding conformity, further research is needed to confirm this speculation.

Our second prediction, that the inclusion of eye contact would increase the level of conformity with the agents, was not confirmed. This result replicates the outcome of another non-IVR research (Davey and Taylor, 1968), in which the authors attribute it to the fact that eye contact is only effective when combined with other social cues such as posture changes, gestures, and facial expressions. On the contrary, we confirmed our hypothesis that the eye contact manipulation can affect the sense of Social Presence. Specifically, participants in the EC condition stated significantly higher social presence than the participants in the NEC condition. However, the higher sense of social presence did not translate into a higher conformity level. Literature suggests that a higher sense of social presence leads to higher social influence (Oh et al., 2018), but it did not occur in this study on conformity, contrary to our prediction. An explanation of that is relevant to the type of conformity, which depends on the motives that led the participants to conform. Specifically, as mentioned above, the conformity in this case was informational, as the participants adopt the agents' opinions in order to fulfill their desire to be correct, rather than to fit in, which is the case of normative conformity. An interpretation could be that informational conformity with agents does not depend on the humanization of the computer (which the case of social presence), but on the belief that the agents are reliable regardless to the extent on which they are perceived as social entities. In this case, the type of the task (line length comparison) may contribute, as computers are considered to be reliable in these types of tasks (Weger et al., 2015). This could be studied by testing the impact of social presence on conformity with agents in task that humans are considered as more reliable than computers (e.g., moral judgment). In that case, we believe that the sense social presence could affect the level of conformity.

Interestingly enough, the inclusion of eye contact as a social cue appears to influence the participants' overall subjective experience. Participants expressed more doubts (lower responses confidence) about their responses when social pressure was exerted by agents who made eye contact. This finding suggests that although eye contact had some influence on participants' decision-making process, it was probably not strong enough as there was no impact on their final responses. This finding can also be explained by the stronger sense of social presence that it is associated with social influence (Oh et al., 2018).

Nonetheless, no significant differences regarding the evaluation of the agents' behavioral realism emerged between the two conditions. Participants in the EC condition, even though they stated that the agents felt more socially present, did not rate them as more realistic than the participants in the NEC condition. It is important to note here that the manipulation of the agents' gaze behavior did not affect the perceived realism of the agents. Should the opposite have occurred, we would not be able to properly compare the two experimental conditions.

Some additional findings regarding the participants' response times emerged between the two experimental conditions. Even though participants' response time did not appear to be influenced by the eye contact manipulation, in the EC condition it was found to be associated with participants' confidence. Specifically, participants in the EC condition who stated lower confidence in their responses took significantly longer to respond. This result is in line with the literature that suggests that post-decisional confidence is negatively correlated with choice latency (e.g., Zakay and Tuvia, 1998).

An unexpected setback of the study was the difference that arose in the reported Self-Esteem between the two experimental conditions. Participants in the NEC reported higher Self-Esteem than participants in the EC condition. In order to exclude the possibility that the results were biased due to these baseline differences, a correlation analysis was performed between self-esteem and each dependent variable under investigation. The analysis showed no correlations between Self-Esteem and any other dependent variable (e.g., conformity), except for the case of the two variables related to participants' gaze behavior (Look-At Duration and Mutual Gaze Duration). Given this, the possibility that the results (e.g., for conformity) could be attributed to the difference in Self-Esteem can be rejected and safely attributed to the different condition. Regarding the measures related to the participants' gaze behavior mentioned above (Look-At Duration and Mutual Gaze Duration) that were found correlated with Self-Esteem, it was shown that, in the EC condition, participants with higher self-esteem tended to turn their gaze more frequently toward the agents and performed more mutual gaze with the agents than did participants with lower self-esteem. This association is supported in the literature (Fugita et al., 1977; Vandromme et al., 2011).

The impact of participants' Self-Esteem on their gaze behavior, observed in EC condition, consists of an interesting result that needs further investigation. Participants' gaze behavior was not the focus of this study, and the data collected was not very accurate compared with data provided by an eye tracking HMDs available (e.g.,[1]). Hence, a more in-depth analysis of the participants' gaze behavior was not possible. However, this study shows that the use of IVR and virtual agents can be ideal for this kind of experiments, thanks to its ability to provide a high level of experimental control between multiple experimental sessions.

This study confirms previous findings on the importance of designing artificial agents with realistic behavior toward the users in order to enhance one's experience in IVEs. More interestingly, the findings suggest that the agents' behavior may influence lower levels of conformity, by affecting the user's decision-making. Further, the agents' non-verbal behavior, such as eye contact as employed here, can have an impact on the sense of social presence, which has been shown to affect in turn the overall experience of the user.

This study showed that the creators or the moderators of IVR applications can use agents in order to influence and direct the users' decision-making, through conformity. Social conformity is not limited to simple perceptual tasks, as in this experiment, but extends to other forms of behaviors and attitudes. The use of agents for indirect influence for the user could be used in various ways, such as directing the users of an IVR game in order enhance their game experience, or to influence them in a transaction, for example, in an immersive e-shop.

The findings of this study are very promising and highlight the need for further investigation in order to understand the factors that affect conformity with agents in IVEs. A factor that should be explored in a future study is whether agency (the extent to which the user believes that a virtual human is controlled by a real human rather than the computer) affects conformity. In this study, we showed that agents can elicit conformity; however, we do not know if the conformity will be greater with the use of avatars. Unfortunately, in this study we did not collect such data and we cannot know if the participants perceived the agents being controlled by a computer or by other real humans. Another important factor is the type of the task. In this study, we used a simple objective-perceptual task. The impact of the agents' opinion on more social-objective tasks is also an interesting avenue to be explored in a future study.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

[1] vive.com/eu/product/vive-pro-eye

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CK made substantial contribution in the conception and design of the study, in the data collection, in the analysis and interpretation of the data analysis, and in writing the article. DM-G substantially contributed in the conception and design of the study, in the interpretation of the data analysis, and in revising critically the manuscript, and she supervised and coordinated all the steps of the study. DB made substantial contribution in the design of the study, in data analysis, and in drafting and revising critically the manuscript. MC made substantial contribution in the acquisition of the data and in drafting and revising critically the article. All authors contributed to the article and approved the submitted version.

FUNDING

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 739578 and the Government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination, and Development.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.02254/full#supplementary-material

REFERENCES

Asch, S. (1955). Opinions and social pressure. *Sci. Am.* 193, 31–35. doi: 10.1038/scientificamerican1155-31

Asch, S. E. (1951). "Effects of group pressure upon the modification and distortion of judgments," in *Group Leadership and Men*, ed. H. Guetzkow (Pittsburgh, PA: Carnegie), 177–190.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr. Gen. Appl.* 70, 1–70. doi: 10.1037/h0093718

Bailenson, J. N., Blascovich, J., Beall, A. C., and Loomis, J. M. (2003). Interpersonal distance in immersive virtual environments. *Pers Soc. Psychol. Bull.* 29, 819–833. doi: 10.1177/0146167203029007002

Bailenson, J. N., Patel, K., Nielsen, A., Bajscy, R., Jung, S. H., and Kurillo, G. (2008a). The effect of interactivity on learning physical actions in virtual reality. *Media Psychol.* 11, 354–376. doi: 10.1080/15213260802285214

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., and Jin, M. (2008b). The use of immersive virtual reality in the learning sciences: digital transformations of teachers, students, and social context. *J. Learn. Sci.* 17, 102–141. doi: 10.1080/10508400701793141

Bailey, J. O., Bailenson, J. N., Obradoviæ, J., and Aguiar, N. R. (2019). Virtual reality's effect on children's inhibitory control, social compliance, and sharing. *J. Appl. Dev. Psychol.* 64:101052. doi: 10.1016/j.appdev.2019.101052

Bente, G., Eschenburg, F., and Aelker, L. (2007). "Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication," in *Proceedings of the 10th Annual International Workshop on Presence*, (Barcelona: International Society for Presence Research), 207–214.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., and Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103–124. doi: 10.1207/S15327965PLI1302_01

Bombari, D., Schmid Mast, M., Canadas, E., and Bachmann, M. (2015). Studying social interactions through immersive virtual environment technology: virtues, pitfalls, and future challenges. *Front. Psychol.* 6:869. doi: 10.3389/fpsyg.2015.00869

Coleman, J. F., Blake, R. R., and Mouton, J. S. (1958). Task difficulty and conformity pressures. *J. Abnorm. Soc. Psychol.* 57, 120–122. doi: 10.1037/h0041274

Davey, A. G., and Taylor, L. J. (1968). The role of eye-contact in inducing conformity. *Br. J. Soc. Clin. Psychol.* 7, 307–308. doi: 10.1111/j.2044-8260.1968.tb00575.x

Deutsch, M., and Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *J. Abnorm. Soc. Psychol.* 51:629. doi: 10.1037/h0046408

Fugita, S. S., Agle, T. A., Newman, I., and Walfish, N. (1977). Attractiveness, self-concept, and a methodological note about gaze behavior. *Pers. Soc. Psychol. Bull.* 3, 240–243. doi: 10.1177/014616727700300217

Galanou, C., Galanakis, M., Alexopoulos, E., and Darviri, C. (2014). Rosenberg self-esteem scale Greek validation on student sample. *Psychology* 5, 819–827. doi: 10.4236/psych.2014.58093

Gavish, B., and Gerdes, J. H. Jr. (1998). Anonymous mechanisms in group decision support systems communication. *Decis. Support Syst.* 23, 297–328. doi: 10.1016/S0167-9236(98)00057-8

Gergen, K. J., and Bauer, R. A. (1967). Interactive effects of self-esteem and task difficulty on social conformity. *J. Pers. Soc. Psychol.* 6, 16–22. doi: 10.1037/h0021216

Hertz, N., and Wiese, E. (2016). Influence of agent type and task ambiguity on conformity in social decision making. *Proc. Hum. Fact. Ergon. Soc.* 60, 313–317. doi: 10.1177/1541931213601071

Jenness, A. (1932). The role of discussion in changing opinion regarding a matter of fact. *J. Abnorm. Soc. Psychol.* 27, 279–296. doi: 10.1037/h0074620

Kraemer, P. (2013). Do as we do, not as you think: the effect of group influence on individual choices in a virtual environment. *J. Virt. Worlds Res.* 6, 1–11. doi: 10.4101/jvwr.v6i1.7002

Kyrlitsias, C., and Michael, D. (2016). "Influence by others' opinions: social pressure from agents in immersive virtual environments," in *Proceedings of the 2016 IEEE Virtual Reality (VR)*, (Piscataway, NJ: IEEE), 213–214. doi: 10.1109/VR.2016.7504729

Kyrlitsias, C., and Michael-Grigoriou, D. (2018). Asch conformity experiment using immersive virtual reality. *Comput. Anim. Virt. Worlds* 29:e1804. doi: 10.1002/cav.1804

Lea, M., Spears, R., and de Groot, D. (2001). Knowing me, knowing you: anonymity effects on social identity processes within groups. *Pers. Soc. Psychol. Bull.* 27, 526–537. doi: 10.1177/0146167201275002

Midden, C., Ham, J., and Baten, J. (2015). "Conforming to an artificial majority: persuasive effects of a group of artificial agents," in *Persuasive Technology. Lecture Notes in Computer Science*, Vol. 9072, eds T. MacTavish and S. Basapur (Cham: Springer), 229–240. doi: 10.1007/978-3-319-20306-5_21

Nazligul, M. D., Yilmaz, M., Gulec, U., Gozcu, M. A., O'Connor, R. V., and Clarke, P. M. (2017). "Overcoming public speaking anxiety of software engineers using virtual reality exposure therapy," in *Proceedings of the European Conference on Software Process Improvement*, (Cham: Springer), 191–202. doi: 10.1007/978-3-319-64218-5_15

Oh, C., Bailenson, J., and Welch, G. (2018). A systematic review of social presence: definition, antecedents, and implications. *Front. Robot. AI* 5:114. doi: 10.3389/frobt.2018.00114

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Sherif, M. (1935). A study of some social factors in perception. *Arch. Psychol.* 27, 1–60.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 3549–3557. doi: 10.1098/rstb.2009.0138

Slater, M., and Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Front. Robot. AI* 3:74. doi: 10.3389/frobt.2016.00074

Slater, M., Usoh, M., and Steed, A. (1994). Depth of presence in virtual environments. *Presence Teleoperat. Virt. Environ.* 3, 130–144. doi: 10.1162/pres.1994.3.2.130

Stavroulia, K. E., Baka, E., Lanitis, A., and Magnenat-Thalmann, N. (2018). "Designing a virtual environment for teacher training: enhancing presence and empathy," in *Proceedings of Computer Graphics International 2018*, (New York, NY: ACM), 273–282.

Tsikerdekis, M. (2013). The effects of perceived anonymity and anonymity states on conformity and group think in online communities: a Wikipedia study. *J. Am. Soc. Inform. Sci. Technol.* 64, 1001–1015. doi: 10.1002/asi.22795

Vandromme, H., Hermans, D., and Spruyt, A. (2011). Indirectly measured self-esteem predicts gaze avoidance. *Self Ident.* 10, 32–43. doi: 10.1080/15298860903512149

Von der Pütten, A. M., Krämer, N. C., Gratch, J., and Kang, S. H. (2010). It doesn't matter what you are! Explaining social effects of agents and avatars. *Comput. Hum. Behav.* 26, 1641–1650. doi: 10.1016/j.chb.2010.06.012

Weger, U. W., Loughnan, S., Sharma, D., and Gonidis, L. (2015). Virtually compliant: immersive video gaming increases conformity to false computer judgments. *Psychon. Bull. Rev.* 22, 1111–1116. doi: 10.3758/s13423-014-0778-z

Zakay, D., and Tuvia, R. (1998). Choice latency times as determinants of post-decisional confidence. *Acta Psychol.* 98, 103–115. doi: 10.1016/S0001-6918(97)00037-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Computer Science

ORIGINAL RESEARCH
published: 23 December 2020
doi: 10.3389/fcomp.2020.536880

Focusing on Emotional and Social Intelligence Stimulation of People With Dementia by Playing a Serious Game—Proof of Concept Study

Rakel Berenbaum[1], Chariklia Tziraki[1], Reem Baum[2], Adi Rosen[1], Tuvia Reback[1], Judith Abikhzer[1], Daphna Naparstek[1] and Boaz M. Ben-David[2,3,4]**

[1] MELABEV—Community Clubs for Eldercare, Research and Development Department, Jerusalem, Israel, [2] Baruch Ivcher School of Psychology, Interdisciplinary Center (IDC) Herzliya, Herzliya, Israel, [3] Toronto Rehabilitation Institute, University Health Networks (UHN), Toronto, ON, Canada, [4] Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada

OPEN ACCESS

Edited by:
Carlos Vaz De Carvalho,
Polytechnic Institute of Porto, Portugal

Reviewed by:
Alejandro Catala,
University of Santiago de Compostela, Spain
Maria Chiara Pino,
University of L'Aquila, Italy

***Correspondence:**
Rakel Berenbaum
melabev1@gmail.com
Boaz M. Ben-David
boaz.ben.david@idc.ac.il

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science

Received: 21 February 2020
Accepted: 16 November 2020
Published: 23 December 2020

Citation:
Berenbaum R, Tziraki C, Baum R, Rosen A, Reback T, Abikhzer J, Naparstek D and Ben-David BM (2020) Focusing on Emotional and Social Intelligence Stimulation of People With Dementia by Playing a Serious Game—Proof of Concept Study. Front. Comput. Sci. 2:536880. doi: 10.3389/fcomp.2020.536880

Context: Dementia is one of the top five chronic diseases, which has an overwhelming impact on patients' life quality, family, and caregivers. Currently, research relating to people with dementia (PwD) focuses on the deterioration of cognitive abilities. A more innovative approach, and the one taken by this paper, is focusing on methods to maintain and improve functionality, communication and quality of life of PwD by building on remaining capacities in the yet unexplored domain of emotional and social intelligence (ESI). The use of serious games for PwD (SG4D) aimed at building social and emotional capacity is a budding field of research.

Objectives: Proof of concept that the, low cost, easy to deploy SG4D, called "My Brain Works" (MBW), co-designed with PwD, enhances ESI, based on the Bar-On ESI model.

Methods: 27 PwD, clients at MELABEV dementia day center, participated in a mixed methods 12 weeks pilot, proof of concept study using a tablet SG4D co-designed with PwD. Quantitative performance data was collected automatically by the tablet during game sessions. In this paper we focus on the analysis of the qualitative and quantitative data related to ESI, observed by 10 different researchers, during each game session.

Results: Quantitative data revealed: both the PwD with high and low MoCA scores had similar average ESI scores. Qualitative analysis revealed that the PwD demonstrated 9 sub-components of the Bar-On ESI Model.

Conclusion: While there is no drug to stop cognitive decline associated with dementia, interventions related to ESI, on the other hand, may improve functioning and quality of life. Despite declines in cognitive abilities, our study shows that a tablet based SG4D can stimulate their ESI and evoke responses in self-awareness, empathy, social and communication capacities. Using SG4D to exercise and maintain social skills is an area that may be promising in the future and may help counter the negative effects of social isolation and loneliness. Such games, while not focusing on cognitive improvement, may also impact on cognitive functioning and help bridge the gap between caregiver and PwD. More research is needed with larger sample sizes.

Keywords: serious games for dementia (SG4D), emotional intelligence, social intelligence, caregiving, communication, quality of life, human computer interaction, human human interaction

INTRODUCTION

Dementia is a neurodegenerative illness, which has an overwhelming impact on the patient, family and caregivers (World Health Organization, 2019). Most research relating to people with dementia (PwD) focuses on the deterioration of cognitive abilities. Rarely does it take note of the remaining capacities of emotional and social intelligence (ESI), though many researchers suggest that these intelligences are much more accurate determinants of successful communication and relationships than mental intelligence (Goleman, 2006). This is true especially in dementia, leading the current study to focus on emotional and social intelligence and dementia. Specifically, we aimed to show a proof of concept for a novel serious game for dementia SG4D, entitled "My Brain Works" (MBW) (Tziraki et al., 2017). We focused on whether PwD, even with severe dementia, can engage with a computerized game, enjoy the game and socially interact with it in a meaningful way.

Cognitive functioning in several domains may change in healthy aging (e.g., executive functions, see Hasher and Zacks, 1988), while functioning in others are retained (e.g., speech processing in ideal listening conditions, see Ben-David et al., 2011) or even increase (e.g., vocabulary, see Ben-David et al., 2015). Compared to cognitive performance, emotional processing is taken to be relatively reserved in healthy aging (even if changed; Ben-David et al., 2019) and in unhealthy aging. This leads the current literature to recommend that care of people with dementia should be focused on the stimulation of emotional function (e.g., sympathy and empathy), rather than relying solely on the stimulation of cognitive function (Hirazakura et al., 2008; Fujii et al., 2014a).

Not all emotional processes are preserved in dementia. PwD may have emotion processing deficits in several domains, along with their cognitive deficits, for example reduced perception of emotional facial expressions (Hoffmann et al., 2010; Klein-Koerkamp et al., 2012a,b; Kumfor et al., 2014). This ability plays a significant role in communication and is one of the most important aspects of social cognition (León et al., 2011; Torres et al., 2015; Torres Mendonça De Melo Fádel et al., 2018). Social cognition allows individuals to partake in social situations, enabling them to perceive and recognize the thoughts, emotions, and behaviors of others (Shany-Ur and Rankin, 2011). An intact association between implicit and explicit cognitive functions are required in order to successfully decipher and interact with the social environment around us. As PwD show changes in social cognition (Snowden et al., 2003; Poveda et al., 2017) it can further impair their quality of life. Indeed, PwD may display difficulties in understanding social cues or recognizing emotions (Phillips et al., 2010). For instance, some PwD have trouble adapting to change, are unconcerned with others' feelings and are unable to control emotions. These social cognition deficits are independent of cognitive dysfunction and may increase over time (Cosentino et al., 2014). Impairment in emotion recognition by PwD has a large effect on the caregiver. These impairments can cause failures to modify behavior that is offensive to others (Blair, 2003).

The psychological and behavioral symptoms of dementia are reported to occur in ∼90% of people with dementia. These symptoms can include such things as: anxiety, apathy, emotional lability, aggressiveness, disinhibition, and irritability. This can lead to breakdowns in communication and increased conflict in relationships, and has implications for caregiver burden and depression (Martinez et al., 2018). Apathy and withdrawal in PwD can be as distressing to caregivers as agitation and aggression (Burns and Iliffe, 2009). Taken together, these factors play an important role in caregivers' decisions to institutionalize the PwD and are detrimental to their quality of life (McLellan et al., 2008).

Empathy is essential for social interaction as well, and a crucial trait to understand the intentions and behaviors of others and to react accordingly. Empathy results from the interaction of four components: shared neural representation, self-awareness, mental flexibility, and emotion regulation. All these capacities can also be affected in dementia (Bartochowski et al., 2018; Fischer et al., 2019). This may relate to a deficit in perception of positive emotions. Perceiving positive emotions in others helps people feel connected to one another. Failure to recognize and respond to positive emotions can make people with dementia seem cold and distant, making it harder for others to connect to them.

On top of all the other difficulties, PwD tend to be socially isolated and therefore lack opportunities to exercise various aspects of their remaining emotional and social intelligence (ESI). They are unable to express themselves clearly, so they spend most of their time alone, doing little and not being engaged in meaningful activity. The health and well-being consequences of social isolation and loneliness are increasingly recognized (Cacioppo and Hawkley, 2009; Cacioppo and Cacioppo, 2014; Cole et al., 2015). Engaging a PwD in meaningful activity is therefore important and is a priority when providing care for this population in order to improve well-being through an increase in positive emotions (Jones et al., 2015).

Unfortunately, most efforts related to PwD focus on cognitive changes from dementia while ignoring the fact that there is remaining emotional intelligence. That while there is a growing call in the recent literature to focus on the remaining capacities of emotional intelligence (Hirazakura et al., 2008; Fujii et al., 2014a). Indeed, the most common, and recognized, cognitive evaluations used with PwD are the Montreal Cognitive Assessment Scale (MoCA) (Wallace et al., 2019), and the Mini-Mental State Examination (MMSE) (Creavin et al., 2016); while there is very little evaluation of emotional and social abilities related to PwD. However recently, researchers in Japan (Fujii et al., 2014b) have developed and tested a Mini-Emotional State Examination (MESE) for the examination of emotional functions in people with dementia, to accompany the MMSE. They found that the difference in distribution of MMSE and MESE scores show that cognitive and emotional functions are independently affected in dementia, and therefore assessment of both these functions should be taken into consideration in the care and management of PwD. In fact, they encourage working with the remaining emotional intelligence of PwD (Hirazakura et al., 2008; Fujii et al., 2014a). Following this, the goal of the current study is to test whether PwD can engage with our novel serious game, "MBW," in a meaningful way (social and emotional) accompanied by positive emotions.

The Current Study: Can PwD Engage Emotionally and Socially With MBW SG4D?

There are currently three major conceptual models relating to emotional and social intelligence: (a) the Mayer-Salovey-model (Mayer and Salovey, 1997) which defines this construct as the ability to perceive, understand, manage and use emotions to facilitate thinking, measured by an ability-based measure (Mayer et al., 2002); (b) the Goleman model (1998) which views this construct as a wide array of competencies and skills, measured by multi-rater assessment (Goleman, 2006); and (c) the Bar-On (2000) model which describes a cross-section of interrelated emotional and social competencies, skills and facilitators that impact intelligent behavior, measured by self-report (Bar-On, 1997, 2004) or multi-rater assessment (Bar On and Handley, 2003). The latter model was chosen as the theoretical basis for this study, as it was considered to be the most suited for observational analyses and training of the PwD population.

Table 1 presents three of the core elements in the Bar-On model of Emotional-social intelligence (ESI), (Bar-On, 2006) that were the focus of the current study. ESI, as conceptualized by the Bar-On model, is a multi-factorial array of interrelated emotional and social competencies, skills and facilitators that influence one's ability to understand and express themselves, understand others and interact with them, as well as cope with daily demands and challenges and pressures (Cherniss, 2004; Bar-On, 2006). It contains five main components (divided into 15 sub-components): intrapersonal, interpersonal, adaptability, stress management, and general mood. These core elements are at the basis for effective social interactions. Indeed, to accurately perceive, understand and express our feelings, and control our emotions (so they work for us and not against us) are essential qualities for effective communication, social adeptness, adaptability, self-actualization and happiness (Bar-On and Parker, 2000). Emotionally and socially intelligent behavior can be enhanced in school, the workplace and in the clinical setting in order to improve performance, self-actualization and subjective well-being (Bar-On, 2006). Much more research is needed in this area (Birks and Watt, 2007). As well, little has been done in the area of exercising and enhancing the emotional intelligence capabilities in those with dementia (McLellan et al., 2008). While there is decline in emotional social intelligence capacities associated with dementia, there are still many ESI abilities remaining that might not be apparent to the caregivers and could be enhanced (Fujii et al., 2014a).

TABLE 1 | Three components of Bar-On's model of emotional social intelligence.

Components	Sub-components	
Intrapersonal		Self-awareness and self-expression:
	Self-regard	To accurately perceive, understand and accept oneself.
	Emotional self- awareness	To be aware of and understand one's emotions.
	Assertiveness	To effectively and constructively express one's emotions and oneself.
	Independence	To be self-reliant and free of emotional dependency on others.
	Self-actualization	To strive to achieve personal goals and actualize one's potential.
Interpersonal		Social awareness and interpersonal relationship:
	Empathy	To be aware of and understand how others feel.
	Social responsibility	To identify with one's social group and cooperate with others.
	Interpersonal relationship	To establish mutually satisfying relationships and relate well with others.
Adaptability		Change management:
	Reality testing	To objectively validate one's feelings and thinking with external reality.
	Flexibility	To adapt and adjust one's feelings and thinking to new situations.
	Problem solving	To effectively solve problems of a personal and interpersonal nature.

Serious games, developed especially for PwD (SG4D), in order to engage them through ESI, may help to exercise these remaining abilities, which may even also have some positive effects on cognitive functioning. Serious games may offer the promise of low cost easy to deploy interventions in the care of PwD (Astell, 2010; Robert et al., 2014). In addition, they require minimal professional supervision (i.e., by an occupational therapist) and can be played with the assistance of formal or informal caregivers.

Our novel SG4D, "My Brain Works" (MBW) was designed to answer these specific needs. In our original study, we presented MBW, a low cost, easy to deploy serious game, co-designed with and developed especially for PwD. Our major goal was to bridge the transfer gap between "game designers" practice and knowledge and neuro-psychosocial scientific knowledge of aging and dementia (Tziraki et al., 2017). The overall aim of

FIGURE 1 | Depiction of game screen.

our gaming approach was to facilitate people with moderate and advanced dementia to arrive at an increased sense of self efficacy, which, according to recent research in neuropsychology, directly contributes to psychological, cognitive, and physical health, and thus serves as a key enabler in augmenting and prolonging functionality (Choi and Twamley, 2013). The target was to train a set of functional simple daily tasks, essential and culturally relevant to daily life. For example, choosing current utensils for dinner (see **Figure 1**). Each task was then divided into subtasks, utilizing an occupational therapy methodology, primarily adapted from neuro-rehabilitation. The game screens were designed in a visually engaging way, considering several design issues related to sensory degradation (Ben-David and Schneider, 2009, 2010), topic (Ben-David and Icht, 2017, 2018) and goal changes (for discussion, see Ben-David et al., 2018). Users are asked to sort, find, drag, and move items on a tablet, using the simple and easy to use qualities of a touch screen. The original study (Tziraki et al., 2017) was designed to answer the following questions: (1) Are serious computer games acceptable, accessible and engaging for people with moderate and advanced dementia? (2) Are people with moderate and advanced dementia able to use a tablet? and (3) Can PwD improve the speed of performing a task with practice, indicating their ability to learn? That study demonstrated that "MBW" is (1) accessible, acceptable, and enjoyable for the target population; (2) PwD were able to use the tablet, enjoyed the game; and (3) were able to improve their speed of performance in "MBW."

As a follow-up for the original findings, the current study aimed to investigate whether use of an improved version of the novel game can engage PwD, stimulate and exercise their remaining emotional and social capacities? To meet these objectives, 10 researchers were asked to observe PwD playing several sessions on the "MBW" SG4D to collect data on interactions, engagement both emotional and social. We present qualitative and quantitative analyses of their observations. The data was collected using a novel tool designed specifically for this study, the "MBW Emotional and Social Intelligence Evaluation Form" (see **Appendix A**). This observation form is an adaptation of the ACIS, occupational therapy assessment (Forsyth et al., 1999), to explore the emotive and social capacities of PwD's interactions with the MBW SG4D. This assessment tool includes the key domains of the Bar-On Model of Emotional and Social Intelligence.

The current study had several hypotheses tested in qualitative and quantitative analysis of observations. (1) PwD will interact with the SG4D; (2) PwD will be able to operate the game; (3) PwD will enjoy using the SG4D and will not be frustrated; and most importantly, (4) the use of a serious game may be able to act as an augmentative alternative communication for PwD and their caregivers, allowing the caregiver to enter the world of the PwD and learn about their hidden capacities, especially their emotional and social abilities (Berenbaum et al., 2011). This was tested by interactions with the tablet and by gauging effective communication with the care giver present during the game play.

METHODS

Design of the New SG4D "MBW"

Goals and General Design

For the current proof of concept pilot study, we improved the English version of the SG4D "MBW" and built a user

FIGURE 2 | Depiction of "sensory interaction screen".

management and data collection system around it. We used the development process theory described at length previously (Tziraki et al., 2017), aiming to create a theory-based SG4D, with input from a multi-disciplinary team familiar with aging, dementia, user experience design, gaming theory, and technology, as well as direct input from end users (using the iterative process, see Valdez et al., 2015) and data collected automatically by the system.

In the second stage of development of "MBW" we focus further on user interface and design, aiming to increase social and emotional engagement—the act of being occupied or involved with external stimuli. It is not always clear how to effectively engage PwD in activities and derive the desired benefits, as well as how to measure engagement, especially since there has been a dearth of research examining engagement in PwD (Trahan et al., 2014; Perugia et al., 2018). There is evidence though, to suggest that interventions that support the sense of self, by using tasks or objects with specific meaning to the person, are more likely to engage PwD (Cohen-Mansfield et al., 2009; Zimmerman et al., 2011). Thus, explicit models of emotional identity offer an attractive mechanism for developing more appropriate and effective technologies. This is of specific importance, as PwD are much more focused on emotional and social issues than on abstract problems (Carstensen and Mikels, 2005; Mikels et al., 2005; Blanchard-Fields, 2007). But how exactly does one turn a flat screened tablet into a serious game that can evoke emotions and social interaction in PwD? This is truly challenging.

The methodology that we followed began with the choice of a multidisciplinary team familiar with aging, dementia, user experience design, gaming theory, and technology. Multidisciplinary collaboration is often encouraged in user centered design (UCD) including partnerships between designers and clinicians (Robert et al., 2014; Cornet et al., 2020). The synergies of the team helped to translate appropriate game ideas that came from the aging and dementia experts into emotionally and socially stimulating experiences designed with input of the user experience (UX), gaming and technology experts. We gave special consideration to design aspects that would trigger the PwD emotionally and socially, keeping in mind the importance of emotional alignment (Ienca et al., 2017; Robillard et al., 2018). We used the inclusive participatory design method which ensures benefits to end-users, promotes engagement and ensures usability (Robillard et al., 2018). We also considered more personalized aspects of the game screens (i.e., culturally appropriate, clear, welcoming pictures and design choices to enhance PwDs engagement with the SG4D).

Great care was devoted to the choice of each picture used, in order to increase emotional and social interaction with the SG4D. First, we chose to use real life photos as opposed to animations to make them more realistic and understandable. We chose pictures that would be familiar to a broad range of cultures, not childish, but that includes pictures of children, and all the generations. These intergenerational and family images with clear unambiguous emotional expressions were chosen to make it easier for PwD to recognize the emotions of the characters and empathize, sympathize and interact with them (McAllister et al., 2020), and give a feeling of warmth and belonging (see **Figure 2**). Nature scenes were included as well, to elicit emotions of calmness and enjoyment (Gamble et al., 2014; Thake et al., 2017).

The interdisciplinary team worked together to ensure instructions are clear, friendly and inviting for a person with dementia (Di Rosa et al., 2014; Mondini et al., 2014; Lam et al., 2015). Another aspect included throughout was appropriate and respectful humor (Berenbaum, 2003). Being able to understand humor is one of the capacities remaining for many in the middle stages of dementia. Since most people are not able to laugh spontaneously, the SG4D acts as a stimulus to trigger laughter. Indeed, laughing has positive psychological, physical and social benefits. It releases tensions, dilutes stress, and induces relaxation (Eng, 2000). Throughout, game tasks were designed to match the abilities of PwD in order to increase self-efficacy (Tziraki et al., 2017).

Each game screen was person-centered (Kitwood, 1997), and culturally appropriate to engage them in an interactive way. **Figure 1** is a depiction of a sample game screen where the PwD was instructed to follow written and oral instructions to sort, find, drag, and move items on the tablet touch screen. Game screens targeted the following skills: eye hand coordination, language skills (reading, comprehension), understanding and following instructions, praxis, memory, sustained attention, and object recognition. They also targeted other physical, cognitive and emotional social skills such as gnosis, association, mental rigidity, recognition, abstraction, association, inhibition, basic math skills, logic, language and communication responses. All tasks included in game screens were in order to maintain appropriate challenges and spark interest throughout gameplay.

The system was designed in a modular way so that the game screens sequence can be interchanged, and various timings in the game are configurable. This allows for testing a variety of game experience scenarios.

The game's length, interaction timers and difficulty levels are all configurable in the system and may entail further experimentation. Indeed, the design, playfulness, feel, and looks of the system and game, were all inspired by the field of dementia care, as well as input of the MELABEV staff. We had a co-design approach to the game development based on the clients' feedback in terms of usability, and clarity for each screen of the game. We also included a training session for the PwD, before commencing the pilot, so they could learn how to use the technology. At the end of the iterative development stage, we had a prototype of a tablet-based game for PwD with 23 game screens including 15 "physical interaction screens" —where the PwD must sort, find, drag, and move items on the tablet touch screen, and eight "sensory interaction screens" —not requiring physical interaction with the tablet but where the PwD looks at pictures on the screen and interacts sensorially, emotionally and socially. This prototype was used for the pilot proof of concept study described here.

Technology Readiness Levels (TRL) Scale

The design of the novel "MBW" SG4D can be examined using an adaptation of NASA's Technology readiness levels (TRL) scale, similar to the EU's adaptation of the scale methodology (Earto, 2014). We also look at the methodology described by a publication by the LEITAT organization (Jamier et al., 2018). MBW's readiness level can be described by the matrix (see **Table 2**) using a similar adaptation to the TRL scale, as described by the LEITAT organization. We denote five components of the system, aiming to ultimately provide a competitive marketable solution, which encompasses the many aspects of therapy and user interaction described by this article, is robust and scalable, and is also coupled by credible research data and analysis.

Our (1) basic unit is the game itself, an App which can run on a tablet and shows the various game screens in succession. Next, we denote the (2) Integrated System, coupling the game App with a user login base and a server back-end to collect data. All development and coding having to do with the accompanying research are denoted (3) Sub System. Lastly, the penultimate and ultimate components are the (4) Operational System, a market deployment ready prototype, and (5) Application, the final software deployment solution, a robust and scalable technology product, encompassing all stages of physical site deployment.

The different components of the system are at various readiness levels ranging from "technology concept formulated" (TRL 2) to "technology demonstrated in a relevant environment" (TRL6; i.e., an active dementia day care center) as seen in **Table 2**. The basic unit and integrated system of "MBW" (column 1, 2 of **Table 2**) have been demonstrated to be at TRL level 6. We have successfully completed an initial system pilot, testing the various aspects of the system in one of MELABEV's 4 day-care centers, which can be regarded as a relevant environment for this game.

Proof of Concept Pilot Study

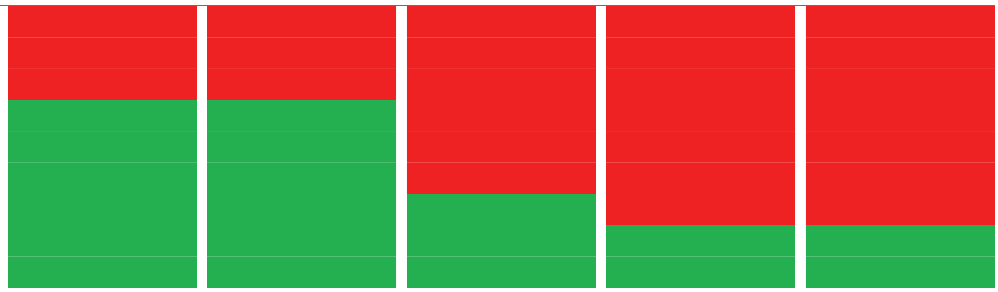
Setting

The present proof of concept pilot study took place at MELABEV, an English speaking, dementia day center, Jerusalem, Israel (Berenbaum, 2010). MELABEV has 4 day-care centers attended by ∼500 PwDs, ranging from people with moderate cognitive impairment (MCI) to advanced dementia. While the literature reports on it being difficult to engage PwD in activities, the activity staff at MELABEV do their utmost to find the appropriate activities for each client (Berenbaum et al., 2017a,b). MELABEV's professional staff routinely uses computer games on a one-to-one basis for cognitive stimulation gaming (Berenbaum et al., 2011), as well as reminiscence therapy at the computer (Douglas et al., 2004). Primary family caregivers who enroll the PwD in the day care program consent to the participation of their family member with these kinds of technology, as well as all other activities in the day care center.

Meaningful informed consent for people with dementia is challenging. Thus, for our pilot study, we utilized the participatory consent process (Dewing, 2007, 2008). Before each gaming session, the research assistant asked the PwD if he/she agreed to participate in the gaming session. Upon agreement, the PwD voluntarily got up and was guided by the research assistant to a designated space (one of four quiet rooms) to play the SG4D. If the PwD did not agree to participate, or if he/she said or acted as if he/she didn't want to continue, the game session was promptly terminated, and the PwD was led to the regular activity room at the center, with no consequences whatsoever to the services they received in the center.

TABLE 2 | "MBW" technology readiness matrix.

Panel A		Timeline →				
		Basic unit	**Integrated system**	**Sub system**	**Operational system**	**Application**
System component		Game UI	Pilot Deployment	Research Planning and Goals	UX	Market ready deployment
Description		Playable game App on a tablet for PwD	Integration of the game with a user base and login	Define and implement how to measure (data collection) success and progress of PwD by means of technology	A "user first" system, ready for deployment, including all relevant UI's for all types of users (i.e., multiple daycare centers, or home use)	System ready for scalable robust site deployment and/or App public on vendor stores
TRL phase	**Sum TRL**	6	6	3	2	2
Competitive manufacturing	TRL 9					
	TRL 8					
Product demonstration	TRL 7					
	TRL 6					
Technology research	TRL 5					
	TRL 4					
Preliminary new technology research	TRL 3					
	TRL 2					
	TRL 1					



Conceptually, we split development to specific research and development components, ranging from abstract ideas, to operational physical systems, and place them on a timeline, from left to right. The components and timeline are described by the columns of the Matrix. The various TRL's, according to the chosen scale, are placed on the Matrix's row. A table cell then links any component with any TRL. In the Matrix a green/red cell describes a completed/ uncompleted TRL, for the corresponding system component. The TRL phase is highlighted in blue.

Participants

Twenty-seven PwD, clients at the MELABEV dementia day center, were recruited to the proof of concept study. During recruitment, we excluded clients with aggression, delusional behavior, a history of alcohol or substance abuse, depression, severe auditory, and/or visual or motor deficits, as assessed by the professional staff at MELABEV. From the 27 who started the pilot, eight were removed from analysis for the following reasons: (a) reached ceiling in performance (the task was apparently not challenging as evident in their higher scores on cognitive assessment, $N = 2$); (b) did not complete more than one session ($N = 4$); (c) extremely low affect ($N = 2$). Performance of nine-teen participants was analyzed (14 women and five men, ages ranging from 65 to 97 years; M age = 87 years, SD = 8 years). Fifty-six percent were from the US, 26% from Eastern European countries, 9% from South Africa, and 9% from England. The participants had worked in various professions in the past including: programing, beautician, architecture, accounting, social work, teaching, and business.

The 19 participants included in the analysis came from two different groups at the Melabev center: Higher Functioning (13 participants) and Lower Functioning (six participants). PwD are placed in these groups based on various criteria set by the Melabev staff, which include group dynamics as well as cognitive abilities. To assess participants' preserved cognitive abilities, Melabev center uses the Montreal Cognitive Assessment Scale (MoCA; Wallace et al., 2019). Participants in the Higher Functioning group scored between nine and 28 on the MoCA scale, and participants in the Lower Functioning groups scored below six.

Gaming Sessions

Ten researchers (all graduate or undergraduate interns from the fields of psychology, nursing, social work, pharmacology, pre-med, and speech therapy), who were not involved in the development process of the game, participated in the proof of concept study. They accompanied the PwD during game play. Their main task was to observe the sessions and document the PwDs engagement with the game, focusing on emotional, and social abilities. During each 20–30 min gaming session, the research assistant, working one-on-one with each client, filled in two different assessment forms (as described in the Tools section). Researchers were also instructed to help with any technological issues that might arise during game play (i.e., no internet connection; Manera et al., 2017).

Tablet Setup

Attention was paid to tablet set up before each session. The sound level was set to the client's most comfortable level, noting the impact of auditory sensory degradation in aging (for example, see Ben-David et al., 2016). Before starting, each PwD was reminded by the researchers to use only one hand on the tablet and to keep the other on the table or on their lap. Research assistants were encouraged to let the client interact with the game independently and not complete the tasks for them.

Training. Before commencing with the pilot, each PwD was presented with a tutorial on how to use the tablet properly (Cunnah et al., 2019). The research assistant informed the PwD:

> *"In case you have never used a tablet before—we will now do something to learn how to use this one, since each tablet is a bit different. You will use your finger to select items by tapping them with a finger and dragging them (left or right or up or down) on the screen. We will learn how hard or soft to touch the screen, and how to drag things from place to place on the screen."*

Gaming Sessions

Following instructions, participants commenced with the game session. In each gaming session, the PwD could play with all 23 game screens. After the completion of each game session, the PwD was asked for their feedback which they gave both verbally to the researchers and by entering it into the tablet. When mistakes (or inactivity) were recorded by the system, it provided cues/prompts to guide the participants toward expected activity (usually, by repeating the instructions). After two cues, participants were shown the correct answer, and the game advanced to the next game screen. The duration of each game session was between 20 and 30 min. All sessions took place at approximately the same time of day in a quiet room. In every game session, each PwD had the opportunity to play the complete game of 23 game screens.

Data Collection and Tools

This was a mixed methods study with both quantitative and qualitative data collected (Creswell and Creswell, 2017).

MBW Emotional and Social Intelligence (ESI) Evaluation Form

All researchers described their observations, during the game play, on an ESI evaluation form, designed especially for this study. This ESI form, presented in **Appendix A**, was modified from the ACIS (Assessment of Communication and Interaction Skills, used by occupational therapists) (Forsyth et al., 1999; D'Amico, 2017) by a multidisciplinary team (gerontologist, MD, OT, speech therapist, psychologist). The ACIS is a structured observation tool with 20 discrete skills concerning physicality, information exchange and relations. The ESI tool is designed to capture, in detail, a person's social interactional ability while he or she is participating in a meaningful social context. It identifies strengths as well as problem areas. We included social interactional abilities from the ACIS that were feasible for the research assistants to observe during game play and that corresponded to three components of the Bar-On emotional social intelligence model: interpersonal, intrapersonal, and adaptability (Birks and Watt, 2007).

The "MBW ESI Evaluation Form" includes 11 parameters related to communication and interpersonal social skills rated on a 5-point Likert scale. If the research assistants witnessed any of the following 11 capacities in the reactions of the PwD to the tablet during any of the game screens they would rate it from 1 to 5 (one being the lowest frequency, five being the highest). The form also included one other parameter. This was filled in if at any time during the gaming session, the PwD initiated communication with the research assistant who sat next to them throughout the gaming session. This was rated on a 5-point Likert scale. This form was filled out by the researchers as the PwD was using the game. A total ESI score was derived, in order to assess reliability of the ESI Form, we conducted a Cronbach's Alpha test which indicated that the evaluation has an acceptable reliability ($\alpha = 0.92$).

Subjective Observations

In addition to the ESI form, the "MBW Game Screen Subjective Evaluation Form" (see **Appendix B**) was used which prompted researchers to report on their subjective feelings related to the PwDs' performance on each of the 23 game screens during each game play session. The form included both close and open-ended questions, such as: (1) "Did the client enjoy the game?" (2) "Document specific quotes that the PwD shared during the game session", respectively. The open-ended part of the form had no specific guiding questions.

Correct Engagement Score

A login-based system was used to monitor game plays by the PwD and saved on a centralized anonymized database for analyses. The analysis in the current study focused on "physical interaction screens." These are 15 of the MBW game screens, where the PwD was asked to actively engage with the game—e.g., choose a specific object on the screen and drag it to a designated place on the screen.

Accuracy of performance was found to be very low in the first three sessions, as expected. Likewise, the time users spent with each game screen varied significantly between participants, which can be explained by the inability of the PwD to focus on a specific task. As a result, an aggregate "correct engagement" score was calculated. Specifically, in each game screen, choosing and/or dragging the correct object was considered correct engagement with the tablet, irrespective of success in the game screen. The maximum score for each session was 15 (as 15 physical interaction screens were analyzed) and minimal score was 0.

Qualitative Data Analysis

Subjective observations and quotation data from MBW game screen subjective evaluation form, was collected on an Excel document and analyzed using grounded theory (Chun-Tie et al., 2019). Five different researchers analyzed the data for themes, three are co-authors on this paper, and the other two had not been involved in the game development or pilot study. The group spanned different fields of expertise (medicine, occupational therapy, gerontology, social work, and neurobiology) and levels of expertise (from interns to emerita). Each researcher read the Excel data a number of times on their own searching for themes. Then they discussed all the themes found by all the researchers with about 92% of original agreement between them and came to a consensus on the themes. Then two of the researchers returned to the data to collect the quotes they felt best portrayed the themes that had been chosen that related to emotional and social intelligence. They came up with a long list of quotes which together they narrowed down to about 20 quotes for each sub-category of the Bar-On Model of Emotional Social Intelligence (total of 180 quotes). This list of quotes was given to four different professionals (three social workers and a speech-language therapist) who were asked to, each on their own, choose the three most representative quotes from the list of 20 demonstrating each sub-category of the Bar-On ESI. The quotes that best represent the different themes are presented in **Table 3**.

RESULTS

Quantitative Data

In the following analyses, we tested for significant differences between the two subgroups of PwD participants, higher functioning ($n = 13$) and lower functioning ($n = 6$).

As expected, the average number of "correct engagements" was significantly higher for the higher functioning group compared to the lower functioning group, $t_{(17)} = 3.5$, $p = 0.003$ (see **Figure 3A**; $M = 9.7/15$, $SE = 2.3$, and $M = 5.4/15$, $SE = 2.5$; for the higher and lower functioning subgroups, respectively). However, no significant difference was found between the two subgroups on the ESI scores, $t_{(17)} = 1.6$, $p = 0.131$, (see **Figure 3B**; $M = 3.7/5$, $SE = 0.9$, and $M = 2.9/5$, $SE = 0.8$; for the higher and lower functioning subgroups, respectively).

In the next step, we tested the link between correct engagement performance and ESI scores. The analysis, conducted across both PwD groups, suggested that engagement performance was significantly correlated with ESI, $F_{(1, 17)} = 5.1$, $p = 0.038$, with performance in the SG4D explaining 23% of the ESI variance, $r^2 = 0.23$ (See **Figure 4**). When correlation analysis was conducted separately for each of the PwD subgroups, the link was found to be significant only for the higher functioning subgroup, $F_{(1, 11)} = 5.1$, $p = 0.046$, $r^2 = 0.31$, but not for the lower functioning subgroup, $F_{(1, 4)} = 0.47$, $p = 0.53$. This could hint on the higher relation between performance and ESI in higher functioning group, or simply be an artifact of the smaller number of participants in the lower functioning group (and hence, smaller variance in performance).

In sum, higher functioning PwD participants (as assessed by the MoCA) performed better on the game (as assessed by correct engagement) than their lower-functioning peers. However, both groups were not differently assessed on ESI. Thus, PwD's cognitive functioning was predictive of performance on the game, but not of ESI. Interestingly, ESI and performance on the game were significantly correlated for the higher functioning PwD group, suggesting that with residual cognitive capacity, ESI could be possibly used to improve performance on the game.

Qualitative Data

Three general overarching themes emerged from the grounded theory analysis: (1) Manifestations: manifestations of aging and dementia symptoms, and the types of associated mistakes; (2) Learning: learning techniques and signs of learning; and (3) Emotional Social Intelligence. These three themes are interweaved with three categories of interactions: interactions with the technology, interactions with the research assistants, and interactions with the objects on the screen.

TABLE 3 | Quotes demonstrating selected sub-components of Bar-On's model of emotional social intelligence.

Components Sub-components	Quotes of PwD while using MBW
Intrapersonal	
Self-regard	"This is too easy for me." "I know what to do but can't get it to move." "I should be wearing my glasses" "I really wanted to put the fork here but knew it wouldn't go" "I'm a slow poke", "I've had enough", *she took a break and then continued till the end.*
Emotional self- awareness	"I'd really enjoy that I'm sure" "this is not my cup of tea I must tell you" nodding, "I'd be happy here" "I'd like to have some too"
Assertiveness	"This picture is disgusting" "The flowers are pretty, **not** the lady" "I think I'm getting a little fed up with this you know..."
Self-actualization	*After realizing that she didn't get the answer correctly* "wait, can we start over?" "something more?" *wanting to know what else she could do*
Interpersonal	
Empathy	"they're enjoying themselves" "she's having a good time" "She looks a bit worried" "Let's leave her alone" "Now you feel better"
Social responsibility	"Thank you" *Wishes the characters* "Bon appetite" "Good morning how are you today?" "When should we come again?" *While raising cup towards research assistants says* "Cheers!" "Congratulations" "he finished his soup give him something else"
Interpersonal Relationship	"Hello" "Oh, my, what family is this?" "A lovely family" "Beautiful" "You're so sweet and beautiful" *started laughing* "you can eat the soup because I love you."
Adaptability	
Reality testing	"I've seen this before" "This is funny, those are the same faces from yesterday." "This must be you and your Dad?" "Where is that?" — nature screen "Where is she talking from?"
Problem solving	Asks clarifying questions "is this what we are supposed to do?" Reread instructions "Who goes first?" -trying to clarify instructions "I can't see it", "this goes here I guess"

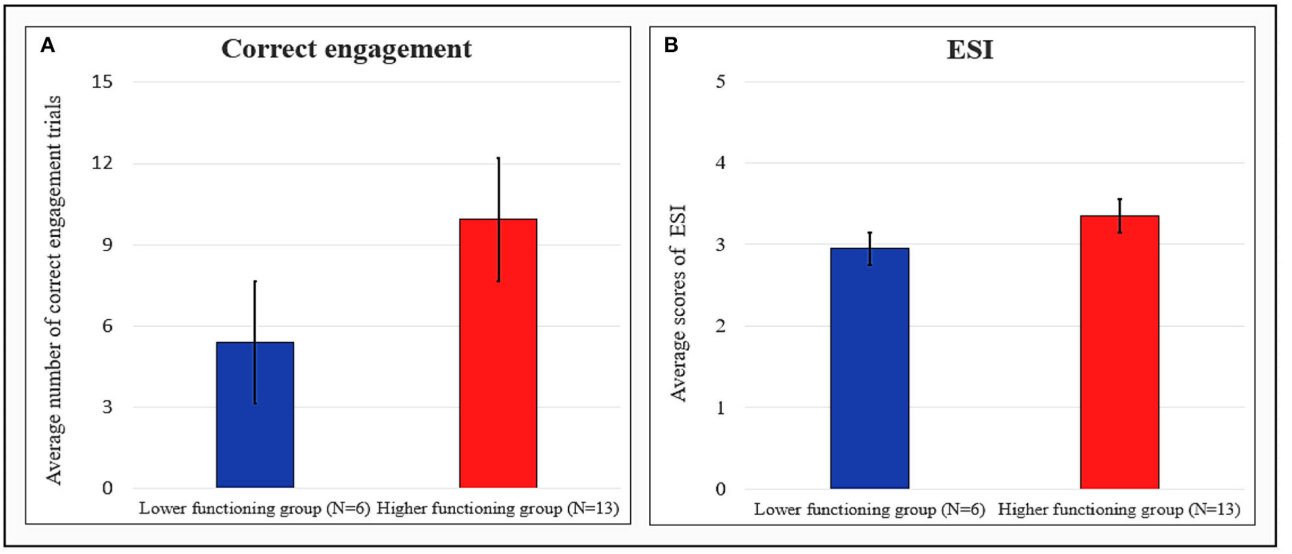


FIGURE 3 | (A) presents aggregate correct engagement scores, and **(B)** presents average ESI assessments for the lower (blue) and higher (red) functioning PwD subgroups.

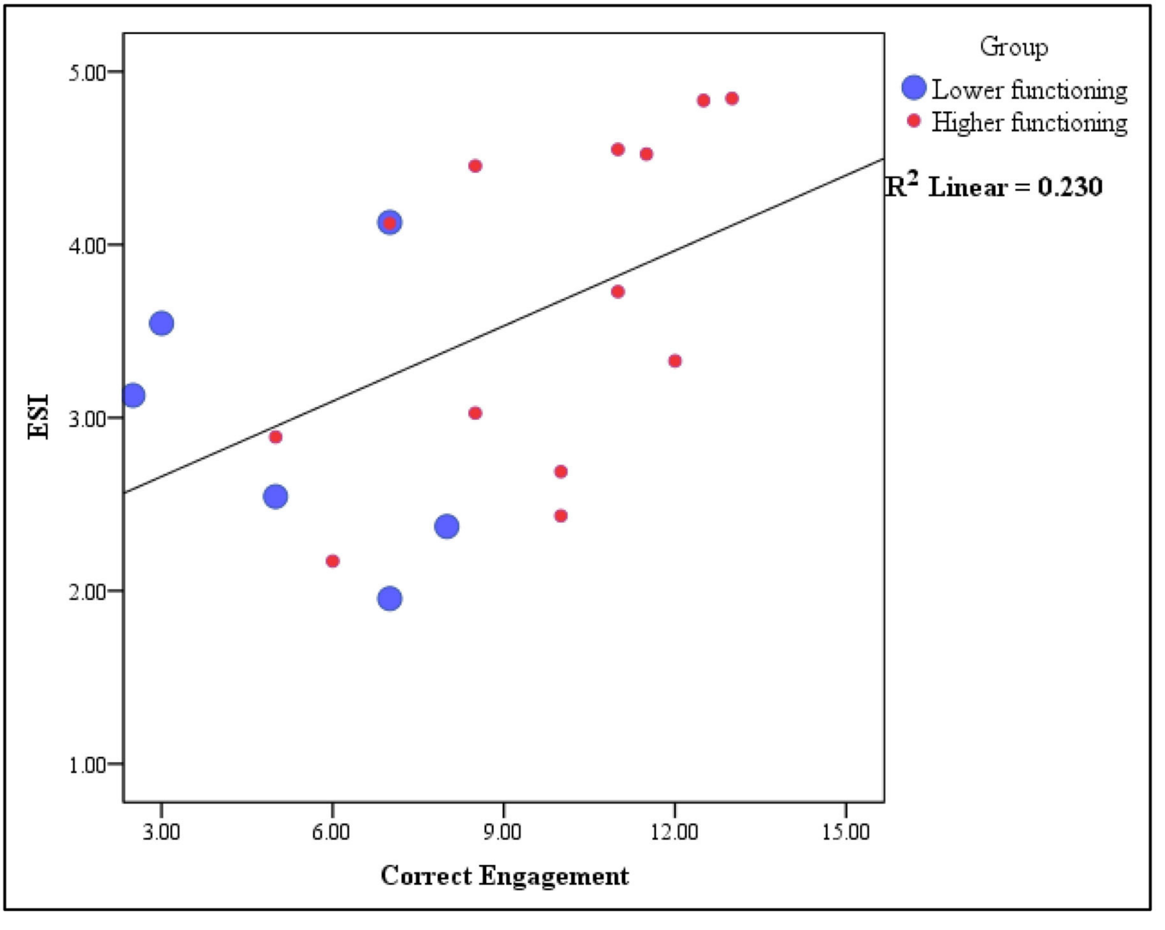


FIGURE 4 | Simple regression model analysis for lower (blue) and higher (red) functioning PwD. Correlation between the aggregate correct engagement scores and the average ESI assessments.

Aging and Dementia Symptoms and Types of Mistakes

Symptoms of dementia appear to cause difficulties for the PwD when playing the SG4D. Some issues that arose were related to disorientation, distractions in their surroundings, stress, and slow speed of processing related to figuring out answers and implementing them by dragging on the screen. Motor impairments, such as physical rigidity, poor fine motor skills and sleepiness, were also mentioned as a factor. The types of mistakes mentioned by research assistants were related to mental rigidity (i.e., mentally stuck on the previous screen), apraxia (inability to initiate drag movement), poor spatial orientation (i.e., dragging in wrong direction), disinhibition (i.e., circling aimlessly on the screen), and impaired abstraction (i.e., understanding instructions literally).

Learning Techniques and Signs of Learning

The research assistants were also clear in their reports on demonstrated learning by PwD. They also reported that in 89% of the game plays participants appeared to enjoy playing the game. Some were even able to anticipate the instructions to the game, after practice. Observations indicated the use of various learning techniques in order to complete the game. Some of the PwD used the trial and error method. It was sometimes evident that they used a self-correcting mechanism when they felt that their first attempt was incorrect. Others asked the research assistants for help with specific tasks, or with their general confusion. Some reread the instructions out loud or to themselves and consciously contemplated the instructions. For example, sometimes they counted out loud to keep track of what they were doing.

Emotional and Social Intelligence

In our grounded theory analysis of the research assistants' written observations of the PwD playing the SG4D, and their quotes, we found the following 9 sub-components of the Bar-On Model of ESI, as divided to three categories: (1) Intrapersonal: self-regard, emotional self-awareness, assertiveness, self-actualization; (2) Interpersonal: Empathy, social responsibility, interpersonal relationship; and (3) Adaptability: Reality testing, Problem solving. **Table 3** shows selected quotes as examples demonstrating each sub-component: (words in italics are to explain the context to the reader).

In sum, the qualitative analysis suggests that PwD encountered difficulties related to dementia in playing the game. However, they demonstrated the ability to overcome these difficulties, to learn new techniques and to interact in a meaningful emotional and social way with the research assistants.

DISCUSSION

The world of people with dementia (PwD), as well as their family and caregivers, is filled with isolation. This is due both to the neurodegenerative aspects of the disease, as well as the challenges caregivers face on how best to communicate with a PwD. The disease makes it hard for the PwD to interact with their surroundings and social settings. Social exchange and communication become difficult. This, in turn, creates a social barrier and a social distance between those around the PwD and the PwD. Technological tools such as serious games could facilitate elimination of barriers, allowing the PwD and their caregiver to communicate in a more meaningful and positive way. Technologies for evaluating and exercising emotional and social capacities are not well-studied in the dementia population (Goodkind et al., 2015; Dourado et al., 2019; Nazareth, 2019).

The current proof of concept pilot study suggests that PwD with moderate to severe dementia can engage with our novel tablet-based SG4D "MBW" and found it accessible and acceptable. This was demonstrated by the observed enthusiasm and participation of PwD with the game system, and by specific data collected by the researchers and tablet that document Bar-On's markers of emotional and social intelligence (Bar-On and Parker, 2000).

Data collected in our pilot study hints that PwD may be able to recognize emotions in the characters on the screens of the SG4D, and react and express their own emotions. They even expressed empathy toward characters in the game. They were able to respond appropriately to social interactions and initiate interactions with both the objects on the screen and the research assistants. These are emotive and social capacities that have not been studied in PwD, especially with moderate and advanced dementia (MoCA scores as low as 4). In fact, there is still a paucity of research on emotional and social intelligence in PwD, and how it can be used to improve interactions with caregivers (Calabria et al., 2009). The current study suggests that the use of a serious game designed with the unique needs of a PwD in mind can facilitate expression of emotions and thus open doors for communication for PwD and their caregivers.

Our preliminary findings also suggest that emotional and cognitive abilities may not be similarly affected by dementia. Cognitive abilities, as assessed by the (routinely used) MoCA test were not found to have an effect on PwD's emotional and social intelligence assessments by research assistants, even though cognitive scores significantly affected accuracy of performance in the game. These results relate to Fujii et al.' (2014a) assertion that cognitive and emotional functions are independently affected in dementia. Our results further highlight their suggestion that assessment of both cognitive and ESI should be taken into consideration when testing PwD. Taken together, findings indicate that remaining ESI skills in dementia may be used as scaffoldings to improve functioning. Technology may play a role in improving these remaining capacities.

Finally, it seems that the use of a touch-screen, tablet-mediated, serious game can exercise ESI capacities that may be relatively preserved but typically remain hidden in dementia. As well, such a SG4D, may be used as a more accurate tool to assess reserved cognitive functioning since it is done when the PwD is relaxed, playful and even laughing, rather than stressed and threatened by a formal cognitive battery. As Plato said "You can discover more about a person in an hour of play than in a year of observation."

More research should be done to better understand how to maintain, exercise and utilize remaining ESI capacities in dementia and to explore further technologies in various settings. We hope to be able to do this in the process of moving to TRL6-7 (prototype demonstration) and indeed in turn eight and nine (operational system) which will define in detail our planned public use cases. Our aim is to expand the use to all of MELABEV's daycare centers, other daycare centers in Israel and other countries, as well as home use for the client with his/her personal aid. Use cases of the system can vary greatly between daycare center and home use deployment. While deploying the system in a daycare center aids the accompanying research, we believe that many other aspects of use and user interaction with the game system can be contrived and observed from analysis of home use, see suggestions for adaptation of paradigms to remote acess, given COVID-19 social restrictions, in Ben-David et al. (2020). Our aim is to further investigate use cases in the future, using even more advanced analytics. Indeed, a publicly available App will serve a greater therapeutic purpose for a larger population, than focusing on daycare center deployment.

We believe that remaining ESI capacities might even be used to improve, or at least slow the deterioration of, cognitive abilities, functionality and communication between the PwD and their caregivers and families. Further research in this area may impact on the quality of life of PwD and their caregivers.

LIMITATIONS

While this study has benefits due to it being a mixed methods study, including rich qualitative data, as well as many research assistants involved in doing the testing, it has limitations as well. As in many studies related to care of PwD, the sample size is small, yet no different than common in published research in the field (see **Appendix C**). Another limitation is the fact that the study was carried out in a dementia day care center with a full schedule, which at times interested the people with dementia more than being involved in the pilot study. We hope in future studies to test the effect of playing "MBW" in a home setting as well.

CONCLUSION

An earlier research from our team (Tziraki et al., 2017) documented that the use of a serious game, designed with the input from unique needs of a PwD in mind, is acceptable accessible and engaging for people with moderate and advanced dementia. We also found the serious game may be helpful in improving cognitive function, such as speed of processing. The current study focused on assessing how and if PwD can interact in a meaningful emotional and social manner with a technology-based serious game, and whether playing the game will be accompanied by meaningful interactions with caregivers present. Our preliminary results suggest that the use of a serious game can facilitate the expression of ESI capacities that typically remain hidden in dementia. We further suggest that the serious game developed for this specific aim, SG4D "MBW," may be able to facilitate the stimulation of these capacities in ways that caregivers can also recognize and thus contribute to improved social connectivity.

There is yet no drug to stop the cognitive decline associated with dementia, thus we suggest that interventions related to PwD and ESI may improve functioning and quality of life. Despite declines in cognitive abilities, our two studies may indicate that PwD can interact with a tablet SG4D using social and communication capacities. Using SG4D with PwD in order to help them exercise and maintain social skills is an area that may be promising in the future and may help counter the negative effects of social isolation and loneliness. Such games may also have an impact on cognitive functioning and even on improving the quality of life for both PwD and their caregivers by facilitating communication. Further research should explore whether SG4D may be used as evaluation for reserved cognitive, and ESI abilities on one hand, and as a gauge for functional capacities that could possibly be improved with training, on the other. More research is needed in this area with larger sample sizes.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. All identifiable images used in this manuscript have been obtained from a stock image repository.

AUTHOR CONTRIBUTIONS

RBe and CT planned the research, including the pilot, the qualitative analysis, and its interpretation. RBa and BB-D conducted the quantitative analysis and its interpretation. TR participated in the qualitative analysis. JA helped plan the pilot and gave input to writing the paper. DN and AR wrote the section on design and technology considerations. The manuscript was written in collaboration of BB-D, RBe, CT, and RBa. All authors contributed to the article and approved the submitted version.

FUNDING

Yad Hanadiv and Arison foundations partially funded this research and development. The corresponding author was partially supported by a grant from the Israeli Science Foundation (ISF grant number 861/18).

ACKNOWLEDGMENTS

We would like to thank MELABEV Talpiot staff, for making it possible for the study to take place in their center (Dvora Marcus, Nancy Brown, Noach Cohen, Jackie Diamond, Frayda Vizental-Shomron, Tamar Berger, and Sharon Finkel). As well as J-interns, Skilled volunteers for Israel for sending us interns to participate in this research. Dvora Katsev, speech therapist, for helping with ACIS adaptation and research assistants Rachel Feldstein, Skylar Dixon, Raquel Deviance, Esther Chaya, Daniella Yarrow, Naomi Ehrenkranz, Riva Kent, Daniella Spero, and Tova Leon for helping with the pilot. We wish to thank Maya Mentzel, CANlab manager for all of her help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp.2020.536880/full#supplementary-material

REFERENCES

Astell, A. (2010). Developing computer games for people with dementia. *Gerontechnology* 9:189. doi: 10.4017/gt.2010.09.02.289.00

Bar On, R., and Handley, R. (2003). *The Bar On EQ-360*. Toronto: Multi-Health Systems.

Bar-On, R. (1997). *The Emotional Quotient Inventory (EQ-i): A Test of Emotional Intelligence*. Toronto: Multi-Health Systems.

Bar-On, R. (2000). "*Emotional and social intelligence: insights from the Emotional Quotient Inventory (EQ-i),*" In Handbook of emotional intelligence. San Francisco: Jossey-Bass.

Bar-On, R. (2004). "The bar-on emotional quotient inventory (EQ-i): rationale, description and summary of psychometric properties," in *Measuring Emotional Intelligence: Common Ground and Controversy*, Hauppauge, NY: Nova Science Publishers. 115–145.

Bar-On, R. (2006). The Bar-On model of emotional-social intelligence (ESI). *Psicothema* 18, 13–25.

Bar-On, R., and Parker, J. D. A. (2000). *The Handbook of Emotional Intelligence: Theory, Development, Assessment, and Application at Home, School and in the Workplace*. SanDiego, CA: Jossey-Bass.

Bartochowski, Z., Gatla, S., Khoury, R., Al-Dahhak, R., and Grossberg, G. T. (2018). Empathy changes in neurocognitive disorders: a review. *Ann. Clin. Psychiatry* 30, 220–232.

Ben-David, B. M., Avivi-Reich, M., and Schneider, B. A. (2016). Does the degree of linguistic experience (native versus nonnative) modulate the degree to which listeners can benefit from a delay between the onset of the maskers and the onset of the target speech? *Hear. Res.* 341, 9–18. doi: 10.1016/j.heares.2016.07.016

Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., and Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *J. Speech Lang. Hear. Res.*? 54, 243–262. doi: 10.1044/1092-4388(2010/09-0233)

Ben-David, B. M., Erel, H., Goy, H., and Schneider, B. A. (2015). "Older is always better": age-related differences in vocabulary scores across 16 years. *Psychol. Aging* 30, 856–862. doi: 10.1037/pag0000051

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H., and Shakuf, V. (2019). Age-related differences in the perception of emotion in spoken language: The relative roles of prosody and semantics. *J. Speech Lang. Hear. Res.* 62, 1188–1202. doi: 10.1044/2018_JSLHR-H-ASCC7-18-0166

Ben-David, B. M., and Icht, M. (2017). Oral-diadochokinetic rates for Hebrew-speaking healthy ageing population: non-word versus real-word repetition. *Int. J. Lang. Commun. Disord.* 52, 301–310. doi: 10.1111/1460-6984.12272

Ben-David, B. M., and Icht, M. (2018). The effect of practice and visual feedback on oral-diadochokinetic rates for younger and older adults. *Lang. Speech* 61, 113–134. doi: 10.1177/0023830917708808

Ben-David, B. M., Malkin, G., and Erel, H. (2018). "Ageism and neuropsychological tests," *in Contemporary Perspectives on Ageism, (Cham: Springer)*. 277–297. doi: 10.1007/978-3-319-73820-8_17

Ben-David, B. M., Mentzel, M., Icht, M., Gilad, M., Dor, Y. I., Ben-David, S., et al. (2020). Challenges and opportunities for telehealth assessment during COVID-19: iT-RES, adapting a remote version of the test for rating emotions in speech. *Int. J. Audiol.* doi: 10.1080/14992027.2020.1833255. [Epub ahead of print].

Ben-David, B. M., and Schneider, B. A. (2009). A sensory origin for color-word stroop effects in aging: a meta-analysis. *Aging Neuropsychol. Cogn.* 16, 505–534. doi: 10.1080/13825580902855862

Ben-David, B. M., and Schneider, B. A. (2010). A sensory origin for color-word Stroop effects in aging: simulating age-related changes in color-vision mimics age-related changes in Stroop. *Aging Neuropsychol. Cogn.* 17, 730–746. doi: 10.1080/13825585.2010.510553

Berenbaum, R. (2003). The humorous side of dementia. *J Dement. Care*. 11:12. Available online at: https://careinfo.org/jdc-archive/the-humorous-side-of-dementia/

Berenbaum, R. (ed.). (2010). Alzheimer's and Other Dementias. Treating and Coping with Mentally Frail in the Community. *Jerusalem: Melabev and Eshel.*

Berenbaum, R., Lange, Y., and Abramowitz, L. (2011). "Augmentative alternative communication for Alzheimer's patients' families using SAVION," in *Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments*, (Crete). doi: 10.1145/2141622.2141677

Berenbaum, R., Tziraki, C., and Cohen-Mansfield, J. (2017b). The right to mourn in dementia: to tell or not to tell when someone dies in dementia day care. *Death Stud*. 353–359. doi: 10.1080/07481187.2017.1284953

Berenbaum, R., Tziraki, C., and Mazuz, K. (2017a). The enduring self: personhood, autonomy and compassion in the context of community-based dementia daycare centers. *J. Compassion. Health Care*, 4:8. doi: 10.1186/s40639-017-0037-z

Birks, Y., and Watt, I. (2007). Emotional intelligence and patient-centred care. *J. R. Soc. Med* 100, 368–374. doi: 10.1258/jrsm.100.8.368

Blair, R. (2003). Facial expressions, their communicatory functions and neuro–cognitive substrates. *Philos. Trans. Roy. Soc. London Series B Biol Sci.* 358, 561–572. doi: 10.1098/rstb.2002.1220

Blanchard-Fields, F. (2007). Everyday problem solving and emotion: an adult developmental perspective. *Curr. Direct. Psychol. Sci.* 16, 26–31. doi: 10.1111/j.1467-8721.2007.00469.x

Burns, A., and Iliffe, S. (2009). Clinical review: dementia. *Br. Med. J.* 338:b75. doi: 10.1136/bmj.b75

Cacioppo, J., and Cacioppo, S. (2014). Older adults reporting social isolation or loneliness show poorer cognitive function 4 years later. *Evid. Based Nursing* 17, 59–60. doi: 10.1136/eb-2013-101379

Cacioppo, J., and Hawkley, L. (2009). Perceived social isolation and cognition. *Trends Cogn. Sci.* 13, 447–454. doi: 10.1016/j.tics.2009.06.005

Calabria, M., Cotelli, M., Adenzato, M., Zanetti, O., and Miniussi, C. (2009). Empathy and emotion recognition in semantic dementia: a case report. *Brain Cogn.* 70, 247–252. doi: 10.1016/j.bandc.2009.02.009

Carstensen, L. L., and Mikels, J. A. (2005). At the intersection of emotion and cognition. *Aging and the positivity effect. Curr. Direc. Psychol. Sci.* 14, 117–121. doi: 10.1111/j.0963-7214.2005.00348.x

Cherniss, C. (2004). "Emotional intelligence," in *Encyclopedia of Applied Psychology, Vol. 2,* ed C. Spielberger (Oxford: Elsevier). 315–321. doi: 10.1016/B0-12-657410-3/00565-1

Choi, J., and Twamley, E. W. (2013). Cognitive rehabilitation therapies for Alzheimer's disease: a review of methods to improve treatment engagement and self-efficacy. *Neuropsychol. Rev.* 23, 48–62. doi: 10.1007/s11065-013-9227-4

Chun-Tie, Y., Birks, M., and Francis, K. (2019). Grounded theory research: a design framework for novice researchers. *Sage Open Med.* 7:2050312118822927. doi: 10.1177/2050312118822927

Cohen-Mansfield, J., Dakheel-Ali, M., and Marx, M. (2009). Engagement in persons with dementia: the concept and its measurement. *Am. J. Geriatr. Psychiatry* 17, 299–307. doi: 10.1097/JGP.0b013e31818f3a52

Cole, S., Capitanio, J., Chun, K., Arevalo, J., Ma, J., and Cacioppo, J. (2015). Myeloid differentiation architecture of leukocyte transcriptome dynamics in perceived social isolation. *Proc. Natl. Acad. Sci. U.S.A*. 112, 15142–15147. doi: 10.1073/pnas.1514249112

Cornet, V. P., Toscos, T., Bolchini, D., Rohani Ghahari, R., Ahmed, R., Daley, C., et al. (2020). Untold stories in user-centered design of mobile health: practical challenges and strategies learned from the design and evaluation of an app for older adults with heart failure. *JMIR Mhealth Uhealth* 8:asssese17703. doi: 10.2196/17703

Cosentino, S., Zahodne, L., Brandt, J., Blacker, D., Albert, M., Dubois, B., et al. (2014). Social cognition in Alzheimer's disease: a separate construct contributing to dependence. *Alzheimers Dement.* 10, 818–826. doi: 10.1016/j.jalz.2013.12.021

Creavin, S., Wisniewski, S., Noel-Storr, A., Trevelyan, C., Hampton, T., Rayment, D., et al. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database Syst. Rev*. 13:CD011145. doi: 10.1002/14651858.CD011145.pub2

Creswell, J. W., and Creswell, J. D. (2017). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications.

Cunnah, K., Howe, D., Thorpe, J., Dunn, R., Platt, R., White, C., et al. (2019). Training people with dementia/cognitive impairment and their carers in the use of web-based supportive technologies (Innovative practice). *Dementia*. doi: 10.1177/1471301219887592

D'Amico, M. (2017). *A review of occupation and impairment based assessments used in occupational therapy*. *OCCUPATION: A Medium of Inquiry for Students, Faculty and Other Practitioners Advocating for Health through Occupational Studies* 2:4. Available online at: https://nsuworks.nova.edu/occupation/vol2/iss1/4

Dewing, J. (2007). Participatory research: a method for process consent with persons who have dementia. *Dementia* 6, 11–25. doi: 10.1177/1471301207075625

Dewing, J. (2008). Process consent and research with older persons living with dementia. *Res. Ethics* 4, 59–64. doi: 10.1177/174701610800400205

Di Rosa, E., Hannigan, C., Brennan, S., Reilly, R., Rapčan, V., and Robertson, I. (2014). Reliability and validity of the Automatic Cognitive Assessment Delivery (ACAD). *Front. Aging Neurosci.* 6:34. doi: 10.3389/fnagi.2014.00034

Douglas, S., James, I., and Ballard, C. (2004). Non-pharmacological interventions in dementia. *Adv. Psychiatr. Treat.* 10, 171–177. doi: 10.1192/apt.10.3.171

Dourado, M., Torres Mendonça de Melo Fádel, B., Simões Neto, J., Alves, G., and Alves, C. (2019). Facial expression recognition patterns in mild and moderate alzheimer's disease. *J. Alzheimers Dis.* 69, 539–549. doi: 10.3233/JAD-181101

Earto (2014) *The TRL Scale as a Research and Innovation Policy Tool, EARTO Recommendations*. Available online at: https://www.earto.eu/wp-content/uploads/The_TRL_Scale_as_a_R_I_Policy_Tool_-_EARTO_Recommendations_-_Final.pdf (accessed October, 2020).

Eng, M. (2000). The appropriate useof humour in Alzheimer's care. *Alzheimer's Care Guide*, 8, 1–2.

Fischer, A., Landeira-Fernandez, J., Sollero de Campos, F., and Mograbi, D. (2019). Empathy in Alzheimer's Disease: Review of Findings and Proposed Model. *J. Alzheimers Dis.* 69, 921–933. doi: 10.3233/JAD-180730

Forsyth, K., Lai, J., and Kielhofner, G. (1999). The Assessment of Communication and Interaction Skills (ACIS): measurement properties. *Br. J. Occup. Ther*. 62, 69–74. doi: 10.1177/030802269906200208

Fujii, M., Butler, J., Hirazakura, A., and Sasaki, H. (2014b). Mini-emotional state examination for dementia patients. *Geriatr. Gerontol. Int.* 14, 508–513. doi: 10.1111/ggi.12127

Fujii, M., Butler, J., and Sasaki, H. (2014a). Emotional function in dementia patients. *Psychogeriatrics* 14, 202–209. doi: 10.1111/psyg.12059

Gamble, K., Howard, J., and Howard, D. (2014). Not just scenery: viewing nature pictures improves executive attention in older adults. *Exp. Aging Res.* 40, 513–530. doi: 10.1080/0361073X.2014.956618

Goleman, D. (2006). *Emotional Intelligence*. New York, NY: Random House.

Goodkind, M., Sturm, V., Ascher, E., Shdo, S., Miller, B., Rankin, K., et al. (2015). Emotion recognition in frontotemporal dementia and Alzheimer's disease: a new film-based assessment. *Emotion* 15, 416–427. doi: 10.1037/a0039261

Hasher, L., and Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. *Psychol. Learn Motivation* 22, 193–225. doi: 10.1016/S0079-7421(08)60041-9

Hirazakura, A., Hatakeyama, R., Fukuoka, Y., Satoh, A., Kobayashi, K., Fujii, M., et al. (2008). Emotional therapy for patients with dementia. *Geriatr. Gerontol. Int.* 8, 303–306. doi: 10.1111/j.1447-0594.2008.00486.x

Hoffmann, M., Cases, L., Hoffmann, B., and Chen, R. (2010). The impact of stroke on emotional intelligence. *BMC Neurol.* 10:103. doi: 10.1186/1471-2377-10-103

Ienca, M., Fabrice, J., Elger, B., Caon, M., Scoccia Pappagallo, A., Kressig, R., et al. (2017). Intelligent assistive technology for Alzheimer's disease and other dementias: a systematic review. *J. Alzheimers Dis.* 56, 1301–1340. doi: 10.3233/JAD-161037

Jamier, V., Irvine, B., and Aucher, C. (2018). *Demystifying TRLs for Complex Technologies. LEiTaT Project Blogs. Available* online *at: https://projects.leitat.org/demystifying-trls-for-complex-technologies/* (accessed October 20, 2019).

Jones, C., Sung, B., and Moyle, W. (2015). Assessing engagement in people with dementia: a new approach to assessment using video analysis. *Arch. Psychiatr. Nursing* 29, 377–382. doi: 10.1016/j.apnu.2015.06.019

Kitwood, T. M. (1997). *Dementia Reconsidered: The Person Comes First*. London: Open University Press.

Klein-Koerkamp, Y., Baciu, M., and Hot, P. (2012b). Preserved and impaired emotional memory in alzheimer's disease. *Front. Psychol.* 3:331. doi: 10.3389/fpsyg.2012.00331

Klein-Koerkamp, Y., Beaudoin, M., Baciu, M., and Hot, P. (2012a). Emotional decoding abilities in alzheimer's disease: a meta-analysis. *J. Alzheimers Dis.* 32, 109–125. doi: 10.3233/JAD-2012-120553

Kumfor, F., Sapey-Triomphe, L., Leyton, C., Burrell, J., Hodges, J., and Piguet, O. (2014). Degradation of emotion processing ability in corticobasal syndrome and Alzheimer's disease. *Brain* 137, 3061–3072. doi: 10.1093/brain/awu246

Lam, K. J., Dijkstra, T., and Rueschemeyer, S. A. (2015). Feature activation during word recognition: action, visual, and associative-semantic priming effects. *Front. Psychol.* 6:659. doi: 10.3389/fpsyg.2015.00659

León, F. G., Hernández, L. M., and Martínez, J. M. A., and Meilán, J. J. G. (2011). Emotional processing of facial expressions in normal and pathological *Aging* 48, 33–38. doi: 10.5538/1137-1242.2011.48.33

Manera, V., Ben-Sadoun, G., Aalbers, T., Agopyan, H., Askenazy, F., Benoit, M., et al. (2017). Recommendations for the use of serious games in neurodegenerative disorders: 2016 Delphi Panel. *Front. Psychol.* 8:1243. doi: 10.3389/fpsyg.2017.01243

Martinez, M., Multani, N., Anor, C., Misquitta, K., Tang-Wai, D., Keren, R., et al. (2018). Emotion detection deficits and decreased empathy in patients with alzheimer's disease and parkinson's disease affect caregiver mood and burden. *Front. Aging Neurosci.* 10:120. doi: 10.3389/fnagi.2018.00120

Mayer, J. D., and Salovey, P. (1997). *What is Emotional Intelligence. Emotional Development and Emotional Intelligence: Educational Implications. New York, NY: Harper Collins,* 3–34.

Mayer, J. D., Salovey, P., and Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT).* Toronto: Multi-Health Systems, Inc.

McAllister, M., Dayton, J., Oprescu, F. I., Katsikitis, M., and Jones C. M. (2020). Memory keeper: a prototype digital application to improve engagement with people with dementia in long-term care (innovative Practice). *Dementia* 19, 1287–1298.

McLellan, T., Johnston, L., Dalrymple-Alford, J., and Porter, R. (2008). The recognition of facial expressions of emotion in Alzheimer's disease: a review of findings. *Acta Neuropsychiatr.* 20, 236–250. doi: 10.1111/j.1601-5215.2008.00315.x

Mikels, J., Larkin, G., Reuter-Lorenz, P., and Carstensen, L. (2005). Divergent trajectories in the aging mind: changes in working memory for affective versus visual information with age. *Psychol. Aging* 20, 542–553. doi: 10.1037/0882-7974.20.4.542

Mondini, S., Arcara, G., and Jarema, G. (2014). Semantic and syntactic processing of mass and count nouns: data from dementia. *J. Clin. Exp. Neuropsychol.* 36, 967–980. doi: 10.1080/13803395.2014.958437

Nazareth, D. S. (2019). "Emotion recognition in dementia: advancing technology for multimodal analysis of emotion expression in everyday life," Presented at the *2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW).*

Perugia, G., Van Berkel, R., Díaz-Boladeras, M., Català-Mallofré, A., Rauterberg, M., and Barakova, E. (2018). Understanding engagement in dementia through behavior. the ethographic and laban-inspired coding system of engagement (ELICSE) and the evidence-based model of engagement-related behavior (EMODEB). *Front. Psychol*. 9:690. doi: 10.3389/fpsyg.2018.00690

Phillips, L., Scott, C., Henry, J., Mowat, D., and Bell, J. (2010). Emotion perception in Alzheimer's disease and mood disorder in old age. *Psychol. Aging* 25, 38–47. doi: 10.1037/a0017369

Poveda, B., Osborne-Crowley, K., Laidlaw, K., Macleod, F., and Power, K. (2017). Social cognition, behaviour and relationship continuity in dementia of the alzheimer type. *Brain Impair*. 18, 175–187. doi: 10.1017/BrImp.2016.35

Robert, P., König, A., Amieva, H., Andrieu, S., Bremond, F., Bullock, R., et al. (2014). Recommendations for the use of serious games in people with alzheimer's disease, related disorders and frailty. *Front. Aging Neurosci.* 6:54. doi: 10.3389/fnagi.2014.00054

Robillard, J., Cleland, I., Hoey, J., and Nugent, C. (2018). Ethical adoption: a new imperative in the development of technology for dementia. *Alzheimers Dement.* 14, 1104–1113. doi: 10.1016/j.jalz.2018.04.012

Shany-Ur, T., and Rankin, K. (2011). Personality and social cognition in neurodegenerative disease. *Curr. Opin. Neurol.* 24, 550–555. doi: 10.1097/WCO.0b013e32834cd42a

Snowden, J., Gibbons, Z., Blackshaw, A., Doubleday, E., Thompson, J., Craufurd, D., et al. (2003). Social cognition in frontotemporal dementia and Huntington's disease. *Neuropsychologia* 41, 688–701. doi: 10.1016/S0028-3932(02)00221-X

Thake, C., Bambling, M., Edirippulige, S., and Marx, E. (2017). A psychoevolutionary approach to identifying preferred nature scenes with potential to provide restoration from stress. *HERD: Health Environ. Res. Design J.* 10, 111–124. doi: 10.1177/1937586717705085

Torres Mendonça De Melo Fádel, B., Santos De Carvalho, R., Belfort Almeida Dos Santos, T., and Dourado, M. (2018). Facial expression recognition in Alzheimer's disease: a systematic review. *J. Clin. Exp. Neuropsychol.* 41, 192–203. doi: 10.1080/13803395.2018.1501001

Torres, B., Santos, R., Sousa, M., Simões Neto, J., Nogueira, M., Belfort, T., et al. (2015). Facial expression recognition in Alzheimer's disease: a longitudinal study. *Arquivos de Neuro-Psiquiatria* 73, 383–389. doi: 10.1590/0004-282X20150009

Trahan, M., Kuo, J., Carlson, M., and Gitlin, L. (2014). A systematic review of strategies to foster activity engagement in persons with dementia. *Health Educ. Behav.* 41, 70S–83S. doi: 10.1177/1090198114531782

Tziraki, C., Berenbaum, R., Gross, D., Abikhzer, J., and Ben-David, B. (2017). Designing serious computer games for people with moderate and advanced dementia: interdisciplinary theory-driven pilot study. *JMIR Serious Games* 5:e16. doi: 10.2196/games.6514

Valdez, R., Holden, R., Novak, L., and Veinot, T. (2015). Transforming consumer health informatics through a patient work framework: connecting patients to context. *J. Am. Med. Inform. Assoc.* 22, 2–10. doi: 10.1136/amiajnl-2014-002826

Wallace, S., Donoso Brown, E., Simpson, R., D'Acunto, K., Kranjec, A., Rodgers, M., et al. (2019). A comparison of electronic and paper versions of the montreal cognitive assessment. *Alzheimer Dis. Assoc. Disord.* 33, 272–278. doi: 10.1097/WAD.0000000000000333

World Health Organization (2019). *Dementia*. Available online at: https://www.who.int/news-room/fact-sheets/detail/dementia (accessed October 15, 2019).

Zimmerman, S., Hasher, L., and Goldstein, D. (2011). "Cognitive ageing: a positive perspective," in *Paradoxical Brain*, ed N. Kapur (Cambridge: Cambridge University Press), 130–150. doi: 10.1017/CBO9780511978098.009

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 21 June 2021
doi: 10.3389/fpsyg.2021.572877

Comparing the Impact of Heart Rate-Based In-Game Adaptations in an Exergame-Based Functional High-Intensity Interval Training on Training Intensity and Experience in Healthy Young Adults

Anna Lisa Martin-Niedecken[1]*, Tiziana Schwarz[2] and Alexandra Schättin[2]*

[1] Subject Area in Game Design, Department of Design, Zurich University of the Arts, Zurich, Switzerland, [2] Motor Control and Learning, Institute of Human Movement Sciences and Sport, Department of Health Sciences and Technology, ETH Zürich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Carina Soledad González González,
University of La Laguna, Spain

Reviewed by:
Anna Pribilova,
Slovak Academy of Sciences (SAS), Slovakia
Jan David Smeddinck,
Newcastle University, United Kingdom

***Correspondence:**
Anna Lisa Martin-Niedecken
anna.martin@zhdk.ch
Alexandra Schättin
alexandra.schaettin@hest.ethz.ch

Specialty section:
This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Psychology

Received: 15 June 2020
Accepted: 05 March 2021
Published: 21 June 2021

Citation:
Martin-Niedecken AL, Schwarz T and Schättin A (2021) Comparing the Impact of Heart Rate-Based In-Game Adaptations in an Exergame-Based Functional High-Intensity Interval Training on Training Intensity and Experience in Healthy Young Adults. Front. Psychol. 12:572877. doi: 10.3389/fpsyg.2021.572877

Physical inactivity remains one of the biggest societal challenges of the 21st century. The gaming industry and the fitness sector have responded to this alarming fact with game-based or gamified training scenarios and thus established the promising trend of exergaming. Exergames—games played with the (whole) body as physical input—have been extolled as potential attractive and effective training tools. Simultaneously, researchers and designers are still exploring new approaches to exploit the full potential of this innovative and enjoyable training method. One way to boost the attractiveness and effectiveness of an exergame is to individualize it with game adaptations. A physiological parameter that is often used to balance the physical challenge and intensity of exergames to the player's fitness skills is the heart rate (HR). Therefore, researchers and designers often rely on age-based, maximum HR (HR_{max}) formulas originating from performance diagnostics. In combination with the player's assessed real-time HR during an exergame session, the pre-determined HR_{max} is used to adapt the game's challenge to reach a pre-defined HR and physical intensity level (in-exergame adaptations), respectively. Although the validity and reliability of these age-based HR_{max} formulas were proven in heterogeneous target populations, their use is still often criticized as HR is an individual parameter that is affected by various internal and external factors. So far, no study has investigated whether the formula-based pre-calculated HR_{max} compared to a standardized individually pre-assessed HR_{max} elicits different training intensities, training experiences, and flow feelings in an exergame. Therefore, we compared both variants for in-exergame adaptation with the ExerCube – a functional high-intensity interval training exergame – in healthy young adults. Comparing the results of the two conditions, no significant differences were found for HR parameters and perceived physical and cognitive exertion, nor for overall flow feelings and physical activity enjoyment. Thus, the formula-based in-exergame adaptation approach was suitable in the presented study population, and the ExerCube provided an equally reliable in-exergame adaptation and comparable exergame play experiences. We discuss our findings in the context of related work on exergame adaptation approaches and draw out some implications for future adaptive exergame design and research topics.

Keywords: exergame, game balancing, heart rate, effectiveness, attractiveness, assessement, ExerCube, in-game adaptation

INTRODUCTION

Numerous guidelines preach the urgent necessity of regular physical activity to maintain a physically and mentally healthy lifestyle at all ages. According to the American College of Sports Medicine, a program of regular exercise, including cardiorespiratory, resistance, flexibility, and neuromotor training in addition to the activities of daily living, is essential for most adults to improve and maintain their physical fitness and health (Garber et al., 2011). However, surveys by the World Health Organization continuously reveal that physical inactivity remains the greatest public health problem of the 21st century (Trost et al., 2014). In addition to a lack of motivation and time, changing behavioral and environmental factors as well as a number of common exercise barriers are the main reasons for this persisting problem (Trost et al., 2002). Therefore, stakeholders from various fields have called for new concepts for attractive and effective training alternatives to reduce entry barriers and help to maintain training adherence for a wide range of people over a period of several years (Marshall and Linehan, 2020).

Exergames, which require physical effort and are controlled by (whole) body movements (Oh and Yang, 2010), have been promoted as suitable tools for providing attractive and effective training alternatives or supplements by the interdisciplinary research and development (R&D) community. In more than 10 years, R&D work has proved that exergames have the potential to be training tools that are both effective (i.e., increasing physical-cognitive fitness, endurance, strength, and coordination) (Staiano and Calvert, 2011a,b; Sween et al., 2014; Best, 2015; Benzing et al., 2016; Kari, 2017; Mura et al., 2018; Stojan and Voelcker-Rehage, 2019; Xiong et al., 2019) and attractive (e.g., increasing training adherence, motivation, flow, and engagement) (McRae et al., 2012; Valenzuela et al., 2018; Martin-Niedecken et al., 2019b; Tondello et al., 2019). Furthermore, exergamging promotes physical activity and training in different target populations (Lu et al., 2013; Kappen et al., 2019; Martin-Niedecken and Schättin, 2020). However, the majority of the evaluated exergames, which were not necessarily specifically designed for the purpose of obtaining certain training results or to be used as a motivating alternative to traditional training methods, did not meet the required intensity or effectiveness, nor did they induce the intended training adherence or long-term motivation (e.g., Marshall and Linehan, 2020).

A promising approach towards boosting the effectiveness (e.g., training intensity and outcomes) and attractiveness (e.g., flow, immersion, enjoyment, and motivation) of exergames is the personalization through system adaptations, also known as dynamic game balancing (Mueller et al., 2012; Altimira et al., 2016), game difficulty adjustment (Adams, 2010), dynamic difficulty adjustments (Hunicke, 2005), or multiplayer game balancing (Gerling et al., 2014). Besides various pre-exergame and real-time in-exergame adaptation parameters (e.g., game speed, frequency of in-game tasks, and increasing input-movement intensity and range), in-exergame adaptations based on the player's heart rate (HR) have been proposed and explored by various researchers and designers. These HR-based concepts were proven to be feasible and beneficial approaches to balance the player's abilities and the challenges of the exergame (Sinclair et al., 2009; Stach et al., 2009; Mueller et al., 2012; Hoffmann et al., 2015; Ketcheson et al., 2015; Martin-Niedecken and Götz, 2016, 2017; Martin-Niedecken, 2018; Martin-Niedecken and Mekler, 2018; Muñoz et al., 2018; Martin-Niedecken et al., 2019a,b). At the same time, these formulas are often criticized because HR is an individual parameter that is influenced by various internal and external factors (Stach et al., 2009; Mueller et al., 2012; Hoffmann et al., 2015; Ketcheson et al., 2015). However, no study has yet investigated whether the formula-based pre-calculated maximum HR (HR_{max}) compared to a standardized individually pre-assessed HRmax elicits different training intensities, training experiences, and flow feelings in an exergame. Furthermore, exergames and their underlying technologies have the potential to serve as assessment tools for in-exergame adaptation parameters, and could thus replace strenuous and unmotivating traditional test procedures with playful assessments (e.g., Konstantinidis et al., 2015). Nonetheless, this has not yet been explored in much detail. To sum up, further interdisciplinary R&D work is needed to fill these gaps and to further explore the full potential of (adaptive) exergames as an innovative training and assessment tool.

Our work explores these R&D gaps, with the goal of better understanding the design requirements and potential of attractive and effective exergames. Based on an overview of different pre- and in-exergame adaptation approaches, we present a study that compares two different play conditions of a newly developed, adaptive functional high-intensity interval training (fHIIT) exergame, the so-called ExerCube, using objective and subjective measures for training intensity (effectiveness) and experience (attractiveness) in healthy young adults. The ExerCube automatically adapted to the targeted HR range (defined as percentage of HR_{max}) by either (i) the individually pre-assessed HR_{max} or (ii) the formula-based pre-calculated HR_{max}. Furthermore, we explored the potential utility of the ExerCube as a HR_{max} assessment tool by comparing the

Abbreviations: HR, heart rate; R&D, research and development; HR_{max}, maximum heart rate; HR_{avg}, average heart rate; fHIIT, functional high-intensity interval training; HCI, human-computer interaction; min, minutes; sec, seconds.

ExerCube procedure to a standard ergometer protocol. Based on our results, we discuss how individual in-exergame adaptation can influence the effectiveness and attractiveness of exergames.

RELATED WORK

During the last few years, interdisciplinary researchers and designers have started exploring different exergame adaptation approaches in various target populations. These approaches were based on theories and models from disciplines such as human movement science, human–computer interaction (HCI), game research, and psychology.

Multi-Level Pre-exergame Adaptations

Pre-exergame adaptations based on various design levels of an exergame have so far been considered only by a limited number of R&D studies. Following, we present selected studies that used pre-exergame adaptations in different application areas.

Hardy et al. (2014) tested the dependence of motivation, perceived difficulty, and performance on specific level features as well as goal-setting in a balance exergame to find out whether the effectiveness of the training can be increased intentionally by changing level features or setting personal goals. They found a significant influence of single features on psychological (motivation and diffculty) and physiological constructs (performance and play time).

Altimira et al. (2013, 2014) investigated how two traditional balancing approaches (i.e., playing with the non-dominant hand and using a pre-determined head-start) affect players' experiences in traditional table tennis and in game-based table tennis with the Nintendo Wii, respectively. Their studies showed that playing with the non-dominant hand discouraged players in traditional table tennis and that having a score disadvantage discouraged them in the digital version. Building on this, Altimira et al. (2017) studied how digital technology (i.e., altering the sports equipment: playing with a smaller bat-head or a smaller table) can be used as a resource for game balancing in an augmented table tennis exergame. They showed that dynamic adjustments enhanced engagement more than static adjustments.

Jensen and Grønbæk (2016) developed and evaluated the effects of three balancing schemes based on a physical, an implicit-digital, and an explicit-digital approach, implemented into a ball-controlled exergame. They demonstrated that all three game balancing approaches were feasible and gave equal chances to win while all players enjoyed the balanced gameplay.

Gerling et al. (2014) examined different game adjustments such as score multipliers, the precision of the input movements, and the number of movements implemented in a dancing game. They found that obvious game balancing can reduce players' self-esteem in comparison to hidden game balancing. Score balancing reduced the appearance of extreme performance gaps between players. The adjustment of input movement precision reduced small differences in players' performances and in asymmetric physical input, e.g., a player in a wheelchair.

Siegel and Smeddinck (2012) examined an approach to dynamic difficulty adjustments in an exergame for Parkinson's disease using patients' range of motion as well as movement speed and accuracy as adaptation parameters. They found that this approach was viable and appreciated by therapists. However, the system might benefit from increased flexibility.

HR-Based In-Exergame Adaptations

Considerably more widespread is R&D work in the area of in-exergame adaptations, especially focusing on the player's HR.

Mueller et al. (2012) presented an urban jogging system that used HR data and spatialized sound to create an equitable, balanced experience between joggers of different fitness levels and geographical locations. They demonstrated that real-time HR-based balancing positively affected players' experiences in a remote jogging application because they performed in their own training zone, while still engaging with another person.

Stach et al. (2009) used HR scaling in a multiplayer cycling exergame. Players' in-exergame performances were based on their effort relative to their fitness level. They demonstrated that HR scaling reduced the performance gap of different fitness levels. Moreover, engagement was not significantly affected during gameplay.

Ketcheson et al. (2015) evaluated HR power-ups to encourage vigorous training intensities in a cycling exergame intervention. This real-time game mechanism provided in-exergame rewards when players reached the targeted HR level (e.g., the avatar may be more powerful). The pedaling was used to control the avatar's movement while a standard controller was used to navigate the avatar in different directions and to release in-game actions. Researchers concluded that HR power-ups enhanced exertion levels while also increasing players' enjoyment levels.

Hoffmann et al. (2015) implemented and tested an algorithm that controlled the physical load of an endurance cycling exergame to approach and maintain a pre-defined HR. They used the pedaling frequency (cadence) as a game-controlling parameter and the resistance (Watt) as an adaptive control mechanism. The evaluation indicated that the developed algorithm was a feasible approach for controlling an individual adaptive training load in the cycling exergame.

Muñoz et al. (2018) investigated the effectiveness and attractiveness of a Pong-like floor-projected exergame in older adults. They showed that real-time HR-based in-exergame adaptations (e.g., speed and training zone) increased the time the older adults spent in the recommended exertion level by approximately 40% compared to conventional functional training. Furthermore, this exergame training provided a controlled, safe, joyful, and effective cardiovascular training in older adults.

Physical–Cognitive In-Exergame Adaptations

There are two exergame approaches that extend the HR-based approach by exploring HR-based (physical) and performance-related (cognitive) in-exergame adaptations. Sinclair et al.'s (2007, 2009) dual flow model proposes certain design strategies to balance players' gaming (cognitive) and fitness (physical) abilities with the actual required skills to successfully play and enjoy an exergame in real-time. Based on this model, Martin-Niedecken and Götz (2016, 2017), Martin-Niedecken (2018) designed and evaluated the adaptive fitness game environments Plunder Planet for children and ExerCube for adults (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019b). Among other things, they experimented with performance-related and physiologically-based real-time in-exergame adaptation mechanics and implemented them as algorithms.

Plunder Planet (Martin-Niedecken and Götz, 2016, 2017; Martin-Niedecken, 2018) is a single- and two-player exergame that can be played in two versions. One variation includes a full-body motion controller providing physical guidance via six big buttons distributed on two racks (height and weight adjustable) on the player's left- and right-hand side, requiring cognitive and coordinative skills as well as haptic interactions. The other variation includes the gesture-based Kinect sensor, which allows more natural and intuitive input movements and more freedom of movement. In both variations, the player navigates a flying pirate ship along a deserted racing track and has to overcome virtual obstacles and avoid collisions with sandworms. The player's HR is tracked via a chest strap. The exergame can be manually or automatically adapted in real-time to the player's physical and cognitive performance. Thus, the speed of the ship and the frequency of virtual obstacles (physical challenge) are increased or decreased based on the player's HR. The track characteristics (e.g., flatter or curved) and the number of options for overcoming an obstacle (one to three option(s), i.e., hard to easy, respectively) are adapted to the player's performance (cognitive challenge). A study in children proved the fundamental functionality and usability (attractiveness and effectiveness) of Plunder Planet (Martin-Niedecken and Götz, 2016). Martin-Niedecken and Götz (2017) found that the implementation of adaptive game mechanics provided players with an enjoyable and effective (moderate intensity) exergame experience, and that playing the game with the different controllers resulted in different spatial presence and gameplay experiences depending on the player's preferences as well as play and sports skills. An advanced study (Martin-Niedecken and Götz, 2017) demonstrated that the adaptive Plunder Planet version was significantly better than the non-adaptive one in relation to game flow, dual flow, motivation, enjoyment, and spatial presence, as well as the children's physiological responses.

The ExerCube (**Figure 1**; Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019a,b) is an immersive mixed-reality fitness game for single or multiple players. The player is surrounded by three walls that serve as projection screens and haptic interfaces for energetic bodily interactions. A customized motion-tracking system tracks players' movements via HTC Vive trackers (attached to their wrists and ankles). In the functional fitness game scenario Sphery Racer, which is projected onto the walls of the ExerCube, the player races along a fast-paced sci-fi underwater race track via an avatar on a hoverboard. The motion-tracking system transfers the executed movements (based on a functional workout) onto the avatar and thus onto the virtual racing track. Along the race, players are challenged by functional whole body exercises (e.g., squats, lunges, and burpees) and

FIGURE 1 | Functional high-intensity interval training in the ExerCube.

by an additional cognitive challenge as players have to quickly process track information and react accordingly (i.e., reaction, planning, and coordination challenges). To ensure an attractive and effective workout experience for a wide spectrum of players with different skill sets, this fitness game continuously adapts game difficulty to players' individual fitness and cognitive skills. Training intensity is measured via continuous HR tracking (i.e., players wear a HR-sensor chest strap). Depending on the targeted training intensity (e.g., moderate or high), tracking is set to an individually pre-defined HR range, defined as a percentage of HR_{max}, where HR_{max} is manually inserted or automatically calculated based on the following formula (Nes et al., 2013):

$$HR_{\max} = 211 - \text{age} \times 0.64 \qquad (1)$$

Based on this training range and the player's HR, the exercise difficulty, gaming speed, frequency of obstacles, and track characteristics are increased or decreased within the training session. Cognitive skills are measured by in-exergame performance (reacting to visual stimuli at the right time) and are balanced by the display timing of the next movement direction. In a previous empirical study, the first early stage prototype was found to be on par with personal training in terms of immersion, motivation, and flow (Martin-Niedecken et al., 2019b). A second study with the redesigned prototype aimed to compare the objective and subjective physiological training intensity in the ExerCube to that induced by conventional fHIIT (Martin-Niedecken et al., 2020). This study demonstrated that the ExerCube is a feasible tool for inducing fHIIT intensity. While the average HR was significantly lower than in the conventional functional HIIT condition, maximal HR was at the same high level for both trainings, and the percentage of HR_{max} calculation (based on Nes et al., 2013) showed medium to high training intensities. Furthermore, the ExerCube training yielded significantly better results for flow, enjoyment, and motivation.

Formula-Based In-Exergame Adaptations

In addition to some relatively unexplored parameters for pre- and in-exergame adaptations (e.g., controller adaptations), the player's HR has been proposed and explored by various researchers and designers. HR-based adaptation was proved to be a feasible and beneficial approach to balance player's physical abilities and the exergame challenge and thus to enhance training experience (e.g., engagement, perceived effectiveness, and motivation) and training effects (e.g., performance and outcomes) (Sinclair et al., 2009; Stach et al., 2009; Mueller et al., 2012; Hoffmann et al., 2015; Ketcheson et al., 2015; Martin-Niedecken and Götz, 2016, 2017; Martin-Niedecken, 2018; Martin-Niedecken and Mekler, 2018; Muñoz et al., 2018; Martin-Niedecken et al., 2019a,b). This can be attributed to the fact that in the area of sports and fitness training, HR has already been thoroughly investigated (Ludwig et al., 2018). Furthermore, HR is relatively easy to implement technically in an exergame by means of various wearable devices that allow sending real-time HR data via Bluetooth to the game engine.

So far, only some studies have explored the feasibility of implemented HR prediction models to incorporate HR-based adaptation mechanisms into an exergame (Hoffmann et al., 2015, 2016; Hoffmann and Wiemeyer, 2017b). However, most researchers and designers still rely on age-based formulas to calculate HR_{max} with which to approach and maintain a pre-defined HR range and physical load in exergames. Two examples are given by Robergs and Landwehr (2002) and Nes et al. (2013):

$$HR_{\max} = 220 - \text{age} \tag{2}$$

$$HR_{\max} = 211 - \text{age} \times 0.64 \tag{3}$$

These formulas are then implemented in combination with a targeted HR range (defined as percentage of HR_{max}) to set the training intensity of an exergame (e.g., beyond 80% of HR_{max} corresponds to high-intensity levels (MacInnis and Gibala, 2017)).

Although numerous validation studies are available (e.g., Nes et al., 2013), these formulas are often criticized because HR and HR_{max} are highly individual parameters that, besides age, can be influenced by various internal (e.g., gender, training status (Nes et al., 2013), genetics (Wang et al., 2009), and mood (Petrov et al., 2014)) and external factors (e.g., environmental conditions, nutrition, and water supply) (Ludwig et al., 2018). Thus, some researchers pre-assessed the individual HR_{max} and targeted range of HR of players with standardized ergometer tests and used this in combination with pre-defined cycling training protocols to implement a pre-defined physical load in a cycling exergame (Barathi et al., 2018; Farrow et al., 2019).

Exergame-Based Assessments

Another promising approach is the application of exergames as assessment tools (Konstantinidis et al., 2015). This could be another interesting step towards user-friendly and motivating training scenarios and exploiting further in-exergame adaptation mechanisms and strategies. There are hardly any exergame-based assessments to measure cognitive and physical fitness or mental state. Such assessments, although, would allow a more holistic classification of the personal skills, deficits, and mood, and thus would provide a more individualized and detailed default setting of in-exergame adaptation parameters. So far, traditional assessment batteries are still performed manually and outside the (exer)game setting. Thinking about HR_{max} standardized assessments, these testing protocols often involve a player going right up to or even beyond their limits in a way that is not necessarily a pleasant situation to experience. Integrated into an exergame, HR_{max} assessments might become more pleasant since it is known from related work that the immersive and motivating nature of a game helps to shift the focus from one's own body and the physical strain to the gaming experience and the cognitive level (Martin-Niedecken et al., 2019b). Furthermore, this could increase the overall usability of exergames as attractive and effective training tools on the market, and support trainers and therapists.

MATERIALS AND METHODS

To fill the aforementioned gaps, we conducted a comparative study aiming to explore whether there are objective and subjective differences in training intensity (effectiveness) and experience (attractiveness) when the ExerCube is automatically adapted to the targeted HR range (defined as percentage of HR_{max}) by either the individually pre-assessed or the formula-based pre-calculated HR_{max}. Furthermore, this study aimed to gain some early indications about the usage of the ExerCube to determine individual HR_{max} by comparing the procedure to a standardized ergometer protocol (Maier et al., 2016).

Because of these objectives, the project consisted of two parts to investigate both study questions (**Figure 2**). The study ran from November 2019 to February 2020. Measurements were made either at the research group laboratory (ETH Zürich, Hönggerberg, Zurich, Switzerland) or at the Sphery gym (Asylstrasse 64, Zurich, Switzerland). All measurements were made by one investigator who was familiar with the study set-up. Participants were continuously supervised by the experienced investigator and the instructions were provided in a standardized manner for each participant. The ethics committee of the ETH Zürich, Switzerland (EK 2019-N-137), approved the study protocol. Before any measurements were carried out, all eligible participants had to sign a form giving their informed consent according to the Declaration of Helsinki.

Participants

For this study, the minimal intended sample size of 20 healthy (self-reported by health questionnaire) young adults aged 18–35 years was based on a previous study that examined similar parameters (e.g., HR) (Martin-Niedecken et al., 2020) in the same training setting, as well as on the possibility of losses or refusals. Participants were excluded from the study if one of the following exclusion criteria was presented: (1) history of cardiovascular issues that would prevent training participation, (2) asthma (not

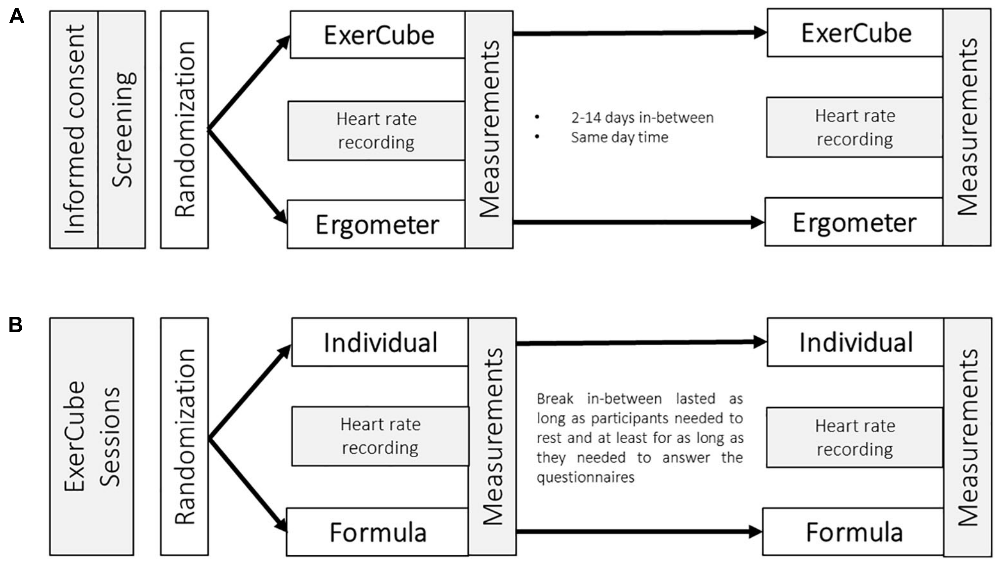


FIGURE 2 | Study procedure. **(A)** Testing sessions to assess individual maximal heart rate. **(B)** Comparison of in-exergame adaptation. In-between **(A)** and **(B)** are 2–14 days.

controllable), (3) musculoskeletal injuries that would prevent training participation, (4) pain that would be reinforced by sports activities, and (5) pregnancy. For recruitment, different methods were used, such as word-of-mouth and emailing (ETH and company Sphery Ltd, Zurich, Switzerland), without offering any financial compensation for attendance. Prior to the first measurement, all participants were fully informed about the procedure, benefits, and risks of the study.

Procedure

Testing Sessions to Assess Individual Maximal Heart Rate

Two separate testing appointments were carried out to assess the individual HR_{max} via an ergometer and an ExerCube protocol. Both testing protocols are presented in **Figure 3**. At the beginning of the ergometer and the ExerCube session, the resting HR was assessed in a sitting position for 5 minutes (min). HR measurements were continued during the HR_{max} testing sessions. At the end of each session, participants rated their physically and cognitively perceived exertion. The sequence of the testing sessions was randomized to minimize subsequent effects and an intervening period of 2–14 days was set.

Comparison of In-Exergame Adaptations

The last appointment included two ExerCube sessions. At the beginning of the appointment, resting HR was assessed in a sitting position for 5 min. HR measurements were continued during both ExerCube sessions. Participants performed the "dual flow protocol" of the Sphery Racer for 11 min (20 seconds (sec) onboarding and calibration scene, 10 min training, and 40 sec pit stops) at a training intensity of 80% HR_{max}. The respective HR_{max} of the sessions was defined in two different ways:

- Individual: higher value of both testing sessions
- Formula-based (Nes et al., 2013):

$$HR_{\max} = 211 - \text{age} \times 0.64 \tag{4}$$

The ExerCube training protocol started with a short onboarding and calibration scene (20 sec) and contained five exercise intervals on the virtual racing-track, each accessed from an intermediate short pit stop (10 sec).

- Interval 1 (1 min): Touch, Touch low, Touch high (left (L) /right(R))
- Interval 2 (2 min): + Squat, Jumping, Punch (L/R)
- Interval 3 (2 min): + Lunge (L/R)
- Interval 4 (2 min): + Skipping
- Interval 5 (3 min): + Burpee

Exercises started with low-to-moderate intensity (in terms of both physical and cognitive load) and gradually increased over time to high-intensity. The physical and cognitive challenge were gradually adapted independently over the whole training session on a 10-point difficulty scale, where one level was defined as one step on the 10-point scale (e.g., from 5 to 6). In both ExerCube sessions, the game aimed to keep players at 80% HR_{max}. A lower HR led to an increase in physical challenge, i.e., speed and exercise frequency (one level per check), while a higher HR led to a decrease (once 100% HR_{max} was reached, this decrease was sped up by three levels to ensure players' safety). The system employed a strategy for increasing players' HR, i.e., when 80% HR_{max} has not been reached, it checked actual HR every 20 sec and every 10 sec when HR was above 90% HR_{max}. The cognitive challenge increased by one level (resulting in a delayed display of the direction of the next exercise) if the player performed error-free for 20 sec. If the player made three mistakes within 20 sec, the difficulty decreased by one level (resulting in an earlier display of the direction of the next exercise). Thus,

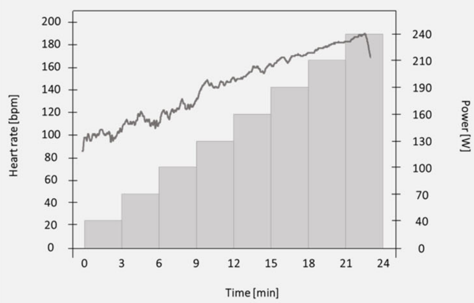
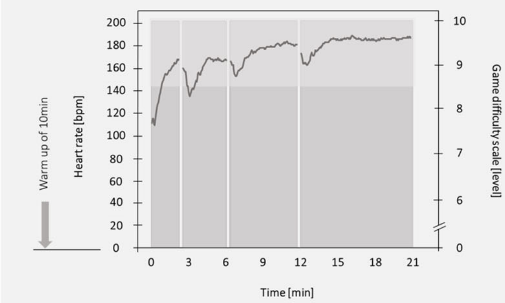
	Ergometer	Exergame
Device	Bicycle	ExerCube
Exercise	Repetitive cycling movements	Functional whole-body movements
Warm-up	No, protocol started at a low level	Yes, 10 min dual-flow protocol of the Sphery Racer (with tutorial), training intensity at 80% HRmax for familiarization
Protocol	Swiss Olympic protocol • **Start load:** Woman: 40W & Man: 70W • **Load steps:** 30W after 3 min • **Cadence:** Favorable, but must be maintained, recommendation: 80-100 revolutions per min • **Interval:** No breaks, continuous increase	Upper body protocol of the Sphery Racer, training intensity at 100% HRmax (formula calculation) • **Start level:** 5 • **Load:** Intended between level 8.5 and 10, controlled by study investigator (light grey area in the illustration) • **Interval:** – Interval 1 (2 min): Touch, Touch low, Touch high (left (L)/right (R)) – Pit stop/Break (20 sec) – Interval 2 (4 min): + Punch (L/R) – Pit stop/Break (20 sec) – Interval 3 (5 min): + Jumping – Pit stop/Break (20 sec) – Interval 4 (9 min): + Skipping
Termination	• Individual termination by participant • Cadence is clearly decreasing	• After 21 min or individual termination by participant • After five consecutive errors
Testing destination	ETH laboratory	Sphery gym
Exemplary illustration		

FIGURE 3 | Comparison of the different protocols to determine individual maximal heart-rate. Reference for Swiss Olympic protocol (Maier et al., 2016).

physical and cognitive game challenges were controlled each by an independent algorithm. Nevertheless, physical and cognitive load controls were interconnected in the gameplay. For example, when the HR decreased (physical algorithm), the game sped up and the acceleration of the game speed increased the number of exercises and therefore increased the probability of mistakes on the game track (cognitive algorithm).

Both ExerCube sessions were carried out at one appointment. The sequence of the sessions was randomized to balance pre-fatigue effects and minimize subsequent effects. The break in-between lasted as long as the participant needed to rest (self-evaluated) and at least for as long as they needed to answer the questionnaires (about 8 min). After each ExerCube session, the participants rated their general, physically, and cognitively perceived exertion and answered questions about their training experience.

Assessment

Heart rate data recordings were performed in the resting state, during HR_{max} assessment (ergometer and ExerCube), and during the ExerCube sessions (formula and individual HR_{max}) to measure the average HR (HR_{avg}) and/or HR_{max}. Resting HR was determined using the average over the three measurements. For each measurement, resting HR was the average over 5 min under the condition of a steady state being reached before starting the measurement. Participants wore a HR-receiving chest belt of the brand Wahoo (Wahoo Fitness 2014, Atlanta, GA, United States) for HR data collection. The chest belt was either connected (via Bluetooth) to the ExerCube, or the compatible "Wahoo Fitness" App (installed on an android mobile phone).

The Borg 6- to 20-point (6 = very, very light, 20 = very, very hard) and modified 10-point (1 = very weak, 10 = very, very strong) rating scales were selected to assess perceived exertion (Borg, 1982). The 6- to 20-point scale was used to assess the general perceived exertion (Borg) and the 10-point scale was used to assess both physically ($Borg_{physical}$) and cognitively ($Borg_{cognitive}$) perceived exertion. Training experience was assessed by three questionnaires: (a) Flow Short Scale (FSS), (b) Flow State Scale (FStS), and (c) Physical Activity Enjoyment Scale (PACES). The FSS and the FStS assessed participants' flow experience (Jackson and Marsh, 1996; Rheinberg et al., 2003). The FSS consists of 13 items and the FStS consists of nine items (short version). In the FSS, the flow experience is measured overall and as three factors: fluency of performance, absorption by activity, and perceived importance. The FStS measures flow experiences during physical activity. Further, participants' enjoyment of the training was assessed via the PACES, consisting of 18 items (Kendzierski and DeCarlo, 1991; Motl et al., 2001). The FSS and PACES questionnaires were rated on a 7-point Likert scale (FSS: 1 = not at all, 7 = very much; PACES: bipolar statements,

TABLE 1 | Baseline characteristics.

Age (years)	25.3 ± 1.5
Gender	Men: n = 12, Women: n = 9
Years of education	17.4 ± 1.1
Fitness status (self-rated)[1]	3.8 ± 1.0
Activity (hours per week)	3.9 ± 2.8
Resting heart-rate (bpm)	76.5 ± 11.8
Exergame experience	61.9% (yes), 38.1% (no)
ExerCube experience	23.8% (yes), 76.2% (no)

N = 21. Data are mean values (standard deviation) as indicated. [1]Fitness status (self-rated): 1 = poor, 2 = satisfactory, 3 = average, 4 = good, 5 = very good, 6 = competitive sports level.

1 = disagree a lot, 7 = agree a lot) and the FStS was rated on 5-point Likert scale (1 = not at all, 5 = very much).

Analysis

Statistical analysis was conducted in SPSS (IBM SPSS 26). Since the criteria for a parametric analysis were not given, comparisons of the HR values, rating scales, and questionnaires were performed using the Wilcoxon signed-rank test. The level of significance was set at $p < 0.05$. Effect sizes were calculated using the formula (Cohen, 2013):

$$r = \frac{z}{\sqrt{N}} \quad (5)$$

where z = z-score and N = number of participants.

An effect size $0.1 \leq r < 0.3$ is considered a small effect, $0.3 \leq r < 0.5$ a medium effect, and $r \geq 0.5$ a large effect.

RESULTS

The study was performed with 21 participants (9 women, 12 men) aged 25.3 ± 1.5 years. The participants' baseline characteristics, fitness status, and exergame experience are presented in **Table 1**.

Assessment of Maximal Heart Rate

Results of the testing sessions are presented in **Table 2**. No significant difference was measured between the ergometer and the ExerCube testing sessions for HR_{max} (z = -0.444, p = 0.657, r = 0.07). Significant higher values for the ExerCube testing session were measured for HR_{avg} (z = -4.017, $p < 0.001$, r = 0.62), time to HR_{max} (z = -3.563, $p < 0.001$, r = 0.55) and $Borg_{cognitive}$ (z =-3.984, $p < 0.001$, r = 0.61). The average training level of the ExerCube was between 8.6 ± 0.5 (mean ± standard deviation) over all participants. The ExerCube testing session was finished by 20 of 21 participants.

Training Experience Comparing Individual Versus Formula HR_{max}

One participant had to be excluded from the analysis due to technical difficulties. No significant differences were measured for HR and perceived exertion parameters (**Table 3**). For the questionnaire data, a significant difference resulted for the item "unambiguous feedback" in favor of the individual HR_{max} condition (z = -2.121, p = 0.034, r = 0.34). All the other questionnaire data showed no significant differences. Results of the questionnaire data are presented in **Table 4**. **Table 5** shows an overview of the perceived feelings between the formula and the individual HR-based ExerCube sessions.

DISCUSSION

The aim of this study was to explore whether there are objective and subjective differences in training intensity (effectiveness) and experience (attractiveness) when the ExerCube is automatically adapted to the targeted HR range (defined as percentage of HR_{max}) by the individually pre-assessed or formula-based pre-calculated HR_{max}. Furthermore, this study aimed to gain some early indications about the usage of the ExerCube to determine individual HR_{max} by comparing the procedure to a standardized ergometer protocol. The following sections discuss the results in the context of related work and knowledge in the area of (individualized) in-exergame adaptations and exergame-based assessment using HR. Furthermore, implications are illustrated for future adaptive exergame design and research topics.

In-Exergame Adaptations: Training Experience and Intensity Comparing Individual and Formula Based HR_{max}

Comparing the individual and formula-based HR_{max} ExerCube conditions, one significant difference was assessed for the item "unambiguous feedback" in favor of the individual HR_{max} condition. This significant difference, however, must be considered with caution since a significant difference does not always imply a (clinically) relevant difference (Page, 2014). Furthermore, this questionnaire item was already at a high level for both conditions.

All the other questionnaire items, the short survey on player's feelings, and the assessed subjective and objective training intensity showed favorable values independently of the HR_{max} condition. The reason for this might be that the comparison of the pre-defined individual HR_{max} values (formula and individual), used for defining the training intensity in the ExerCube, revealed no significant difference, indicating that the ExerCube allowed a reliable in-exergame adaptation as well as gameplay experience. These positive experiences, including high feelings of flow and enjoyment, as well as the favorable training intensity are in line with previous ExerCube studies (Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019a,b). Thus, in-exergame adaptations via HR seem to be feasible for triggering an individually attractive and effective gameplay experience, as demonstrated in previous studies (Stach et al., 2009; Mueller et al., 2012; Hoffmann et al., 2015; Martin-Niedecken and Götz, 2016, 2017; Martin-Niedecken, 2018; Muñoz et al., 2018).

Furthermore, findings of this study indicate that the formula concept may be a good alternative to the individually determined HR_{max} for in-exergame adaptation. Nevertheless, the presented

TABLE 2 | Comparison of maximal heart rate testing.

	Ergometer	ExerCube	z	p	r
HR_{max} (bpm)	192.0 (185.0, 196.0)	189.0 (184.0, 199.0)	–0.444	0.657	0.07
HR_{avg} (bpm)	149.0 (142.0, 156.0)	172.0 (165.0, 181.0)	–4.017	<0.001*	0.62
Time to HR_{max} (min)	19.1 (17.2, 24.7)	16.1 (15.2, 19.1)	–3.563	<0.001*	0.55
Borg cognitive (1–10)	2.0 (1.0, 2.5)	6.0 (5.0, 6.0)	–3.984	<0.001*	0.61
Borg physical (1–10)	8.0 (8.0, 9.0)	8.0 (7.0, 9.0)	–1.067	0.286	0.16

*N = 21. Data are median (interquartile range) as indicated. Data comparison was analyzed using Wilcoxon-signed rank test. *p < 0.05, p-values are two tailed. Effect size r: 0.1 ≤ r < 0.3 small effect, 0.3 ≤ r < 0.5 medium effect, r ≥ 0.5 large effect. HR_{max} = maximal heart rate, HR_{avg} = average heart rate.*

TABLE 3 | Comparison of ExerCube conditions (formula vs. individual) for heart rate and perceived exertion.

	Formula	Individual	z	p	r
Pre-defined HR_{max} (bpm)	195.0 (194.5, 195.6)	193.0 (186.8, 200.8)	–0.766	0.444	0.12
HR_{max} (bpm)	178.0 (170.5, 187.5)	181.0 (171.0, 185.0)	–0.065	0.948	0.01
HR_{avg} (bpm)	151.0 (141.3, 157.3)	150 (139.3, 163.0)	–0.497	0.619	0.08
Borg (6–20)	14.0 (13.0, 15.0)	15.0 (13.0, 15.0)	–1.035	0.301	0.16
Borg cognitive (1–10)	4.0 (3.0, 5.0)	4.0 (3.0, 5.0)	–1.071	0.284	0.17
Borg physical (1–10)	5.0 (4.3, 6.0)	5.5 (5.0, 6.0)	–0.263	0.793	0.04

*N = 20. Pre-defined HR_{max} was used to determine training intensity (defined as percentage of HR_{max}) for in-exergame adaptation. Data are median (interquartile range) as indicated. Data comparison was analyzed using Wilcoxon-signed rank test. *p < 0.05. p-values are two tailed. Effect size r: 0.1 ≤ r < 0.3 small effect, 0.3 ≤ r < 0.5 medium effect, r ≥ 0.5 large effect. HR = heart rate, HR_{max} = maximal heart rate, HR_{avg} = average heart rate.*

TABLE 4 | Comparison of ExerCube conditions (formula vs. individual) for enjoyment and flow.

	Formula	Individual	z	p	r
PACES	6.0 (5.3, 6.3)	5.9 (5.3, 6.4)	–0.047	0.962	0.01
FSS	5.6 (4.9, 6.2)	5.9 (5.2, 6.1)	–0.468	0.640	0.07
Absorption	5.5 (5.3, 6.3)	5.6 (5.3, 6.0)	–0.263	0.793	0.04
Fluency	5.6 (4.8, 6.1)	5.8 (5.3, 6.3)	–0.787	0.431	0.12
Perceived importance	2.5 (2.0, 3.3)	2.3 (1.8, 3.6)	–0.682	0.495	0.11
FStS	4.3 (3.9, 4.5)	4.3 (4.0, 4.6)	–0.969	0.333	0.15
Challenge-skill balance	5.0 (4.3, 5.0)	5.0 (5.0, 5.0)	–1.414	0.157	0.22
Action-awareness merging	4.5 (3.0, 5.0)	4.0 (4.0, 5.0)	–0.758	0.449	0.12
Clear goals	5.0 (4.0, 5.0)	5.0 (5.0, 5.0)	–1.150	0.132	0.18
Unambiguous feedback	4.0 (3.0, 5.0)	4.0 (4.0, 5.0)	–2.121	0.034*	0.34
Concentration on the task	5.0 (4.0, 5.0)	5.0 (4.0, 5.0)	–1.134	0.257	0.18
Sense of control	4.0 (3.0, 4.8)	4.0 (3.0, 4.0)	–0.535	0.539	0.08
Loss of self-control	5.0 (5.0, 5.0)	5.0 (5.0, 5.0)	–1.414	0.157	0.22
Transformation of time	4.0 (4.0, 5.0)	4.0 (3.3, 5.0)	–1.732	0.083	0.27
Autotelic experience	4.0 (3.3, 5.0)	4.0 (3.0, 5.0)	–0.905	0.366	0.14

*N = 20. Data are median (interquartile range) as indicated. Data comparison was analyzed using Wilcoxon-signed rank test. *p < 0.05. p-values are two tailed. Effect size r: 0.1 ≤ r < 0.3 small effect, 0.3 ≤ r < 0.5 medium effect, r ≥ 0.5 large effect.*

study included a rather homogeneous group of young adults, resulting in an individual HR_{max} that was well covered by the formula. A huge fitness study, however, showed that the implemented formula adequately explained HR_{max} by age, considering an age range of 19–89 years (Nes et al., 2013). Thus, the formula concept might also be suitable for older age groups. Nontheless, one must consider that the formula is an approximation of the real value and (maximal) HR can (daily) be influenced by various internal and external factors such as gender, circadian cycle, blood pressure, lifestyle factors, physical activity, and mental status (Londeree and Moeschberger, 1984; Valentini and Parati, 2009). Therefore, this study provided some early indications that have to be substantiated with more studies considering different player attributes (e.g., activity levels and mental status). Furthermore, it might be interesting to explore which application area (e.g., rehabilitation, fitness, and prevention) might benefit the most from a HR-based training.

TABLE 5 | Survey of different perceived feelings between formula and individual session.

	Formula	Individual	No difference
Which session was more exhausting?	*n* = 11 (55%)	*n* = 7 (35%)	*n* = 2 (10%)
Which session felt more pleasant?	*n* = 9 (45%)	*n* = 8 (40%)	*n* = 3 (15%)

N = 20.

Exergame-Based Assessment: ExerCube as a HR_{max} Assessment Tool

In terms of HR_{max} assessment, results demonstrated that both testing protocols triggered comparable HR_{max} as no significant difference was present. This result gave a first impression that it seems feasible to assess individual HR_{max} using the ExerCube. To our best knowledge, this study is one of the first investigations that assessed individual HR_{max} using an exergame.

Furthermore, results showed a significantly higher HR_{avg} and significantly shorter time to reach HR_{max} for the ExerCube assessment protocol. This might be explained by the fact that both protocols had different design attributes. The ExerCube started directly with a fast increase of the intensity and the warm-up was not part of the testing protocol. Interestingly, 20 out of 21 participants finished the whole ExerCube session lasting 21 min, even if the HR_{max} was reached before the protocol ended. This circumstance might be due to the fact that the ExerCube was not only a physical challenge, but also stimulated cognitive processes that presented in a significantly higher cognitive load for the ExerCube training protocol. In combination with the previously decribed game flow and enjoyment, this higher cognitive engagement might have distracted the player's focus away from their physical exhaustion and therefore let participants perform longer on a high intensity (Bertollo et al., 2015; Bigliassi et al., 2016). It may be that the participants could have done their testing in the ExerCube even longer at this high level as no participant was stopped because of continuous performance errors.

Overall, the results of this comparison led to several observations that might be useful concerning exergaming as HR_{max} assessment. A familiarization phase seems to be mandatory for exergames, as conventional HR_{max} assessments usually have a less complex environment, exercise performance, and/or movement patterns, respectively. This familiarization is an important control process to ensure that overexertion is not caused by incorrect performance or misunderstanding of the performance. The length of this phase should be determined depending on the complexity of the exergame, ensuring that the participant has understood the exergame control and play mechanism. A warm-up phase could be used for familiarization, and therefore precede the testing session. Still, a warm-up phase can also be included in the assessment, as is usually the case for conventional HR_{max} testing protocols, by starting at a low-intensity level. Regarding termination criteria, overexertion can be determined by the participant (subjective), as in conventional performance tests or, particularly for exergames, by a pre-defined number of failures, movement precision, accuracy, and power (objective) or by performance worsening (subjective). Moreover, these overexertion parameters as well as the familiarization and warm-up phase are important precautions to ensure the safety of participants.

A special feature of exergames is the unique combination of physical, cognitive, and mental load. The nature of HR_{max} testing protocols is to increase the HR via high physical load. In the ExerCube protocol, the physical load was increased by exercise frequency (racing speed) and physically intensive exercises (e.g., skipping). Furthermore, the ExerCube included cognitive stimuli via information processing of the virtual track and the required in-game actions (e.g., reaction, planning, and coordination). In this study, the cognitive load of the ExerCube was more or less at the same level throughout the testing session. Cognitive stimulation, as mentioned before, could be supportive as it might distract from the physical exhaustion (Bertollo et al., 2015; Bigliassi et al., 2016). However, an overloaded cognitive stimulation could have opposite effects as fewer resources might be available for the physical performance (MacMahon et al., 2014). On the other hand, an increasing cognitive load could also be a part of a HR_{max} testing protocol as a high cognitive load seems to increase HR (Mehler et al., 2009; Rudolf et al., 2016). Next to physical and cognitive load, mental load can also have an effect on the HR as excitement and stress can initiate biological responses (Valentini and Parati, 2009). Nevertheless, how physical, cognitive, and mental load should be combined in a HR_{max} testing protocol is part of future studies because further research is needed to understand, strengthen, and complement the interaction of these loads.

Parameters for Real-Time In-Exergame Adaptations

These study results initiate the discussion of how HR or further parameters (e.g., insights from eye tracking) could be used (exclusively or in combination) for real-time in-exergame adaptation, allowing an individually tailored exergame experience. Individual training and game adaptations based on user requirements may increase training/gaming motivation (attractiveness) and success (effectiveness) (Sinclair et al., 2007, 2009; Martin-Niedecken and Götz, 2016, 2017; Martin-Niedecken, 2018; Martin-Niedecken and Mekler, 2018; Martin-Niedecken et al., 2019a,b).

Knowledge from different research fields and disciplines (e.g., sport science and HCI) should be used to examine different parameters assessing physical (Wallace et al., 2014; Burgess, 2017; Coyne et al., 2018; McLaren et al., 2018), cognitive (Solovey et al., 2014; Grassmann et al., 2016; Oschlies-Strobel et al., 2017; He et al., 2019; Hughes et al., 2019; Zhou et al., 2020), and mental (Schrader et al., 2017; Mostefai et al., 2019) load. Objective parameters could be defined by physiological and performance-related factors (Vasilyev et al., 2019). Physiological parameters could be measurements, outcomes, and variables related to HR (e.g., heart rate variability), respiration, eye tracking, facial expression, skin conductance, and brain (e.g., functional near-infrared spectroscopy and electroencephalogram) and muscle (e.g., electromyography) activity. Performance-related parameters could be reaction time and failure rate as well as movement execution, acceleration, deceleration, and accuracy. In addition to the objective parameters, subjective parameters (e.g., rating scales) could be used to determine the different loads (Smith et al., 2014; Saw et al., 2016; Meckel et al., 2018).

An essential starting point is that the parameters should suit the exergame mechanics, components, and (training) goal. Exergames could record performance-related data but also integrate devices or game mechanics that measure and assess objective (e.g., physiological and performance-related factors) and subjective (e.g., rating scales) parameters to determine the player's physical, cognitive, and mental load. In real-time in-exergame adaptations, the parameters can be implemented at different levels of the exergame such as controller (e.g., sensitivity of tracking), game (e.g., audio-visual appearance) and player (e.g., range of motion) (Martin-Niedecken and Götz, 2017; Martin-Niedecken, 2018; Martin-Niedecken et al., 2019a; Wiemeyer, 2019). However, due to the complex combination of physical, cognitive, and mental components, further conceptual thoughts have to be considered in future studies.

Studies that look at the individual components of an exergame and their interdependencies seem promising to determine the influence of the above-mentioned loads and thus to support the integration of suitable objective and subjective parameters for real-time in-exergame adaptations (Gutjahr et al., 2019; Martin-Niedecken et al., 2020). For example, one should keep in mind whether the game scenario, and thus the processed information, is coupled (e.g., running in an engaging environment and catching a robber) or uncoupled (e.g., running and solving calculation tasks) to the player's physically performed movements because this further affects the interplay of the exergame components (Herold et al., 2018; Martin-Niedecken et al., 2019a).

Exergame-based assessments might support the determination of parameter(s) under a maximal (e.g., HR_{max}), optimal or standardized testing situation. Result(s) of performance tests can then be used to individually determine the starting load and/or the individually targeted training intensity (e.g., 80% HR_{max}), and that in turn can be used to control real-time in-exergame adaptations (Hardy et al., 2015). These individual in-exergame adaptations can be controlled by specific algorithms and may even be improved by the inclusion of artificial intelligence. By requiring and storing player information in internal models, AI might allow dynamic modeling and prediction of an exergame track (Wenger, 2014; Streicher and Smeddinck, 2016; Hoffmann and Wiemeyer, 2017a; Ludwig et al., 2018; Gang et al., 2019).

Overall, the usability and feasibility of these parameters have to be considered in proportion to the potential gain for the exergame's attractiveness and effectiveness.

Limitations

In the context of this study, some limitations have to be mentioned. One is the homogeneous study population of fairly fit younger adults, allowing only a limited generalization of the study results as HR in particular is an individual parameter that depends on several internal and external factors. Therefore, future studies should consider including participants from different age ranges and different fitness levels to check whether these results can be replicated or if different considerations or further adjustments have to be made for different conditions.

Furthermore, two limitations in the context of the HR_{max} testing session should be mentioned. The testing protocols used differed in their structure and this might have influenced the study results. Nevertheless, the study results gave early indications of how far the ExerCube could already be used to determine HR_{max} in its existing design. Moreover, the speed of the ExerCube during the HR_{max} testings was subjectively regulated (maximal performance for each participant) by the observing study investigator, allowing maximal speed adaptation that would not otherwise be possible due to automatic in-exergame adaptation of the ExerCube. A next step would be to elaborate standardized testing protocol(s) for the ExerCube to regulate the subjective components down to a minimum.

In addition, the use of a treadmill instead of a bicycle might be even more appropriate since the movements in the ExerCube were performed in a standing position. Nevertheless, the difference of HR_{max} between bicycle and treadmill ergometer tend to be small at young ages but become larger with increasing age (Londeree and Moeschberger, 1984). Furthermore, this testing protocol already gave an early indication of how the ExerCube can be used to assess HR_{max}.

CONCLUSION

Given the urgent demand for attractive and effective training tools, exergames represent a promising and innovative approach. Nonetheless, exergames need to fulfill certain design and training aspects to be a real alternative to conventional training methods. A promising approach towards boosting the effectiveness and attractiveness of exergames is the personalization through in-exergame adaptations. Among other things, HR has often been used to balance the physical load and training intensity of exergames with the player's fitness skills. The most common way was to implement an age-based HR_{max} formula into the exergame, allowing the exergame to reach a targeted training intensity (e.g., percentage of HR_{max}). To contribute to this promising topic, we explored different HR conditions (a standardized individually pre-assessed HR_{max} and a formula-based pre-calculated HR_{max}) in the exergame ExerCube in healthy young adults and compared the impact on training intensity and experience. Comparing the results of the two conditions, no significant differences were found for HR parameters and perceived exertion (physical and cognitive), nor for overall flow feelings and enjoyment. Thus, the formula-based in-exergame adaptation approach was suitable in the presented study population, and the ExerCube provided an equally reliable in-exergame adaptation and comparable exergame play experience. Furthermore, we investigated the usage of an ExerCube protocol to determine HR_{max} by comparing the procedure to a standard ergometer protocol. Results indicated that the ExerCube seems to be a feasible tool for assessing individual HR_{max}. Finally, we derived some implications for future adaptive exergame design and research topics.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the ETH Zürich, Switzerland (EK 2019-N-137). The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AM-N and AS conceptualized, designed, and drafted the manuscript. TS conducted the study (supervised by AS and AM-N). TS and AS led the data analysis and interpretation. AM-N also contributed to the latter. All authors created the study design, compiled the training protocols, carefully selected the assessment methods, and critically reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

We kindly thank Sphery Ltd., for providing the infrastructure and technical support throughout the study duration, which ensured a smooth implementation of the study. We also thank all participants for participating in our study.

REFERENCES

Adams, E. (2010). *Fundamentals of Game Design*. Berkeley, CA: New Riders.

Altimira, D., Billinghurst, M., and Mueller, F. (2013). "Understanding handicapping for balancing exertion games," in *Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems*, Paris, 1125–1130.

Altimira, D., Clarke, J., Lee, G., Billinghurst, M., and Bartneck, C. (2017). Enhancing player engagement through game balancing in digitally augmented physical games. *Int. J. Hum. Comput. Stud.* 103, 35–47. doi: 10.1016/j.ijhcs.2017.02.004

Altimira, D., Mueller, F. F., Clarke, J., Lee, G., Billinghurst, M., and Bartneck, C. (2016). "Digitally augmenting sports: an opportunity for exploring and understanding novel balancing techniques," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose, CA, 1681–1691.

Altimira, D., Mueller, F. F., Lee, G., Clarke, J., and Billinghurst, M. (2014). "Towards understanding balancing in exertion games," in *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, Funchal, 1–8.

Barathi, S. C., Finnegan, D. J., Farrow, M., Whaley, A., Heath, P., Buckley, J., et al. (2018). "Interactive feedforward for improving performance and maintaining intrinsic motivation in VR exergaming," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, QC, 1–14.

Benzing, V., Heinks, T., Eggenberger, N., and Schmidt, M. (2016). Acute cognitively engaging exergame-based physical activity enhances executive functions in adolescents. *PLoS One* 11:e0167501. doi: 10.1371/journal.pone.0167501

Bertollo, M., di Fronso, S., Lamberti, V., Ripari, P., Reis, V. M., Comani, S., et al. (2015). To focus or not to focus: is attention on the core components of action beneficial for cycling performance? *Sport Psychol.* 29, 110–119. doi: 10.1123/tsp.2014-0046

Best, J. R. (2015). Exergaming in youth. Effects on physical and cognitive health. *Z. Psychol.* 221, 72–78. doi: 10.1027/2151-2604/a000137

Bigliassi, M., Silva, V. B., Karageorghis, C. I., Bird, J. M., Santos, P. C., and Altimari, L. R. (2016). Brain mechanisms that underlie the effects of motivational audiovisual stimuli on psychophysiological responses during exercise. *Physiol. Behav.* 158, 128–136. doi: 10.1016/j.physbeh.2016.03.001

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Med. Sci. Sports Exerc.* 14, 377–381.

Burgess, D. J. (2017). The research doesn't always apply: practical solutions to evidence-based training-load monitoring in elite team sports. *Int. J. Sports Physiol. Perform.* 12, S2–S136.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Academic press.

Coyne, J. O., Haff, G. G., Coutts, A. J., Newton, R. U., and Nimphius, S. (2018). The current state of subjective training load monitoring—a practical perspective and call to action. *Sports Med. Open* 4:58.

Farrow, M., Lutteroth, C., Rouse, P. C., and Bilzon, J. L. (2019). Virtual-reality exergaming improves performance during high-intensity interval training. *Eur. J. Sport Sci.* 19, 719–727. doi: 10.1080/17461391.2018.1542459

Gang, P., Zeng, W., Gordienko, Y., Rokovyi, O., Alienin, O., and Stirenko, S. (2019). Prediction of physical load level by machine learning analysis of heart activity after exercises. *arXiv* [Preprint] arXiv:1912.09848.

Garber, C. E., Blissmer, B., Deschenes, M. R., Franklin, B. A., Lamonte, M. J., Lee, I.-M., et al. (2011). American college of sports medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med. Sci. Sports Exerc.* 43, 1334–1359. doi: 10.1249/mss.0b013e318213fefb

Gerling, K. M., Miller, M., Mandryk, R. L., Birk, M. V., and Smeddinck, J. D. (2014). "Effects of balancing for physical abilities on player performance, experience and self-esteem in exergames," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, ON, 2201–2210.

Grassmann, M., Vlemincx, E., von Leupoldt, A., Mittelstädt, J. M., and Van den Bergh, O. (2016). Respiratory changes in response to cognitive load: a systematic review. *Neural Plast.* 2016:8146809.

Gutjahr, M. O., Ellermeier, W., Hardy, S., and Wiemeyer, J. (2019). The pupil response as an indicator of user experience in a digital exercise game. *Psychophysiology* 56:e13418.

Hardy, S., Dutz, T., Wiemeyer, J., Göbel, S., and Steinmetz, R. (2015). Framework for personalized and adaptive game-based training programs in health sport. *Multimed. Tools Appl.* 74, 5289–5311. doi: 10.1007/s11042-014-2009-z

Hardy, S., Kern, A., Dutz, T., Weber, C., Göbel, S., and Steinmetz, R. (2014). "What makes games challenging? considerations on how to determine the "challenge" posed by an exergame for balance training," in *Proceedings of the 2014 ACM International Workshop on Serious Games*, Orlando, FL, 57–62.

He, D., Donmez, B., Liu, C. C., and Plataniotis, K. N. (2019). High cognitive load assessment in drivers through wireless electroencephalography and the validation of a modified N-back task. *IEEE Trans. Hum. Mach. Syst.* 49, 362–371. doi: 10.1109/thms.2019.2917194

Herold, F., Hamacher, D., Schega, L., and Müller, N. G. (2018). Thinking while moving or moving while thinking–concepts of motor-cognitive training for cognitive performance enhancement. *Front. Aging Neurosci.* 10:228.

Hoffmann, K., Sportwiss, D., Hardy, S., Wiemeyer, J., and Göbel, S. (2015). Personalized adaptive control of training load in cardio-exergames—a feasibility study. *Games Health J.* 4, 470–479. doi: 10.1089/g4h.2014.0073

Hoffmann, K., and Wiemeyer, J. (2017a). Predicting short-term HR response to varying training loads using exponential equations. *Int. J. Comput. Sci. Sport* 16, 130–148. doi: 10.1515/ijcss-2017-0011

Hoffmann, K., and Wiemeyer, J. (2017b). "Statistical models for predicting short-term hr responses to submaximal interval exercise," in *Proceedings of the*

International Symposium on Computer Science in Sport, (Cham: Springer), 57–68. doi: 10.1007/978-3-319-67846-7_6

Hoffmann, K., Wiemeyer, J., and Hardy, S. (2016). "Prediction and control of the individual heart rate response in exergames," in *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, (Cham: Springer), 171–178. doi: 10.1007/978-3-319-24560-7_22

Hughes, A. M., Hancock, G. M., Marlow, S. L., Stowers, K., and Salas, E. (2019). Cardiac measures of cognitive workload: a meta-analysis. *Hum. Factors* 61, 393–414. doi: 10.1177/0018720819830553

Hunicke, R. (2005). "The case for dynamic difficulty adjustment in games," in *Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology*, Valencia, 429–433.

Jackson, S. A., and Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *J. Sport exerc. Psychol.* 18, 17–35. doi: 10.1123/jsep.18.1.17

Jensen, M. M., and Grønbæk, K. (2016). "Design strategies for balancing exertion games: a study of three approaches," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, Brisbane, QLD, 936–946.

Kappen, D. L., Mirza-Babaei, P., and Nacke, L. E. (2019). Older adults physical activity and exergames: a systematic review. *Int. J. Hum. Comput. Interact.* 35, 140–167. doi: 10.1080/10447318.2018.1441253

Kari, T. (2017). "Promoting physical activity and fitness with exergames: updated systematic review of systematic reviews," in *Transforming Gaming and Computer Simulation Technologies Across Industries*, ed. B. Dubbels, (Pennsylvania PA: IGI Global), 225–245. doi: 10.4018/978-1-5225-1817-4.ch013

Kendzierski, D., and DeCarlo, K. J. (1991). Physical activity enjoyment scale: two validation studies. *J. Sport Exerc. Psychol.* 13, 50–64. doi: 10.1123/jsep.13.1.50

Ketcheson, M., Ye, Z., and Graham, T. N. (2015). "Designing for exertion: how heart-rate power-ups increase physical activity in exergames," in *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, London, 79–89.

Konstantinidis, E. I., Antoniou, P. E., and Bamidi, P. D. (2015). "Exergames for assessment in active and healthy aging-emerging trends and potentialities," in *Proceedings of the International Conference on Information and Communication Technologies for Ageing Well and e-Health*, (Lisbon: Scitepress), 325–330.

Londeree, B., and Moeschberger, M. (1984). Influence of age and other factors on maximal heart rate. *J. Card. Rehabil.* 4, 44–49.

Lu, A. S., Kharrazi, H., Gharghabi, F., and Thompson, D. (2013). A systematic review of health videogames on childhood obesity prevention and intervention. *Games Health Res. Dev. Clin. Appl.* 2, 131–141. doi: 10.1089/g4h.2013.0025

Ludwig, M., Hoffmann, K., Endler, S., Asteroth, A., and Wiemeyer, J. (2018). Measurement, prediction, and control of individual heart rate responses to exercise—basics and options for wearable devices. *Front. Physiol.* 9:778. doi: 10.3389/fphys.2018.00778

MacInnis, M. J., and Gibala, M. J. (2017). Physiological adaptations to interval training and the role of exercise intensity. *J. Physiol.* 595, 2915–2930. doi: 10.1113/jp273196

MacMahon, C., Schücker, L., Hagemann, N., and Strauss, B. (2014). Cognitive fatigue effects on physical performance during running. *J. Sport Exerc. Psychol.* 36, 375–381. doi: 10.1123/jsep.2013-0249

Maier, T., Gross, M., Trösch, S., Steiner, T., Müller, B., Bourban, P., et al. (2016). *Manual Leistungsdiagnostik*, (Bern: Swiss Olympic Association).

Marshall, J., and Linehan, C. (2020). Are exer-games exercise? A scoping review of the short term Effects of exertion games. *IEEE Trans. Games* 99:1. doi: 10.1109/tg.2020.2995370

Martin-Niedecken, A. L. (2018). "Designing for bodily interplay: engaging with the adaptive social exertion game" plunder planet"," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, Trondheim, 19–30.

Martin-Niedecken, A. L., Árquez Segura, E. M., Rogers, K., Niedecken, S., and Turmo Vidal, L. (2019a). "Towards socially immersive fitness games: an exploratory evaluation through embodied sketching," in *Proceedings of the Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, Barcelona, 525–534.

Martin-Niedecken, A. L., and Götz, U. (2016). "Design and evaluation of a dynamically adaptive fitness game environment for children and young adolescents," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, Austin TX, 205–212.

Martin-Niedecken, A. L., and Götz, U. (2017). "Go with the dual flow: evaluating the psychophysiological adaptive fitness game environment "Plunder Planet"," in *Proceedings of the Joint International Conference on Serious Games* (Cham: Springer), 32–43. doi: 10.1007/978-3-319-70111-0_4

Martin-Niedecken, A. L., Mahrer, A., Rogers, K., De Bruin, E. D., and Schättin, A. (2020). "HIIT" the Exercube: comparing the effectiveness of functional high-intensity interval training in conventional vs. exergame-based training. *Front. Comput. Sci.* 2:33. doi: 10.3389/fcomp.2020.00033

Martin-Niedecken, A. L., and Mekler, E. D. (2018). "The exercube: participatory design of an immersive fitness game environment," in *Proceedings of the 4th Joint International Conference on Serious Games*, (Cham: Springer), 263–275. doi: 10.1007/978-3-030-02762-9_28

Martin-Niedecken, A. L., Rogers, K., Turmo Vidal, L., Mekler, E. D., and árquez Segura, E. M. (2019b). "Exercube vs. personal trainer: evaluating a holistic, immersive, and adaptive fitness game setup," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, 1–15.

Martin-Niedecken, A. L., and Schättin, A. (2020). Let the body 'n'brain games begin: toward innovative training approaches in esports athletes. *Front. Psychol.* 11:138. doi: 10.3389/fpsyg.2020.00138

McLaren, S. J., Macpherson, T. W., Coutts, A. J., Hurst, C., Spears, I. R., and Weston, M. (2018). The relationships between internal and external measures of training load and intensity in team sports: a meta-analysis. *Sports Med.* 48, 641–658. doi: 10.1007/s40279-017-0830-z

McRae, G., Payne, A., Zelt, J. G., Scribbans, T. D., Jung, M. E., Little, J. P., et al. (2012). Extremely low volume, whole-body aerobic–resistance training improves aerobic fitness and muscular endurance in females. *Appl. Physiol. Nutr. Metab.* 37, 1124–1131. doi: 10.1139/h2012-093

Meckel, Y., Zach, S., Eliakim, A., and Sindiani, M. (2018). The interval-training paradox: physiological responses vs. subjective rate of perceived exertion. *Physiol. Behav.* 196, 144–149. doi: 10.1016/j.physbeh.2018.08.013

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Trans. Res. Rec.* 2138, 6–12. doi: 10.3141/2138-02

Mostefai, B., Balla, A., and Trigano, P. (2019). A generic and efficient emotion-driven approach toward personalized assessment and adaptation in serious games. *Cogn. Syst. Res.* 56, 82–106. doi: 10.1016/j.cogsys.2019.03.006

Motl, R. W., Dishman, R. K., Saunders, R., Dowda, M., Felton, G., and Pate, R. R. (2001). Measuring enjoyment of physical activity in adolescent girls. *Am. J. Prev. Med.* 21, 110–117. doi: 10.1016/s0749-3797(01)00326-9

Mueller, F., Vetere, F., Gibbs, M., Edge, D., Agamanolis, S., Sheridan, J., et al. (2012). "Balancing exertion experiences," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Austin, TX, 1853–1862.

Muñoz, J. E., Cameirão, M., Bermúdez i Badia, S., and Gouveia, E. R. (2018). "Closing the loop in exergaming-health benefits of biocybernetic adaptation in senior adults," in *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*, Melbourn, VIC, 329–339.

Mura, G., Carta, M. G., Sancassiani, F., Machado, S., and Prosperini, L. (2018). Active exergames to improve cognitive functioning in neurological disabilities: a systematic review and meta-analysis. *Eur. J. Phys. Rehabil. Med.* 54, 450–462.

Nes, B., Janszky, I., Wisløff, U., Støylen, A., and Karlsen, T. (2013). Age-predicted maximal heart rate in healthy subjects: the hunt fitness study. *Scand. J. Med. Sci. Sports* 23, 697–704. doi: 10.1111/j.1600-0838.2012.01445.x

Oh, Y., and Yang, S. (2010). "Defining exergames & exergaming," in *Proceedings of the Meaningful Play*, (East Lansing, MI: Michigan State University), 1–17.

Oschlies-Strobel, A., Gruss, S., Jerg-Bretzke, L., Walter, S., and Hazer-Rau, D. (2017). "Preliminary classification of cognitive load states in a human machine interaction scenario," in *Proceedings of the 2017 International Conference on Companion Technology (ICCT)*, (Ulm: IEEE), 1–5.

Page, P. (2014). Beyond statistical significance: clinical interpretation of rehabilitation research literature. *Int. J. Sports Phys. Ther.* 9:726.

Petrov, L., Bozhilov, G., Alexandrova, A., Mugandani, S., and Djarova, T. (2014). Salivary alpha-amylase. heart rate and heart-rate variability in response to an experimental model of competitive stress in volleyball players: sport science. *Afr. J. Phys. Health Educ. Recreation Dance* 20, 308–322.

Rheinberg, F., Vollmeyer, R., and Engeser, S. (2003). "Die erfassung des flow-erlebens. [The assessment of flow experience]," in *Diagnostik Von Selbstkonzept, Lernmotivation und Selbstregulation [Diagnosis of Motivation and Self-Concept]*, eds J. Stiensmeier-Pelster, and F. Rheinberg, (Göttingen: Hogrefe).

Robergs, R. A., and Landwehr, R. (2002). The surprising history of the "HRmax= 220-age" equation. *J. Exerc. Physiol. Online* 5:2002.

Rudolf, K., Grieben, C., Achtzehn, S., and Froböse, I. (2016). "Stress im esport– ein einblick in training und wettkampf," in *Proceedings of the Esport Conference Professionalisierung Einer Subkultur*.

Saw, A. E., Main, L. C., and Gastin, P. B. (2016). Monitoring the athlete training response: subjective self-reported measures trump commonly used objective measures: a systematic review. *Br. J. Sports Med.* 50, 281–291. doi: 10.1136/bjsports-2015-094758

Schrader, C., Brich, J., Frommel, J., Riemer, V., and Rogers, K. (2017). "Rising to the challenge: an emotion-driven approach toward adaptive serious games," in *Serious Games and Edutainment Applications*, eds M. Ma, and A. Oikonomou, (Cham: Springer), 3–28. doi: 10.1007/978-3-319-51645-5_1

Siegel, S., and Smeddinck, J. (2012). "Adaptive difficulty with dynamic range of motion adjustments in exergames for parkinson's disease patients," in *Proceedings of the International Conference on Entertainment Computing*, (Berlin: Springer), 429–432. doi: 10.1007/978-3-642-33542-6_45

Sinclair, J., Hingston, P., and Masek, M. (2007). "Considerations for the design of exergames," in *Proceedings of the 5th international conference on Computer graphics and interactive techniques in Australia and Southeast Asia*, Perth, WA, 289–295.

Sinclair, J., Hingston, P., and Masek, M. (2009). "Exergame development using the dual flow model," in *Proceedings of the Sixth Australasian Conference on Interactive Entertainment*, Sydney, NSW, 1–7.

Smith, M., Fransen, J., and Coutts, A. (2014). *Inducing and Assessing Cognitive Fatigue*. Amsterdam: European College of Sport Science Annual Congress.

Solovey, E. T., Zec, M., Garcia Perez, E. A., Reimer, B., and Mehler, B. (2014). "Classifying driver workload using physiological and driving performance data: two field studies," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Toronto, ON, 4057–4066.

Stach, T., Graham, T., Yim, J., and Rhodes, R. E. (2009). "Heart rate control of exercise video games," in *Proceedings of the Graphics Interface 2009*, (Kelowna, BC: Canadian Information Processing Society), 125–132.

Staiano, A. E., and Calvert, S. L. (2011a). Exergames for physical education courses: physical, social, and cognitive benefits. *Child Dev. Perspect.* 5, 93–98. doi: 10.1111/j.1750-8606.2011.00162.x

Staiano, A. E., and Calvert, S. L. (2011b). He promise of exergames as tools to measure physical health. *Entertain. Comput.* 2, 17–21. doi: 10.1016/j.entcom.2011.03.008

Stojan, R., and Voelcker-Rehage, C. (2019). A systematic review on the cognitive benefits and neurophysiological correlates of exergaming in healthy older adults. *J. Clin. Med.* 8:734. doi: 10.3390/jcm8050734

Streicher, A., and Smeddinck, J. D. (2016). "Personalized and adaptive serious games," in *Entertainment Computing and Serious Games*, eds R. Dörner, S. Göbel, M. Kickmeier-Rust, M. Masuch, and K. Zweig, (Cham: Springer), 332–377. doi: 10.1007/978-3-319-46152-6_14

Sween, J., Wallington, S. F., Sheppard, V., Taylor, T., Llanos, A. A., and Dams-Campbell, L. L. A. (2014). The role of exergaming in improving physical activity: a review. *J. Phys. Act. Health* 11, 864–870. doi: 10.1123/jpah.2011-0425

Tondello, G. F., Mora, A., Marczewski, A., and Nacke, L. E. (2019). Empirical validation of the gamification user types hexad scale in English and Spanish. *Int. J. Hum. Comput. Stud.* 127, 95–111. doi: 10.1016/j.ijhcs.2018.10.002

Trost, S. G., Blair, S. N., and Khan, K. M. (2014). Physical inactivity remains the greatest public health problem of the 21st century: evidence, improved methods and solutions using the '7 investments that work'as a framework. *Br. J. Sports Med.* 48, 169–170. doi: 10.1136/bjsports-2013-093372

Trost, S. G., Owen, N., Bauman, A. E., Sallis, J. F., and Brown, W. (2002). Correlates of adults participation in physical activity: review and update. *Med. Sci. Sports Exerc.* 34, 1996–2001. doi: 10.1097/00005768-200212000-00020

Valentini, M., and Parati, G. (2009). Variables influencing heart rate. *Progr. Cardiovas. Dis.* 52, 11–19. doi: 10.1016/j.pcad.2009.05.004

Valenzuela, T., Okubo, Y., Woodbury, A., Lord, S. R., and Delbaere, K. (2018). Adherence to technology-based exercise programs in older adults: a systematic review. *J. Geriatr. Phys. Ther.* 41, 49–61. doi: 10.1519/jpt.0000000000000095

Vasilyev, V., Borisov, V., and Syskov, A. (2019). "Biofeedback Methodology: a Narrative Review," in *Proceedings of the 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, (Novosibirsk: IEEE), 0011–0016.

Wallace, L., Slattery, K., and Coutts, A. J. (2014). A comparison of methods for quantifying training load: relationships between modelled and actual training responses. *Eur. J. Appl. Physiol.* 114, 11–20. doi: 10.1007/s00421-013-2745-1

Wang, X., Ding, X., Su, S., Li, Z., Riese, H., Thayer, J. F., et al. (2009). Genetic influences on heart rate variability at rest and during stress. *Psychophysiology* 46, 458–465. doi: 10.1111/j.1469-8986.2009.00793.x

Wenger, E. (2014). *Artificial Intelligence and Tutoring Systems: Computational and Cognitive Approaches to the Communication of Knowledge*. (Los altos, CA: Morgan Kaufmann).

Wiemeyer, J. (2019). "Towards a generic framework for serious games," in *Proceedings of the International Symposium on Computer Science in Sport*, (Cham: Springer), 193–200. doi: 10.1007/978-3-030-35048-2_23

Xiong, S., Zhang, P., and Gao, Z. (2019). Effects of exergaming on preschoolers executive functions and perceived competence: a pilot randomized trial. *J. Clin. Med.* 8:469. doi: 10.3390/jcm8040469

Zhou, B., Ghose, T., and Lukowicz, P. (2020). Expressure: detect expressions related to emotional and cognitive activities using forehead textile pressure mechanomyography. *Sensors* 20:730. doi: 10.3390/s20030730

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be interpreted as a potential conflict of interest. Besides their academic career, AM-N and AS are also working for Sphery. AM-N is the Co-Founder and CEO of the startup company Sphery Ltd., that developed the ExerCube. AS has been working as Senior Research and Development Manager at Sphery since November 2019. No revenue was paid (or promised to be paid) directly to AM-N, to AS, to Sphery or the research institutions.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership