Contents

Viewpoint

Extended Reality for Enhanced Telehealth During and Beyond COVID-19: Viewpoint (e26520)
Triton Ong, Hattie Wilczewski, Samantha Paige, Hiral Soni, Brandon Welch, Brian Bunnell. ................................................. 3

Original Papers

The co.LAB Generic Framework for Collaborative Design of Serious Games: Development Study (e28674)
Dominique Jaccard, Laurent Suppan, Eric Sanchez, Audrey Huguenin, Maxence Laurent. .......................................................... 20

A Therapeutic Game for Sexually Abused Children and Adolescents (Vil Du?!): Exploratory Mixed Methods Evaluation (e26062)
Joyce Endendijk, Henny Tichelaar, Menno Deen, Maja Dekovi ................................................................. 37

The Effects of Exergaming on Sensory Reweighting and Mediolateral Stability of Women Aged Over 60: Usability Study (e27884)
Mariann Sápi, Anna Fehér-Kiss, Krisztina Csernák, Andrea Domján, Sándor Pintér ................................................................. 50

Development of a Search Task Using Immersive Virtual Reality: Proof-of-Concept Study (e29182)
Samuel Knobel, Brigitte Kaufmann, Stephan Gerber, Prabitha Urvyler, Dario Cazzoli, René Müri, Tobias Ne, Thomas Nyffeler. ................................................. 62

An Interactive Computer Game for Improving Selective Voluntary Motor Control in Children With Upper Motor Neuron Lesions: Development and Preliminary Feasibility Study (e26028)
Annina Fahr, Andrina Kláy, Jeffrey Keller, Hubertus van Hedel ................................................................. 76

Use of a Virtual Reality Simulator for Tendon Repair Training: Randomized Controlled Trial (e27544)
Tsz-Ngai Mok, Junyuan Chen, Jinhua Pan, Wai-Kit Ming, Qiyu He, Tat-Hang Sin, Jialin Deng, Jieruo Li, Zhengang Zha. ................................................. 107

Medical Students’ Perceptions of Play and Learning: Qualitative Study With Focus Groups and Thematic Analysis (e25637)
A Van Gaalen, A Jaarsma, J Georgiadis. ................................................................. 121

Using Visual Guides to Reduce Virtual Reality Sickness in First-Person Shooter Games: Correlation Analysis (e18020)
Kwang-Ho Seok, YeolHo Kim, Wookho Son, Yoon Kim. ................................................................. 148
Reviews

Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review (e29080)
Sandra Bartelt, Lucia Lanfermann, Till Bärnighausen, Florian Neuhaan, Claudia Beiersmann. ................................. 90

Association of Extensive Video Gaming and Cognitive Function Changes in Brain-Imaging Studies of Pro Gamers and Individuals With Gaming Disorder: Systematic Literature Review (e25793)
Eunhye Choi, Suk-Ho Shin, Jeh-Kwang Ryu, Kyu-In Jung, Yerin Hyun, Jiyea Kim, Min-Hyeon Park. ............................ 136
Extended Reality for Enhanced Telehealth During and Beyond COVID-19: Viewpoint

Triton Ong1, PhD; Hattie Wilczewski1, BSc; Samantha R Paige1, PhD, MPH, CHES; Hiral Soni1, PhD, MBA, MSc; Brandon M Welch1,2, PhD, MS; Brian E Bunnell1,3, PhD

1Doxy.me, LLC, Rochester, NY, United States
2Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, United States
3Department of Psychiatry, University of South Florida, Tampa, FL, United States

Abstract

The COVID-19 pandemic caused widespread challenges and revealed vulnerabilities across global health care systems. In response, many health care providers turned to telehealth solutions, which have been widely embraced and are likely to become standard for modern care. Immersive extended reality (XR) technologies have the potential to enhance telehealth with greater acceptability, engagement, and presence. However, numerous technical, logistic, and clinical barriers remain to the incorporation of XR technology into telehealth practice. COVID-19 may accelerate the union of XR and telehealth as researchers explore novel solutions to close social distances. In this viewpoint, we highlight research demonstrations of XR telehealth during the COVID-19 pandemic and discuss future directions to make XR the next evolution of remote health care.

(JMIR Serious Games 2021;9(3):e26520) doi:10.2196/26520

KEYWORDS
extended reality; virtual reality; augmented reality; mixed reality; telehealth; telemedicine; COVID-19; telepresence

Introduction

In-person health care became limited during the COVID-19 pandemic. Social distancing, travel restrictions, and lockdowns forced many to perform and receive health care remotely via telecommunication (ie, telehealth) [1]. Telehealth emerged as a widely effective and accepted solution to support continuity of care throughout the ongoing pandemic [2-4]. Consensus among providers, patients, and policymakers indicates that the shift to telehealth will likely continue even after the pandemic ends [5-7]. To maintain current uptake and support delivery of the best possible care in the future, telehealth needs to supplement and transcend traditional models of in-person care.

As the health care industry adapts to telehealth, aspects of in-person treatment must be optimized for remote care (eg, conversational flow, physical evaluation, therapeutic presence). Some patients report reluctance to self-advocate during typical telehealth sessions because of poor eye contact and audio interference if more than one person speaks at a time [8]. Providers may find it difficult to build rapport and express empathy with patients via telehealth due to limited visibility of body language and the unavailability of physical presence [9]. Both providers and patients can encounter distractions as telehealth is conducted from within their homes [10]. Although telehealth appears set to become the new norm, novel approaches are needed to optimize patient outcomes by restoring critical aspects of in-person health care for remote formats and expanding clinical options to deliver health services at a distance.

Technologies that evoke presence—the perception, feeling, and interaction with simulations as if they were real [11]—can meaningfully impact the practice and outcomes of telehealth. Immersion into fully simulated virtual reality (VR) [12], simulated objects or overlays superimposed onto users’ real sight in augmented reality (AR) [13], or direct interaction between simulations and the physical world in mixed reality (MR) [14] each afford new ways to support, extend, and enhance health care practice in the shift to remote delivery. These VR,
AR, and MR technologies—collectively referred to as extended reality (XR)—have been demonstrated for inpatient and outpatient psychiatric, medical, and rehabilitative applications with equal or greater effectiveness than their non-XR standard treatments [13,15-17]. However, research on XR as an extension of remote health care is comparatively recent and has yet to be synthesized.

The need to explore XR for telehealth has never been greater than in the fallout of COVID-19. Postponement of regular and preventive medical care, the psychological and developmental impacts of the extended pandemic, and escalating reports of provider burnout are harrowing signs on the health care horizon [18-26]. Research on the combination of XR and telehealth embodies recent events, addresses current limitations of conventional telehealth, and paves the path for the health care of tomorrow. We believe that XR telehealth research performed since the onset of COVID-19 will set the tone for research and innovations in the coming years. In this viewpoint, we provide a narrative review of current XR telehealth research and highlight future directions to address remaining barriers.

**XR for Telehealth Before COVID-19**

The potential for reality-altering technologies in remote health care has been heralded since the earliest days of VR. Early VR required complex and costly computing hardware that kept the technology localized, for use by a single individual, and prohibitively expensive until the proliferation of the internet and affordable computing hardware [27,28]. Once the internet became widely available in the mid-1990s, surgical applications of VR and AR expanded to include multiuser supervision by remote experts, and detailed simulations to plan and practice surgical procedures [29-31]. Growing interest in interactive therapy drove MR technologies for at-home telerehabilitation, sometimes using off-the-shelf video game console hardware [32,33]. XR gradually matured with consumer-oriented hardware and software packages that led to interdisciplinary developments such as online VR for psychiatric treatments and sophisticated medical training simulations [34,35]. However, costs and technical complexity remained barriers to the wide deployment of XR for telehealth [36].

Trends in home computing and entertainment made user-friendly, robust, and polished XR equipment available for personal ownership in the 2010s, which facilitated the rapid growth in XR telehealth research and development [37]. Since then, XR telehealth has expanded to a wide variety of remote health care applications in AR telesurgery and telesupervision, MR training and simulations, VR telerehabilitation, and telemonitoring of fully remote interventions and specialized medical equipment with VR/AR [13,38,39]. Similar growth occurred in consumer markets with reported use of entertainment VR apps and video games for therapeutic purposes such as mental wellness, identity exploration, healthy aging, and social anxiety [40-42].

Advances in internet infrastructure, computer technology, and portable consumer entertainment gradually decreased costs and increased consumer interest in XR devices. Increasing commercial availability of these devices and continued innovative research placed XR telehealth on a mainstream trajectory in 2019 [43].

**XR for Telehealth During COVID-19**

**Resource Constraints and the Rise of Telehealth**

In early 2020, the World Health Organization declared the COVID-19 pandemic and promoted social distancing to limit the spread [44]. Soon after, mass consumption of personal protective equipment (PPE) such as N95 face masks, medical-grade sanitizers, and disposable gloves led to global supply shortages for both health care workers and the general population [45]. The health care industry’s primary response to COVID-19 and PPE shortages was a rapid and widespread shift to remote services [4,46]. By June 2020, in-person health care visits were down by 30% while telehealth visits increased by up to 2013% [47]. This shift was even more pronounced for mental health services, which saw 70% fewer patients in person while telehealth sessions increased by up to 6558% [48,49].

XR telehealth was reported to be an effective solution that enhanced safety and reduced PPE consumption in COVID-19 health care settings. Two frontline case studies showed how medical specialists, unable to travel during quarantine, used AR to provide remote consultation and ventilator management for COVID-positive patients [50,51]. Intelligent AR information displays enhanced the workflows of frontline hospital staff to increase clinical efficiency, improve remote team communications, reduce COVID exposure by 51.5%, and decrease PPE consumption by 83% [52]. VR simulation systems for trauma and emergency medicine offered effective and high-fidelity alternatives to traditional supervision with less need to consume PPE for on-site medical training [53]. XR telehealth alternatives were also demonstrated for patient therapy. A VR group-singing intervention as respiratory therapy for spinal cord injury was found to be feasible, acceptable, enjoyable, and reported as less socially inhibited than the in-person prospects of the same intervention [54]. General population users in another feasibility study favored telemental health in VR over the standard webcam format [55]. Multiuser VR in analog telehealth conditions was found to be an ideal environment to conduct evidence-based cognitive behavioral therapy (CBT) in a space that felt immersive, expressive, private, anonymous, and free from judgment [56]. Overall, this published research shows that XR technologies complemented telehealth solutions to support frontline health care workers and maintain social distancing for critical evidence-based treatments during COVID-19.

**Access to Medical and Mental Health Care**

In addition to social distancing, local and state governments imposed travel restrictions to limit the transmission of COVID-19 and reduce strain on health care systems [57]. However, these restrictions entailed collateral costs to the public’s health. Reduced public transportation disproportionately impacted lower-income and ethnic groups in urban areas, and further destabilized access to health care in rural regions with few local specialists [58-60]. Extended pandemic conditions also increased the global risk of psychological distress, impacted
people’s daily habits and life plans, and subjected vulnerable populations to prolonged social isolation [57,61].

Barriers to health care access intensified during the pandemic and XR telehealth emerged as a responsive option. XR telehealth had a particularly significant impact to increase the immediacy of care and access among medically and geographically isolated populations who required continual rehabilitation services [62,63]. In addition to providing immersive and accessible care, XR telehealth connected people in virtual spaces to combat social isolation and maintain health-promoting social relationships over distances [64]. For example, location-based AR video games provided a protective effect for social, physical, and mental health during the pandemic [65-69]. Experts also promoted XR telehealth developments as a potential solution to address the downstream developmental impacts of the prolonged pandemic upon children and adolescents who eagerly take to new technologies [25,70].

**Burnout and Contagion Exposure Among Health Care Workers**

The psychological distress of COVID-19 was particularly burdensome for health care workers. Sudden increases in workload, overcrowding, medical supply shortages, exposure to the virus, and the suffering of patients led to extreme physical and emotional strain upon frontline and hospital staff [22,71-73]. Health care workers were seven times more likely to exhibit severe COVID-19 symptoms than other workers due to their frequent and extreme exposure to contagion environments [74]. It is estimated that more than 3500 US frontline health care workers have died from COVID-19 contracted during their health care service [75]. Burnout among health care workers proved to be another contagion that spread within hospital wards with cascading staff turnover, compassion fatigue, and secondary traumatic stress [76,77]. Experts anticipate severe downstream impacts upon health care workers, and urge for evidence-based therapeutic interventions responsive to the impacts of COVID-19 and for methods to reduce health care workers’ exposure to the virus [78-81].

CBT for mindfulness is known to alleviate burnout and improve overall mental wellness among health care workers [71,82-84]. XR virtual visits have emerged as a promising technology for stress reduction using CBT and other evidence-based telenental health approaches [85,86]. With the advent of telehealth, VR and AR for COVID-related stress and trauma therapies have been promoted for distribution among frontline health care workers [51,87,88]. Although these studies are ongoing, XR telehealth for posttraumatic stress among survivors of COVID-19 has shown promising effects [89,90], and it is reasonable to expect these effects to generalize to providers from these same environments and traumatic experiences [91]. In addition to targeting burnout among health care workers, XR has been used to innovate health care workflows with remote, intelligent, and burnout-reducing solutions. A preliminary application of AR telesurgical consultations allowed remote specialists to provide real-time expertise for COVID-positive patients without travel or exposure risks, paving a way for large-scale future implementation [50]. Interconnected AR systems improved infection control, increased access to specialist remote supervision, reduced time spent in contagion environments, and enhanced clinical workflows in frontline health care environments [51,52].

**Economic and Professional Pressures on Health Care Providers**

The accumulated effects of social distancing, chronic resource shortages, travel restrictions, hospital surges, and pandemic stress created instability for current and future health care providers. By August 2020, more than 16,000 private practices had permanently closed with 41% of their peers facing the same fate with unsustainable loss of staff, patients, and income [20]. The subsequent low viability of for-profit clinics exacerbated concerns with reduced health care support [92]. This new fragility in health care networks was particularly straining for already underserved rural regions and ethnic populations [93,94]. Prolonged pandemic conditions further inhibited traditional pathways to hands-on health care experience and clinical supervision, which delayed professional development of the next generation of health care providers [95-97]. Governments, hospitals, and health care providers rallied to support public health during COVID-19, but extended pandemic conditions created a clear need for remote health solutions to sustain health care practice, improve access to health care, and provide quality health care education.

Many aspects of health care and education were ready for XR and telehealth before the pandemic but remained underutilized due to equipment costs, unresponsive legislation, and limited health insurance coverage [98,99]. COVID-19 produced the conditions necessary to accelerate change, and now provides ample opportunity for those who embrace telehealth and complementary technologies. XR telehealth allowed providers to deliver services into patients’ own homes and naturalistic environments, which has long been a limitation of traditional clinical treatment [100]. Low-cost, off-the-shelf hardware and royalty-free software for therapy and rehabilitation made XR telehealth an economically feasible solution [101,102]. XR telehealth training and education also rose in response to COVID-19. The realistic, interactive capabilities of XR were broadly promoted as a solution to reach and educate patients and trainees over distances [103]. Simulations in VR and AR were common, safe, and repeatable alternatives to risky on-site in-person medical student training [104-106]. A cohort of medical-surgical students set to graduate during COVID-19 rated VR training as realistic for 77% of clinical assessments, 81% of treatment options, and 94% of diagnostics. After exposure to the virtual training, 84% of the cohort reported interest in the future use of VR for medical training and 90% overall satisfaction with virtual learning [107]. The rise of telehealth provided options for health care providers to sustain their practice when in-person visits became unfeasible. As part of telehealth, XR also proved to be a critical solution to provide health services and education amid pandemic conditions.
XR Telehealth After COVID-19

General Prospects

The impact of COVID-19 on the health care industry was sudden, severe, and broad. Longitudinal data are necessary to evaluate XR telehealth as an alternative to traditional in-person treatment and training. Nevertheless, XR telehealth served as a critical solution to the emergent conditions of COVID-19, maintenance of health care systems, and preparation of future providers. Telehealth is likely to become a staple of health care practice, as the majority of patients and more than 90% of providers intend to continue remote care beyond the resolution of COVID-19 [108-113]. Telehealth on its own is broadly effective and accepted but still leaves some patients and providers dissatisfied with their interactions with providers, specifically in their ability to feel present and build therapeutic relationships [114-116]. This lack of communicative nuance creates a vagueness in non-XR telehealth interactions that can be interpreted as awkward or even malicious [117]. Continued research and development of XR for telehealth can address some of these barriers to enhance therapeutic relationships, expand clinician capacity, and empower patients toward optimal health outcomes.

XR Can Facilitate Telepresence to Strengthen Teletherapeutic Relationships

Therapeutic alliance is one of the best predictors of treatment success and health outcomes [118-120]. Therapeutic alliance is broadly defined by the relationship between the provider and patient, fostered through mutual agreement of clinical goals and the strategies to achieve those goals [121]. Non-XR telehealth is effective, accepted, and sustainable, but can make it difficult to replicate the communicative nuances and rapport building of in-person health care [122,123]. Health care providers who seem rigid, distant, or distracted (ie, typical attributes of non-XR telehealth [124,125]) produce poorer therapeutic alliances [126], which lead to higher chances of dropout, dissatisfaction, and negative health outcomes [127-129]. As a result, a small but important minority of providers believe that their patients do not enjoy telehealth as much as in-person care [116,130].

Preliminary evidence shows that XR can enhance remote interactions to strengthen therapist-patient relationships. Patients who received interactive CBT using VR avatars reported feeling less judged by their physical appearance, that the VR space was somewhere they could be honest and private with their therapist, and that the interaction felt more casual than an in-person clinic visit [56]. Likewise, physicians reported building rapid trust with their patients while jointly viewing patient body scans in VR and AR [131]. VR has also been promoted over non-VR alternatives for benefits such as more comfortable treatment, higher engagement, greater satisfaction, more consistent practice, greater skill transfer, and facilitation of nonverbal communication that improves therapist-client contact [132].

XR facilitates presence, when one perceives that the virtual environment is real [11]; embodiment, when one perceives a virtual body as one’s own real body [16]; and telepresence, when one perceives that they are inhabiting another place with virtual others [133]. Each of these aspects can aid in the establishment, improvement, and maintenance of therapeutic alliances in telehealth. Miloff and colleagues [134] developed an automated AR hologram embedded in VR exposure therapy and demonstrated that therapeutic alliance measures generalized to the virtual therapist. Although patients reported positive perceptions of this audio-only VR therapist, visually and behaviorally realistic VR therapists have been shown to evoke greater perceived presence [135]. Realistic XR avatars and XR interactions tend to evoke stronger physical and emotional closeness and greater confidence in the credibility of the therapist [135-137], which are key factors that influence an alliance with a virtual health care provider. However, XR telehealth is a nascent field, and more research is needed to understand how the two technologies interact to cultivate therapeutic alliances and impact health outcomes.

XR With Telehealth Can Expand the Reach of Clinicians and Researchers

XR technologies were used frequently for health care education and training prior to COVID-19 [138-140], and this practice is expected to become increasingly common as traditional on-site medical education remains limited under the pandemic [86,104]. Simulation training in XR provides highly realistic experiences that deeply immerse learners in clinically realistic scenarios to facilitate skill acquisition and skill transfer for real application [141]. XR simulations can also provide repeatable exposure to important but improbable clinic scenarios, to prepare for states of emergency, and to access otherwise impossible views of medical procedure [103,142]. Remote XR simulations were used to facilitate skill development, prevent contagion spread, and rapidly disseminate COVID-relevant medical education during the pandemic [52,104,105]. Although the relationship between XR simulation training and clinical outcomes remains unclear [138,143-145], further exploration of remote XR training can help health care workers acquire, develop, and maintain cutting-edge skills with limited access to facilities or clinical populations. XR simulations stand to benefit from technologies to enhance realism and transfer of skills such as remote supervision, haptic feedback, anatomical replicas responsive to MR, and artificial intelligence to provide the most flexible and clinically beneficial education of future health care providers [146].

Health care is complex, fluctuating, and high-stakes work that often necessitates coordination of schedules, tasks, and information between multiple providers and teams. Unfortunately, hospitals are notoriously inefficient and error-prone due to a historic lack of human factors considerations in workflows, communications, equipment, user interfaces, and physical environments [147-150]. XR can play an important role in connected collaborative health care. Telesurgery with AR is a prominent example of how the marriage of XR and telehealth can improve health care work environments with surgeons receiving notes from expert consultants directly on their real-time view of the patient, seeing a proctor’s hands directing incisions, and delivering the expertise of medical specialists to regions with few or no local specialists [13,14,50,151]. The benefits of XR telesurgery have recently been demonstrated in nonsurgical medical teams for live...
distanced collaboration for inpatient unit care and coordination [51,152,153]. XR technologies enable immersive learning environments and pervasive sensor-display interfaces in the field. Telehealth enables real-time remote specialist consultation and expert supervision. The combination of XR and telehealth represents a system of potential force multipliers that can support, improve, and extend the capabilities of health care practitioners.

XR telehealth has increased patient access to health care, but this relationship has rich bidirectional potential to explore; clinicians and researchers can use XR telehealth to gain better access to patients and participants. Persky [154] described how controlled programmatic XR experiences could merge with remote clinical trials to minimize researcher and participant travel burdens; streamline and automate data collection; and critically improve engagement, retention, and procedural integrity. The recent popularity of consumer XR entertainment devices such as Facebook Oculus Quest 2, Sony PlayStation VR, and smartphone-based Google Cardboard can function as recruitment, enrollment, and data collection solutions with access to participants in their naturalistic settings. The use of fictionalized XR avatars to represent researcher and participant bodies can provide complete control over social manipulations and single- and double-blind study logistics [155]. Complicated data displays, technical instructions, and study processes such as informed consent are also easy for participants and researchers to visualize and interactively explore together in XR [156-158]. It will be critical to study XR for telehealth as a solution to extend historically localized research practices with mobility deployment to the general public and outreach for remote, underserved populations [159].

**XR Can Empower Patients to Seek Health Care and Improve Outcomes**

Patients are empowered when they are treated as active collaborators in understanding and making decisions about their health care, rather than as passive subjects merely compliant with “doctors’ orders” [160]. Telehealth has already improved patients’ access to health care by removing geographical barriers (eg, travel costs and arrangements); however, remaining social and behavioral barriers to patient empowerment may be addressed with humanizing and engaging XR technologies.

There are widespread cultural stigmas that inhibit health-seeking behavior [161,162]. For example, men tend to avoid medical and mental health care to the point of early death and preventable decline in quality of life [163-165]. Other stigmas of diagnosis, gender, sexuality, ethnicity, body image, criminal history, and others similarly compromise health care utilization and outcomes [166-170]. Telehealth provides a beneficial distance that can make patients with stigmatized conditions feel confident and comfortable seeking services from their own homes [171-173]. XR can further enhance telehealth to include temporary, therapeutic distance from stigmatized bodies or identities. Avatars are 3D computer-generated models used in virtual environments to represent agents (eg, patients, providers, computer-controlled characters) [174]. The simulated nature of XR avatars makes them uniquely flexible for personalized health care approaches and interventions. One’s avatar can resemble their physical likeness in XR therapies faithful to what they would be like during an in-person health care visit [142]. Alternatively, patients may build a fictionalized avatar to provide a more comfortable degree of anonymity, extend embodiment-oriented therapies, and as a clinical enhancement for telehealth providers’ web-side manner [175,176]. Matsangidou and colleagues [56] recently demonstrated the many benefits of fictionalized avatars in VR treatment for both therapists and patients. Therapists tasked patients to build their own avatars, which provided useful clinical insights as to how the patient viewed themselves (ie, avatar appearance compared to real body). Patients attributed a wide variety of subjective benefits to the use of VR avatars, the most important of which were corroborated by therapists who noticed that avatars occasioned remarkable patient openness and trust. Interestingly, the therapists were also depicted with VR avatars in the form of simplistic cubes that were reported to enhance patients’ willingness to discuss difficult topics and engage in other mental health exercises. Telehealth allows patients to access health care with no need to travel, and XR can further enable access to care with no need for concern they will be judged. It will be important to explore avatars in XR telehealth as a solution for stigmas of visible medical conditions (eg, skin disease and burns) [177,178], criminal history or potential (eg, prevention of offense related to sexual preference) [179,180], and provide unprecedented opportunities for mental health and wellness [86,181,182].

The presence of and embodiment facilitated by avatars in responsive XR environments can result in simulated experiences that feel more real than physical reality (ie, hyperpresence) [183]. This hyperpresence may have tremendous clinical potential for remote health care. For example, patient motivation tends to be low for at-home rehabilitation due to the gap between unpleasant exercise and long-term health outcomes [184-186]. XR can boost the salience of physical rehabilitation with fictional but proactive feedback, similar to those in modern entertainment video games. Exaggerated body tracking in XR showed participants’ virtual bodies as stretching further and running faster than their real bodies, which significantly improved performance, enjoyment, and motivation for unsupervised rehabilitation-oriented exercises [187,188]. Hyperpresence in XR can also enable treatment contexts that in-person care and telehealth cannot. Traditional mental health treatments for internal stimuli (eg, emotional states or auditory hallucinations) rely on guided imagination that can alternatively be visualized and engaged with directly in XR [189,190]. Further, counterfactual hyperpresence can make health care more approachable for the shy or therapies that can be socially awkward. Group singing is an effective and cost-efficient intervention to promote respiratory health, but participants report lack of confidence when singing in front of others [191]. The same group-singing intervention in remote VR made participants feel socially uninhibited owing to their manifestation as anonymous and nonhuman VR avatars [54]. Hyperpresence is an emerging concept that merits investigation as a potential path for XR to enhance telehealth patient engagement, retention, and comfort. XR is currently used to alter patients’ sense of where they are and what they are doing but can also enhance patients’ sense of who they are in the future of telehealth [192].
Remaining Barriers and Steps Forward

Telehealth revolutionized health care to meet patient needs at a distance. Although the majority of telehealth adoption was due to the pandemic, it is clear that telehealth will continue to expand beyond the resolution of COVID-19. We believe that XR is the next evolution for remote care built upon decades of foundational research and innovative demonstrations published in response to COVID-19. Toward that future, however, the barriers to XR are broadly similar to those of telehealth. Both technologies involve costly investments in equipment and training, can be abandoned after investment because of poor usability, and rely on broadband internet access that limits access on the basis of socioeconomic status and geographic location [151,193-195]. To realize the teleportant and empathetic future of XR telehealth, key barriers to mainstream adoption of both technologies must be addressed.

XR technologies involve complex electronic sensors, displays, and networks, which make costs a constant prohibitive factor. In 2000, a clinically sufficient VR headset with necessary computers and proprietary software could cost approximately US $17,000 (adjusted for 2021 inflation) [34]. However, high-end VR setups can be purchased today for use in one’s own home for about US $3000 total with a growing variety of free-to-use and open-source software [196]. Low-end XR (eg, Google Cardboard VR, Holokit AR) costs as little as US $15-$50, involves the use of smartphones many people already own, and has been applied with clinically significant treatment success [102,197]. Costs are anticipated to continue decreasing as consumer XR hardware becomes more established [198]. Concerns over costs can be further addressed with formal cost-benefit analyses of comparative treatment costs and impacts on quality of life.

Commercially produced XR hardware removes many barriers for health care providers interested in the technology, but this reliance on proprietary devices and software can be a double-edged sword for telehealth. Privacy is an ongoing concern with increasingly interconnected health care technologies [199]. This issue is particularly tricky with XR, as few devices exist on consumer markets that are compliant with regulatory health policies and the constantly evolving ways people use XR [40,200,201]. Although XR telehealth can feel private, the reality is that many XR devices and applications needlessly collect identifiable information and share user data with third parties. Certain XR devices such as Oculus Quest 2 are inoperable after data logging, and the manufacturer explicitly discourages its use with protected health information [202]. The recent rapid uptake of telehealth and XR continues to highlight the need for privacy and policy focused on health care end users [203]. Health care researchers, clinicians, consumers, and XR developers will need to organize and openly communicate to promote transparent and responsive privacy measures.

As a relatively new area of research, XR telehealth has a growing number of ethical concerns to address. First, the contexts in which XR telehealth are or are not appropriate have not been well established. XR telehealth may not benefit all health conditions or contexts equally. Treatment of high-risk conditions (eg, suicidality) still necessitates in-person responsiveness, while XR remains inaccessible to some (eg, those with chronic neurological conditions) or is unlikely to help others (eg, those with acute delirium). Second, practice competency is unclear with emerging telehealth and XR technologies. The broad foundational principles of competency are expected to be maintained as the public settles into widespread utilization of telehealth [204,205]. However, few telehealth practitioners are also experienced in software and hardware development. This current reality leaves most decisions about XR telehealth features and functions out of the hands of health care providers, making interdisciplinary collaborations a vital need into the future [206]. Third, the unique uses of XR for telehealth carry equal potential for misuse. Immersion, presence, copresence, and embodiment can facilitate remote health care, but it is not yet known how to best utilize these components or when one component should be emphasized over others. Modification of experiential states needs to be transparent and responsible in proportion to the potential risks [207]. Organizations for health care research and practice will need to establish and discuss ethical guidelines for XR telehealth. This is particularly important in light of reports that some are using nonmedical technology resources (eg, apps, games, websites) in lieu of qualified health care and the growing availability of self-help resources with little or no medical oversight [208,209].

In current and coming years, XR content offerings may be the greatest barrier for deployment via telehealth. There are currently about 60 million regular VR users and 91 million regular users of AR in the United States [210]. Major technology companies aim to make XR ubiquitous in the near future, which may make telehealth a more appealing use case for XR [211]. Although there is a growing variety of commercial XR telehealth options, the vast majority of XR consumption is for entertainment and industrial application [212]. XR for telemental health is promising, but uptake has been slow due to lack of usability or easy integration into existing clinical workflows [14,195,213]. It will be vital for researchers, clinicians, and developers to collaboratively assure that telehealth is a priority market in the design of XR hardware and software [214]. Furthermore, there are currently few sources of reputable, evidence-based, comprehensive information for telehealth providers to learn about and make treatment decisions with XR. Scholarly, clinical, and patient advocacy organizations should formally curate XR hardware and software to help navigate emerging treatment options for telehealth providers and patients.

The research literature on XR for telehealth is new, vast, and accelerating. The wide variety of XR hardware and software, study designs and populations, metrics and outcome variables, and vocabularies can be difficult to navigate and synthesize. The parameters of what constitutes VR, AR, and MR are still being explored, leaving overlap and obscurity in terms for practice and literature searches [215,216]. A consistent finding in XR narrative, systematic, and meta-analytic reviews is the variability in approaches that prevent formal comparison between studies [14,143,217]. Toward that end, Birckhead and colleagues [218] have provided recommendations to guide...
progressive and programmatic lines of XR research. Other good practices in this field of research include pretraining to orient participants to XR and minimize error, repeated exposure to detect and control for novelty confounds, and on-demand technical support during XR studies [54,219,220]. Failures with XR telehealth are equally important to publish as successes to accumulate details relevant to application and sustainability. Consistency of language, replicability, long-term effectiveness, and best practices for implementing XR telehealth must be disseminated to establish a comprehensive and conceptually systematic literature.

**Conclusion**

Studies published during the COVID-19 pandemic showed that XR for telehealth helped health care providers stay safe during treatment of COVID-19 patients, improved the way health care was delivered to patients remotely, helped sustain a healthy frontline health care workforce, and supported the professional development of current and future providers. Toward the future of telehealth, we argue that XR can enhance interactive nuances and treatment options for telehealth patients, function as a force multiplier for health researchers and clinicians, and provide new options for at-risk patient populations. Cost, privacy, ethical practice, actionable practice guidelines, and improvements to research approaches must be addressed to fully realize the potential benefits of XR for telehealth. Despite these barriers, XR technologies have unique potential to enhance, extend, and expand the future of telehealth and beyond the COVID-19 pandemic.

**Conflicts of Interest**

BW is a shareholder and all other authors are employees of Doxy.me, LLC, a commercial telemedicine company. The authors declare no other conflicts of interest.

**References**


https://games.jmir.org/2021/3/e26520


148. Ong et al. JMIR Serious Games 2021 | vol. 9 | iss. 3 | e26520 | p.15https://games.jmir.org/2021/3/e26520 (page number not for citation purposes)


Abbreviations

AR: augmented reality
CBT: cognitive behavioral therapy
The co.LAB Generic Framework for Collaborative Design of Serious Games: Development Study

Dominique Jaccard¹, MSc; Laurent Suppan², MD; Eric Sanchez³, PhD; Audrey Huguenin¹, BSc; Maxence Laurent¹, BSc

¹Media Engineering Institute, University of Applied Sciences Western Switzerland, Yverdon, Switzerland
²Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland
³Technologies de Formation et Apprentissage, Educational Technologies, Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland

Corresponding Author:
Dominique Jaccard, MSc
Media Engineering Institute
University of Applied Sciences Western Switzerland
Av. Sports 20
Yverdon, 1400
Switzerland
Phone: 41 244592638
Email: dominique.jaccard@heig-vd.ch

Abstract

Background: Serious games are increasingly used at all levels of education. However, research shows that serious games do not always fulfill all the targeted pedagogical objectives. Designing efficient and engaging serious games is a difficult and multidisciplinary process that requires a collaborative approach. Many design frameworks have been described, most of which are dedicated to the development of specific types of serious games and take the collaborative dimension into account only to a limited extent.

Objective: Our aim was to create a generic serious game design framework that could be adapted to all kinds of serious games and implemented in a collaborative web platform.

Methods: We combined the results of a literature review with our experience in serious game design and development to determine the basic building blocks of a collaborative design framework. We then organized these building blocks into categories and determined the features that a generic design framework should include. Finally, based on the paradigm of complex systems and systemic modelling, we created the co.LAB generic design framework and specifications to allow its implementation in a collaborative web platform.

Results: Based on a total of 10 existing design methodologies or frameworks, 23 building blocks were identified and represent the foundation of the co.LAB framework. These blocks were organized into 5 categories: “context and objectives,” “game design,” “mechanics,” “learning design,” and “assessment.” The arrangement by categories provides a structure that can be visualized in multiple and complementary ways. The classical view links game and learning design while other views offer project, systemic, and process visualizations. For the implementation of the co.LAB framework in a web platform, we propose to convert the building blocks into “cards.” Each card would constitute a collaborative working space for the design of the corresponding block. To make the framework adaptive, cards could be added, adapted, or removed according to the kind of serious game intended. Enhancing the visualization of relationships between cards should support a systemic implementation of the framework.

Conclusions: By offering a structured view of the fundamental design elements required to create serious games, the co.LAB framework can facilitate the design and development of such games by virtue of a collaborative, adaptive, and systemic approach. The different visualizations of the building blocks should allow for a shared understanding and a consistent approach throughout the design and development process. The implementation of the co.LAB framework in a collaborative web platform should now be performed and its actual usability and effectiveness tested.

(JMIR Serious Games 2021;9(3):e28674) doi:10.2196/28674

https://games.jmir.org/2021/3/e28674
KEYWORDS
serious game; educational game; simulation game; design; design framework; methodology; collaborative design; collaborative web platform

Introduction

Background
The term “serious games” is used in many different meanings, and there is still no strict agreement on what it exactly encompasses. In broad definitions, such as Zyda’s [1] or in Michael and Chen [2], serious games may be used for achieving any kind of nonentertainment objectives, including education, health, public policy, or communication. In other definitions, such as those by Abt [3] or Loh et al [4], serious games are restricted to educational and training objectives. In this article, we use the term serious games in the sense of the definition by Loh et al [4], thus encompassing any kind of digital games or simulation created for educational or training purposes.

Serious Games for Educational Purposes
Serious games are effective tools to support learner-centered teaching practices [5-9], and interest in serious games has flourished at all levels of education. The COVID-19 crisis has accelerated the digitalization of education, and the use of digital educational resources such as serious games is expected to increase even further in the coming years [10]. However, research shows that serious games do not always fulfill all the targeted pedagogical objectives [11-13]. To reach its objectives, a serious game needs to successfully integrate gaming and learning aspects and be accepted by the teachers who will use it. Collaborative design of serious games is recognized as a success factor for both this integration and acceptance.

Collaborative Design of Serious Games
The collaborative work of a multidisciplinary team, including game developers, teachers (or trainers), and educational scientists, is required to design and develop efficient serious games [13-16]. This collaboration is recognized as a significant factor in the pedagogical relevance of the resulting development [14,17].

Integration of gaming and learning aspects, and integration of serious games into an overall pedagogical scenario, has been recognized as a key success factor [13,18,19]. Thus, the collaboration within the development team must ensure that partial contributions of different specialists who possess complementary knowledge and expertise will result in a coherent solution integrating pedagogical and playful aspects.

Collaboration in serious game multidisciplinary development teams can however be difficult [20]. Difficulties arise from the intrinsic multidisciplinarity of serious game design and from the challenging balance between game and pedagogical elements. Communication and coordination problems resulting from differences in vocabulary, background, and expectations also arise during the design and development phases [8,16,20-22]. Thus, while mandatory, collaboration in serious game design can be difficult.

Collaborative Web Platforms
Collaboration during serious game design could be facilitated by the use of a collaborative web platform. This platform should support typical collaborative dimensions such as mutual understanding, information pooling, communication management, group problem solving, reaching consensus, and task or time management [23,24].

For a multidisciplinary team, often geographically dispersed, a web platform may offer collaborative functionalities such as shared workspaces with a global design overview (mutual understanding), up-to-date documents (information pooling), discussion threads (communication management), voting systems (reaching consensus), and project management (task and time management). De Troyer [20] emphasized the need for this type of tool to support and stimulate the collaborative development of serious games. Regular tools usually employed for software development are not suitable for the development of serious games [20]. The main reason is that such tools were designed for software developers, not for multidisciplinary development teams incorporating noncomputer scientists [20]. Another reason is that existing platforms do not provide the necessary overview and integration needed for the development of both the serious game and its pedagogical integration [20]. Thus, collaborative web platforms may support collaborative design of serious games, but existing platforms are not suitable for that purpose.

Current Serious Game Design Frameworks
A collaborative platform dedicated to serious game design and development should be based on a design framework. Serious game design frameworks and methodologies are intended to provide development teams with design foundations and guidelines that support collaboration in the development of an integrated solution [8,16,25,26]. The framework implemented in a collaborative platform must allow, on the one hand, the development of the greatest number of different types of serious games and, on the other hand, be compatible and facilitate the implementation of collaborative features.

Most existing frameworks are dedicated to the design and development of specific types of educational games [16,27]. Thus, when beginning a new serious game project, the design team must choose a specific design framework and get used to it. Although not always straightforward, the task of choosing such a framework is usually rather easily feasible. However, using a framework that is too specific and not adaptable enough as the basis for a collaborative platform would force design teams to adapt to the framework, which could cause major problems. To achieve the intended serious game, the framework should be adapted to the project, rather than having to adapt the projected game to the framework.

A collaborative framework should provide the design teams with an overview of all design elements. Some existing frameworks give a broad overview of 3 or 4 categories to be
considered (such as “Play, Pedagogy and Fidelity” in [8] or “Context, Pedagogy, Representation and Learner” in [28]) but do not provide a detailed view of specific design elements of each category. Few frameworks give a more detailed list of game and learning design elements, but do not offer a categorization and structured view that may enable understanding the role of the different experts and the link between elements. Most frameworks do not include the design of the pedagogical scenario into which the serious game will be implemented.

A collaborative framework should also provide a project management perspective. Project management during serious game design and development is challenging because of the difficulty of managing multidisciplinary teams and of the need to adopt an iterative process [16,29]. Project management support (task, time, resource allocation) during serious game design and development is highlighted as needed [30], but not included in existing frameworks. Most existing frameworks include some specific guidelines but do not support collaborative work nor provide practical guidance describing how the different steps of the development process should be carried out.

Thus, existing serious game design frameworks were not designed with the goal to be implemented in a web platform and present some shortcomings in that perspective. If some of the needed qualities are found in each existing framework, none of them include the complete set of necessary qualities. An ideal collaborative serious game design platform should be based on a framework that provides an overall structure with content that can be customized by the end users. It should be an adaptive framework rather than “one framework to rule them all” and thus be considered more as a general methodology (a set of tools and guidelines) rather than as a traditional framework. It should support a collaborative and interprofessional approach, as well as the possibility to view the development process from different angles, offering both a broad overview of design categories and a detailed view of design elements.

The co.LAB Project

The co.LAB project, which is funded by the Swiss National Science Foundation, aims at improving efficiency and relevance in serious game design and development by supporting the collaboration between all members of the multidisciplinary development team. This goal should be achieved by developing a methodological framework associated with a collaborative web platform dedicated to the co-design, co-development, and co-evaluation of serious games.

Objectives

Our main objective was to create a methodological framework suitable for implementation into a collaborative web platform. This framework should enable the design and development of all kinds of serious games. Our secondary objective was to define guidelines and basic collaborative functionalities for the implementation of such a framework in a collaborative web platform.

Methods

To identify the elements of a generic serious game design framework, we combined the results of a literature review with the authors’ experience in serious game development.

The literature review was based on a search in Google Scholar using the terms “serious game design” and “educational game design.” We added articles that were known by the authors and did snowballing searches from references and citations of identified articles. As the aim of the literature review was to identify the essential building blocks needed to develop the basic structure of a generic framework, we considered a systematic review unnecessary; it might have yielded more results but probably not led to the inclusion of more building blocks. We then selected the most relevant frameworks based on their suitability for our purpose. We considered a framework to be relevant when it had actually been used for the development of at least one serious game and had been described in enough detail to allow replication. While citation numbers were used to select the most influential frameworks, those less frequently cited were not excluded if deemed interesting or innovative. We also added articles and books linked to more general concepts related to game-based learning or game design.

After selecting the relevant frameworks, we identified their main design elements.

Design elements were coded using the following steps:

1. Design elements explicitly presented in the selected frameworks were reproduced as is (verbatim).
2. Design elements appearing in texts or in graphics, but not explicitly presented, were added by creating a specific and relevant terminology. The terminology was proposed by one author (DJ) and confirmed by a second author (ML).
3. For frameworks dedicated to the design of serious games in broader fields than training, we assigned generic design elements to the corresponding specific element of the learning domain identified in steps 1 and 2 (eg, objectives were assigned to learning objectives). This was performed by one author (DJ) and confirmed by a second author (ML).
4. The design elements identified during steps 1 and 2 were then reviewed to regroup identical or duplicate items. That was done by one author (DJ) and confirmed by a second author (ML).

Results of steps 1 to 4 were then debated among all authors. Any disagreement was resolved by reaching consensus.

We grouped the identified design elements into categories according to their characteristics, the available literature regarding their use in serious game development, and the experience of the authors. The organization into categories was proposed by an author (DJ) and discussed among all authors and finally validated by the last author (ML).

We then converted design elements into building blocks, which represent the smallest units of the co.LAB framework. When similar or identical design elements were called differently, we decided upon a terminology that was then used to refer to the building blocks. For the sake of readability and consistency,
some elements were renamed or merged. The proposition to rename or merge elements was issued by one author (DJ), discussed among all authors, and finally validated by the last author (ML).

Finally, once the co.LAB generic serious game design framework had been established, we determined the specifications required for its implementation in a collaborative web platform. This was achieved by converting the building blocks into cards: Each building block corresponds to a card with its collaborative functionalities. Each card can either be used or discarded according to its relevance for the design and development of a particular serious game.

During the entire process, we also took into account the authors’ experience in serious game development. Three of the authors (DJ, ML, and AH) belong to the AlbaSim research lab (Media Engineering Institute, University of Applied Sciences of Western Switzerland), which has been developing serious games for more than a decade. This lab has conducted serious game projects from design to implementation in many different educational fields such as emergency triage at hospital, cardiac clinical evaluation, oncology care, project management, computer education, energy management, or crime scene investigation [31-36]. Another author (LS) has conducted serious game projects in fields like resuscitation and COVID-19 infection prevention and control at the Geneva University Hospitals [37,38].

Results

Literature Analysis

The review and analysis of existing design frameworks confirmed that no single model or theory can currently be applied to the design of every kind of serious game. This is best explained by the fact that serious games may be of such different types and used in such different learning paradigms and contexts that a unique design framework may not be possible. This is confirmed by Plass et al [27] in their analysis of theoretical foundations of game-based learning: “It does not appear likely that a single theory will emerge that can guide the design of games for learning in general.” This is also in line with our experience in serious game development.

Another observation was that not all serious game design frameworks have the same objectives. Some are more oriented toward serious game design elements [6,19,25], some more toward the design and development process [39], some cover both design elements and design process [16], and others are more concerned with theoretical foundations of game-based learning [8,27]. While most frameworks are conceived to guide the design of serious games for learning purposes, some of them are intended for the development of serious games for generic purposes (including learning, but not only) [16,40]. Most frameworks focus on game design but do not take into consideration the learning scenario in which the game should be integrated.

A synthesis of the design elements identified in the selected frameworks can be found in Table 1. In this table, frameworks are presented in descending order of number of citations. This may be a sign of the influence of the framework, but not necessarily of its intrinsic quality.

Some elements are present in most frameworks, such as play, interactivity, and feedback. Some interesting elements are however only present in a few frameworks, such as usage context (ie, the context in which the game will be used, designing the simulation model, or defining learner specifications).
Table 1. Design elements extracted from existing serious game design frameworks and methodologies, which are presented in descending order of number of citations (from left to right).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning objectives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning functions</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning foundations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pedagogical scenario</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning mechanics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning incentives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Learning assessment</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game goals and rules</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Structure and progression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game mechanics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decorum, sounds, aesthetics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Game assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Simulation model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fidelity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game incentives</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usage context</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Learners' specifications</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Play, interactivity, feedback</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immersion, motivation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Besides the identification of these design elements, another key component was brought out through our analysis of the literature. Indeed, most frameworks emphasize the importance of developing an integrated system that includes and links gaming and learning aspects [6,8,19,25,26]. The successful alignment between game mechanics and learning mechanics is thus highlighted as an essential feature for the success of serious games [25,27,42,43].

Finally, many frameworks also tackle the development process, which is presented as both multidisciplinary and iterative [16,27,44]. But none of the selected frameworks provide specific project management functionalities.

**Specifications for a Generic Serious Game Design Framework**

Since the aforementioned analysis confirmed that no currently available framework is truly exhaustive and as the diversity of serious game designs must be acknowledged, a single design method can hardly be developed. We therefore concluded that the generic framework we sought to create should be a methodology (ie, a set of methods and best practices at the disposal of a design and development team), rather than a method.

This generic framework should include elements common to most serious game designs but be adaptive to allow for specific designs. The framework should also foster collaboration between the various specialists involved in the design project. This could be achieved by allowing the design team to visualize the links between design elements throughout the game design process. This requires the development of a systemic framework.

**Adaptive Features**

Design specificities result from factors such as the type of serious game (which can include a simulation model, narratives, or neither of these features), usage context (ie, children or professionals, face-to-face, or online), and final objectives (eg, education, education and research, summative assessment).

Design specificities support the idea of developing an adaptive framework that would allow design teams to adapt the general model provided by the framework to their current project. This means that, depending on the design context and objectives, it should be possible to merge, add, or remove building blocks in the model.

The basic framework should therefore provide the main building blocks of most serious game designs (for example, pedagogical...
objectives, pedagogical scenario, or game design), but the design
team should then be able to adapt these building blocks to their
specific design project.

**Systemic Features**

Serious game designs include many elements, all of which are
interconnected. A successful design depends as much on the
quality of each element as on the relevance and adequacy of the
links between these elements [8,19,25,26,42]. It is the
relationships between the elements that give the final product
its coherence. We thus hypothesized that the paradigm of
complex systems and systemic modelling [45,46] would be a
suitable approach for the design of serious games as complex
systems.

This supports the idea of developing a framework that shall
encourage a systemic approach. The systemic features should
provide a vision of the serious game design as a whole made
up of interacting elements. This implies that the framework
should aim to provide both an overview of the building blocks
used for the design of the serious game and of the relationships
between them.

**The co.LAB Framework**

In accordance with our previous findings, we defined the co.LAB
framework as a methodological framework for serious game
design.

In order to bring together in a structured vision all the design
elements identified in Table 1, we defined 5 main categories:
(1) Context and objectives, (2) Game design, (3) Learning
design, (4) Mechanics, and (5) Assessment.

Each design element has been assigned to a category. For the
sake of readability and consistency, we rearranged the design
elements into 23 building blocks. Most design elements were
reproduced verbatim. To be consistent with the category to
which they were assigned, some elements were either renamed
(“structure and progression” became “game structure”) or
merged (fidelity and simulation model). Discussions between
authors led to the addition of one building block (“game
universe”) that was not clearly mentioned in any of the selected
frameworks but found in general game design literature [47].
The proposed building blocks are not intended to represent all
the potential elements that could be used for the design of any
serious game. Rather, these building blocks represent the basic
elements of standard serious game design. They are intended
to be customizable to fit specific projects.

Figure 1 presents the identified building blocks grouped into
categories, which will be described in the following section.
Context and Objectives

The “Context and objectives” category is intended to give an overview of the problem at hand and a first idea of the solution. Context includes a description of the environment in which the serious game will be used: classrooms or other premises, available technology, number of participants, available class hours, and all other initial constraints that should be taken into account [40,48].

Learning goals are used to give a general definition of the knowledge and skills that participants should acquire by playing the serious game. As for the game outline, defining learning goals early in the development process helps give a direction to the project even though they may evolve. A single sentence summarizing these goals can also be used for external communication.

The game outline is a short description of the serious game. It often takes the form of “The player takes the role of a __ which is in a __ (context/situation/environment). His/her objectives are __. For this, he/she must __.” The game outline can evolve during the course of the development, but its definition from the start of the project gives a direction to the development team. The game outline will often be used for communication with people outside of the project (such as stakeholders).

Learning Design

Learning design aims at defining and designing the learning aspects of the serious game.

Defining the profiles of the participants, including their digital literacy [49,50], interest in learning the subject matter, and gaming and simulation experience will help adapt the content of the game and its mode of delivery [16,48].

Regarding learning functions, by definition, serious games are designed for a primary purpose other than pure entertainment [51]. When designed for learning purposes, a clear definition of the learning functions is necessary to achieve the intended goal. Development teams must define whether the serious game will be used as an exercise designed to test or apply existing knowledge or skills, to support knowledge or skill acquisition, or to prepare for future courses [27].

Learning objectives (or learning outcomes) are the results of breaking down learning goals into measurable sub-elements. They define what participants should have learned by the end of the serious game training sessions and are often stated as “at the end of the serious game, participants will be able to...” The learning objectives are the basis for building the serious game structure and content, defining appropriate teaching and learning methods, and designing learning assessment modalities. They can be used to inform students of what they are expected to learn.

For the learning foundations, appropriate learning theories and pedagogical modalities must be chosen. Depending on learning objectives, an appropriate learning theory could be behaviorist, constructivist, or socio-constructivist [8,27]. They can be declined in several pedagogical approaches such as experiential learning or problem-based learning [8]. The choice of appropriate learning theories and pedagogical modalities is a necessary condition for the achievement of learning objectives. For example, if the main learning objective is to develop practical palpation skills for clinical assessment and the pedagogical modalities are “observation,” there will be an inconsistency that may prevent the learner from acquiring the intended skill.

For the knowledge foundations, the objective is to identify and validate the content related to the knowledge and skills participants are expected to acquire. This is the field of professional expertise. For example, evidence-based triage rules and processes should be the knowledge foundations of a serious game designed to teach emergency triage procedures. The identification of relevant knowledge foundations should be performed through a review of the relevant literature or of professional standards in collaboration with subject matter experts. It will also be necessary to define how the serious game will enable knowledge or skill acquisition [16,27].

The serious game should not be a stand-alone intervention but rather be embedded within a pedagogical scenario. The pedagogical scenario is therefore related to the general structure of the course or of the study program. The pedagogical scenario can be made up of a sequence containing game sessions, theoretical lectures, and personal working time. Depending on the pedagogical scenario, a subscenario can be required to support the use of the serious game. This subscenario generally includes 3 phases: prebriefing, orchestration of the game, and debriefing [52-55]. The activities taking place around the game (prebriefing and debriefing) are as important as the game itself.

Mechanics

Mechanics are at the core of the framework. They form the link between learning design and game design. In line with the model by Arnab et al [42], learning objectives should be linked to learning mechanics, which should be linked to game mechanics. Game mechanics should then be linked to game goals, rules, and structure.

The development team must decide upon the main learning mechanics that will be implemented in the serious game. We agree with the definition given by Plass et al [56], who defines learning mechanics as “patterns of behavior or building blocks of learner interactivity, which may be a single action or a set of interrelated actions that form the essential learning activity that is repeated throughout a game.” Learning mechanics can include activities such as remembering, understanding, applying, analyzing, evaluating, or creating. Learning effectiveness increases when learning and game mechanics are aligned with learning objectives [42,56-58]. This leads participants to develop and exercise their cognitive abilities throughout the game to reach its ultimate goal.

Game mechanics are the set of actions repeated by the player throughout the game [59] and are therefore the basic elements of interactivity. A game can include a single game mechanic (such as only shooting, jumping, or answering questions) or an integrated set of game mechanics (for example, moving around freely while answering questions and collecting objects). In serious games, game mechanics have a double objective,
resulting in 2 constraints: (1) engaging participants in taking part in the game and (2) ensuring consistency with learning mechanics. An incorrect choice of game mechanics can therefore quickly lead to failure in serious game implementation.

Learning and game incentives and rewards are used to support participant engagement and motivation. Incentives can be either *intrinsic* or *extrinsic* [58,60]. Intrinsic incentives are linked to game play and learning outcomes, whereas extrinsic incentives are not directly related to these elements. The most commonly used extrinsic incentives are points, badges, and trophies. Intrinsic incentives are more effective than extrinsic ones in achieving the learning objectives or any other intended goal. Indeed, it has been pointed out that gamification mechanisms purely based on rewards and on extrinsic motivators only bring short-term benefits and can be worthless or even harmful in the long run [58]. Intrinsic incentives may be harder to implement but are more beneficial. They can come from 3 sources: (1) mastery (learning to the point of feeling mastership regarding a specific knowledge or skill), (2) autonomy (being able to choose between several paths), and (3) relatedness (not feeling alone, feeling connected to others or to the situation) [61].

It is through interaction with the mechanics of play and learning that participants advance in the game and acquire knowledge and skills. To be successful, the interactions designed by the development team must result in both meaningful play and meaningful learning. Salen et al [59] defined meaningful play as emerging from players’ actions that are discernible (players receive feedback) and integrated into the game play (players understand how their actions influence the course of the game).

Meaningful learning, as opposed to rote learning, is achieved when the learner is actively engaged in the learning process and the newly learned information is connected with previous knowledge. Mayer [62] argued that meaningful learning occurs when learners build knowledge for successful problem solving. In serious games, meaningful learning may be achieved when participants need to acquire new knowledge for solving problems encountered in the game. Meaningful learning may happen either inside the game or outside the game, for example during the debriefing phase.

The successful implementation of meaningful play and meaningful learning leads to what could be called meaningful serious gaming.

**Game Design**

The game design includes the detailed description of all the elements that form the serious game.

Regarding goals and rules, setting a goal is essential for developing the pleasure and motivation to play. The goal of the game should be understandable, concrete, simple, clear, achievable, and rewarding if achieved [47,59]. A game is an artificial conflict to be resolved by the player [59]: The development team will have to decide which activities and interactions the player will be allowed to perform (game rules) to achieve the objective (game goal).

The game universe corresponds to the world in which the game will be played. It may be a fictional world or a simulation of the real world. The game universe should be consistent with the learners’ profiles.

Regarding the fidelity and simulation model, a simulation is a simplified representation of reality that seeks to achieve fidelity. Different kinds of fidelities have been described, all of which are used to enhance realism: sensory fidelity (audio-visual), narrative fidelity (dialogues, story), and cognitive fidelity (reflections that players make in the serious game) [8,63]. The types of fidelity chosen by the development team must be consistent with pedagogical objectives.

User interfaces (UI) and user experience (UX) are related to what the player will see and experience. They will impact the emotional feeling of the game and the pleasure of playing [26,27]. The graphical design and sounds must be aligned with the game universe and the desired fidelity. UX and game usability must be considered according to the context of usage and learners’ profiles.

The game structure should include the description of both the game and the learning progression. If there are prerequisite relationships between knowledge chunks acquired during the game, they will have to be taken into account when defining the game-learning sequences [19,41]. A progression that is too difficult or too slow will decrease the player’s motivation. This is in line with the concept of flow described by Csikszentmihalyi [64]. The progression must be thought of as much from a game perspective by game designers as from a learning perspective by educators [19].

Narratives are the content and structure of the story. They can include information given by a narrator, dialogues between the player and characters, and emails. Writing dialogues and narratives means creating an interactive scenario that will evolve according to the player’s choices. The quality of narratives will depend on the number and quality of choices and on the number and quality of feedback.

**Assessment**

How the game and its objectives will be evaluated is part of the overall design. This may include game assessment by participants, learning assessment within the game itself, or assessments outside the game. If a research project is considered, the research questions should be clearly defined, and research protocols should be established and registered. This will help determine which data and indicators will ultimately be needed and how data processing and visualization should be carried out. Care must also be taken to ensure compliance with personal data protection regulation. Consent mechanisms and the need for ethics approval must also be considered.

**Visual Organization of the Categories**

By grouping design building blocks into 5 categories, the framework aims at providing a structured view of the game design. This view enables all members of the development team to focus on the building blocks on which they are working (pedagogical engineers may focus on learning design, while graphic designers on user interfaces) while simultaneously providing an overview of the project and of the relationships...
between the building blocks. The categories are structured both vertically and horizontally (Figure 2).

**Figure 2.** Structured vision of the 5 serious game design categories.

---

**The Game and Learning Vision**

Traditionally, serious game design is viewed as a blend of learning and game design. In Figure 2, the left side of the framework corresponds to the learning design, and the right side corresponds to the game design. As in most design models [44,56], this vision emphasizes the inclusion of game and learning designs in serious game design. The Mechanics category can be seen as linking them together.

**The Project Vision**

The co.LAB framework can also be viewed from top to bottom. The upper section defines the problem and the overall objectives of the project. The middle section defines the solution: the game and the associated learning concept. The lower section defines how the solution will be evaluated both from the game and learning perspectives. This may in some cases be similar to the “success criteria” defined in project management theories.

---

**Discussion**

**Implementation in a Collaborative Web Platform**

To support the collaborative work of a multidisciplinary development team often geographically dispersed and to enable all partners to have a common and up-to-date vision of the design, the co.LAB framework should be implemented in a collaborative web platform. In this section, we discuss the specifications and requirements for this implementation.

The web platform should be open source and open access, thus allowing any serious game development team to use it freely. At the beginning of a serious game project, the development team should access a visual initial design template based on the co.LAB framework. Figure 3 shows how the main screen of a design project based on the co.LAB framework could look. As team management or project monitoring functionalities would be useful to help allocate resources and quickly check on the project’s advancement, they should be embedded in the web platform.
Cards as a Collaborative Workspace for Each Building Block

In the web platform, each building block (e.g., context, game outline) should be implemented as a Card. By clicking on the card, the development team would access a dedicated collaborative workspace. The workspace of each card should provide teamwork functionalities such as collaborative writings, discussion threads, modification proposals, and document sharing. Each card should be complemented with methodological resources such as definitions, best practices, tools, or theoretical references (Figure 4).

Figure 4. Wireframe of a Card, with collaborative working spaces and access to resources.
Implementation of the Adaptive Features

At the beginning of a serious game development project, the development team would be provided with a basic template including the most frequently used cards. This basic structure covers the conventional elements of standard serious game design and is suitable for use in this form by junior development teams. More experienced development teams could adapt this basic model by adding, modifying, or deleting cards. Cards could either be added from a store of already available optional cards or created from scratch by the development team.

Providing abilities to select the most relevant cards, to discard others, and to create missing cards should allow the initial template to be adapted toward an already existing model or to customize it for a specific serious game design. The framework’s basic template should also be adaptive. Based on the analysis of the traces of use of different development teams, the platform administrators should be able to make the model evolve.

Implementation of the Systemic Features

Figure 3 shows a default view of the model in which cards are grouped by categories. The platform should also allow the development team to develop a systemic view of the project by creating links between cards. Specific views should then be generated to allow team members to visualize such links (Figures 5 and 6).

Figure 5. Activity network diagram view, emphasizing precedence dependency relationships.

![Activity network diagram](image-url)
There are 3 main types of relationships between cards. The first, precedence dependency relationship, is used to create the activity network diagram and for scheduling. The second, causality, is used to inform that something defined in 1 card must be taken into account in another. The third, which may be called interrelationship, indicates that 2 cards should be seen as a coherent whole. This interrelationship can also be described as a bidirectional causality.

Each relationship could also have its own collaborative workspace, with best practices and resources. Some relationships could already be defined in the initial model. For instance, as an interrelationship is mandatory to link game mechanics and learning mechanics, this relationship could already be available from the start of the project. The collaborative workspace attached to this relationship would provide best practices guidelines related to the alignment of game and learning mechanics.

The development team should also be able to create specific relationships, as, for example, a link between the context and user interfaces with a remark that the game should be playable on smartphones.

Supporting all the Project Phases
The web platform should support the development team all along the course of the project, from design to development and evaluation. In Figure 7, we propose a serious game development process that should be implemented in the web platform. This process is a quite classical adaptation of traditional agile project management. Using this kind of agile approach for serious game development is endorsed by Verschueren et al [16] and Alvarez et al [44] and by the authors’ experience.

The co.LAB framework presented in this paper focuses on the Requirements and Design phases of the process. It is however possible to extend the framework to encompass the whole process (Figure 8).

By providing a coherent link between building blocks during all phases of the project, the co.LAB framework could be used throughout the life cycle of the serious game.

The different visualizations of the building blocks (grouped into categories, activities diagram, relationships, project life cycle) correspond to different ways of approaching the same problem of serious game design and development. These various visualizations should allow for a shared understanding and a consistent approach throughout the design and development process.
Figure 8. Extension of the co.LAB framework along the entire serious game life cycle.

Principal Findings
The co.LAB serious game design framework was created by identifying design elements commonly used to design and develop serious games. These design elements were defined and synthesized to create design building blocks, which were grouped in 5 categories: (1) context and objectives, (2) game design, (3) mechanics, (4) learning design, and (5) assessment. The framework recognizes the diversity of serious game design and is conceived to be adapted to specific contexts, by adding or removing building blocks.

The different visualizations of the building blocks (grouped into categories, activity network diagram, causality relationships, project life cycle) correspond to different ways of approaching the problem of serious game design and development. These various visualizations should allow for a shared understanding and a consistent approach throughout the design and development process.

The co.LAB framework is designed to be implemented in a collaborative web platform, with implementation recommendations that should support teamwork and knowledge sharing within a multidisciplinary team and favor an adaptive and systemic approach. The co.LAB framework may be used as a guideline along all project phases, from requirements to design, development, tests, implementation, and evaluation.

Future Work
The co.LAB framework is currently being implemented in a collaborative web platform.
Guidelines related to each card are currently being developed and will gradually be implemented in the platform. Their development is based on a literature review and on the authors’ experience. Their relevance will be evaluated by end users, and they will be updated according to the feedback obtained.

We plan on testing this framework and the web platform on which it is being implemented through the development of different kinds of serious games. This should allow us to troubleshoot the platform and identify the most important areas of improvement.

In a future version, we plan on implementing electronic assistance to help users find the most suitable combination of cards depending on their specific serious game project.

Limitations
The co.LAB framework is not based on a complete systematic review of serious game design frameworks, but rather on a review of the most influential and most relevant frameworks according to the authors’ opinions and experience. However, should any particular design element be missing from the current version of the framework, its adaptive features should allow development teams to include them.

The co.LAB framework is also based on the authors’ experience in serious game design and development. Even though the authors have developed many serious games in different subject matters and contexts, their experience is still limited and does not include all kinds of serious game development. Once again, the adaptive features of the co.LAB framework should mitigate this limitation.

Finally, the co.LAB framework has been neither fully implemented on a web platform nor used to create a full-fledged serious game yet. Even though its development was theory-driven and based on relevant and authoritative references, it should be thoroughly tested before its routine use can be recommended.

Comparison With Prior Work
In comparison with previous work, the co.LAB framework proposes some novelties. First, this framework recognizes the diversity of serious game design and is intended to be adaptable.
to specific contexts. Second, the framework was designed to be implementable in a collaborative web platform. Finally, this implementation in the web platform is based on a systemic approach of the design process.

Conclusion
By offering a structured view of fundamental design elements, the co.LAB framework should facilitate the design and development of serious games through a collaborative, multidisciplinary, adaptive, and systemic approach. The ability to visualize the building blocks and their relationships from different standpoints should allow for a shared understanding and a consistent approach throughout the design and development process. The co.LAB framework was designed to be implemented in a collaborative web platform that is currently under construction. Once fully implemented, the actual usability and effectiveness of this new framework should be thoroughly tested.

Acknowledgments
The co.LAB project is supported by the Swiss National Science Foundation (NRP 77).

Conflicts of Interest
None declared.

References
15. Chaudy Y, Connolly TM, Hainey T. EngAGe: A Link between Educational Games Developers and Educators. 2015
20. De Troyer O. Towards effective serious games. 2017 Presented at: 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games); September 6-8, 2017; Athens, Greece. [doi: 10.1109/vs-games.2017.8056615]


Abbreviations

UI: user interface
UX: user experience
©Dominique Jaccard, Laurent Suppan, Eric Sanchez, Audrey Huguenin, Maxence Laurent. Originally published in JMIR Serious Games (https://games.jmir.org), 02.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
A Therapeutic Game for Sexually Abused Children and Adolescents (Vil Du?!): Exploratory Mixed Methods Evaluation

Joyce Johanna Endendijk¹, PhD; Henny Tichelaar¹, MSc; Menno Deen², PhD; Maja Deković¹, PhD

¹Child and Adolescent Studies, Utrecht University, Utrecht, Netherlands
²Studio Lapp, Utrecht, Netherlands

Corresponding Author:
Joyce Johanna Endendijk, PhD
Child and Adolescent Studies
Utrecht University
Heidelbergaan 1
Utrecht, 3548 CS
Netherlands
Phone: 31 30 253 4896
Email: j.j.endendijk@uu.nl

Abstract

Background: Talking about experiences of sexual abuse in therapy is difficult for children and adolescents. Possible reasons for this difficulty are a lack of vocabulary to describe the situation or feelings of shame, fear, and self-blame associated with sexual abuse. The serious game Vil Du?! was developed to help children open up about their sexual abuse experiences. Vil Du?! is a nonverbal communication game that resembles a dress-up game in which children can show the therapist what happened to them.

Objective: This exploratory evaluation study examines which working elements of the game could be identified in therapy with victims of sexual abuse (aim 1). In addition, this study examines how therapists evaluate the acceptability of the game (aim 2).

Methods: The therapists completed 23 web-based surveys on the use of Vil Du?! In addition, semistructured interviews were conducted with 10 therapists. The data were analyzed in NVivo following previously reported stepwise guidelines.

Results: Regarding aim 1, therapists mentioned various working elements of Vil Du?!; for instance, Vil Du?! puts the child in control of the situation. In addition, Vil Du?! reduces barriers to disclosure because there is no need to talk or have eye contact with the therapist. Regarding aim 2, Vil Du?! was generally evaluated more positively than negatively by the therapists. For instance, therapists indicated that using Vil Du?! is time efficient and might make the treatment process less confronting and difficult for the client. According to therapists, most clients indeed experienced less tension and more positive (or neutral) emotions than negative emotions when using Vil Du?!

Conclusions: The most important working elements of Vil Du?! according to therapists, are that it enables children to regain control over their sexual abuse experiences and reduces barriers to disclosing sexual abuse experiences. The more positive evaluation of Vil Du?! indicates the acceptability of the game for therapists as well as their clients.

(JMIR Serious Games 2021;9(3):e26062) doi:10.2196/26062

KEYWORDS

child sexual abuse; psychotherapy; serious games; evaluation; working elements; acceptability

Introduction

Statement of the Problem

Talking about experiences of child sexual abuse (CSA) is difficult for children and adolescents. Possible reasons for this difficulty are feelings of shame, fear, and self-blame, a lack of understanding or denial of the situation, or a lack of vocabulary to describe the situation [1-4]. However, therapy for child and adolescent victims of CSA often relies heavily on verbal narration and the processing of one’s experiences. This is especially the case for cognitive behavioral therapy, which is the most often used and recommended form of psychotherapy for CSA [5,6]. To increase the suitability of cognitive behavioral therapy for children and adolescents, researchers have suggested the use of content that is tailored to the developmental needs of young clients and to emphasize elements of play [7,8]. Using
play in a therapeutic setting puts the child or adolescent in control, which is known to reduce tension and stress [9]. In addition, play provides an age-appropriate and natural manner for children to express their feelings, which they are often unable to express through language [10]. Furthermore, play makes treatment more fun and engaging [11]. Originating from these ideas, the third author (M Deen) and others developed a serious game, called Vil Du?! [12], to help children open up about their sexual abuse experiences to their therapists. This study provides an exploratory evaluation of Vil Du?! among therapists using the game in therapy for sexually abused children or adolescents. More specifically, we examined the working elements of the game (ie, elements that can explain its effects on the therapy process) therapists identified, as well as therapists’ experiences with using the game (ie, benefits, limitations, and experiences of their clients).

Other Serious Games for Children and Adolescents in Treatment for CSA

To date, there are limited serious game offerings for use in the context of CSA. One example is Orbit, a CSA prevention computer game targeted at students aged 8-10 years [13]. The goal of this adventure game is for the player to do everything they can to help the character Sammy that has suffered from sexual abuse. During several mini-games, children learn about recognizing CSA, perpetrator tactics, barriers to telling, building a healthy self-concept, and the importance of trusted adults who they can turn to. In addition, 2 unpublished master theses describe the development of serious prototype games aimed at helping children disclose CSA experiences [14,15]. Pharshy [14] developed a story-telling game in which children can create new stories of their own experiences, or edit existing stories, by using images, drawings, text, and self-created avatars. Parents or caregivers could monitor the child’s stories for possible CSA experiences. Andersson [15] developed a tool for use in a therapeutic context. The prototype contained different scenes and storylines that therapists could show and play out with children, which could spark conversations about the different (sexual) abuse-related situations children might find themselves in. The 3 existing games have been developed primarily for CSA prevention purposes and not for use in a CSA therapy context, such as Vil Du?! In addition, the effects of these 3 games have not yet been evaluated.

**Vil Du?! Design and Working Elements**

To the best of our knowledge, the serious game Vil Du?! (Danish for “Do you want to [talk about...]?!”) is the first digital game that is used in a treatment context for child and adolescent CSA victims. Vil Du?! is a nonverbal communication game in which children can show the therapist what happened to them (Figure 1). In the game, which resembles a dress-up game, both therapist and child operate a self-chosen character, each on their own tablet (provided by the therapist). The tablets are synchronized to each other, so actions performed on one screen are also visible on the other, enabling digital interaction between the therapist and child (without the necessity of looking each other in the eyes). Both players can perform various actions on the other character by clicking (eg, undress) or dragging explicit icons (eg, mouth, hand, penis, and buttocks; Figure 2) over the character’s body. Each player can express their boundaries or pause or stop the game by pressing a Time-Out button (Figure 3). While playing, the therapist can probe the child to talk about their experiences, thoughts, and feelings. In addition to the Time-Out button, Vil Du?! does not include verbal statements or textual markers. The goal of the game is to give children a voice without the need to talk and to put children in charge of creating their own story, normative structure, and values associated with love, sex, and romance. Vil Du?! has been used by therapists with children and adolescents between the ages of 5 and 17 [16].

**Figure 1.** Overview of the game environment of Vil Du?!
The first phase in evaluating a new tool is to define the working elements of the intervention [17]. Hereby, the goal is to find and define promising practice-based elements or components that could explain what works in the use of a serious game in therapy for CSA victims [18]. These working elements refer to important elements through which serious games (ie, Vil Du?!) might lead to beneficial outcomes in therapy for CSA [19]. Several possible working elements are identified in the game. First, Vil Du?! might reduce barriers to disclosure of CSA experiences because of its nonverbal character and because the use of the game does not require face-to-face communication between the therapist and client. With Vil Du?!, children can show, instead of talk about, what happened to them, which might help to overcome the barrier of a lack of vocabulary to describe the situation [2]. In addition, there is no need for eye contact with the therapist when using Vil Du?! because the client can look at the tablet screen. This might reduce feelings of shame, fear, and self-blame associated with CSA experiences [1,3,4]. Researchers have indeed suggested that serious games “can be used as a ‘third party in the room’, helping to make the therapeutic process less difficult for adolescents by taking some of the emphasis off direct face-to-face conversations” [20].

Second, Vil Du?! provides children with an environment in which they can communicate through play. Vil Du?! resembles a dress-up game in which children can undress or redress a character and perform actions on, or with, the character by dragging icons over the character’s body (eg, use the hand icon to give a high five). The play element of Vil Du?! might have an effect on several components of therapy. For instance, Vil Du?! might be helpful for the cognitive restructuring of incorrect and maladaptive thoughts related to the CSA experience. CSA victims often have incorrect or maladaptive thoughts related to the abuse experiences (eg, “I could have prevented the abuse”). For therapists to restructure such incorrect and maladaptive thoughts, they first need to explicate children’s thoughts for which play (eg, with dolls, by drawing, or with Vil Du?!) can be used as a vehicle [21,22]. These explicit thoughts can be processed or restructured in successive therapy sessions. In addition, the explicit icons and characters of Vil Du?! enable children to engage in play that realistically depicts their traumatic experiences. When therapists encourage such realistic play, clients are gradually exposed to traumatic memories that might otherwise be avoided or suppressed [23]. Moreover, regarding the therapy component of trauma narration, playing
out one’s experiences might help children more comfortably express the details of highly traumatic experiences, while also showing emotions and thoughts that can later be processed by the therapist [23].

Third, with Vil Du?! children are in control of showing their experiences. This control is of particular importance in the context of CSA, in which children were not in control of the situation, and can help them regain a sense of control over their lives [24]. Moreover, children cannot achieve progress in therapy when they feel out of control [25]. With Vil Du?!, children can decide what they will show to the therapist, and they have control over stopping or pausing the game with the Time-Out button. Vil Du?! also puts children in control because the use of icons can have various connotations. For example, moving the hand icon over the back of the character can mean a pat on the back, stroking the back, or a slap on the back, depending on the context the child created or on verbal explanations provided by the child. Thus, it is not the game that guides how to interpret play; it is the player who controls the meaning of the interaction. The nondirective nature of Vil Du?! might be helpful in therapy for restructuring incorrect and maladaptive thoughts, because nondirective play can be considered as a restructuring process of one’s thoughts [26].

Fourth, Vil Du?! gives children the opportunity to share and recreate their stories in a safe and nonnormative environment. With Vil Du?!, children can share the story of what happened to them by having one character performing actions on another character. Vil Du?! does not include statements or textual markers that stipulate whether a sexual relationship or action is right or wrong. As such, the game does not demonize the players or their actions, which may create a safe environment for children to open up about their sexual experiences. Normative statements about the inappropriateness of certain sexual behaviors and how children should behave in certain circumstances could problematize an already difficult CSA experience, which may increase feelings of self-blame and shame [27]. Research also shows that creating a narrative of one’s CSA experiences might be a critical mechanism for producing positive outcomes after CSA [28]. In addition, children themselves often mentioned that creating a narrative of the sexual abuse by drawing or writing was specifically the most helpful part of therapy [28,29]. Stories might enable children to create a mental map of events and ideas and revisit them as and when the stories are narrated or remembered again [30]. In addition, games in which children are invited to create or tell a story have been found to help children express themselves safely, because such games aid children in absorbing complex concepts through play [31] as well as allowing children to express their feelings through an indirect medium [26].

Acceptability of Vil Du?! for its Users

Next, to evaluate the possible working elements of Vil Du?!, it is also important to evaluate the acceptability of the game for its users [17]. Acceptability refers to how intended individual users react to an intervention or tool [32]. Two users are relevant in the context of therapy—the therapist and the client. The therapist’s acceptability of a tool is important in that it determines whether therapists will use the tool. Relevant in this regard is how the therapist feels about the intervention in terms of its benefits and limitations (ie, affective attitude [33]). Client acceptability is clinically important as, if clients prematurely drop out of therapy because they are, for example, dissatisfied with how the therapy is being delivered, therapy cannot be completed and symptoms may persist unnecessarily [34]. Relevant indicators of the acceptability of Vil Du?! by clients might be the emotions, tension, and dissociation (ie, detachment from reality) experienced by clients during the use of Vil Du?! Such experiences are predictive of treatment outcome and adherence [35,36] and can thus provide preliminary information as to whether Vil Du?! is acceptable for clients.

Goal of This Study

Vil Du?! might be a valuable tool for therapy, but the use of the game in therapy is still in its infancy. For instance, a systematic user manual for Vil Du?! is currently being developed. In addition, therapists are still exploring how Vil Du?! could be used in therapy. To evaluate a tool in such an early phase, an exploratory evaluation among therapists seems most appropriate as it offers more flexibility than rigorous effect studies [17]. Such early evaluations can provide valuable indications of the possible clinical impact of new tools, as well as the acceptability of a tool by its users [37]. This information can subsequently be used for further development or refinement of a tool [17]. Therefore, our mixed-methods study aims to provide an exploratory evaluation of Vil Du?! by answering two research questions. First, which working elements can be identified in how Vil Du?! is used by therapists? We examined the following working elements specific to Vil Du?!: no need to talk; functions as a third party in the room or no face-to-face communication necessary; puts the child in control; and playful, safe, and nonnormative environment. Second, how do therapists evaluate the use of Vil Du?! in psychotherapy for CSA? We examined the benefits and limitations of Vil Du?! identified by therapists, as well as therapists’ views of the emotions, tension, and dissociation (ie, detachment from reality) their clients experienced during the use of Vil Du?!

Methods

Participants

A total of 21 therapists using the Vil Du?! app at the time of the study were contacted personally, as well as via email, to fill out a web-based questionnaire every time they used Vil Du?! with a client. We could identify these therapists as they had received a working license for Vil Du?! from the app developers. The 21 therapists were also asked to forward the invitation to other colleagues who might have used Vil Du?! The therapists completed the web-based questionnaire for 23 clients with CSA experience. The mean age of the clients was 11.38 (SD 3.96; minimum=5, maximum=18) years. The majority of the clients were women (14/23, 61%). Therapists mostly used Vil Du?! in the context of client-centered therapy (13/23, 57%), followed by play therapy (8/23, 35%), and cognitive behavioral therapy (4/23, 17%). None of the therapists used Vil Du?! for psychoanalytic therapy or group therapy. The total number of therapists participating in the web-based questionnaire could not be determined because therapists could fill out the...
questionnaire anonymously. A total of 12 therapists provided contact information and indicated that they were willing to participate in the interview part of the study. The characteristics of the 10 therapists who actually participated in the interviews are presented in Table 1.

Table 1. Characteristics of therapists participating in the interview phase of the study.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>Age (years)</th>
<th>Experience in youth care (years)</th>
<th>Therapy with CSA(^a) clients or therapy related to sexuality (hours per week)</th>
<th>Work organization(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>42</td>
<td>20</td>
<td>5</td>
<td>1(^c)</td>
</tr>
<tr>
<td>2</td>
<td>44</td>
<td>15</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>42</td>
<td>17</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>12</td>
<td>32</td>
<td>2(^d)</td>
</tr>
<tr>
<td>5</td>
<td>52</td>
<td>30</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>22</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>28</td>
<td>5</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>17</td>
<td>32</td>
<td>3(^e)</td>
</tr>
<tr>
<td>9</td>
<td>62</td>
<td>10</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>38</td>
<td>17</td>
<td>24</td>
<td>4(^f)</td>
</tr>
</tbody>
</table>

\(^a\)CSA: child sexual abuse.

\(^b\)Participants with the same number in this column work for the same organization.

\(^c\)Youth care organization providing contextual care for child sexual abuse victims.

\(^d\)Specialized youth care organization focusing on adolescent sexuality.

\(^e\)Organization providing specialized family care.

\(^f\)Youth care center for psychotrauma and sexual abuse.

**Design and Procedure**

We used a mixed-methods triangulation design with quantitative and qualitative data collected and analyzed at approximately the same time. Both types of data were merged and given equal emphasis in the interpretation [38]. Data were collected via web-based questionnaires (including quantitative and qualitative questions) and semistructured qualitative interviews. Through both methods, we derived input from therapists on the working elements of Vil Du?! and how they evaluate its acceptability. For both the web-based questionnaire and the semistructured interviews, questions and topics were based on the possible working elements of Vil Du?! identified in the literature. We also included aspects of user acceptability derived from existing measures [20].

Therapists completed the web-based survey via LimeSurvey (version 3.20, LimeSurvey Project; closed format with links sent to participants). They provided web-based informed consent for their participation at the beginning of the questionnaire. For the duration of the study (June 2019 to June 2020), we invited therapists to complete the questionnaire each time they used Vil Du?! in therapy sessions. The completion rate was 100% (23/23), as each question was mandatory. Participants were able to change their answers using a back button. The survey included 14 questions (displayed in a fixed order), of which 6 were used for this study.

Subsequently, a trained research assistant conducted and audiotaped interviews with a selection of the therapists (n=10) that completed the aforementioned web-based questionnaire. We selected the widest possible range in terms of the sexual abuse experiences of clients, therapy type, and background characteristics of therapists. For the interview part of this study, therapists provided written informed consent before the start of the interview. The duration of the interviews was, on average, 45 minutes.

The Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University approved this study (FETC19-025). Personal information collected via the survey or interviews was stored on a protected university server. Only the first and second authors had access to this server.

**Web-Based Questionnaire**

**Background Characteristics**

Therapists first had to fill out some background characteristics of the client (ie, age and gender).

**Working Elements and Therapist Acceptability of Vil Du?!**

Participants answered 1 open question regarding the elements of Vil Du?! they thought were effective in the therapy process (ie, “What are positive elements of using Vil Du?! in therapy? By positive elements, we mean elements of the game that can explain its effects on the therapy process”). This question (and accompanying explanation) is based on the common definition of working elements [17]. Participants also reported on the limitations of using Vil Du?! for either the therapist or the client (ie, “Is there something missing or lacking in the game that limits the use in therapy?”).
Client Acceptability

Client acceptability (according to the therapist) with regard to the use of Vil Du?! was assessed with 3 questions on a 5-point Likert scale [20]:

1. “how pleasant/unpleasant they thought the use of Vil Du?! was for the client” (1=very unpleasant and 5=very pleasant).
2. “how helpful they thought the use of Vil Du?! was for the client” (1=very unhelpful and 5=very helpful).
3. “how enjoyable they thought the use of Vil Du?! was for the client” (1=very unenjoyable and 5=very enjoyable).

Semistructured Interviews

The interviews were structured around a topic list, including the following topics, and were discussed in more detail than in the web-based questionnaire.

1. What are the differences between using Vil Du?! and not using Vil Du?! in the treatment process or outcomes, and to which elements of Vil Du?! these differences could be attributed (ie, working elements);
2. therapist acceptability (eg, limitations, benefits); and
3. client acceptability (eg, emotions, stress levels, and entertainment value).

Analyses

Web-Based Questionnaire

SPSS version 24 (IBM Corp) was used to summarize and describe the qualitative and quantitative data from the web-based questionnaires. Frequencies were computed to determine the percentage of questionnaires in which the working element of Vil Du?! was identified for a specific client. Frequencies were also computed to determine the percentage of questionnaires in which a certain limitation of Vil Du?! was mentioned for clients of therapists. Descriptive statistics (mean and SD) were computed to summarize the acceptability of Vil Du?! for clients.

Interviews and Open Questions From the Web-Based Questionnaire

The stepwise guidelines outlined by Zhang and Wildemuth [39] were followed to increase the efficiency, repeatability, and transparency of our qualitative data analysis of the interview data in NVivo (QSR International). In step 1, we transcribed the answers to all questions literally. Observations during the interview (eg, sounds and pauses) were not coded, because they were not of interest to the research questions. In step 2, we defined the unit of analysis as themes; that is, the working elements of Vil Du?! and the experiences of therapists and clients using Vil Du?! In step 3, a coding scheme was developed, consisting of categories related to the working elements (eg, no face-to-face communication necessary, no need to talk, and child in control), benefits, and limitations of Vil Du?! (the coding scheme is available upon request from the authors). In step 4, the first (JJE) and second author (HT) tested the coding scheme on a sample of text (ie, 2 pages of text selected from a total of 4 interviews) to discover unclarities in the coding scheme. These issues have been discussed and resolved. In step 5, all text was coded by the first author while adding new categories to the coding scheme when necessary. For step 6, a set of randomly selected text fragments (20% of the total number of text fragments) was coded by both the first and second authors. Differences in coding were discussed until a consensus was reached. Changes were made to the coding scheme when necessary. The first author recoded the other 80% of the transcripts on the basis of this changed coding scheme. In step 7, we explored the properties and dimensions of the categories as well as the relations between the categories in the full range of data. We merged categories that reflected the same content. We also specified category names based on the content included in a certain category. Finally, we separated the single categories into multiple categories when a category contained different types of information. Answers to the open questions of the web-based questionnaire were coded using the same codes as the interview data.

Merging Quantitative and Qualitative Data From Web-Based Questionnaires and Interviews

For all instances in which data from interviews were merged with data from the web-based questionnaires, data from the interviews were used to further elaborate on the often short answers that were given in the web-based questionnaire.

For aim 1 of this study about the working elements of Vil Du?!, data from the open question about the working elements of the web-based survey were merged with the data from the interviews. More specifically, the percentage of questionnaires that mentioned a certain working element for a client was combined with the percentage of therapists who mentioned the same working element in the interview.

For aim 2 of the study regarding therapists’ evaluations of Vil Du?!, the interviews were used to determine the benefits according to the therapists. In addition, for the limitations, the percentage of questionnaires that mentioned a certain limitation was combined with the percentage of therapists who mentioned the same limitation in the interview. Finally, for client acceptability, quantitative client satisfaction data from the web survey were combined with the percentage of therapists that described in the interviews the expression of certain emotions or levels of tension in their clients.

Results

Working Elements of Vil Du?!

Table 2 lists the working elements of Vil Du?! that could be identified in the questionnaires and interviews with therapists. In 30% (7/23) of the questionnaires and in 90% (9/10) of the interviews, therapists recounted experiences in which the child was in control in the game. Sometimes, control was inherent to the game. For instance, clients had control over the use of the Time-Out button. They also had control over what to show and what not to show to the therapist. Finally, they had control over the interpretation of the icons (eg, by moving the hand icon over the back of the character, they could indicate a pat on the pack, stroking the back, or a slap on the back). At other times, the therapist created the control in the game environment. For instance, some therapists invited their clients to choose characters or start up the game. Other therapists gave clients...
control over the choice to use the game or just talk about CSA experiences.

Most therapists (in 8/23, 35\% of the questionnaires and 8/10, 80\% of the interviews) further recounted that the game reduced barriers to disclosing CSA experiences for their clients. For example, one therapist mentioned, “I noticed that the tablet helped her to disclose and also a bit to break the ice.” Related to lowering barriers to disclosure, most therapists (in 8/23, 35\% of the questionnaires and 9/10, 90\% of the interviews) explicitly mentioned that the game could function as a third party in the room by creating emotional and physical distance between the therapist, the client, and the client’s CSA experiences. According to a therapist, a client described the emotional distance as follows:

*It is very strange, it [the character] is not me, but yet it is me, and I never thought I could tell this much about it.*

Regarding the physical distance, a therapist reported as follows:

*It is not face-to-face, it is not direct, you talk in a triangle. You talk about something [the tablet] that is on the table and about that what has happened, which creates distance.*

### Table 2. Questionnaire and interview data about the working elements of Vil Du?!

<table>
<thead>
<tr>
<th>Working elements</th>
<th>Questionnaires (n=23), n (%)</th>
<th>Interviews (n=10), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puts child in control</td>
<td>7 (30)</td>
<td>9 (90)</td>
</tr>
<tr>
<td>Reduces barriers to disclosure</td>
<td>8 (35)</td>
<td>8 (80)</td>
</tr>
<tr>
<td>By creating a third party in the room</td>
<td>8 (35)</td>
<td>9 (90)</td>
</tr>
<tr>
<td>Because there is no need to talk</td>
<td>8 (35)</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Congruence with children’s digital experience</td>
<td>—a</td>
<td>7 (70)</td>
</tr>
<tr>
<td>Playful environment</td>
<td>—</td>
<td>6 (60)</td>
</tr>
<tr>
<td>Safe and nonnormative environment</td>
<td>2 (9)</td>
<td>3 (30)</td>
</tr>
<tr>
<td>Interactive</td>
<td>—</td>
<td>1 (10)</td>
</tr>
</tbody>
</table>

*aNot available.*

Another working element related to reducing barriers to disclosure was that clients do not need to talk when using the game, which was brought up by therapists in 70\% (7/10) of the interviews and in 35\% (8/23) of the questionnaires. As one therapist recounted, “Some children find it really difficult to find the words and then it [Vil Du?!] is a very nice tool to use.” Another therapist explained, “‘Show me’ might be easier than ‘tell me’.”

In 70\% (7/10) of the interviews, therapists also described that the game was congruent with children’s experience in the digital world. For instance, one therapist mentioned: “They think the app is interesting. It is, of course, very much targeting their digital experience.”

The playful environment of the game was mentioned by therapists in 60\% (6/10) of the interviews as a working element. For example, one therapist explained, “What I really like about the game is that in a playful manner you can make very difficult or shameful experiences discussable.”

In 30\% (3/10) of the interviews and 9\% (2/23) of the questionnaires, therapists brought up the safe and nonnormative environment of the game. As an example, one therapist explained:

*In that sense I really liked it and especially that it was very safe for him to tell his story [...], but also without judgement. That screen does not judge.*

Finally, in 10\% (1/10) of the interviews, the interactive nature of the game was mentioned as a working element. More specifically, the two characters in Vil Du?! can interact with each other, and both the therapist and the client have the possibility to perform actions on the characters and respond to these actions.

**Therapists’ Evaluation of Vil Du?!**

**Therapist Acceptability: Benefits and Limitations**

In terms of benefits, time efficiency was mentioned most by therapists in the interviews (8/10, 80\%). As an example, one therapist explained: “The benefit is in the rapidness. You are there faster, the thing you want to talk about.” Interestingly, one therapist reported the rapidness of the game as a possible disadvantage, at least if you were not prepared:

*If I use it I need to be very alert, very alert. Watch the non-verbal communication [...]. You need to sort of pull out all the stops and ‘am I not missing something?’ [...] But it is, it is not a slow thing. It is very fast. So, if somebody who is going to work with it, who has not used it before, I think ‘pay attention’!*

Further, in 50\% (5/10) of the interviews, therapists mentioned the benefit that by using Vil Du?! you could obtain more information and details about the client’s CSA experience than without the game. Benefits that were only mentioned in 10\% (1/10) to 20\% (2/10) of the interviews were as follows: better attention of the client during disclosure of CSA experiences, the possibility to stop the game immediately with the Time-Out button when necessary, the fact that the game is highly structured, the user-friendliness and accessibility of Vil Du?!, and that Vil Du?! is useful for many treatment components. However, of 10 therapists, 1 (10\%) thought it was difficult to...
envisage how the game could be used for other treatment components than creating a narrative of the clients’ sexual abuse experiences.

The most mentioned limitation by therapists in 90% (9/10) of interviews and 57% (13/23) of questionnaires was that the game lacks complexity in surroundings and in the characters’ movement, dimensionality, emotion expression, and appearance. For example, one therapist explained that the game lacked complexity in:

- **the surroundings.** For instance, then we went outside, then inside, then to school [...] So, that it can become more of a story instead of one incident. That fits with our clients, that is, rarely one incident or one place.

Regarding the characters, one therapist reported as follows:

- **It would be nice if the characters could move a bit more.** At least standing, sitting, and lying down. I noticed with the girl I used it for the first time, the girl with intellectual difficulties, that it was very difficult for her, because it always happened on his bed and he was lying down. So, it was very difficult for her to show what happened.

Other limitations were only brought up by a few therapists in the interviews and questionnaires. For instance, that it was difficult to visualize masturbation, penetration, or an erection, which was mentioned in 22% (5/23) of the questionnaires and 30% (3/10) of the interviews. In addition, the feeling thermometer (currently a static picture; Figure 4) cannot be manipulated by the client to indicate their levels of distress during the use of Vil Du?! (3/23, 13% of questionnaires and 3/10, 30% of interviews). Two therapists further mentioned in the interviews that the game could be boring or childish for older clients.

Finally, some limitations were recounted by single therapists only in either the questionnaire or the interview; therefore, these limitations might be subjective. For example, the game included limited content for psychoeducation or normalizing sexual behavior (mentioned in the questionnaire). Another therapist mentioned technical problems with the game during the interview. An additional limitation was that it was not possible to change the explicitness of the icons (interviews). Relatedly, the vagina icon might not be clear enough for young children (interviews). Finally, one therapist mentioned that the game is difficult to use for older therapists (interviews).

**Figure 4.** The feeling thermometer in Vil Du?!
**Client Acceptability: Emotion Expression, Level of Tension, and Dissociation According to Therapists**

Questionnaire data showed that all aspects of client satisfaction (according to therapists) yielded mean scores above the neutral midpoint of the scale. The highest scores were given for the usefulness of Vil Du?! for the client’s therapy progress. Therapists thought the use of Vil Du?! was between somewhat and highly useful for their clients (mean 4.13, SD 1.14). Therapists also thought the use of Vil Du?! was, as far as possible in the context of CSA, somewhat of a pleasant experience for their clients (mean 3.57, SD 1.47) as well as a bit enjoyable (mean 3.30, SD 1.40).

Similarly, during the interviews, most therapists (9/10, 90%) recounted their clients’ positive emotions when using Vil Du?! or shortly after using Vil Du?! Examples of such positive emotions are fun, joy, happiness, and pride. In 50% (5/10) of the interviews, therapists also described clients who experienced negative emotions, such as fright, irritation, clenching, and anger. Finally, one therapist described 2 clients who were rather neutral in terms of their emotional expression.

With regard to clients’ level of tension, in 70% (7/10) of the interviews, therapists described clients experienced lower or shorter bouts of tension when using Vil Du?! than by simply talking about their experiences. In 40% (4/10) of the interviews, therapists mentioned clients who experienced high levels of tension or signs of dissociation, but they did not link these experiences specifically to the use of Vil Du?! One therapist described the possibility that Vil Du?! elicits higher levels of tension because of its explicit nature, but the presence of the Time-Out button makes tension manageable.

**Discussion**

**Principal Findings**

The goal of this study is 2-fold. The first aim is to identify the working elements in how Vil Du?! was used by therapists. By combining data from questionnaires and semistructured interviews, we found that therapists identified several working elements of Vil Du?! More specifically, the therapists mentioned that Vil Du?! puts the child in control. They also described that the game reduces barriers to disclosure as there is no need to talk and no need for eye contact with the therapist. Furthermore, they thought that the game was congruent with the children’s digital experience. In addition, they mentioned that the game presents a playful as well as a safe and nonnormative environment. Finally, they mentioned the interactivity of the game as a working mechanism.

The second aim is to investigate how therapists evaluated the use of Vil Du?! in psychotherapy for CSA. With regard to this aim, Vil Du?! was generally evaluated more positively than negatively by therapists, indicating the acceptability of the tool by therapists. Therapists mentioned benefits such as time efficiency and the ability to obtain more information and details about clients’ CSA experiences. Limitations were more technical in nature, such as the lack of complexity of surroundings and characters in the game environment, certain sexual behaviors that could not be visualized properly, and a feeling thermometer that could not be manipulated. With regard to client acceptability, according to therapists, most clients experienced less tension and more positive (or neutral) emotions than negative emotions when using Vil Du?! to discuss CSA experiences.

**Working Elements of Vil Du?!**

A working element of Vil Du?! that most therapists identified was that the game reduced the barriers to disclosure. This is not surprising considering that many individuals having experienced CSA are faced with barriers to disclosure, such as lack of verbal abilities, shame, guilt, avoidance, and tension. Talking about CSA experiences is known to be difficult for these reasons (among others) [3]. Several other working elements of the game could explain why and how the game reduces which specific barriers. For example, the element that it is not necessary to talk about CSA experiences might specifically reduce the barrier with regard to lack of verbal abilities or vocabulary. Furthermore, the game does not require eye contact between therapist and client and, as such, functions as a *third party in the room* [20], allowing children to express their feelings through an indirect medium [26]. Eye contact might be particularly intolerable for clients who experience guilt, shame, and avoidance [40]; thus, with Vil Du?! sharing their shameful experiences could be more tolerable. Moreover, Vil Du?! presents children with a safe and nonnormative environment that does not dictate right or wrong, which may reduce feelings of self-blame and shame [1,27]. In addition, the game presents the client with a playful environment, and play is known to reduce tension and stress [9].

The play element of the Vil Du?! app might have an effect on several components of the therapy. For example, individuals with CSA experience often have incorrect or maladaptive thoughts related to abuse experiences (eg, “I could have prevented the abuse”) that need to be restructured in therapy. The type of nondirective play that is possible in Vil Du?! itself might be considered a restructuring process [26]. Playing out one’s experience through Vil Du?! might also help to make thoughts related to one’s CSA experiences explicit [21,22]. These explicit thoughts can be processed or restructured in successive therapy sessions. In addition, play that realistically depicts children’s traumatic experiences, which might be possible with the explicit icons and the characters in Vil Du?!, can be used to achieve gradual exposure of clients to traumatic memories that might otherwise be avoided or suppressed [23]. Finally, playing out one’s experiences through Vil Du?! might be the first step toward creating a narrative of one’s CSA experience [28]. Creating such a narrative is most often done by drawing or writing and is experienced by children as the most helpful part of therapy [28,29].

Therapists also identified other working elements not directly linked to the barriers to disclosure. The first was that the child was in control in the game. Therapists acknowledged the importance of children regaining control over their CSA experiences. Giving children control in therapy sessions is an important factor for achieving progress in therapy [25] and helps children regain a sense of control over their lives [24]. Giving children control through Vil Du?! could be accomplished by
aspects inherent to the game, such as the client’s use of the Time-Out button or control over what to show to the therapist. Therapists also actively invited clients to take control in operating the game. Giving the clients control in operating the game fits with children’s experience and frequent involvement with apps and video games [41].

Congruence of Vil Du?! with children’s digital experience was another working element that was mentioned by the therapists. Vil Du?! was played on the 2 tablets. Especially for young children, the touchscreen technology of tablets offers a mode of interactive experience that is congruent with children’s natural ways of learning and exploring (ie, touch, repeat, trial, and error) [42]. In addition, for adolescents, the use of hand-held devices, such as tablets, mirrors the way they are used to communicating with their peers through social media [43].

Finally, one therapist mentioned the interactive nature of the game as a working element. Vil Du?! enables interactivity between the therapist and the client. Both therapist and client play the game on their own tablet. The tablets are synchronized to each other, so actions performed on one screen are also visible on the other. More specifically, both the therapist and client can represent a character in Vil Du?! and each character can perform actions on the other character, as well as respond to the actions of the other character. Researchers have suggested that interactivity in games “can bring a reciprocal effect to other participants of the communication process by turn-taking, feedback, and choice behavior” [44]. Instant reactions from another player might create quick feedback loops to provoke deeper thinking and learning, as well as making sense of previous experiences [45].

Acceptability of Vil Du?! by Therapists and Clients
In terms of the evaluation of Vil Du?! by therapists, findings show that the game might be acceptable for therapists as well as clients. For therapists, the game increased the efficiency of the therapy process and simultaneously led to the discovery of more details and information about the clients’ CSA experiences. The time efficiency of the use of serious games in psychotherapy for children has also been suggested by other scholars [46,47]. Time efficiency of Vil Du?! could be because of the visual nature (ie, showing is faster than telling) or because of the safe environment Vil Du?! created by allowing children to express their feelings through an indirect medium [26]. The large amount of details and information gathered with Vil Du?! is similar to research showing that children provided more detail about emotionally laden events when they could draw or re-enact the event compared with when they were simply asked to tell [48]. An explanation could be that the icons in Vil Du?! may have provided the client with additional and effective retrieval cues for their CSA-related memories. According to therapists, the most important benefit for clients of discussing CSA experiences with Vil Du?! is a possible reduction in negative emotions and tension compared with just talking about one’s experiences. Less negative experiences during the treatment process are associated with more favorable treatment outcomes and better treatment adherence [35,36].

These benefits seem to outweigh the limitations of the game mentioned with regard to the lack of complexity of characters, surroundings, and icons. These limitations were mainly technical in nature and could be incorporated into an updated version of the game. Other limitations might be specific to certain therapists or clients. For example, as mentioned by one older therapist, the game might be difficult to use mainly for older therapists, because another younger therapist recounted the user-friendliness of the game, particularly as a benefit. For older therapists or less technically skilled therapists, a user manual for Vil Du?! is particularly pertinent and could increase usability.

Limitations of the Study and Directions for Future Research
An important limitation of this study was that we focused mainly on the therapists’ perspective on the use of Vil Du?!; even though we also asked therapists about their clients’ experiences. However, clients themselves might evaluate the use of the game in therapy differently. At the start of the study, we aimed to include the clients’ perspective more by videotaping and coding therapy sessions in which Vil Du?! was used. Even though this part of the study was ethically approved, we were unable to obtain consent from enough clients and their parents to include this method in our study. Future research could include clients’ perspectives by asking clients to fill out short, age-appropriate evaluation questionnaires following the use of a serious game in psychotherapy.

A second limitation relates to the questionnaire we used to assess client acceptability, which was composed of only 3 items. Although these items have been used in previous research [20], 3 items may have been limited in their ability to fully assess client acceptability. Future evaluations of Vil Du?! could use more extensive measures of client acceptability, such as the Child Evaluation Inventory [49].

A third limitation is that we were not able to include therapists that represented a wide range in terms of the therapy type they used and the organization they worked for. The majority of participating therapists were working for the same organization and, thus, with the same treatment protocols. This may have limited the ways in which therapists used the Vil Du?! app and, subsequently, the possible working elements that were identified.

It is important to note that there were some differences in the working elements and limitations identified in the questionnaire data and the interview data. Overall, the percentages of working elements and limitations mentioned by therapists were lower in the questionnaires than in the interviews. These differences could be because of the different formats of the 2 methods. In the semistructured interviews, therapists could freely and extensively discuss their experiences with using Vil Du?! In the web-based questionnaire, therapists had to write down their experiences and, consequently, their answers were generally short. However, the relative importance of specific working elements and limitations were very similar across methods. The working elements and limitations that were most often mentioned in the questionnaires were also the ones most often mentioned in the interviews.

Finally, we identified several working elements of how Vil Du?! was used by therapists. However, we could only speculate about
the reasons for the effectiveness of these elements in the treatment of CSA victims. Future experimental or longitudinal research should examine the effectiveness of incorporating Vil Du?! in CSA treatment. These studies could also examine the specific working elements underlying the effects Vil Du?! may have on the therapy process and client outcomes. However, to take this next step in evaluating the effects of Vil Du?! the findings of this study will be incorporated into a user manual for Vil Du?! With this manual, therapists can use the game in a more systematic way (access to the manual and Vil Du?! can be arranged through the authors).

Conclusions

This exploratory evaluation of Vil Du?! provided promising results for the incorporation of serious games in therapy sessions with children and adolescents. The case study on Vil Du?! clarifies how a game specifically designed to help children open up about CSA experiences might improve psychotherapy for child and adolescent CSA victims. Using Vil Du?! appears to be time efficient, and the game appears to make the treatment process less confronting and difficult for the client. In this study, several working elements of Vil Du?! were identified, which could explain possible improvements in the therapeutic process. First, Vil Du?! enables clients to regain control over their CSA experiences. Second, the game offers a nonverbal communication tool that empowers clients with additional vocabulary (e.g., interactions on the screen). Disclosure of clients’ experiences is furthermore enabled by creating a safe environment in which the client can work on their own tablet and does not need to make eye contact with the therapist. The increased amount of vocabulary, control, and safety might result in a significant time reduction in psychotherapy. This benefits both the client and health care system in general. Rigorous experimental effect studies are now necessary to test whether implementing the serious game Vil Du?! in therapy is effective in reducing barriers to disclosing CSA experiences and in regaining control over these experiences.

Acknowledgments

This research was supported by the Fonds Wetenschappelijk Onderzoek Seksualiteit (18.015) and the KF Hein fund. The funding organizations were not involved in the review and approval of the final manuscript.

Authors’ Contributions

JJE, M Deen, and M Dekovic conceptualized the study. HT conducted the interviews. JJE and HT analyzed and interpreted the data. JJE drafted the manuscript. HT, M Deen, and M Dekovic critically revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

M Deen is the designer of Vil Du?! The other authors declare no conflicts of interest.

References


15. Andersson M. What key design features can be identified in creating a tool/game for helping children open up about sexual abuse through the collaboration with health care professionals?. Skövde, Sweden: University of Skövde; 2019.


Abbreviations

CSA: child sexual abuse

©Joyce Johanna Endendijk, Henny Tichelaar, Menno Deen, Maja Deković. Originally published in JMIR Serious Games (https://games.jmir.org), 03.08.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
The Effects of Exergaming on Sensory Reweighting and Mediolateral Stability of Women Aged Over 60: Usability Study

Mariann Sápi1*, BSc, MA; Anna Fehér-Kiss2, MSc; Krisztina Csernák3, MA; Andrea Domján4*, PhD; Sándor Pintér1, MD, PhD

1Doctoral School of Clinical Medicine, University of Szeged, Szeged, Hungary
2Physiotherapy Center, Albert Szent-Györgyi Health Center, University of Szeged, Szeged, Hungary
3Department of Psychiatry and Psychiatric Clinic, Bács-Kiskun County Hospital Kecskemét, Kecskemét, Hungary
4Department of Physiotherapy, Faculty of Health Sciences and Social Studies, University of Szeged, Szeged, Hungary
*these authors contributed equally

Corresponding Author:
Andrea Domján, PhD
Department of Physiotherapy
Faculty of Health Sciences and Social Studies
University of Szeged
Temesvári krt. 31
Szeged, 6726
Hungary
Phone: 36 62 54 6411
Email: domjan.andrea.agnes@szte.hu

Abstract

Background: Older adults tend to experience difficulties in switching quickly between various reliable sensory inputs, which ultimately may contribute to an increased risk of falls and injuries. Sideward falls are the most frequent cause of hip fractures among older adults. Recently, exergame programs have been confirmed as beneficial tools for enhancing postural control, which can reduce the risk of falls. However, studies to explore more precisely which mechanism of exergaming directly influences older women’s ability to balance are still needed.

Objective: Our aim was to evaluate, in a single-group pretest/posttest/follow-up usability study, whether Kinect exergame balance training might have a beneficial impact on the sensory reweighting in women aged over 60.

Methods: A total of 14 healthy women (mean age 69.57 [SD 4.66] years, mean body mass index 26.21 [SD 2.6] kg/m²) participated in the study. The volunteers trained with the commercially available games of Kinect for Xbox 360 console 3 times (30 minutes/session) a week over a 6-week period (total of 18 visits). Participants’ postural sway in both the anteroposterior (AP) and mediolateral (ML) directions was recorded with NeuroCom Balance Master 6.0. To assess and measure postural sensory reweighting, the Modified Clinical Test of Sensory Interaction in Balance was used, where volunteers were exposed to various changes in visual (eyes open or eyes closed) and surface conditions (firm or foam surface).

Results: In the ML direction, the Kinect exergame training caused a significant decrease in the sway path on the firm surface with the eyes open ($P<.001$) and eyes closed ($P=.001$), and on the foam surface with the eyes open ($P=.001$) and eyes closed ($P<.001$) conditions compared with baseline data. The follow-up measurements when compared with the baseline data showed a significant change in the sway path on the firm surface with the eyes open ($P<.001$) and eyes closed ($P<.001$) conditions, as well as on the foam surface with the eyes open ($P=.003$) and eyes closed ($P<.001$) conditions. Besides, on the firm surface, there were no significant differences in sway path values in the AP direction between the baseline and the postraining measurements (eyes open: $P=.49$; eyes closed: $P=.18$). Likewise, on the foam surface, there were no significant differences in sway path values in the AP direction under both eyes open ($P=.24$) and eyes closed ($P=.84$) conditions.

Conclusions: The improved posturography measurements of the sway path in the ML direction might suggest that the Kinect exergame balance training may have effects on sensory reweighting, and thus on the balance of women aged over 60. Based on these results, Kinect exergaming may provide a safe and potentially useful tool for improving postural stability in the crucial ML direction, and thus it may help reduce the risk of falling.
Introduction

Slipping, tumbling, or any other kind of an unintentional loss of balance, which results in a fall and subsequent hospitalization due to injury, is a serious global concern for people over the age of 60 according to the World Health Organization [1]. It has been shown that age-related deficits can manifest in cognitive function [2], in neuromuscular control mechanisms [3,4], and in the following 3 sensory systems: the visual [5], the somatosensory [6], and the vestibular [7]. Various studies have shown that older individuals have a tendency to use proprioception rather than visual and vestibular cues for postural motor control. This dependence on the proprioceptive system also increases with age [3,8]. In direct contrast to this, Haibach et al [9] found that older adults tend to rely more heavily upon their visual input rather than the other sensory systems to compensate for age-related deficiencies.

According to previous studies [10,11], adults tend to experience difficulties in switching quickly between various reliable sensory inputs, which ultimately may contribute to an increased risk of falls. However, Allison et al [12] suggest that this particular process is not impaired among the target population as a direct result of aging. Regardless of whether sensory reweighting deteriorates or remains unchanged with age, therapists should aim to plan programs that can develop these previously mentioned sensory systems and thus decrease the risk of falls.

It has been confirmed that community-dwelling women over the age of 65 are at least two times as likely to suffer hip fractures due to a fall when compared with men [13]. In one study [14], osteoporosis-related fractures in Hungary were investigated and offered incidence data not only on hip, but also on several fractures between 1999 and 2003, when the total population was approximately 10 million inhabitants. According to the data reported in this 5-year period, 404,380 Hungarian women and 206,009 men over the age of 50 had at least one fracture. A possible reason behind this phenomenon might be attributed to the difference between each gender’s change in the level of sex hormones during various stages in life. The changes may contribute to older women having a more significant decrease in bone mineral density [15]. Besides age-related hormonal changes, multitasking increases women’s gait variability, and this has a direct relationship to the prevalence of falls [16]. Furthermore, elderly women with an abnormal balance while walking are more likely to fall [17]. According to the findings of Qazi et al [18], a static posturography test demonstrated that the mediolateral (ML) component of postural sway is most strongly associated with long-term fracture risk in postmenopausal women. In addition to that, sideward falls are the most frequent cause of hip fractures among older adults [19], meaning that it is of key importance to detect with posturography the quantifiable information on body sway that cannot be visible to the clinicians’ naked eyes [20]. Signs of instability are sometimes not immediately apparent in the clinical setting, but sensitive measurements, such as postural sway, can predict the likelihood of falls [18]. For this reason, it is essential to implement training programs that improve sensorimotor control in the critical ML direction.

Recently, exercise games that are played in a virtual, but realistic environment (exergames) have become popular in various fields of research. The use of different types of virtual reality (VR) systems has been considered a beneficial method to improve health gains in different populations and pathological conditions [21]. According to current systematic reviews, video game–based trainings help support physical health [22-25] and cognitive functions among older adults [26-28]. In the last decade nonimmersive VR (without the use of a head-mounted device) exergame trainings with the Kinect system have been proven to be favorable in improving postural control among older adults [29-33]. A recent study revealed significant effects on balance in older adults who had VR exercise training versus an inactive control group [34], as well as a conventional exercise training group [35]. However, the exact mechanism of action of exergaming in improving the balance ability of older adults is a complex process that remains unclear [26]. Thus, in order to provide sound recommendations for their clinical use, the authors suggested conducting further studies to explore more precisely which mechanism of exergaming directly influences an older individual’s ability to balance (in other words, what are the causes of the observed changes or what are the improvements from exergame interventions).

It has been suggested that one of the underlying effects of exergames might originate from sensory reweighting. Body sway–based assessments such as the Sensory Organization Test or the Modified Clinical Test of Sensory Interaction in Balance (m-CTSIB) are sensitive tools for measuring sensory feedback reactions and processes during static stance. These measurements can confirm changes in sensory reweighting following exergaming in patients with Parkinson disease [36], in healthy and young adults [37,38], in older adults [39], in healthy women [40], and in women with fibromyalgia [41].

In the past 3 years, the effect of exergaming on sensory reweighting among older women has received little attention despite its clinical importance for physiotherapists. Because of the limited number of studies available on this topic, this usability study is focused on examining the potential effects of a Kinect exergame training on sensory reweighting and balance in the ML direction in healthy older women.

Methods

Participants

For the purpose of this study, healthy, community-dwelling older women above the age of 60 were recruited via local announcements in the senior centers within the city of Szeged, Hungary.
Hungary. Exclusion criteria included self-reported comorbidities (such as cognitive impairment; disorders of the heart; circulatory, musculoskeletal, and respiratory ailments; autoimmune diseases; and neurological conditions), hearing or vision loss, prosthetics or artificial limbs, wounds or corns on lower extremities, and the use of medication that could affect balance or participation in other organized physical training exercise programs. Twenty active, community-based volunteers signed up for the training program; however, due to the exclusion criteria, only 14 of them could participate in the study. This study was performed according to the Declaration of Helsinki and was approved by the Ethics Committee of the University of Szeged, Hungary (registration No. 125/2015 SZTE). All participants gave their signed informed consent before participating in the training program.

Training Protocol
The applied equipment consists of a motion-sensing RGB camera named Kinect (v1), an Xbox 360 console, and video games developed by Microsoft. During the training, pictures of the game’s scene and a player’s avatar were projected onto the wall via the camera’s full-body 3D motion capture. Before the training program commenced, volunteers had not had any experience with exergames or any of the previously mentioned devices, and so it was important to have an introductory meeting prior to the first training session where instructions were given on how to play the gesture-controlled video games and an opportunity to experience them first-hand. The training took place at Albert Szent-Györgyi Clinical Center’s Physiotherapy Department 3 times a week over a 6-week period (total of 18 visits). These sessions were assisted by physiotherapists. Participants were instructed to wear a comfortable outfit and safe footwear for the 30-minute training. Games were chosen based on the type of movements their performance required, with the main aspect being that games had to contain patterns of everyday functional movements which modeled usual, frequent natural motions. Commercially available Kinect games were played by the participants which demanded continual displacement of the participants’ center of gravity (COG), transference of weight between lower limbs, and lateral trunk bending and frequent sidesteps. The motor stimulation during gameplay required balanced reactions and continuous postural adjustments associated with fast movement of the legs and arms. During the first half of the training sessions, games that consisted of more foreseeable movements and simple elements (eg, football, skiing) were played. Other more complex games that needed higher cognitive attention and fast reaction (20,000 Leaks, Space Pop, Reflex Ridge, River Rush) were selected to be played in the second half of the training sessions. All participants played the same type of games in pairs, in the same order on every training occasion, but were never allowed to play the same game on 2 consecutive training sessions. During the training, the players’ adaptation and progression, as well as the level of difficulty of the game, were continuously recorded and modified based on each participant’s overall ranking in the game. Between each game, there was approximately a 1-minute transition time so that players could take a short break.

Measurement
In general, in order to assess an individual’s ability to both integrate various senses of balance and compensation, while 1 or more of these senses may be lacking [33], NeuroCom Balance Master 6.0 (Clackamas) and the m-CTSIB [42-44] were used. The posturography measurements were performed at 3 separate intervals: before the first training, after the completion of the training program (posttraining), and 6 weeks after the last training session (follow-up).

The Balance Master 6.0’s software provided the location of both the COG and center of pressure across all tests for the m-CTSIB. The m-CTSIB test was initially developed by Shumway-Cook and Horak [45] to differentiate sensory (somatosensory, visual, and vestibular) inputs involved in postural stability during a steady-state balance assessment, and it explored balance on various surface types, with and without vision, using 4 sensory conditions: (1) firm surface, eyes open; (2) firm surface, eyes closed; (3) foam surface, eyes open; and (4) foam surface, eyes closed. The results provided by the Balance Master 6.0’s software package gave 3 measurements of COG (3 × 10 s) in the anteroposterior (AP) and ML directions [35]. Based on a previous study [43] with elderly females in all 4 sensory conditions, this test had good to excellent reliability of ML (intraclass correlation coefficient 0.88-0.93) and AP path length (intraclass correlation coefficient 0.85-0.90).

For the assessment of balance on the foam surface, a NeuroCom square foam balance assessment pad (size 46 × 46 × 13 cm) was used. During the assessments, the base of support was fixed, and participants stood comfortably barefooted with arms to their side and their feet next to a mark on the platform. The measurements took place in a quiet room away from distractions.

Data Analysis: Sway Path
The following equations were applied to calculate the sway paths in the ML and AP directions

\[
x = \sum_{i=1}^{n} s_x(i) \times \frac{1}{n}
\]

\[
y = \sum_{i=1}^{n} s_y(i) \times \frac{1}{n}
\]

where \( n \) is the total number of samples; \( i \) is the sample number; \( s_x \) is the path length of ML ways; and \( s_y \) is the path length of the AP displacements of COG.

The following statistical analysis was conducted using Statistica 13 software (StatSoft). All sets of data were checked for normal distribution using the Kolmogorov–Smirnov test. Factorial analysis of variance was used to analyze sway data of the m-CTSIB test on firm and unstable (foam) surfaces to evaluate the main effects and the influences of the 2 visual conditions (eyes open and eyes closed) at all 3 time conditions (baseline, after the training, follow-up) as within-subjects factors. All values are given as mean (SD). The post hoc test was the Newman–Keuls test. A level of significance of \( P < .05 \) was applied.
Results

Overall, 14 female volunteers (mean age 69.57 [SD 4.66] years, mean body mass index 26.21 [SD 2.6] kg/m²) participated in the study without any dropouts.

Changes in Sway Path During Quiet Stance in the ML Direction

In the ML direction, the Kinect exergame training caused a significant decrease in the sway path on the firm surface with eyes open (\(P<.001\)) and eyes closed (\(P=.001\)), and on the foam surface with eyes open (\(P=.003\)) and eyes closed (\(P<.001\); Figures 1 and 2). There were no significant differences in sway path values on the firm surface between eyes open and eyes closed conditions during the baseline (\(P=.81\)), after the training (\(P=.30\)), and follow-up (\(P=.48\)) evaluations. However, on the foam surface, results showed a significant interaction of vision \(\times\) time for the sway path (\(F_{2,246}=3.70, P=.02\)). Before the training, the sway path on the foam (unstable) surface with eyes closed was significantly longer (\(P<.001\)), whereas after the training the absence of visual information did not result in a significant increase (\(P=.16\)) of the sway path (Figure 2).

Figure 1. The effect of the Kinect training on sway path (mean [SD]) in the ML direction when standing on the firm surface with open and closed eyes. Statistically significant differences in sway path with eyes open (\(P<.001\)) and eyes closed (\(P=.001\)) postraining conditions compared with the baseline data (asterisk). The follow-up measurements when compared with the baseline data showed statistically significant change in sway path on the firm surface with eyes open (\(P<.001\)) and eyes closed (\(P<.001\)) (asterisk). ML: mediolateral.
Figure 2. The effect of the Kinect training on sway path (mean [SD]) in the ML direction when standing on the foam surface with open and closed eyes. Statistically significant differences in sway path with eyes open ($P=.001$) and eyes closed ($P<.001$) posttraining conditions compared with the baseline data. The follow-up measurements when compared with the baseline data showed statistically significant changes in sway path with eyes open ($P=.003$) and eyes closed ($P<.001$) (asterisk). Statistically significant difference in sway path during baseline measurements with eyes closed ($P<.001$) compared with the eyes open condition (circle). ML: mediolateral.

Changes in Sway Path During Quiet Stance In the AP Direction

On the firm surface, there were no significant differences in sway path values in the AP direction between the baseline and the posttraining measurements (Figure 3; eyes open: $P=.49$; eyes closed: $P=.18$). Likewise, on the foam surface, there were no significant differences in sway path values in the AP direction under both eyes open ($P=.24$) and eyes closed ($P=.84$) conditions. During follow-up measurements, a main effect of vision was noted; in other words, closing the eyes resulted in a significant increase of the sway path ($P<.001$; Figure 3). On the unstable foam surface, a main effect of vision was observed and the absence of visual information significantly increased ($P<.001$) the sway path length in all time conditions (Figure 4).
Figure 3. The effect of the Kinect training on sway path (mean [SD]) in the AP direction when standing on the firm surface with open and closed eyes. No statistically significant differences in sway path values on the firm surface between the baseline and posttraining measurements (eyes open \( P = .49 \) and eyes closed \( P = .18 \)). Statistically significant differences \( (P < .001) \) in comparison with the baseline measurement (asterisk) and in comparison with the open eye condition (circle) \( (P < .001) \). AP: anteroposterior.
Discussion

Principal Findings

Several studies have previously confirmed the beneficial effects of exergames on postural control among older adults [22,23,26,29,31-35]. This usability study shows that a simple Kinect game–based balance training might be beneficial for older women by improving balance in the ML direction. This study also demonstrates that exergaming might have a favorable effect in regards to the specific process of adjusting the sensory contributions to balance control [46], namely, sensory reweighting.

Increased Lateral Stability

Based on our study results, an important finding is that the sway path in the ML direction on firm and foam surfaces, with eyes open and closed, improved statistically significantly, whereas no significant change was detected in the AP direction. However, decreased sway path indicates improved stability in the ML direction, which was concluded by Qazi et al [18] as the strongest component of postural sway predicting fractures in postmenopausal women. According to previous studies in the elderly population [47-49], ML sway can often be associated with risk of falls due to decreased proprioception and lower extremity muscle weakness in the lateral direction [50]. In light of the present findings following the training, improved sway results in the ML direction were observed when participants were standing on the foam surface with their eyes open. Significant decrease of ML sway might also implicate an improvement in proprioceptive function following the Kinect training. This finding is similar to the results of Sadeghi et al [51], which suggest that Kinect exergaming can improve proprioception by providing visual feedback and challenging motor skills and visual coordination.

Improvement in Sensory Reweighting

An important finding of this paper is that the Kinect exergame training program significantly reduced postural sway on the foam surface with the eyes closed. Under this condition of the m-CTSIB, the central nervous system mostly relies on vestibular information [45]. In the review by Tahmosybayat et al [52] no exergame study has been presented that would train and assess sensory integration and sensory reweighting. Moreover, the authors suggested that the elements of sensory integration are too unsafe to be trained by disturbed sensory inputs during exergames. However, Roopchand-Martin et al [39] have examined the changes in m-CTSIB results following the Nintendo Wii Fit balance training in community-dwelling adults aged over 60. They found no significant results on the foam surface with the eyes closed condition after the training. By contrast, a Kinect-based physical exercise balance intervention in women with fibromyalgia has revealed significant improvements in the m-CTSIB with eyes closed on foam.
surfaces [41]. Another study [40] that examined the Wii Fit balance training for healthy women also found similar results: significant sensorimotor improvement in unilateral stance and limb strength. Nitz et al [41] concluded that these findings might not be surprising because the activities included in the Wii Fit training (such as yoga, balance, aerobic, and strength activities) involved considerable single-limb balance requirements and body weight–resistance movements. Although the aforementioned studies [40,41] investigated the effects of exergaming on balance in various sensory conditions with m-CTSIB, sensory reweighting following the trainings has not been proposed in these papers.

In contrast to these results, Yen et al [36] demonstrated that VR balance training significantly improved sensory reweighting in older adults with Parkinson disease when both visual and somatosensory inputs were unreliable. They have suggested that the VR training might be especially beneficial for fall prevention within this high-risk target group, as similar conditions may also occur in reality due to various extrinsic environmental risk factors (inappropriate footwear, poor lighting, slippery surfaces, loose rugs, or uneven steps) [36]. Other studies have found significant improvement in the eyes closed condition on unstable surface following a Wii Fit balance training among young adults [37] and healthy adults [38]. According to these studies [37,38], the reason for the improved vestibular function might be due to the quick displacement of COG in different directions, causing rapid changes in the head position. Similarly, during pretest measurements in this study, closing the eyes on the foam surface resulted in a statistically significant increase in the sway scores in the ML direction; however, posttraining measurements did not show deteriorated sway results. The reason for this might be that after the training, participants shifted to the remaining, accurate source of sensory information, and mainly relied on the vestibular system. Another possible explanation might be that during exergaming, participants had to complete several tasks containing movements such as reaching out and lateral steps while they needed to often change their head position.

Santos et al [53] have suggested that VR therapy enables patients to become immersed in an imaginary world, in which environmental perception is altered by artificial stimuli, thus resulting in a sensory conflict that can act on the vestibulo-ocular reflex (VOR). The central nervous system reacts to vestibular stimuli by reflexes such as the VOR, which stabilizes vision during head motion, and the vestibulospinal reflex, which induces a compensatory body motion to stabilize the head and body, and prevent falls [54]. Thus, types of exergames that require head movements in particular (rotation, lateral flexion, flexion) while players’ eyes are focusing (gazing) at one point can function as VOR training. Based on this study’s results, the applied Kinect games might improve sensory reweighting in favor of relying on vestibular inputs. In this study, while participants were playing the exergames, they had to keep their eyes on the screen while performing various head and limb movements.

Lessons Learned From Kinect Exergames

Games such as 20,000 Leaks, River Rush, Reflex Ridge, Super Saver Football mini-game, Space Pop, and Skiing might especially challenge the VOR because they require frequent head displacement movements. Additionally, these games could also improve stability in the lateral direction because of frequent weight shifting and sideward stepping. According to Swanenburg et al [55], exergaming that requires active stepping movements and that involves moving game projection is usable and facilitates gaze stability during head movements, which resulted in improved gait in healthy older adults. As balance is determined by various factors and maintained by complex processes, designing a balance training program requires precisely defining which target components or systems ought to be trained. Health care professionals might use exergames that could display participants’ changes of postural sway, reaction time, and limit of stability in various directions. Games which can train VOR by gaze stability during head movements should be provided for continuous monitoring to track players’ head movements. As falls occur mostly during activities of everyday life, exergames should be designed to involve functional movements that represent motions from daily life: alternately raising the feet from the ground (eg, stair stepping, stepping out of the bathtub), or reaching movements forward and sideways (eg, taking an item off a shelf below and above shoulder height, cleaning a window, hanging out the washing).

Limitations

This usability study has encountered certain limitations as no sample size calculation was performed, and due to the lack of a control group and the relatively small sample size, the results should be interpreted cautiously. Therefore, these findings are not conclusive. Recruiting volunteers via local paper announcements for exergaming was not sufficient to get the attention we had hoped for. We believe that to attract more participants for future balance training programs, other types of advertisements should be used to generate interest. Posts on social media with video demonstrations and trials could raise interest especially among the youth, who could encourage their older relatives to participate. In this study, only older women participated, but to examine whether there is a gender difference in sensory reweighting following exergaming, future studies should also include a group of male participants. Investigating the effects of exergaming in older individuals with vestibular dysfunctions could also be beneficial, as this population is especially at risk of falling. Although there are studies that have described the positive effects of exergaming on balance ability [56,57] and on higher-order cognitive functions [58] when training independently at home, this study could not have been performed using a home-based exergame program. The reason for that is that the applied commercially available Kinect games are in English and no Hungarian translation is available. Therefore, participants needed assistance with starting and setting up the games, as well as technical help.

Conclusions

In this usability study, women’s sway path decreased in the ML direction not only on the firm surface with eyes open, but also on the foam surface with eyes closed as a result of following
the Kinect exergame training. These findings might support the idea that although the Kinect exergame training did not specifically contain direct challenging sensory tasks (e.g., tilting or unstable surface or closed eyes exercises), the reduced sway results suggest that exergames could additionally result in sensory reweighting. The reason for this might be that the games contained tasks that needed constant gaze stabilizing and frequent head displacements. Therefore, this study’s improved sway results in the ML direction might contribute to decreased risk of falls among older women.

Acknowledgments
We thank our first volunteer Mrs Margit Rigó-Nagy for the initial tryout of Kinect exergaming and her kind support and encouragement during the whole research. The authors are grateful for the help provided in the original study by physiotherapy students Mártá Borsics and Vivien Mészáros, as well as senior volunteers for participating and consenting to be a part of this study. We are thankful to Mónika Szűcs for her help with the statistical analysis. We thank Candace Smith, Katalin Balázs, and Roman Pentsa for their valuable comments and linguistic corrections. The authors declare that the measurements reported in this paper fully comply with all the present Hungarian laws and regulations. This study was supported by the EU-funded Hungarian grant EFOP3.6-1-16-2016-00008.

Authors’ Contributions
SP and AD share senior authorship.

Conflicts of Interest
None declared.

References


Abbreviations  
AP: anteroposterior  
COG: center of gravity  
m-CTSIB: Modified Clinical Test of Sensory Interaction in Balance  
ML: mediolateral  
VOR: vestibulo-ocular reflex
VR: virtual reality

Edited by N Zary; submitted 11.02.21; peer-reviewed by M Jordan-Marsh, P Backlund; comments to author 16.04.21; revised version received 30.04.21; accepted 29.05.21; published 21.07.21.

Please cite as:
Sápi M, Fehér-Kiss A, Csernák K, Domján A, Pintér S
The Effects of Exergaming on Sensory Reweighting and Mediolateral Stability of Women Aged Over 60: Usability Study
JMIR Serious Games 2021;9(3):e27884
URL: https://games.jmir.org/2021/3/e27884
doi:10.2196/27884
PMID:34287215

©Mariann Sápi, Anna Fehér-Kiss, Krisztina Csernák, Andrea Domján, Sándor Pintér. Originally published in JMIR Serious Games (https://games.jmir.org), 21.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Original Paper

Development of a Search Task Using Immersive Virtual Reality: Proof-of-Concept Study

Samuel Elia Johannes Knobel¹, MD; Brigitte Charlotte Kaufmann²,³, PhD; Stephan Moreno Gerber¹, PhD; Prabitha Uwyler⁴, PD; Dario Cazzoli¹,³, PD; René M Müri¹,²,⁴, Prof Dr; Tobias Nef¹,⁴,⁵, Prof Dr; Thomas Nyffeler¹,²,³, Prof Dr

¹Gerontechnology & Rehabilitation Group, University of Bern, Bern, Switzerland
²Perception and Eye Movement Laboratory, Departments of Neurology and BioMedical Research, Inselspital, Bern University Hospital, Bern, Switzerland
³Neurocenter, Luzerner Kantonsspital, Lucerne, Switzerland
⁴Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
⁵ARTORG Center for Biomedical Engineering Research, University of Bern, Bern, Switzerland

Corresponding Author:
Thomas Nyffeler, Prof Dr
Neurocenter
Luzerner Kantonsspital
Spitalstrasse
Lucerne, 6004
Switzerland
Phone: 41 205 56 86
Email: thomas.nyffeler@luks.ch

Abstract

Background: Serious games are gaining increasing importance in neurorehabilitation since they increase motivation and adherence to therapy, thereby potentially improving its outcome. The benefits of serious games, such as the possibility to implement adaptive feedback and the calculation of comparable performance measures, can be even further improved by using immersive virtual reality (iVR), allowing a more intuitive interaction with training devices and higher ecological validity.

Objective: This study aimed to develop a visual search task embedded in a serious game setting for iVR, including self-adapting difficulty scaling, thus being able to adjust to the needs and ability levels of different groups of individuals.

Methods: In a two-step process, a serious game in iVR (bird search task) was developed and tested in healthy young (n=21) and elderly (n=23) participants and in a group of patients with impaired visual exploration behavior (ie, patients with hemispatial neglect after right-hemispheric stroke; n=11). Usability, side effects, game experience, immersion, and presence of the iVR serious game were assessed by validated questionnaires. Moreover, in the group of stroke patients, the performance in the iVR serious game was also considered with respect to hemispatial neglect severity, as assessed by established objective hemispatial neglect measures.

Results: In all 3 groups, reported usability of the iVR serious game was above 4.5 (on a Likert scale with scores ranging from 1 to 5) and reported side effects were infrequent and of low intensity (below 1.5 on a Likert scale with scores ranging from 1 to 4). All 3 groups equally judged the iVR serious game as highly motivating and entertaining. Performance in the game (in terms of mean search time) showed a lateralized increase in search time in patients with hemispatial neglect that varied strongly as a function of objective hemispatial neglect severity.

Conclusions: The developed iVR serious game, “bird search task,” was a motivating, entertaining, and immersive task, which can, due to its adaptive difficulty scaling, adjust and be played by different populations with different levels of skills, including individuals with cognitive impairments. As a complementary finding, it seems that performance in the game is able to capture typical patterns of impaired visual exploration behavior in hemispatial neglect, as there is a high correlation between performance and neglect severity as assessed with a cancellation task.

(JMIR Serious Games 2021;9(3):e29182) doi:10.2196/29182

KEYWORDS

virtual reality; serious game; search task; stroke; neglect; usability; development; immersion; concept; gaming

https://games.jmir.org/2021/3/e29182

JMIR Serious Games 2021 | vol. 9 | iss. 3 | e29182 | p.62
Introduction

A crucial component of neurorehabilitation and its success is therapy adherence and repetition. One promising possibility to improve patients’ therapy adherence, and thereby its outcome, is to enhance training motivation by means of serious games [1] and immersive virtual reality (iVR) [2,3]. The combination of serious games and iVR allows for the investigation of visual exploration behavior, which is highly relevant in activities of daily living and therefore frequently used in the diagnosis and rehabilitation of patients suffering from a neurodegenerative disease [4-7].

The primary purpose of serious games is not to be fun, but to teach, train, or assess skills in an entertaining way [8]. Due to technological achievements in recent years, their importance in education [9] rehabilitation [10], and medical training [11] is growing, as they have several advantages. First, serious games are standardized, which means that each user will experience the same task. Second, different game settings offer the possibility to automatically adapt the task difficulty to the user’s individual skills and performance level. Finally, serious games offer the possibility to track the user’s game performance (eg, achieved scores) and game behavior as a measure for activities of daily living (eg, reaction times) [12].

The different settings and measured performance values have been used to implement adaptive difficulty scaling, which has shown to be an important element of serious games leading to 2 key elements [13-16]. First, a balance point (ie, when game difficulty is still challenging, yet does not exceed the player’s abilities) can be achieved. Second, changes in game mechanics should allow adapting the difficulty throughout the game, to continuously and optimally match the increasing skill level of the player [17]. Importantly, both elements are essential to keep the player motivated [18].

The benefits of serious games can be further improved by the use of iVR [19-21]. iVR presents computer-generated artificial, but interactable (ie, hand-held controllers), 360° environments or pre-rendered 360° videos inside a head-mounted display (HMD). With technological improvements over the last decade, iVR is now also increasingly used in clinical applications such as in motor rehabilitation for gait and balance [22,23], surgery training [24-26], or anxiety treatment [27]. Particularly in tasks that involve any kind of motor activities, iVR has some advantages, such as the possibility to objectively measure progressive improvement in trained skills, perform task-oriented repetitive training, and apply multisensory feedback and task variation [28,29]. Rizzo et al [12] summarized evidence showing that skills gained in the iVR environment can be transferred to activities of daily living (eg, crossing the street [30]) reflecting the ecological validity of tasks in iVR. The high ecological validity can be explained due to fewer distractions from external stimuli and the intuitive interaction with the virtual environment, whereas the interaction and thus behavior in the virtual environment can be tracked by recording head and hand movements. Conclusively, tasks in iVR tend to feel more naturalistic, and several studies have shown that the naturalistic feeling of a task correlates with higher enjoyment, better performance, and better motivation [31,32]. This naturalistic feeling is created by the so-called immersion [33-35] (ie, a situation in which the real world is ignored in favor of the virtual environment [36]).

Visual exploration behavior is a crucial element of activities of daily living (eg, crossing the street, grocery shopping) [37] and corresponds to purposefully looking around in the present environment (ie, actively acquiring visual information through coordinated movements of the eyes and head [38]). Therefore, an impaired visual exploration behavior could lead to a reduction of performance in activities of daily living and thus in quality of life [39,40]. Impairment of the visual search behavior is also one landmark of patients suffering from hemispatial neglect [37,41]. Hemispatial neglect is a visuospatial attention disorder that frequently occurs after a right hemispheric stroke. Its characteristic is the inability to attend or respond to stimuli presented within the left contralesional space [42].

Therefore, the aim of this proof-of-concept study was to develop a serious game using iVR in which participants perform a visual search task that encourages the exploration of their environment. We hypothesized that the given task has high usability and limited or no side effects and that the performance can be adapted dynamically to the skills of the participants.

Methods

The main goal of this study was to develop a gamified search task that encourages players to explore their visual environment. We named it the bird search task, and development and evaluation were divided into 2 steps.

First, the 2D game “Crazy Chicken” (ak tronic Software & Services gmbh, Saerbeck, Germany) was used as inspiration, and the game mechanics were transferred to a 3D iVR environment and then modified and tested with healthy young and elderly participants. Based on the findings of the first step, in the second step, the task was further adapted and tested with patients with hemispatial neglect after a right hemispheric stroke.

Game Development and Apparatus

Bird Search Task for Healthy Participants

The 2D game “Crazy Chicken” was identified as a suitable gamified task because it encourages players to explore their visual environment. In the original “Crazy Chicken” game, the chickens (visual targets) appear at random locations on the computer screen and fly at a constant velocity in random directions. The player, by constantly scrolling left and right on a scrollable, 2D screen, has to search and tag the targets before they disappear after a constant time delay.

To transfer the task into 3D, a simple virtual environment was designed using the gaming development platform Unity3D [43]. The environment was built as a wide circular area surrounded by trees and mountains, as shown in Figure 1 and Figure 2. A mobile gaming laptop was used to render the virtual environment (HP-Omen, graphic-card NVIDIA GTX1050 and CPU Intel i7).

https://games.jmir.org/2021/3/e29182
Figure 1. Immersive virtual reality game, with (A) the participant’s field of view (yellow) within the head-mounted display, which moved if the player (blue) turned his or her head, and the area where the target could appear (dashed line; spawn area), which was locked to the midsagittal plane; and (B) a schematic representation of a participant wearing the head-mounted display and performing the task.

Figure 2. Exemplary scene of the gameplay, (A) as implemented in step 1 (healthy individuals) and (B) after the modifications performed in step 2 (patients with right hemispheric stroke and hemispatial neglect).

The iVR hardware consisted of an HMD and a hand-held controller (HTC Vive, High Tech Computer Corporation, Taoyuan, Taiwan; resolution of 2160 x 1200 pixels; full HD; horizontal and vertical field of view of 110 degrees [yellow area in Figure 1A]; frame rate of 90 Hz). The x-y position of the handheld controller was continuously recorded. Using iVR hardware, 360° of the virtual environment could be explored by moving the eyes and turning the head. The targets (eg, birds) could be tagged by aligning the handheld controller with the target and simultaneously pressing a trigger on the controller. Target appearance was set to randomly take place within a restricted area in front of the player (± 60° horizontally and ± 50° vertically, as defined with respect to the trunk’s midsagittal plane; see dashed lines in Figure 1A). In order to promote exploration of the whole extent of the virtual environment, the horizontal area in which the targets could appear was 10° larger than the player’s field of view (ie, additional 5° on either side; Figure 1A).

To alert the player of a new appearing target, a short (1 second) auditory signal (chicken cackle) was presented binaurally via headphones (XQISIT oE400, Strax Americas Inc, Miami, FL). Then, the target appeared at a position randomly determined within the spawn area and flew horizontally towards the right or left (direction randomly selected) with a constant velocity of 2 °/second. If the target was detected and successfully tagged, it fell vertically and disappeared. In case the target was not tagged within the maximum presentation time of 15 seconds, it disappeared. In either way, after a fixed interstimulus interval of 2 seconds, a new target appeared, again with an alerting cackle.

The game was played in rounds. Each of the 10 played rounds consisted of 30 targets. For each target, the time until it was found was measured, and the percentage of found targets was calculated for each round.

We modified the task further by integrating several difficulty levels, to be able to adapt the task to participants with different needs and skills.

We implemented 15 difficulty levels, in which changes in difficulty were achieved by manipulating both target behavior and task mechanics. For each of the 15 levels, the values of the settings changed stepwise, according to the “Change Per Level” threshold presented in Table 1. Concerning target behavior, both the maximum lifetime and speed of the targets were manipulated across levels. For instance, in the easiest level (Level 1), the targets moved with a speed of 2 °/second and were presented for 15 seconds before they disappeared; in the most difficult level (Level 15), the targets moved with a speed of 35 °/second and were presented for 4 seconds before they disappeared. Hence, the more difficult the level, the faster the participants had to explore the visual environment in order to find the targets before they disappeared. Concerning task mechanics, the task adapted its difficulty (ie, changed the difficulty level) automatically, based on performance in the previous round. Each level had a level-up threshold (ie, if the threshold value was reached, the next round started in a higher difficulty level) and a level-down threshold (ie, if the threshold
value was not reached, the next round started in a lower difficulty level), defined according to the percentage of found targets. If the percentage of found targets in a particular round was not higher than the level-up threshold and not lower than the level-down threshold, the difficulty in the next round did not change.

For example, in a round at level 6, the targets move with a speed of 13.8 °/seconds, and the maximum lifetime of the targets is 11.07 seconds. Assuming that the participant would find 25 of the 30 targets in this particular round (ie, 83.3% found targets), then the next round would start at level 7 since the percentage of found targets is above the level-up threshold of level 6 (ie, 77.1%). If the player finds only 15 of the 30 targets (ie, 50% found targets), then the next round would start at level 5, since the percentage of found targets would be below the level-down threshold of level 6 (ie, 67.1%).

Table 1. Change per level algorithm.

<table>
<thead>
<tr>
<th>Task and target parameters</th>
<th>Level 1</th>
<th>Change per levela</th>
<th>Level 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lifetime (seconds)</td>
<td>15</td>
<td>−0.786</td>
<td>4</td>
</tr>
<tr>
<td>Speed (°/second)</td>
<td>2</td>
<td>+2.36</td>
<td>35</td>
</tr>
<tr>
<td>Level-up threshold (%)</td>
<td>&gt;70</td>
<td>+1.42</td>
<td>&gt;90</td>
</tr>
<tr>
<td>Level-down threshold (%)</td>
<td>&lt;60</td>
<td>+1.42</td>
<td>&lt;80</td>
</tr>
</tbody>
</table>

aThe stepwise change per level between Level 1 and Level 15.

**Bird Search Task for Patients With Hemispatial Neglect**

In a second step, based on the findings collected in healthy participants and on recommendations of the clinicians, the version for patients was developed. As neglect’s clinical picture is very heterogeneous across patients, the task needs to be easily adaptable to the patients’ individual needs. The aim was that patients would not get overwhelmed and frustrated due to too great a difficulty but also not get bored by a task that was too easy. For this purpose, we implemented an easy-to-use setting file, in which the total number of targets, number of targets per round, and starting difficulty level could be set individually for each patient. Additionally, several adaptations in design and task mechanics were performed. First, as patients with neglect process visual information slower [44,45], the landscape of the gameplay would be too complex; therefore, in order to reduce the number of distractive elements, the trees in the background and the clouds in the sky were removed from the game scenery (see Figure 2). Second, based on the recommendations of clinicians, the minimal speed of the targets was lowered from 2 °/second to 0.1 °/second; this manipulation generally lowered the difficulty level as well as reduced the need for fast head movements that could promote side effects.

**Participants**

The study was approved by the Ethics Committee of the Cantons of Bern and the Ethics Committee of north-west and central Switzerland and was conducted in accordance with the latest version of the Declaration of Helsinki. All participants gave written informed consent before participation.

In step one, the feasibility of the task was assessed with 21 younger healthy participants recruited at the University of Bern (10 women; mean age, 28.1, SD 5.5 years) and with 23 older healthy participants recruited during a chess tournament for seniors (1 woman; mean age, 71.3, SD 6.3 years). All participants had no history of neurological nor psychiatric disorders. Previous VR experience was reported by 16 of the younger participants and 1 of the older participants.

In step two, 11 inpatients with hemispatial neglect (5 women; mean age, 69.6, SD 13.0 years) after right hemispheric, subacute stroke were recruited at the Neuorehabilitation Clinics of the Inselspital, Bern University Hospital (sites Bern and Riggisberg) and of the Kantonsspital Luzern, Switzerland. Demographic characteristics for each patient are presented in Table 2. The study was always conducted with a mobile setup in the place where the patient was currently hospitalized. All patients showed significant left-sided neglect in activities of daily living, as assessed with the Catherine Bergego Scale (CBS, Range 0-30, 0 = normal) [46], and had normal or corrected-to-normal vision. One of the patients reported previous VR experience.
Table 2. Individual demographical and neuropsychological data from step 2 for patients with hemispatial neglect.

<table>
<thead>
<tr>
<th>Patient code</th>
<th>Age range (years)</th>
<th>Gender</th>
<th>Lesion type</th>
<th>Time since stroke (days)</th>
<th>CBSa</th>
<th>CoCb (SNTc single)</th>
<th>Number of played chickens</th>
<th>Play duration (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_26</td>
<td>70-75</td>
<td>Male</td>
<td>Ischemic</td>
<td>85</td>
<td>6</td>
<td>0.184</td>
<td>195</td>
<td>16.5</td>
</tr>
<tr>
<td>P_27</td>
<td>55-60</td>
<td>Male</td>
<td>Hemorrhagic</td>
<td>145</td>
<td>2</td>
<td>0.208</td>
<td>100</td>
<td>10.4</td>
</tr>
<tr>
<td>P_28</td>
<td>80-85</td>
<td>Male</td>
<td>Bleeding</td>
<td>49</td>
<td>8</td>
<td>0.746</td>
<td>50</td>
<td>9.4</td>
</tr>
<tr>
<td>P_29</td>
<td>80-85</td>
<td>Female</td>
<td>Ischemic</td>
<td>52</td>
<td>9</td>
<td>0.998</td>
<td>44</td>
<td>8.1</td>
</tr>
<tr>
<td>P_30</td>
<td>75-80</td>
<td>Male</td>
<td>Ischemic</td>
<td>42</td>
<td>8</td>
<td>0.293</td>
<td>80</td>
<td>9.2</td>
</tr>
<tr>
<td>P_31</td>
<td>50-55</td>
<td>Female</td>
<td>Ischemic</td>
<td>106</td>
<td>3</td>
<td>0.414</td>
<td>80</td>
<td>13.2</td>
</tr>
<tr>
<td>P_32</td>
<td>65-70</td>
<td>Female</td>
<td>Ischemic</td>
<td>59</td>
<td>2</td>
<td>−0.067</td>
<td>80</td>
<td>10.2</td>
</tr>
<tr>
<td>P_33</td>
<td>70-75</td>
<td>Female</td>
<td>Ischemic</td>
<td>58</td>
<td>3</td>
<td>0.824</td>
<td>80</td>
<td>9.3</td>
</tr>
<tr>
<td>P_34</td>
<td>85-90</td>
<td>Male</td>
<td>Ischemic</td>
<td>29</td>
<td>18</td>
<td>0.998</td>
<td>80</td>
<td>8.4</td>
</tr>
<tr>
<td>P_35</td>
<td>50-55</td>
<td>Male</td>
<td>Hemorrhagic</td>
<td>39</td>
<td>10</td>
<td>0.046</td>
<td>150</td>
<td>13.0</td>
</tr>
<tr>
<td>P_36</td>
<td>60-65</td>
<td>Female</td>
<td>Ischemic</td>
<td>38</td>
<td>17</td>
<td>0.191</td>
<td>100</td>
<td>10.1</td>
</tr>
</tbody>
</table>

aCBS: Catherine Bergego Scale (0-30).
bCoC: Center of Cancellation (~1 to 1).
cSNT: Sensitive Neglect Test.

In addition to CBS, where a value ≥1 means the patient has a neglect, the objective neglect severity was assessed on the day of the task by means of the paper-and-pencil Sensitive Neglect Task (SNT), single version [47]. The SNT is a cancellation task in which patients are asked to mark 40 targets among 240 distractors. Based on the distribution of the marked targets, the Center of Cancellation (CoC) [48] was computed, representing an objective measure of neglect severity. The CoC reflects the mean deviation of the marked targets from the center and is normalized to values ranging from −1 to 1. Zero indicates no spatial bias, where a CoC ≥0.081 represents a significant rightward shift (ie, left-sided neglect).

Outcome Measures

Questionnaires

To evaluate the feasibility and usability of the newly implemented task, several questionnaires were used to assess the participants’ individual gaming experience.

To assess acceptance, usability, and participant’s perception of the visual search task and of the VR system, 3 questions from the System Usability Scale (SUS) [49] were used, as previously reported by Gerber et al [50] and Knobel et al [51]. The questions were answered using a 5-point Likert-scale, ranging from “fully disagree” to “fully agree.” Each question represented a different aspect of the perception of the task (ie, motivation, frustration, how challenging it is, entertainment). Therefore, no mean score was calculated across questions; instead, each score was considered independently.

Furthermore, to assess immersion and presence, questions from the Igroup Presence Questionnaire (IPQ) [56] were used. The question subset was already used in another iVR study by Gerber et al [50] and was answered on a 5-point Likert-scale.

The questions from each questionnaire are shown in Table 3.

The Simulator Sickness Questionnaire (SSQ) [52] was used to assess side effects such as cybersickness, oculomotor problems, and disorientation [53,54]. In order to reduce the workload of the study participants, a subset of 7 questions from the SSQ was used, as previously reported by Gerber et al [50] and Knobel et al [51]. The questions were answered using a 4-point Likert-scale (ie, “None,” “Slight,” “Moderate,” “Severe”). Again, the mean score across all questions was calculated for each participant.

Additionally, in order to assess the enjoyment of the task, the Perception of Game Training Questionnaire (PGTQ) [55] was administered. The PGTQ consists of 4 questions that are answered using a 7-point Likert-scale, ranging from “fully disagree” to “fully agree.” Each question represented a different aspect of the perception of the task (ie, motivation, frustration, how challenging it is, entertainment). Therefore, no mean score was calculated across questions; instead, each score was considered independently.
Table 3. Exact formulations that were asked in the questionnaires.

<table>
<thead>
<tr>
<th>Questionnaire, number</th>
<th>Question</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUS&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>I thought the system was easy to use</td>
<td>Usability</td>
</tr>
<tr>
<td>2</td>
<td>I think that I would like to use this system frequently</td>
<td>Usability</td>
</tr>
<tr>
<td>3</td>
<td>I felt very confident using the system</td>
<td>Usability</td>
</tr>
<tr>
<td>SSQ&lt;sup&gt;c,d&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>General discomfort</td>
<td>Sickness</td>
</tr>
<tr>
<td>5</td>
<td>Stomach awareness</td>
<td>Sickness</td>
</tr>
<tr>
<td>6</td>
<td>Sweating</td>
<td>Sickness</td>
</tr>
<tr>
<td>7</td>
<td>Nausea</td>
<td>Sickness</td>
</tr>
<tr>
<td>8</td>
<td>Headache</td>
<td>Oculomotor problems</td>
</tr>
<tr>
<td>9</td>
<td>Eye strain</td>
<td>Oculomotor problems</td>
</tr>
<tr>
<td>10</td>
<td>Dizziness</td>
<td>Disorientation</td>
</tr>
<tr>
<td>IPQ&lt;sup&gt;e,f&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>In the virtual world, I had a sense of “being there.”&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Immersion</td>
</tr>
<tr>
<td>12</td>
<td>Somehow, I felt that the virtual world surrounded me.&lt;sup&gt;h&lt;/sup&gt;</td>
<td>Presence</td>
</tr>
<tr>
<td>PGTQ&lt;sup&gt;i,j&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>I was motivated for a good performance.</td>
<td>Motivation</td>
</tr>
<tr>
<td>14</td>
<td>The game was frustrating.</td>
<td>Frustration</td>
</tr>
<tr>
<td>15</td>
<td>The game was challenging.</td>
<td>Challenge</td>
</tr>
<tr>
<td>16</td>
<td>The game was entertaining.</td>
<td>Entertainment</td>
</tr>
</tbody>
</table>

<sup>a</sup>SUS: System Usability Scale.

<sup>b</sup>Fully disagree to fully agree; score range, 1-5; midpoint, 3; scored as the mean of Q1-Q3.

<sup>c</sup>SSQ: Simulator Sickness Questionnaire.

<sup>d</sup>None to severe; score range, 1-4; midpoint, 2.5; scored as the mean of Q4-Q10.

<sup>e</sup>IPQ: Igroup Presence Questionnaire.

<sup>f</sup>Score range, 1-5; midpoint, 3; each question is scored individually.

<sup>g</sup>Not at all to very much.

<sup>h</sup>Fully disagree to fully agree.

<sup>i</sup>PGTQ: Perception of Game Training Questionnaire.

<sup>j</sup>Fully disagree to fully agree; score range, 1-7; midpoint, 4.5; each question is scored individually.

Performance Indicators During the Task

The presented task allowed us to measure several performance indicators, namely the changes in difficulty levels over time and the mean search time to detect the targets.

Based on the mean search time of the targets, the mean search time per participant was calculated, representing the mean time until a participant tagged a target (not-found targets were excluded). Moreover, the controller position of the VR setup was recorded over the entire task, which allowed us to track the participants’ hand positions over time, hence delivering information concerning their spatial search behavior.

Evaluation of the Difficulty Scaling

There are different possibilities to assess whether an adaptive difficulty scaling is successful. An easy indirect, but less objective, possibility is to simply evaluate the results of the questionnaires concerning entertainment and frustration. A more elaborate and objective approach is to consider the number of level changes over the task rounds with respect to the starting level. A population in which the starting level is much easier than the average balance point should show a greater increase in difficulty (ie, more upward level switches) in the initial phases of the task (ie, when they increase to their balance point) and then a smaller increase over time (ie, when the balance point is reached, but the participants still gradually get better at the task due to practice). The better the starting level matches the abilities of the group, the smaller the difficulty changes should be in the initial phases of the task. Nevertheless, due to practice (eg, better aiming, better search strategies), participants are expected to get better in the task, and there should thus be at least a small difficulty level increase over time.
A plateau reflects the balance point at which the level did not change anymore between rounds. This indicates that participants reached their optimal task difficulty level (ie, their performance was not at the ceiling in a particular difficulty level so that the algorithm would increase it in the next round and not at the floor that the algorithm would decrease the difficulty level in the next round).

**Statistical Analyses**

The mean SUS and SSQ scores, reflecting the usability and number of side effects, respectively, were computed for each group (ie, young, elderly, and stroke). The 3 and 7 items of the SUS and SSQ, respectively, were averaged.

The means of the groups for the PGTQ and the IPQ questions were calculated per question and were displayed as histograms.

The change in level for every participant was computed by subtracting the starting level from the levels they were in the consecutive rounds. This change of level was used as a performance measure and visualized as the level change over time relative to the starting level. The position over time of the hand was illustrated by plotting the controller x-y position over time.

The difference in search time between the young group and elderly group was calculated using a 2-sided $t$ test. For the neglect group, a Pearson correlation was calculated between neglect severity (CoC in stroke group) and the mean search time of targets in the VR game. For the statistical analysis, the alpha was set to 0.05.

All analyses and visualizations were performed with R [57] and MATLAB [58].

**Results**

**Questionnaires**

In step one, the feasibility and usability of the task were assessed in 2 groups: young and elderly healthy participants. The analysis of the SUS scores (mean of the 3 questions; see Table 3) revealed high usability in both groups (Table 4). Both young and elderly participants reported that they would even like to play the task frequently (mean scores: young, 3.95, SD 0.74; elderly, 4.04, SD 1.26).

In general, almost no side effects were reported, as assessed with the SSQ score (mean of the 7 domains; see Table 3). In the young group, the mean score reflected minimal side effects (Table 4). More precisely, only 3 young participants reported severe side effects in 1 of the 7 domains (2 cases of stomach awareness and 1 case of dizziness). In the elderly group, the mean score was similarly low (Table 4). Only 1 elderly participant reported a severe side effect (sweating).

The PGTQ consisted of 4 questions (for details, see Table 3) concerning motivation, frustration, challenge, and entertainment. The participants of both groups were very motivated (mean scores: young, 6.38, SD 1.12; elderly, 6.73, SD 0.46), entertained (mean scores: young, 6.28, SD 0.56; elderly, 5.83, SD 1.70), and not frustrated (mean scores: young, 2.57, SD 0.90; elderly, 1.83, SD 1.44). The extent to which the task was challenging was rated above midline in both groups (mean scores: young, 4.29, SD 0.90; elderly, 5.61, SD 1.23).

The question for immersion (mean scores: young, 3.81, SD 0.75; elderly, 3.96, SD 1.02) and presence (mean scores: young, 4.24, SD 0.70; elderly, 4.48, SD 0.59) were rated high in both healthy groups (Figure 3).

In step two, the feasibility of the task was assessed in patients with neglect. Patients with neglect rated the task as highly usable (Table 4), and almost no side effects were reported on the SSQ (Table 4). More precisely, none of the patients reported any severe side effect. Only 3 patients reported side effects. One patient reported moderate sweating, while 2 others reported mild sweating and mild headache.

According to the PGTQ scores (Figure 3), the patients were highly motivated (mean 6.18, SD 1.17), entertained (mean 6.27, SD 0.79), and not frustrated (mean 1.91, SD 0.70). Furthermore, the degree of challenging score was around the midline (mean 3.82, SD 1.33; ie, most patients rated the task as neither too difficult nor too easy). The patients felt high immersion (mean 3.64, SD 1.43) and presence (mean 4.09, SD 0.83)

---

**Table 4.** Results of the System Usability Scale (SUS), where 1 means “unusable” and 5 means “very usable,” and the Simulator Sickness Questionnaire (SSQ), where 1 means “None” and 4 means “Severe” side effects.

<table>
<thead>
<tr>
<th>Study group</th>
<th>SUS, mean (SD)</th>
<th>SSQ, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young</td>
<td>4.41 (0.49)</td>
<td>1.42 (0.45)</td>
</tr>
<tr>
<td>Elderly</td>
<td>4.46 (0.73)</td>
<td>1.27 (0.24)</td>
</tr>
<tr>
<td>Stroke</td>
<td>4.73 (0.44)</td>
<td>1.05 (0.10)</td>
</tr>
</tbody>
</table>
Figure 3. Igroup Presence Questionnaire results from the healthy participants (young, elderly) and the patients (stroke).

Adaptive Difficulty Scaling

In all 3 groups, the qualitative illustration of the increase in the level differences shows a plateau over time (Figure 4).

In young participants, a steep increase in the level difference (ie, the levels get more difficult) was found at the beginning of the task, reaching a plateau (optimum difficulty) at a level that was close to the possible maximum of 7 increases. There were only 7 increases possible because they started at level 8, meaning they could level up to 15 by increasing their level 7 times. In elderly participants and stroke patients, the increase in the level difference revealed a more moderate and heterogeneous change over time. Both groups needed more rounds to reach their optimum; compared visually, this optimum was lower than in the young group.

Figure 4. Level difference (and standard error) in the number of levels over the round compared to the starting level. In the neglect-group for the bars, only the 6 patients that played for 10 rounds were included. The raw values are included for the other 3 patients. The level difference is the mean of the difference between the level in round x minus the starting level.

In-Game Measures

Controller Position Over Time

During the task, the controller position was continuously recorded and could be extracted for offline analysis. Figure 5 presents the exemplary data of the controller position (highlighted in blue) for one participant per group.

In the exemplary participants of the young group (Figure 5A) and the elderly group (Figure 5B), the controller movements were symmetrically distributed. In the exemplary participants of the stroke group (Figure 5 C), a narrowing of the movement as well as a clear rightward shift can be observed. The patients’ hand movements mainly took place within the right hemispace due to left neglect.
**Search Time**

The mean search time in young (mean 1.533, SD 0.342) and elderly (mean 1.694, SD 0.275) participants did not significantly differ ($t_{38.4}=1.71$, $P=.095$). There was a significant, strong correlation between the mean search time and neglect severity (ie, the CoC in the SNT; $r=0.70$, $P=.037$) [59]. An additional analysis investigating the correlation between search time per hemispace (ie, targets appearing within the left vs right hemispace) and neglect severity revealed a hemispace-dependent effect (Figure 6). A significant correlation was found for targets appearing within the left hemispace ($r=.809$, $P=.008$; strong correlation [59]). However, for targets appearing within the right hemispace, no significant correlation was found ($r=.353$, $P=.351$).

**Discussion**

For this study, we developed and evaluated a dynamic visual search task in iVR. As hypothesized, the task could easily be adapted to different skills, impairments, and experience levels. Further, the developed system has high usability and acceptance and resulted in only slight or no side effects in healthy participants and in right-hemispheric stroke patients with hemispatial neglect.

**Questionnaires and Feedback**

The ratings for usability and acceptance of the bird search task were high in both healthy participants and stroke patients with hemispatial neglect. This result is in line with previous findings from studies investigating the acceptance of HMD-VR in elderly [60,61], studies using VR for neglect diagnostics [51], and studies in which a similar VR setup was used to apply audio-visual stimulation in critically ill patients [62].

Regarding side effects, overall, only slight to none were reported. Furthermore, the nature and intensity of side effects
were in line with the literature [51,54,62]. In the group of patients with neglect, no severe side effects were reported (only 1 participant reported moderate sweating). This is potentially due to the game adaptations (eg, slower targets require slower head movements), simplification, and individual adjustments performed for the patients in the step two. This low rate of side effects is also in line with the high values for immersion and presence, as those correlate negatively [63].

The game experience was assessed by means of a questionnaire with 4 parts (frustration, degree of challenge, entertainment, motivation). Indeed, motivation, entertainment, and low frustration levels are crucial aspects for any possible future application [64]. In the bird search task, all participants were highly motivated, entertained, and even though not all participants found all targets, they did not feel frustrated. In particular, the group of patients with neglect had high levels of motivation and entertainment and low frustration levels due to 2 possible reasons. First, this might be the result of the adaptive difficulty scaling; second, this could be the result of the ability to individualize the starting conditions to the patients’ needs and abilities. According to the literature, high motivation is also in line with the high values for immersion the participants reported [31,32,65,66].

Interestingly, the question as to whether the task was challenging was answered very differently among the 3 groups. While elderly healthy participants reported the task to be very challenging, the young healthy participants did find it rather easy. This difference between the elderly and young groups is not surprising, as most of the participants in the elderly group did not have any VR experience, whereas the young group was rather experienced in VR. The distribution of answers in patients with neglect was rather spread, indicating that, for some patients, the task was not particularly challenging, while for others, it was. A possible explanation for this result might be found in the individual settings adapted for the group of patients with neglect. As the individual performance level was estimated, some might have been under- or overestimated. Nevertheless, even if the performance level of some patients with neglect would have been under-or overestimated, according to their ratings, they were still highly entertained and motivated.

**Difficulty Levels and Adaptive Difficulty Scaling**

The descriptive analysis of the change in difficulty levels over time evidenced that the participants in the young group showed a steep increase in the initial phases of the task, suggesting a higher balance point than the one they started with. In the elderly group, the initial increase was less steep, but it was still clearly observable and gradual over time. In the neglect group, the initial increase seems to be delayed, but over time, there is a level increase relative to the starting level. This might be due to the attentional impairments and other cognitive deficits, resulting in a need for more rounds to get better at the game. Nevertheless, based on this measure, we could show that the individualized task was able to adapt to the individual level of impairment of the patients over time and thus keep them motivated and not overwhelmed [18].

Therefore, as the feedback regarding motivation and frustration assessed by means of the questionnaires shows clearly, the adaptive difficulty scaling was able to address the issue of very different skill levels and different progression speeds across participants.

**In-Game Measures**

VR tasks offer several opportunities to individually evaluate participants’ task performance using in-game measures. One of these possibilities is the analysis of the hand position over time, reflecting the spatial search behavior of the participants. As patients with neglect typically fail to explore the contralesional space, the assessment of the individual hand movements may be a valuable parameter to characterize neglect manifestations and severity. Indeed, in exemplary data sets, we were able to show typical neglect patterns (ie, healthy participants move their hand in the peripersonal space symmetrically during the search, whereas for patients with neglect, these movements are limited within the left, contralesional side).

Another parameter to estimate the ecological validity of the task in patients with neglect is the average search time relative to the objective neglect severity. Our results revealed the typical neglect pattern [37,67]; the worse the neglect — as measured by the CoC — the more time the stroke patients needed to find targets appearing within their contralesional, left side.

**Limitations**

The main limitation of the present study is the small sample size of the neglect group and the fact that the patients were not assessed with a comprehensive test battery (ie, including other measures of neglect severity like behavior in free visual exploration [6], other cancellation tasks [68], or line bisection tasks [69]). Furthermore, even though the group was age-matched, there were mainly male participants in the group of healthy elderly participants, and as they were all chess players, the generalizability of the results in this group are limited. For this, further research with a more representative sample would be needed.

Due to the 2-step process, no direct comparison of the in-game performance between participants with normal and impaired visual exploration behavior was possible.

**Outlook and Conclusion**

The presented bird search task was shown to be entertaining, motivating, and even immersive. Due to the implemented difficulty levels, it adapted well to different populations with different skills and previous VR experience and even in patients with cognitive disturbance after stroke. In particular, the bird search task seems to be able to pick up on typical patterns of neglect and to correlate with the results of established instruments.

Future studies should investigate and evaluate these aspects for potential application in diagnosis or therapy. For the further evaluation of the potential diagnostic or rehabilitative value, the frequency and duration of playing the task should be investigated in a longitudinal randomized clinical trial. Furthermore, the tool should be compared with the standard care of patients with neglect.
Acknowledgments

T Nyffeler was supported by SNF grant 32003B_196915.

Authors’ Contributions

SEJK, SMG, DC, BCK, RMM, T Nef, and T Nyffeler designed the study. SEJK, SMG, and T Nef developed the tool and the setup. SEJK, BCK, and PU recruited the participants. SEJK, BCK, T Nef, and RMM analyzed the data. SEJK, DC, BCK, and PU wrote the manuscript. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References


42. Unity Technologies. URL: https://unity.com/ [accessed 2021-06-13]


47. Reinhart S, Leonhard E, Kerkhoff G. Sensitive Neglect Test (SNT) single and dual task. Saarland University. URL: https://www.uni-saarland.de/fileadmin/upload/lehrauftr/kerkhoff/Materi/é_1%C3%BC_Diagnostik_Therapie_Instructions_SNT_single_and_dual_task.pdf [accessed 2021-06-13]


49. Brooke J. SUS - A quick and dirty usability scale. 1996. URL: https://hell.meiert.or (accessed 2021-06-13)


57. R Core Team. R: A Language and Environment for Statistical Computing. 201. URL: https://www.r-project.org [accessed 2021-06-13]


Abbreviations

CBS: Catherine Bergego Scale
CoC: Center of Cancellation
HMD: head-mounted display
IPQ: Iggroup Presence Questionnaire
iVR: immersive virtual reality
PGTQ: Perception of Game Training Questionnaire
SNT: Sensitive Neglect Test
SSQ: Simulator Sickness Questionnaire
SUS: System Usability Scale

©Samuel Elia Johannes Knobel, Brigitte Charlotte Kaufmann, Stephan Moreno Gerber, Prabitha Urwyler, Dario Cazzoli, René M Müri, Tobias Nef, Thomas Nyffeler. Originally published in JMIR Serious Games (https://games.jmir.org), 02.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
An Interactive Computer Game for Improving Selective Voluntary Motor Control in Children With Upper Motor Neuron Lesions: Development and Preliminary Feasibility Study

Annina Fahr¹,²,³, MSc; Andrina Kläy¹,², MSc; Jeffrey W Keller¹,², PhD; Hubertus J A van Hedel¹,², PhD, PT

¹Swiss Children’s Rehab, University Children’s Hospital Zurich, Affoltern a.A., Switzerland
²Children’s Research Center, University Children’s Hospital Zurich, Zurich, Switzerland
³Institute for Biomechanics, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland

Corresponding Author:
Annina Fahr, MSc
Swiss Children’s Rehab
University Children’s Hospital Zurich
Mühlebergstrasse 104
Affoltern a.A., 8910
Switzerland
Phone: 41 44 762 52 97
Email: annina.fahr@kispi.uzh.ch

Abstract

Background: Computer game–based interventions are emerging in pediatric neurorehabilitation, as they can provide two key elements for motor learning—motivating environments that enable long-term compliance, which is particularly relevant for children, and augmented feedback for improving movement performance.

Objective: The overall aim of this study is to develop an interactive computer play for children with upper motor neuron lesions to train selective voluntary motor control and give particular attention to motivation and feedback. We also aim to determine features that make games engaging, investigate which sensory feedback modality is noticed the fastest during play, develop an interactive game, and evaluate its feasibility.

Methods: We identified engaging game features by interviewing 19 children and adolescents undergoing rehabilitation. By using a test version of the game, we determined the response times of 10 patients who had to react to visual, auditory, or combined feedback signals. On the basis of the results of these two subprojects, we developed and designed a game environment. Feasibility was studied in terms of the practicability and acceptability of the intervention among 5 children with upper motor neuron lesions.

Results: The game features deemed the most important by pediatric patients were strategic gameplay (13/29, 45% of answers) and choice (6/29, 21%). While playing the game, an acoustic alarm signal (reaction time: median 2.8 seconds) was detected significantly faster (P=.01) than conditions with other feedback modalities (avatar velocity reduction: median 7.8 seconds; color desaturation: median 5.7 seconds). Most children enjoyed playing the game, despite some technical issues.

Conclusions: The careful identification of game features that increase motivation and feedback modalities that inform children quickly led to the development of an interactive computer play for training selective voluntary motor control in children and adolescents with upper motor neuron lesions.

(JMIR Serious Games 2021;9(3):e26028) doi:10.2196/26028

KEYWORDS
virtual reality; game therapy; rehabilitation; augmented feedback; motivation; mobile phone

Introduction

Background

Patients with upper motor neuron lesions (ie, due to traumatic brain injury, stroke, or spastic cerebral palsy [CP]) exhibit a variety of motor impairments. These symptoms are typically categorized as positive or negative motor signs [1,2]. Positive motor signs are related to excessive muscle activity or movement, such as increased muscle tone and spasticity. Negative motor signs describe insufficient (control of) muscle activity or movement, such as weakness or loss of selective
motor control. Reduced selective voluntary motor control (SVMC) is defined as “the impaired ability to isolate the activation of muscles in a selected pattern in response to demands of a voluntary posture or movement” [2]. Loss of SVMC can clinically manifest as impaired movement control and a multitude of involuntary movements (ie, unintended movements that co-occur with the performance of a voluntary task [3]), such as mass flexion or extension patterns, synergies of muscle activation (ie, obligatory grouped multi-joint movements), or mirror movements (ie, simultaneous, identical movements on the contralateral side) [4,5].

Although impairments in SVMC can be categorized according to the International Classification of Functioning, Disability, and Health into the domain of body function, they can interfere with movements of daily life activities and limit the patients’ independence. For example, children with unilateral spastic CP who exhibit mirror movements require more time for bimanual activities of daily life than peers without mirror movements, independent of their unimanual capacities [5]. In children with CP and lower limb impairments, gross motor function abilities are more strongly related to SVMC than other common impairments, such as muscle weakness or spasticity [6,7]. This relationship between SVMC and gross motor function was not only observed in cross-sectional evaluations but also in longitudinal studies, where an impairment in SVMC was found to be the strongest predictor of a less favorable course of gross motor function [8,9]. When focusing on ambulation, impairments in SVMC were associated with a reduction in gait speed due to decreased knee extension at initial contact, resulting in a reduced step length [10].

Training Methods of SVMC

Multi-disciplinary rehabilitation programs aim to reduce children’s limitations in daily life activities to improve their participation. Despite the known importance of SVMC, only a few interventions have specifically aimed at improving selective control [11]. Robot-assisted ankle movement training implemented in a computer game was used to train the graded ankle dorsiflexion and plantarflexion [12]. After 18 training sessions over 6 weeks, children with spastic CP significantly improved their ankle SVMC. In another pilot trial, 4 children with spastic CP played a commercial video game, which was controlled by surface electromyography (sEMG) signals. It aimed to reinforce activity in a desired muscle and to reduce cocontraction of an agonist-antagonist muscle pair, thus training selective muscle activation [13]. In total, 3 of the 4 children could reduce muscle cocontraction during the game. A therapy program by Adler et al [14] specifically addressed mirror movements interfering with bimanual activities in children with unilateral spastic CP. However, exercises to suppress mirror movements emphasizing the independent use of both hands led to improved bimanual function without reducing the occurrence of mirror movements.

Although these interventions targeted certain aspects of SVMC (eg, improving joint movement control or reducing mirror movements), no approach covered all features of reduced selective motor control. We became interested in designing an intervention that would include training both aspects, improving accurate motor control while simultaneously reducing any sort of involuntary movement. On the basis of our experience with the application of rehabilitation technologies to improve upper and lower extremity functions in children and adolescents with neuromotor disorders [15,16], we started considering the design of a technology-assisted system combined with a game played in a virtual environment. Many technology-assisted interventions based on interactive computer play (ICP), exergames, and virtual reality (VR) have emerged to complement conventional therapies during rehabilitation [15,17,18]. The distinction between these terms describing similar technologies is not always easy, and VR is often used improperly [18,19]. ICP is defined as “any kind of computer game or VR technique where the individual can interact and play with virtual objects in a computer-generated environment” [20]. What distinguishes VR from more general ICP technologies is the immersive component of VR systems [21]. The purpose of exergames is to provide physical activity or exercise through interactive play [22]. An advantage of ICP is that it can provide key aspects relevant for motor learning during rehabilitation; for example, a playful environment increases motivation and thereby enables large numbers of repetitions, whereas the performance of movements can be improved based on augmented feedback [23].

Objectives

Thus, the project’s overall aim was to develop an intervention for children with upper motor neuron lesions to improve SVMC by exploiting the advantages of ICP technology. Our basic idea was to develop an ICP that should train accurate joint movement control while providing immediate feedback when involuntary movements would occur.

Previous studies have already identified the properties of video games that are relevant when designing games because they affect motivation and player engagement, as follows: reward, optimal challenge, feedback, choice or interactivity, clear instructions, and socialization [24]. In addition to the theoretical background of game design principles and motivation, we wanted to consider the game design preferences of the target users. Augmented feedback should make the player aware of the occurrence of involuntary movements while playing an attention-demanding ICP. Therefore, special attention should be given to the question of how such warning feedback is ideally provided. Finally, we aimed to follow up the ICP design process with a small feasibility trial. Therefore, we formulated the following specific research questions for the development of an ICP to train SVMC: (1) which features of commercial games do children and adolescents undergoing neuro-orthopedic rehabilitation like, (2) which modality of feedback indicating a negative occurrence is perceived most easily by these patients while playing an ICP, and (3) is it feasible to use the new ICP as an intervention for children and adolescents with impaired SVMC due to an upper motor neuron lesion?

Methods

Study Overview

To develop an ICP that should motivate children and adolescents with impairments in SVMC to practice for a longer time, we conducted three subprojects: (1) identifying motivating game
features by conducting interviews; (2) identifying which warning feedback signal is detected the quickest while playing the game; and (3) based on the results from these subprojects, the ICP is developed and its preliminary feasibility is evaluated.

This game development study did not fall under the Human Research Act, which was confirmed by the Ethics Committee of the Canton Zurich through two clarifications of responsibility (Req-2019-01005 and Req-2019-00623).

Motivating Game Features

For the survey, we recruited children and adolescents regularly playing games on smartphones or tablets undergoing neuro-orthopedic inpatient rehabilitation at the Swiss Children’s Rehab. The participants and their parents were verbally informed about the interview. Verbal consent was obtained from both before the structured interviews were conducted. In the interview, we gathered ideas and explored the participants’ opinions with the following open-ended questions (in German):

- What are your favorite games?
- What do you like about them? Several answers were possible for each game.
- What feature is the most important for you and would need to be included if you could design your own game?

We asked participants to think of games for smartphones or tablets rather than for game consoles. The latter have extensive control options (eg, many keys or joysticks) that allow complex user interactions. Smartphone or tablet games usually include simpler control modes, which are in line with our idea of training SVMC using basic isolated joint movements to control the ICP.

If participating children or adolescents had difficulties in naming game features spontaneously, we provided suggestions including graphics, actions, shooting, the story, music, being challenged, puzzles, different strategies, character development, large game world, and multiplayer.

Feedback Modalities

Participants

For the second subproject, we again recruited children and adolescents undergoing neuro-orthopedic inpatient rehabilitation. They had to be able to use the arrow keys and press the spacebar on a regular computer keyboard. The exclusion criteria were severe visual or auditory impairments. All participants and their parents were informed orally and written about the protocol. Parents and adolescents aged ≥14 years provided written informed consent. All participants provided verbal informed consent.

Test Game

We created a reaction time (RT) task to investigate in which modality a warning feedback signal is noticed most easily while playing the game. Thus, in this experiment, the augmented feedback signal did not inform the player about undesired comovements, but the signal was simulated and participants had to detect it while playing. The game environment was a preliminary version of the ICP programmed in Unity (Figure 1). The goal of the game was to move an avatar up and down on the screen to collect coins and avoid obstacles. The avatar was controlled with arrow keys on a commercial computer keyboard to provide similar conditions for all participants.

Figure 1. Preliminary game environment and schematic illustration of feedback modes for testing different feedback signal modalities. (A) Game environment in which the avatar had to be navigated up and down. (B) Baseline mode where the avatar’s speed was reduced. (C) In the acoustic mode, the feedback signal was an alarm sound of increasing sound volume. (D) Illustration of the conversion from color to gray scale in the background color mode.
We investigated three modes of gradable feedback signals. Feedback 1, Baseline: no additional signal apart from reducing the avatar’s speed, which had the aim of facilitating the game to enable focusing on the purpose of the game again; Feedback 2, Acoustic: playing an alarm sound; and Feedback 3, Background color: converting the background from color to grayscale (Figure 1). We tested four conditions: (1) Feedback 1 alone, (2) Feedback 1 and 2, (3) Feedback 1 and 3, and (4) Feedback 1, 2, and 3. The feedback signal was introduced randomly between 6 and 12 seconds after a trial had started and gradually increased from 0% to 100% feedback signal intensity (ie, speed reduction from normal game speed to completely static, increase in sound volume of the alarm, and color desaturation to complete gray scale) in 10 seconds. It remained for 2 seconds at 100% signal intensity before the test automatically continued with the next trial. The participants’ task was to press the space bar to turn off the signal as soon as they had detected the feedback signal while continuing to play the game. In this case, the time remaining until the next trial started was normal game play, such that feedback detection success had no influence on the duration of the game. A schematic illustration of the feedback simulation trial design is provided in Figure 2.

**Figure 2.** The upper part shows the organization of one trial, illustrating the start of the trial, the feedback initiation (ON) and increment, and the start of the next trial. The lower part of the figure illustrates the changes that occurred as soon as the participant responded to the feedback signal and how the reaction time was defined. RT: reaction time.

We used the R statistical package (version 3.4.4; R Foundation for Statistical Computing). Normality of the data was evaluated using Shapiro-Wilk tests and by visual inspection of the Q-Q plots. As not all data were normally distributed, conditions were compared using a nonparametric Friedman test with pairwise Wilcoxon signed-rank tests. The significance level \( \alpha \) was set at .05, with a Bonferroni correction for multiple comparisons.

**Game Design and Feasibility Trial**

**SVMC Runner Game**

The game environment was programmed in Unity (version 2019.3.0f6, Unity Technologies) and included sounds and background music. The game design was based on the results of the two previous subprojects, hence the survey on favored game features and the investigation of feedback modalities. A video of the game can be seen in Multimedia Appendix 1. The goal of the game was to move an avatar up and down on the screen to collect coins and avoid certain obstacles (Figure 3). The avatar’s horizontal position was fixed while the game environment moved at a steady pace of shifting by one screen.
width in approximately 6.5 seconds (ie, an object appearing on
the right side of the screen required 6.5 seconds until it
disappeared on the opposite side). As soon as involuntary
movements occurred, an auditory augmented feedback signal
made the player aware of the occurrence of involuntary
movements (based on results from the second subproject). The
signal volume was graded according to the extent of the
involuntary movements. Furthermore, the avatar’s speed
gradually slowed, relative to the involuntary movements. This
response to involuntarily occurring movements was
implemented to make game play temporarily easier. It should
enable the player to focus again on the game and reduce the
occurrence of comovements, because these often appear in
conditions of increased effort [3].

Figure 3. Game environment. The purpose of the game was to collect stars or coins and avoid the obstacles by moving the avatar up and down on the screen. (A) This screenshot from a game level shows a shield power-up in the top-left corner. After collecting it, the avatar was protected from obstacles
for the next 10 seconds. (B) This level had a different theme, although the game elements stayed the same. (C) An example of challenges in one level. Each challenge required a different focus. (D) Players could use the collected coins to personalize the avatar.

The game included six thematically different game environments
(worlds). The six worlds covered 54 levels of increasing
difficulty (eg, from shorter to longer duration; from a few, small,
static obstacles to more, larger and moving obstacles; and from
no enemy to multiple enemies). Each player had their own
profile. Initially, only the first level was playable, whereas the
others remained locked. After successfully finishing a level, the
next higher level became unlocked. At the onset of each level,
certain challenges (eg, collecting a certain amount of coins) that
the player needed to achieve were presented. Players were
provided three challenges per level that required them to adopt
(varying) strategies (Figure 3). Advancing to a new world
needed fulfillment of at least 80% of these challenges. The
players experienced choice as they could freely decide which
challenge or challenges they wanted to tackle. Variability was
further enhanced by adding power-ups (ie, objects that add
temporary extra abilities to the game character, such as
increasing speed, a shield protecting against enemies; Figure
3) to the environment. Once enough coins had been collected,
they could be used to personalize the avatar (Figure 3).

In addition, we considered the visual presentation of the game
to be important. The appealing and motivating effect of visual
features had to be balanced against the risk of confusion or
distraction from the basic task, which can be particularly
important for children with a recently acquired brain injury.
With glasses simulating a strong visual impairment (provided
by the Swiss National Association of and for the Blind), we
verified that contrasts between essential gaming elements were
high, as more than 50% of the children with CP experience
visual impairments [26].

Game Controllers
We pursued the following two control strategies for the ICP to
improve SVMC: (1) a strategy based on muscle activity and (2)
a strategy based on joint movements (Figure 4). The first
approach trained the selective activation of a target muscle
(group) while reducing activity in another muscle (group) that
should remain inactive (similar to Rios et al [13] and Yoo et al
[27] but not restricted to agonist-antagonist muscle pairs). The
sEMG signals of these two selected muscle groups were
recorded using a varioport-e device (Becker Meditec) at a
sampling frequency of 1000 Hz. The sEMG signals were
transferred via Bluetooth to the ICP computer, where they were
filtered (exponential smoothing with a smoothing factor of
0.0015, followed by taking a moving average with a window
size of 10 frames) and fed into the game. The avatar was steered
up by increasing the activation of the target muscle (group),
while lowering the sEMG activity caused a downward movement of the avatar. The \textit{warning} feedback signal was triggered by activity in the muscle (group) that should remain inactive.

For the second approach, we used the ArmeoSenso system (version 1.0; Hocoma AG). This motion capture system was developed for arm rehabilitation and included three wearable inertial measurement units to detect movements of one arm and trunk movements. Although it was originally intended for unilateral upper extremity rehabilitation, it could also be used to track lower extremity movements (unilateral). Joint angles delivered by the inbuilt software were transferred via the user datagram protocol to the ICP computer and served as input signal for the game to train selective movement of the chosen target joint and record involuntary movements in another joint. The upward and downward movements of the avatar were caused by moving the target joint in an upward and downward direction, respectively. Movements of the joint that should not move activated the \textit{warning} feedback signal.

\textbf{Figure 4.} Game controllers. (A) Electromyography-based control of the game via activation of the right tibialis anterior muscle (green cables). Mirror activity on the left side (red cables) would trigger the feedback signal. (B) ArmeoSenso with its three sensors (chest, upper arm, and lower arm). In this example, the game was controlled by elbow flexion and extension without simultaneously moving the shoulder (ie, shoulder abduction would lead to a warning feedback signal).

\textbf{Pilot Tests}

Before testing the game in children and adolescents with impairments in SVMC, we performed pilot sessions in healthy adults and children to identify and resolve bugs in the game design and explore different hardware configurations.

\textbf{Participants}

For the feasibility trial, we recruited children and adolescents with impaired SVMC due to upper motor neuron lesions. Improving SVMC was identified by their physical or occupational therapist as a reasonable rehabilitation goal of their inpatient stay at the Swiss Children’s Rehab. All participants and their parents were informed verbally and in writing about the protocol and provided written informed consent.

\textbf{Study Design}

The study encompassed six sessions of 40 minutes scheduled over 2 weeks. In the first two sessions, we tested each control system (one system per session in randomized order). In the following four sessions, we used the more appropriate system. Together with the participants’ physical or occupational therapist, who treated each participant, we determined the muscle group or movement that had to be trained (target joint) and the involuntary muscle activation or movement that needed to be suppressed.

\textbf{Outcomes}

Study participants were characterized by age, sex, diagnosis, and SVMC impairment measured using the Selective Control Assessment of the Lower Extremity or Selective Control of the Upper Extremity Scale [28-30]. We evaluated the domains’ practicability and acceptability, as described in the model of clinical utility by Smart [31]. We recorded any technical issues related to setting up the ICP, playing the game, and whether the game’s purpose was clear to the participant. To compare the hardware approaches and hints on which strategy to pursue further, participants decided on their preferred system and the therapist indicated which system was more appropriate in terms of reaching the therapy goal. In view of using the ICP in an intervention study in the future, we assessed the time the participant was actively playing. Therefore, the game logged the time when levels were played, excluding all breaks.

After each session, participants were verbally asked whether they enjoyed the game (three-point Likert-scale with answers “no,” “a little,” and “a lot of fun”), whether they would like to play again, and if they experienced any discomfort or harm.

\textbf{Procedures}

Each session started by attaching the sensors or electrodes and calibrating the game, which was accomplished by moving the joint once through the active range of motion or maximally contracting the muscle. In the sEMG mode, the limb was then strapped to a custom-made board with loops for Velcro straps to prevent any movement of the target joint. We decided to control the game by isometric muscle activation after pilot tests, which revealed that when movement was allowed, the joint was mostly held in a position at the end of the range of motion. This led to pain only after a short period of playing. We used standardized, tutorial-like, step-by-step instructions to guide participants in the game environment. Participants decided on
their own level of play and which challenge to tackle. In the first two sessions, played with the different control systems, participants started at level 1 with separate game profiles. This enabled participants to start at the easiest level with both control systems. Sessions 3-6 continued progressively with the system the therapist deemed more appropriate (also considering the participant’s opinion).

**Results**

**Motivating Game Features**

We interviewed 19 inpatients from our rehabilitation center (9 males; mean age 14.3, SD 2.5 years; range 8.0-17.0 years). The most common diagnosis was traumatic brain injury (n=6). Participants mentioned 29 games, of which Minecraft was mentioned three times, followed by 4 Elements, Clash of Clans, and Fortnite, each mentioned twice. They considered requiring a strategy to play the game as the most important feature of the game (13/29, 45% of all answers), followed by having multiple options (6/29, 21%). The term options encompasses that the game allows the player to decide between multiple possibilities to achieve the goal. Further characteristics that participants liked were the possibility of playing in a multiplayer mode, power-ups, and character evolution (Table 1). A full list of answers can be found in Multimedia Appendix 2.

<table>
<thead>
<tr>
<th>Game feature</th>
<th>Answers, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplayer</td>
<td>13 (45)</td>
</tr>
<tr>
<td>Strategic gameplay</td>
<td>13 (45)</td>
</tr>
<tr>
<td>Power-ups</td>
<td>11 (38)</td>
</tr>
<tr>
<td>Multiple options</td>
<td>8 (28)</td>
</tr>
<tr>
<td>Character evolution</td>
<td>7 (24)</td>
</tr>
<tr>
<td>Creativity</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Fighting</td>
<td>3 (10)</td>
</tr>
<tr>
<td>Animations</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Shooting</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Speed</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Timing challenge</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Machines</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Simple idea</td>
<td>1 (3)</td>
</tr>
<tr>
<td>Sport</td>
<td>1 (3)</td>
</tr>
</tbody>
</table>

**Feedback Modalities**

A total of 10 children and adolescents (male: n=9; age: mean 12.6 years, SD 3.5 years; range 7.3-16.3 years) participated. The participants’ diagnoses were bilateral spastic CP (n=5; Gross Motor Function Classification System I-IV and Manual Ability Classification System I-II), sensory-motor neuropathy (n=1), epilepsy (n=1), congenital hemiplegia (n=1), anti-N-Methyl-D-aspartate receptor encephalitis (n=1), and intracranial hemorrhage (n=1). The RTs are shown in Figure 5. They differed significantly between the conditions ($\chi^2 = 26.2; P<.001$). Pairwise comparisons revealed that the RT was significantly longer ($P=.01$) in conditions without the acoustic feedback signal (conditions 1 and 3) than in conditions that included the alarm sound (conditions 2 and 4). RT tended to be longer ($P=.06$) in the baseline condition (1), where only the velocity was reduced, compared with the background color condition (3). RTs in conditions 2 and 4 did not differ ($P=.99$).

Although none of the participants missed responding in any acoustic feedback signal trial, 14% (7/50) of responses in total were missed in the baseline condition of only velocity reduction, whereas 6% (3/50) were missed in the background color condition.
Figure 5. Boxplots of the reaction times for each feedback mode (1–4). In the baseline condition, only the speed of the avatar was reduced. In the other conditions, velocity reduction was combined with an additional feedback signal, that is, an acoustic alarm or a change of the background color. The colors represent individual participants. The small circles show the reaction time of each individual trial, whereas diamonds show the mean reaction time over all 5 trials per participant.

Feasibility Trial
The ICP was piloted in 11 sessions with healthy participants; 5 children and adolescents with impaired SVMC were recruited, and their characteristics are shown in Table 2. One participant (ID 2) stopped the study after two sessions owing to the premature end of rehabilitation because of the COVID-19 pandemic.

Table 2. Characteristics of participants in the feasibility study (n=5).

<table>
<thead>
<tr>
<th>ID number</th>
<th>Age (years)</th>
<th>Sex</th>
<th>Diagnosis</th>
<th>SVMC total (points)</th>
<th>SVMC target joint (points)</th>
<th>System</th>
<th>Target joint or muscle group</th>
<th>Involuntary movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.8</td>
<td>Female</td>
<td>Bilateral spastic CP, GMFCS III</td>
<td>5&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0&lt;sup&gt;c&lt;/sup&gt;</td>
<td>sEMG&lt;sup&gt;f&lt;/sup&gt;</td>
<td>Ankle dorsiflexion</td>
<td>Ipsilateral knee extension</td>
</tr>
<tr>
<td>2</td>
<td>13.0</td>
<td>Male</td>
<td>Bilateral spastic CP, GMFCS II</td>
<td>8&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0&lt;sup&gt;e&lt;/sup&gt;</td>
<td>ArmeeSenso</td>
<td>Knee flexion and extension</td>
<td>Trunk flexion and extension</td>
</tr>
<tr>
<td>3</td>
<td>12.7</td>
<td>Female</td>
<td>Bilateral spastic CP, GMFCS III</td>
<td>7&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1&lt;sup&gt;e&lt;/sup&gt;</td>
<td>sEMG</td>
<td>Knee extension</td>
<td>Contralateral knee extension</td>
</tr>
<tr>
<td>4</td>
<td>12.5</td>
<td>Female</td>
<td>Traumatic brain injury</td>
<td>16&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1&lt;sup&gt;e&lt;/sup&gt;</td>
<td>sEMG</td>
<td>Ankle dorsiflexion</td>
<td>Contralateral ankle dorsiflexion</td>
</tr>
<tr>
<td>5</td>
<td>16.1</td>
<td>Male</td>
<td>Unilateral spastic CP, MACS&lt;sup&gt;g&lt;/sup&gt; III</td>
<td>20&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2&lt;sup&gt;b&lt;/sup&gt;</td>
<td>ArmeeSenso</td>
<td>Shoulder abduction and adduction</td>
<td>Ipsilateral elbow flexion and extension</td>
</tr>
</tbody>
</table>

<sup>a</sup>SVMC: selective voluntary motor control; SVMC total: total Selective Control Assessment of the Lower Extremity (SCALE) score (maximum of 2 points per joint; total of 20 points) or Selective Control of the Upper Extremity Scale (SCUES; maximum of 3 points per joint; total of 30 points); SVMC target score: SCALE or SCUES item scores for the target joint that was practiced with the game.

<sup>b</sup>Refers to the target joint when the ArmeeSenso system was used and to the target muscle group when the electromyography-based system was used.

<sup>c</sup>CP: cerebral palsy.

<sup>d</sup>GMFCS: gross motor function classification system (it standardizes the classification of gross motor function, emphasizing the trunk control and walking ability of children with cerebral palsy [32]. Level 1 means that children perform all the activities of neurologically intact children of the same age, allowing for slight limitations in speed and quality of movements. Level 5 means that children exhibit difficulties in head and trunk control in most positions or in achieving any voluntary control of movement at all).

<sup>e</sup>Selective voluntary motor control assessed with the Selective Control Assessment of the Lower Extremity.

<sup>f</sup>sEMG: surface electromyography.

<sup>g</sup>MACS: manual ability classification system (it classifies how children with cerebral palsy handle objects in daily activities [33]. Level 1 means that the child can handle objects easily and successfully, whereas children at level 5 do not handle objects at all).

<sup>h</sup>Selective voluntary motor control assessed with the Selective Control of the Upper Extremity Scale.
In one case, the ArmeoSenso was favored (ID 5) and also once, the participant and therapist clearly preferred the electromyography-based system (ID 1), because steering the game with the ArmeoSenso proved to be challenging in the case of a small active range of motion. No comparison of systems was possible for the other participants. In two cases, because the unilateral ArmeoSenso system is not able to track mirror movements (IDs 3 and 4), we could only use the sEMG system; in the third case, the patient withdrew (ID 2), but the motion capture system was working reliably until then.

It took more time than expected to set up the ICP until game control functioned reliably, because we encountered difficulties in conducting and standardizing the calibration. Recalibration was necessary more frequently than expected, because problems occurred with steering. Participants were actively playing for 18.3 (SD 5.4) minutes per session on average (range 9.8-30.4 minutes) with training time increasing over the course of the study (15.8 minutes in first session vs 20.3 minutes in fifth session).

Only one participant answered having experienced “no fun at all” (ID 5), while having “a lot of fun” represented the majority of answers. ID 5 remained uninterested over all sessions, whereas two that were a little skeptical in the beginning became more enthusiastic over time (Figure 6). None of the participants reported any discomfort or harm.

**Figure 6.** Participants’ ratings of enjoyment. Scores indicated how much they liked the game and whether they would like to play it again. Only sessions with the more appropriate system are shown.

Discussion

**Principal Findings**

This study describes the development of an ICP to specifically enhance SVMC in children and adolescents. The game-based intervention trains movement control while simultaneously providing feedback to reduce involuntary movements. We designed the game environment based on the favorite game features that children reported in a survey. Subsequently, we tested various feedback conditions to identify the auditory signal as the modality that patients perceived the fastest. An ensuing investigation of usability revealed that children enjoyed the game despite some technical issues.

**Motivating Game Features**

Interestingly, the features mentioned most frequently, strategic gameplay, and having multiple options depend on each other. If a game provides many options, the player needs a strategy to succeed, and vice versa, players can only develop a strategy if they are provided with more than one option. These features are comparable with the interdependent construct choice and interactivity, which are considered key principles of game design to increase motivation [24]. Strategic gameplay was also one of the motivators of video games for therapy mentioned by children with hemiparesis in a focus group [34]. Furthermore, variability, challenge, and competition were identified as common elements that increased motivation in a study investigating the effect of several ICP environments on motivational behaviors in children with CP [35]. Although
variability was not a specific response in this study, the features of multiple options, power-ups, and character evolution introduce variability and help to avoid the fact that a game looks the same every time it is played.

Multiple options were incorporated into the game design in several ways. The game did not follow one particular trajectory with the game avatar but allowed different ways through the environment. Furthermore, players could choose the level to play and which challenge to tackle on their own. Presenting three challenges for each level forced the player to develop and use varying strategies to fulfill each of them. Finally, we included power-ups, the possibility to personalize the game character, and varied the visual appearance of the game with a thematic design of the worlds.

**Feedback Modalities**

Unlike common simple RT tasks, this study investigated a dual-task situation. While participants focused on collecting coins and avoiding obstacles, the feedback signals had to attract their attention. The game already provided many visual inputs, which might have complicated the detection of other visual stimuli (such as reducing speed or changing background color) due to overloading visual perception [36]. In particular, during parts of the game with many obstacles, the visual load could have been high. Although the game sounds and background music were not relevant for playing the game, the alarm sounds were easy to distinguish. The RTs and their variability were considerably lower in conditions with an auditory feedback signal than in conditions with only visual signals. This suggests that the alarm sound had the highest contrast to the game environment and attracted attention most easily.

**Feasibility Trial**

On the basis of most participants expressing moderate-to-high enjoyment and no discomfort, the ICP intervention can be considered acceptable. One child’s spontaneous comment (“I cannot reach this [next] world with only one more session, can I come more often?”) revealed that the design of the game environment created long-term motivation and engagement. These results confirm that game-like interventions can lead to high levels of compliance and motivation [37]. Increased fun and interest in the game-based intervention was also shown in a study directly comparing conventional ankle dorsiflexion training with an ICP protocol [38]. High levels of motivation increase the time dedicated to an intervention and would thereby contribute to the success of a therapy, which depends on adherence over an extended period. We can only speculate why one participant did not like the ICP based on our observations. Nevertheless, we would not expect the results to be different because both steering modes require coordination of one’s movement according to what is displayed in the game. The comparison of hardware systems (game controllers) did not reveal a clear favorite but showed that both systems were required. The use of the sEMG system was obligatory to address mirror movements during training because ArmeoSenso cannot track contralateral movements. The sEMG system was also superior in the case of a small active range of motion, where motion tracking with ArmeoSenso failed. However, ArmeoSenso was particularly useful for recording trunk movements, where the selection of one representative muscle would be difficult. These findings further strengthen our view of an individualized approach to improving SVMC in children and adolescents with upper motor neuron lesions.

**Limitations**

With the survey in the first part of our study, we only assessed the subjective preferences of children and adolescents about game features, assuming that the preferred game elements are motivating. However, it would require a longitudinal study in which these features were systematically manipulated to investigate their influence on long-term engagement. A limitation of the investigation of feedback modalities is that the game was played with the arrow keys and not with the intended way of steering through muscle activation or joint movement. Nevertheless, we would not expect the results to be different because both steering modes require coordination of one’s movement according to what is displayed in the game. The use of the sEMG might have shortened RTs (for all feedback modalities), as this mode of control is likely more common and therefore less challenging than controlling a game by body movements.

Although the overall aim was to develop an intervention for patients with reduced SVMC due to an upper motor neuron lesion, we recruited children with various diagnoses undergoing neuro-orthopedic rehabilitation for the first two subprojects, as we considered these research questions unrelated to SVMC.
advantage of investigating a more general population of children undergoing rehabilitation is the higher generalizability of the results to the ICP design for other pediatric patient groups.

We could not evaluate the preliminary effectiveness of the ICP to determine its appropriateness, although we had intended to do so with an assessment integrated with the ICP, similar to a previously developed assessgame [42,43]. Difficulties in standardizing the calibration prevented comparability between sessions or after recalibration. We are planning a single-subject research design study to evaluate the effectiveness of the game in improving motor control and reducing the occurrence of involuntary movements compared with regular multimodal rehabilitation in children and adolescents with impairments in SVMC. After showing effectiveness in a clinical setting, a future plan could be to implement the intervention in a home-based setting. It would also be interesting to investigate whether potential improvements in SVMC are meaningful to children and translate to improvements in activities of daily life.

Conclusions
This study revealed that strategic gameplay and having multiple options are favorable and likely motivating features of games for children and adolescents undergoing rehabilitation. This indicates that they prefer to choose their own strategies among several options to influence the course of a game. We incorporated these findings in the design of an ICP to train SVMC, as well as the finding that an acoustic warning signal was perceived more quickly than the other modalities that we tested. Feasibility testing revealed that participants enjoyed playing, but some technical issues impeded user experience and should be addressed to optimize practicability. The results of this study could be considered by designers of other interactive computer games if long-term engagement of children, such as during a rehabilitation period, is needed.

Acknowledgments
This research project was funded by the Swiss National Science Foundation (grant 32003B_179471). The funders were not involved in the study design, data collection, analysis, or manuscript preparation. The authors thank all therapists of the Swiss Children’s Rehab that collaborated with this project. We are very grateful to the participating children, adolescents, and their families.

Authors' Contributions
AF conceived the study, performed measurements, conducted the analysis, and wrote the initial draft. JWK conceived the study and performed the measurements. AK conceived the study, was responsible for game software development, and performed measurements. HJAVH supervised the project and acquired funding. All authors critically reviewed the manuscript and agreed with the final version.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Video of the game. This shows the avatar being navigated through the game environment, collisions with obstacles, and the effects of a power-up.

[MP4 File (MP4 Video), 22803 KB - games_v9i3e26028_app1.mp4]

Multimedia Appendix 2
List of all answers during the interview.

[PDF File (Adobe PDF File), 276 KB - games_v9i3e26028_app2.pdf]

References


Abbreviations
- CP: cerebral palsy
- ICP: interactive computer play
- RT: reaction time
- sEMG: surface electromyography
- SVMC: selective voluntary motor control
- VR: virtual reality
Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review

Sandra Barteit¹, MA, PhD; Lucia Lanfermann¹, MSc; Till Bärnighausen¹²³, MSc, MD, PhD; Florian Neuhann¹⁴, MD; Claudia Beiersmann¹, PhD

¹Heidelberg Institute of Global Health, Heidelberg, Germany
²Department of Global Health and Population, Harvard TH Chan School of Public Health, Boston, MA, United States
³Africa Health Research Institute, KwaZulu-Natal, South Africa
⁴School of Medicine and Clinical Sciences, Levy Mwanawasa Medical University, Lusaka, Zambia

Abstract

Background: Augmented reality (AR), mixed reality (MR), and virtual reality (VR), realized as head-mounted devices (HMDs), may open up new ways of teaching medical content for low-resource settings. The advantages are that HMDs enable repeated practice without adverse effects on the patient in various medical disciplines; may introduce new ways to learn complex medical content; and may alleviate financial, ethical, and supervisory constraints on the use of traditional medical learning materials, like cadavers and other skills lab equipment.

Objective: We examine the effectiveness of AR, MR, and VR HMDs for medical education, whereby we aim to incorporate a global health perspective comprising low- and middle-income countries (LMICs).

Methods: We conducted a systematic review according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) and Cochrane guidelines. Seven medical databases (PubMed, Cochrane Library, Web of Science, Science Direct, PsycINFO, Education Resources Information Centre, and Google Scholar) were searched for peer-reviewed publications from January 1, 2014, to May 31, 2019. An extensive search was carried out to examine relevant literature guided by three concepts of extended reality (XR), which comprises the concepts of AR, MR, and VR, and the concepts of medicine and education. It included health professionals who took part in an HMD intervention that was compared to another teaching or learning method and evaluated with regard to its effectiveness. Quality and risk of bias were assessed with the Medical Education Research Study Quality Instrument, the Newcastle-Ottawa Scale-Education, and A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions. We extracted relevant data and aggregated the data according to the main outcomes of this review (knowledge, skills, and XR HMD).

Results: A total of 27 studies comprising 956 study participants were included. The participants included all types of health care professionals, especially medical students (n=573, 59.9%) and residents (n=289, 30.2%). AR and VR implemented with HMDs were most often used for training in the fields of surgery (n=13, 48%) and anatomy (n=4, 15%). A range of study designs were used, and quantitative methods were clearly dominant (n=21, 78%). Training with AR- and VR-based HMDs was perceived as salient, motivating, and engaging. In the majority of studies (n=17, 63%), HMD-based interventions were found to be effective. A small number of included studies (n=4, 15%) indicated that HMDs were effective for certain aspects of medical skills and knowledge learning and training, while other studies suggested that HMDs were only viable as an additional teaching tool (n=4, 15%). Only 2 (7%) studies found no effectiveness in the use of HMDs.

Conclusions: The majority of included studies suggested that XR-based HMDs have beneficial effects for medical education, whereby only a minority of studies were from LMICs. Nevertheless, as most studies showed at least noninferior results when compared to conventional teaching and training, the results of this review suggest applicability and potential effectiveness in LMICs. Overall, users demonstrated greater enthusiasm and enjoyment in learning with XR-based HMDs. It has to be noted that
many HMD-based interventions were small-scale and conducted as short-term pilots. To generate relevant evidence in the future, it is key to rigorously evaluate XR-based HMDs with AR and VR implementations, particularly in LMICs, to better understand the strengths and shortcomings of HMDs for medical education.

**(JMIR Serious Games 2021;9(3):e29080) doi:10.2196/29080**

**KEYWORDS**

virtual reality; augmented reality; global health; income-limited countries; medical education

### Introduction

Augmented reality (AR), mixed reality (MR), and virtual reality (VR)–based technologies open novel ways of teaching and training for medical education, as they allow for immersive experiences that may foster the teaching and learning of complex medical contents. Especially so-called head-mounted devices (HMDs), most often realized as a headset or glasses, seem to be advantageous and adequate to low- and middle-income countries (LMICs) based on their versatile, low-price, and mobile nature [1,2]. Technologies like HMDs make learning content more accessible and engaging, whereas for educators, it broadens their educational impact beyond the classroom and face-to-face teaching [3,4]. Quality education is key to improve health outcomes for all [5], especially in low-resource settings where there is a dire need to strengthen the health workforce [6], particularly today, as health professionals must acquire a great deal of skills and know-how to become competent practitioners [7]. HMDs can potentially be a catalyst for improving educational efforts by increasing the effectiveness of existing medical training programs, as AR-, MR-, and VR-based HMDs enable repeated practice without adverse effects on the patient in various medical disciplines; may introduce new immersive ways to learn complex medical content; and may alleviate financial, ethical, and supervisory constraints on the use of traditional medical learning materials like cadavers and other skills lab equipment [8-11]. Moreover, disruptive technologies such as HMDs can not only help to learn but also prepare medical learners for a highly technologically advanced workplace [12]. Therefore, HMDs hold the promise to be a potential driver in strengthening health systems and the health workforce, which has been key to increasing global life expectancy in recent years [13]. Particularly, LMICs face health worker shortages, skewed distribution of health professionals toward urban areas, and limitations in skill sets and training that do not aptly address the population’s real health needs [13,14]. Regions in Africa and Asia still have alarmingly high health worker shortages [10] despite the Sustainable Development Goals (SDGs) calling for a substantial increase in the recruitment, development, training, and retention of the health workforce in income-limited countries (SDG goal 3C) [15]. To successfully achieve the SDGs for 2030, digital technologies may be a key element, as they bear the potential to enhance health professional performance and training in a rapid and cost-effective way [1,2]. In addition, digital technologies are versatile and are well able today to reflect the varied training needs of health professionals covering a broad field of teaching and training needs, clinical competencies, and skills such as therapeutic and diagnostic skills and communication skills. Nowadays, technologies provide an experience close to reality, without putting the patient at risk during training. Technologies provide a quality standard of technical medical skills, as they are scalable and repeatable until skills are fit for practice. In particular, HMDs are versatile in their use compared to specialized individual simulators already used for medical training and available at low prices, and for providing increased learning space mobility—features particularly valuable in low-resource contexts [1,2,8,9].

Currently, there is a lack of insights on HMDs particularly in LMICs, as most reviews on HMDs have focused on high-income countries, on AR for medical education, or on AR and VR but not in the context of health [16-20]. This systematic review took a global perspective. Technologies are constantly evolving and there is a need for obtaining an overview of current trends in a global context. No other review was found that had considered HMDs for global health professional training considering a recent time frame.

The main objective of this systematic review is to screen the current literature evaluating HMDs using AR, MR, and VR for medical education, and to elucidate the effectiveness of HMDs for medical education in a global context, particularly with regards to LMICs. Two main research questions guided the systematic review: (1) what is the effectiveness of using HMDs for medical education, specifically for knowledge and skills, and (2) what are the strengths and weaknesses of HMDs in medical education?

### Methods

This systematic review was conducted according to the Cochrane Collaboration Handbook for systematic reviews [21] and reported according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; see Multimedia Appendix 1) [22]. There was no review protocol published and the review was not registered.

### Data Source, Search Strategy, and Inclusion and Exclusion Criteria

Seven medical and educational databases for peer-reviewed literature were searched comprising PubMed, Cochrane Library, Web of Science, Science Direct, PsycINFO, Education Resources Information Centre, and Google Scholar (200 first results were extracted, according to Haddaway et al [23]). Gray literature databases that were searched comprised WorldCat (45 results were used from each of the three search concepts XR, medicine, and education) and Global Index Medicus.

In addition, reference lists of selected articles have been hand-searched and included if inclusion and exclusion criteria were met. As relevant HMD technology was first commercially...
introduced in 2014 [24], publications were only included if published after January 1, 2014, until the end of the study period, which was May 31, 2019, and were restricted to the English language. Surveys, editorials, and conference papers were excluded, as well as literature that had no abstract or full-text available.

The PICOS (population, intervention, comparison, outcome, study design) framework guided the inclusion and exclusion criteria of this study [25] (see Table 1). To ensure coverage of all relevant literature to this rather novel topic, search terms were compiled comprehensively and were grouped into the three search concepts of extended reality (XR; which subsumes the concepts of VR, AR, and MR), medicine, and education. The following search terms or keywords were used alone or in combination: virtual reality, augmented reality, mixed reality, medical, health, clinical, education, teaching, training, and learning (see Multimedia Appendix 2 for a detailed overview of the search strategy). The PRISMA flowchart is shown in Figure 1 depicting the screening process.

<table>
<thead>
<tr>
<th>Framework</th>
<th>Description</th>
<th>Inclusion</th>
<th>Exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Health professionals who received medical education</td>
<td>Health professionals, nursing and midwifery professionals, modern health associate professionals, nursing and midwifery associate professionals</td>
<td>Health management and support workers who are not in health service provisioning</td>
</tr>
<tr>
<td>Intervention</td>
<td>Head-mounted displays based on virtual reality, augmented reality, or mixed reality</td>
<td>Head-mounted displays of all kinds that include glasses or goggles</td>
<td>Devices that are not head-mounted displays</td>
</tr>
<tr>
<td>Comparison</td>
<td>Modern vs traditional methods for medical education to evaluate effectiveness of XR tools</td>
<td>Books, pen and paper, chalkboard, face-to-face teaching, traditional lectures</td>
<td>No evaluation of the effectiveness of XR devices</td>
</tr>
<tr>
<td>Outcome</td>
<td>Improved or not improved learning outcome</td>
<td>Concrete learning outcome/evaluation of effectiveness in learning</td>
<td>No concrete outcome</td>
</tr>
<tr>
<td>Study</td>
<td>Literature in English, published between Jan 1 to May 31, 2019</td>
<td>Literature as identified via the search strategy</td>
<td>Literature reviews, meta-analyses, opinion papers; non-English literature; literature published before Jan 1, 2014, and after May 31, 2019</td>
</tr>
</tbody>
</table>

XR: extended reality.
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart for systematic reviews. ERIC: Education Resources Information Centre; HMD: head-mounted display; WHO: World Health Organization; XR: extended reality.

Data Extraction and Study Quality Assessment

Articles were independently screened by two researchers using the eligibility criteria of the PICOS framework. Data were extracted as recommended by the Cochrane Handbook [21] (for detailed data extraction table see Multimedia Appendix 3). Any disagreements were resolved by discussion between the two screening authors. Population characteristics were derived from the studies, including author, year, place, journal, country of study, study design, evaluation methods, number of evaluation methods, type of data analysis, effectiveness, type of study population, medical discipline of study, type of XR, type of HMD, type of learning, study duration, and references.

Quality and risk of bias was assessed with the Medical Education Research Study Quality Instrument (MERSQI), the Newcastle-Ottawa Scale-Education (NOS-E), and “A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies” (ACROBAT-NRSI) [26-28].

Synthesis Method

Data synthesis is reported according to the SWiM (Synthesis Without Meta-analysis) reporting guideline [29]. The research question guided the synthesis groupings, and we focused on the effect of XR HMDs on knowledge and skills gained in relation to the HMD devices used. Furthermore, we reflected each synthesis group according to their medical specialties, as we thought that each medical specialty is different in its focus on knowledge or skills training. No standardization metric was applied. The synthesis method was to extract relevant sections of the studies with regard to knowledge and skills gained, as well as the specific XR HMD. Overall, the risk of bias assessment of nonrandomized study designs showed no critical results (see Multimedia Appendix 4 for details). Accordingly,
the 27 included studies were synthesized with equal weight. We did not restrict the study design. Hence, we did not conduct a meta-analysis, as studies were quite heterogeneous, and the included studies reported different quantities and qualities about knowledge and skills. Effectiveness was synthesized according to the respective study reports. The tables aggregate information about the study characteristics and focal areas of this review (knowledge, skills, XR HMD).

Results

Study Characteristics

A total of 27 studies was included in the review: 17 (63%) VR studies, 7 (26%) AR studies, 2 (7%) MR-focused studies, and 1 (4%) VR and AR study (see Table 2 for study details). Although 24 (89%) studies used only a single HMD, 3 (11%) studies compared two HMDs. All studies were in an academic or hospital setting and mostly compared HMDs to conventional face-to-face training methods [2,30-41]. The included studies used an HMD worn on the head.

Included studies were categorized according to three levels of knowledge (adapted and modified from Górski et al [42], see Multimedia Appendix 5): (1) theoretical knowledge (eg, anatomical atlases and preoperative planning), (2) practical skills (eg, operation trainings and surgical simulators), and (3) attitudes (eg, self-confidence, communication skills, and patient-centeredness).

Overall, the included studies (N=27) comprised a study population of 956 study participants. Sample sizes ranged from 1 participant to 178 participants, with a mean of 35 participants (see Multimedia Appendix 6 for details of the number and type of study participants).

The medical procedures described across studies varied widely. Studies in surgery included training in neurosurgery [43], gastrectomy [2], total hip arthroplasty [35,37], laparoscopy [44], dental surgery [32,45], surgical ophthalmoscopy [33], peg transfer practice [46], or surgical knot training [47] and central line and catheter insertion [48,49]. Anatomy teaching was covered in 4 (15%) studies, 3 of which involved 3D learning structures in neuroanatomy [39,50,51], such as training on the brain cerebrum. Luursema et al [34] focused on the effect of visual ability on anatomical understanding. Ferrandini Price et al [38] trained emergency medicine with a “triage for a mass casualty incident,” Rai et al [31] focused on ophthalmoscopy with “binocular indirect ophthalmoscopy,” Siff and Mehta [52] evaluated “an interactive holographic curriculum” for gynecology training, and Bing et al [53] looked into cervical cancer surgery in Zambia. In the field of urology, Butt et al [40] conducted catheterization. Digital slides were analyzed in the field of pathology [54]; dental implants in dentistry [45] and, in the field of geriatrics, Dyer et al [55] concentrated on neurodegenerative diseases.
Table 2. Characteristics of included studies.

<table>
<thead>
<tr>
<th>Study characteristics</th>
<th>Studies (N=27), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
</tr>
<tr>
<td>2014-2015</td>
<td>4 (15)</td>
</tr>
<tr>
<td>2016-2017</td>
<td>9 (33)</td>
</tr>
<tr>
<td>2018-2019</td>
<td>14 (52)</td>
</tr>
<tr>
<td><strong>Country classification by income-level</strong></td>
<td></td>
</tr>
<tr>
<td>Low-income country</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Middle-income country</td>
<td>2 (7)</td>
</tr>
<tr>
<td>High-income country</td>
<td>25 (93)</td>
</tr>
<tr>
<td><strong>Country of study</strong></td>
<td></td>
</tr>
<tr>
<td>US</td>
<td>10 (37)</td>
</tr>
<tr>
<td>Canada</td>
<td>3 (11)</td>
</tr>
<tr>
<td>UK</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Australia</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Germany</td>
<td>2 (7)</td>
</tr>
<tr>
<td>France</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Ireland</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Spain</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Netherlands</td>
<td>1 (4)</td>
</tr>
<tr>
<td>China</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Taiwan</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Zambia</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Reported study design</strong></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>21 (78)</td>
</tr>
<tr>
<td>Qualitative</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Mixed methods</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Evaluation methods</strong></td>
<td></td>
</tr>
<tr>
<td>Skills tests</td>
<td>18 (67)</td>
</tr>
<tr>
<td>Questionnaires</td>
<td>16 (59)</td>
</tr>
<tr>
<td>Recordings</td>
<td>6 (22)</td>
</tr>
<tr>
<td>Knowledge tests</td>
<td>5 (19)</td>
</tr>
<tr>
<td>Surveys</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Observation</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Self-assessment</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Others</td>
<td>5 (19)</td>
</tr>
<tr>
<td><strong>Number of evaluation methods used</strong></td>
<td></td>
</tr>
<tr>
<td>One method</td>
<td>16 (59)</td>
</tr>
<tr>
<td>Two methods</td>
<td>9 (33)</td>
</tr>
<tr>
<td>Three methods</td>
<td>2 (7)</td>
</tr>
<tr>
<td>More than three methods</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Data analysis conducted in publication</strong></td>
<td></td>
</tr>
<tr>
<td>Inferential statistics</td>
<td>15 (56)</td>
</tr>
<tr>
<td>Descriptive statistics</td>
<td>10 (37)</td>
</tr>
<tr>
<td>Study characteristics</td>
<td>Studies (N=27), n (%)</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Qualitative analysis</td>
<td>1 (4)</td>
</tr>
<tr>
<td>No analysis identified</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Self-concluded effectiveness</strong></td>
<td></td>
</tr>
<tr>
<td>Effective</td>
<td>17 (63)</td>
</tr>
<tr>
<td>Partly effective</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Useful only as additional tool</td>
<td>4 (15)</td>
</tr>
<tr>
<td>No proven effectiveness</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Study population</strong></td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>10 (37)</td>
</tr>
<tr>
<td>Residents</td>
<td>8 (30)</td>
</tr>
<tr>
<td>Physicians/nurses</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Mixed training levels</td>
<td>6 (22)</td>
</tr>
<tr>
<td><strong>Medical discipline</strong></td>
<td></td>
</tr>
<tr>
<td>Surgery</td>
<td>13 (48)</td>
</tr>
<tr>
<td>Anatomy</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Gynecology</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Emergency medicine</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Urology</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Pathology</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Geriatrics</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Dentistry</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Mode of XR</strong></td>
<td></td>
</tr>
<tr>
<td>VR</td>
<td>17 (63)</td>
</tr>
<tr>
<td>AR</td>
<td>7 (26)</td>
</tr>
<tr>
<td>MR</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Combined information</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Type of head-mounted display</strong></td>
<td></td>
</tr>
<tr>
<td>Oculus Rift (Consumer V1/DK2; VR)</td>
<td>8 (30)</td>
</tr>
<tr>
<td>HTC Vive (2016; VR)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>Samsung Gear VR (not specified; VR)</td>
<td>4 (15)</td>
</tr>
<tr>
<td>MS HoloLens (Development Ed; MR)</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Eyesi Indirect System Simulator (Version 1.1.3; AR)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Google Glass (Trial Version; AR)</td>
<td>2 (7)</td>
</tr>
<tr>
<td>No information on brand</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Brother AirScouter (WD-200B)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Daydream View Headset (Not specified; VR)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Epson Moverio (BT-200; AR)</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Sony HMZ (T1 3D Viewer; VR)</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Type of learning and medical discipline</strong></td>
<td></td>
</tr>
<tr>
<td>Practical skills</td>
<td>18 (67)</td>
</tr>
</tbody>
</table>
Studies (N=27), n (%)

### Study characteristics

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surgery</td>
<td>11</td>
<td>41%</td>
</tr>
<tr>
<td>Emergency medicine</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Ophthalmology</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Dentistry</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Urology</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Gynecology</td>
<td>1</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Theoretical knowledge

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anatomy</td>
<td>4</td>
<td>15%</td>
</tr>
<tr>
<td>Surgery</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Pathology</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Gynecology</td>
<td>1</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Attitudes

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geriatrics</td>
<td>2</td>
<td>7%</td>
</tr>
<tr>
<td>Surgery</td>
<td>1</td>
<td>4%</td>
</tr>
</tbody>
</table>

### Duration of Intervention

<table>
<thead>
<tr>
<th>Duration</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1 month</td>
<td>21</td>
<td>78%</td>
</tr>
<tr>
<td>1 month to 6 months</td>
<td>3</td>
<td>11%</td>
</tr>
<tr>
<td>7-12 months</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>1 year to 2 years</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>&gt;2 years</td>
<td>1</td>
<td>4%</td>
</tr>
</tbody>
</table>

---

**Footnotes:**

aResidents were medical doctors in specialized training after completing medical school, under supervision of an experienced doctor. Physicians or nurses were professionals with several years of accredited experience in their field. Mixed training levels indicate that participants of two or more distinct educational levels were part of the study, such as residents and students combined.

bXR: extended reality.

cVR: virtual reality.
dAR: augmented reality.

eMR: mixed reality.
fMultiple types of head-mounted displays may have been used within a study.

---

### Effectiveness of Knowledge, Skills, and Attitudes

In most of the 27 studies, skills were taught with HMDs in the field of surgery (n=18, 67%). Some studies (n=7, 26%) focused on knowledge transfer, mainly in the field of anatomy, and a few studies (n=2, 7%) trained on perceptions and self-confidence. HMDs varied across studies by brand, model, functionality, and type of XR (see Multimedia Appendix 7 for details on reported effectiveness).

**Skills**

Out of the 27 studies, 18 (67%) were identified that focused on skills outcomes [2,30,31,33,35-38,40,41,44-49,53,56].

**Surgery (11 Studies)**

Barré et al [2] explored a VR-based scenario to train medical professionals on sleeve gastrectomy, which study participants described as “realistic and useful to learn surgery” (device: HTC Vive). Sensors for the VR-based scenario were attached to real-life surgical instruments for navigation in a virtually created operating room. The authors found a reduction in cognitive effort and a decrease in stress during prolonged periods of standing during surgical practice [2]. Huang et al [48] focused on “AR simulation of venous catheters” (device: Brother AirScouter), in which AR glasses were used to display instructions on procedural steps of catheter insertion in the form of a digital overlay of information to participants during the medical process. Using AR HMDs was perceived as useful for skills transfer in the operating room [48]. Yoganathan et al [47] found that 360° videos of surgical knot training produced major advantages, as they improved the ability to tie knots during surgery even as stand-alone learning (device: unspecified). Hooper et al [37] taught hip arthroplasty, but the results were unclear as to the impact on medical knowledge (device: Oculus Rift CV1). The authors reported that if the VR training was conducted right before the actual surgery, it may have a positive impact on technical abilities and acknowledged that the impact of VR for surgery would significantly increase in the years to come. Logishetty et al [35] found that the quality of hip surgery performed through AR did not differ from that experienced by in-person training with a physician (device: MS HoloLens). Nonetheless, the authors still perceived AR as a valuable tool for other surgical procedures related to arthroplasty. Peden et
al [33] tested HMDs for the learning of suture skills by medical students, whereby standard face-to-face instruction compared to HMDs showed similar skills outcomes (device: Google Glass). Qin et al [46] compared simulators and a variety of XR devices for peg transfer training (devices: HTC Vive, Samsung Gear). Overall, medical simulation systems contributed to a more immersive and successful training environment for peg transfer [46]. Rochlen et al [49] evaluated the usability and feasibility of AR technology for needle insertion for central venous catheter placement for medical students (device: Google Glass). The authors found that AR technology may constitute an important addition to medical skills training.

Harrington et al [56] discovered that surgeons were more attentive when training in 360° instead of in 2D environments for learning laparoscopic cholecystectomy (device: Samsung Gear VR). The study participants, which were preclinical undergraduate students, found the 360° learning beneficial and entertaining, whereby no significant difference in information retention was found between 360° learning and 2D videos [56]. Huber et al [44] investigated experiences of learning laparoscopic skills (peg transfer, cholecystectomy) with a VR-based HMD (device: HTC Vive) compared to a simulator and found that the use of training with HMD was feasible and that participants were excited about the immersion provided by the HMD. Accuracy scores between the two comparison groups were equal, so the outcome of laparoscopic performance was noninferior [44]. Wu et al [30] tested the feasibility of ultrasound procedures (ultrasound-guided central line procedure) with two groups: one group with the AR-based HMD (device: Google Glass) and the other group using traditional ultrasound [30]. The AR-based HMD projected a digital layer at the corner of the glasses to see whether physicians could increase their focus on the procedure. Overall, the authors found that the Google Glass could be effective in ultrasound training, with the only caveat that study participants needed more time being familiarized with the technology, as well as “more needle redirections, but less head movements” [30].

**Emergency Medicine (2 Studies)**

Azimi et al [41] evaluated whether nursing students and novice learners could improve the learning of the emergency medical procedure of needle chest decompression virtually using an HMD (device: MS HoloLens) [41]. There were no adverse effects of nurses’ emergency training using HMDs, and participants improved in terms of frequency of training and enthusiasm, and recalled and demonstrated proficiency in their training days later [41]. In an emergency scenario with VR HMDs, Ferrandini Price et al [38] tested the stress of study participants of “basic triage in a mass accident event” and found that VR cannot yet replace a clinical simulation. Therefore, they recommended VR as a complementing training method (device: Samsung Gear VR).

**Ophthalmology (2 Studies)**

Leitzritz et al [36] found that inexperienced students practice ophthalmoscopic examination better after HMD training (device: ARO). Rai et al [31] compared AR to face-to-face teaching in the field of ophthalmoscopy and found that AR simulation may be superior for skills learning, especially for novice ophthalmology students (device: EyeSI BIO simulator with headpiece).

**Dentistry (1 Study)**

Lin et al [45] tested an AR-based dental implant placement system (device: Sony HMS-T1) and evaluated the precision of the virtually planned versus the actual prepared implant position. The integration of the AR technology considerably reduced the deviation of placing the implant from planned position, and the accuracy of computer-aided implant surgery increased [45].

**Urology (1 Study)**

Butt et al [40] reviewed VR systems for “urinary catheterization” and found that undergraduate nursing students were enthusiastic about applying it to practical scenarios (device: Samsung Gear).

**Gynecology (1 Study)**

Bing et al [53] replicated hysterectomy training (removal of uterus) in a virtual 3D environment for Zambian surgery trainees (device: Oculus Rift). The authors concluded that “affordable VR might have the potential to enhance access to cancer treatment globally” [53].

### Knowledge

Out of the 27 studies, 7 (26%) studies evaluated the outcome of knowledge with regard to XR interventions [34,39,43,50-52,54].

### Anatomy (4 Studies)

Moro et al [50] evaluated the acquisition of anatomical knowledge (spine anatomy) and compared two different learning modes: desktop-based (device: Oculus Rift) and mobile-based (device: Gear VR). Both groups performed similarly in the knowledge test, although a number of study participants experienced fuzzy vision and malaise using VR. Stepan et al [51] identified neuroanatomical test scores of participants using HMDs (device: Oculus Rift) compared to participants who practiced with conventional anatomy books. The VR-based training led to more engagement, learner motivation, and enjoyment, although there was no difference in exam performance between the two groups. Ekstrand et al [39], evaluated in a randomized controlled study the benefits of VR-based HMD neuroanatomy training and compared it to learning with a paper-based 15-page booklet focusing on anatomy training for medical students. Ekstrand et al [39] mentioned that the HTC Vive in neuroanatomy training can be used as an additional tool to increase knowledge gain but that the VR group did not surpass the control group with respect to learning outcomes. They found similar knowledge test results for VR and learning from books [39]. Luursema et al [34] examined the usability of VR-based HMDs for anatomy learning (cognitive load and problem-solving strategies) for users with respect to stereoscopic depth. They concluded that it was not evident whether depth perception advanced or impeded the uptake of anatomical knowledge. The Luursema et al [34] study stated that digital interventions neither enhanced nor impeded knowledge outcomes in anatomy.

https://games.jmir.org/2021/3/e29080
Surgery (1 Study)
Bairamian et al [43] compared a 3D-printed model to VR-based HMD angiogram models (device: Google Daydream HMD in connection with a smartphone). The HMD offered better resolution and zoom capabilities for study participants (neurosurgical trainees, neurosurgeons), but the 3D-printed model offered better depth perception and manipulation opportunities. The authors concluded that VR-based HMD angiogram may be a viable alternative to 3D-printed models, with untapped educational potential [43].

Pathology (1 Study)
Farahani et al [54] evaluated the feasibility of using VR-based HMDs for reading digital pathology slides of lymph nodes (device: Oculus Rift). Study participants (pathologists) reported that VR pathology slides were limited in their resolution and that they faced difficulties navigating the VR device. Overall, study participants were able to produce accurate diagnoses, and high diagnoses concordance was reached compared between the VR-based HMD and traditional slide system [54].

Gynecology (1 Study)
Siff and Mehta [52] introduced an interactive holographic training module for teaching urogynecologic surgical anatomy (device: MS HoloLens), which involved holograms of female organs, livestreaming of surgical videos, and 3D-projected organs to enhance structural understanding (ligament suspension, sacrospinous ligament fixation). Siff and Mehta [52] observed that the interactive holographic mode of learning was effective in the acquisition of knowledge for surgical anatomy. Study participants ranked the AR-based training as much better when compared to conventional training [52].

Attitudes
Out of the total 27 studies, 2 (7%) studies [32,55] evaluated the effects of HMD training on specific attitudes of health professionals. Dyer et al [55] investigated whether the understanding of neurodegenerative diseases and their impact on patients could be made more transparent to medical students by training with a VR-based HMD (device: Oculus Rift). The results revealed that the impact of simulation on the attitudes of participants was significant [55]. Pulijala et al [32] focused on the impact of “VR surgery on the self-confidence of surgical residents.” They found that the “self-esteem of participants could be increased” [32] when training with the HMD (device: Oculus Rift). This was especially true for participants with little clinical experience. Both interventions were evaluated as effective, since understanding of diseases and self-confidence in surgery increased. The effect on self-confidence was “especially high for inexperienced physicians” [32].

Evaluation Methods
Of the 27 studies, the most common evaluation methods were practical skills tests (n=18, 67%), followed by questionnaires (n=16, 59%), video recordings of procedures (n=6, 22%), knowledge tests (n=5, 19%), surveys (n=4, 15%), observations (n=4, 15%), self-assessments (n=3, 11%), and others (n=5, 19%). Practical skills tests comprised in some cases of observing study participants when repeating learned procedures or repeating practical skills. Questionnaires included general questionnaires, questionnaires combined with follow-up assessments, and Likert scales. Generally, authors used multiple questionnaires throughout the interventions. Videos were generally recorded during practice sessions of medical trainings and were later evaluated. Bing et al [53] measured how participants moved during the simulation and how much time was spent to fulfill the task. The surgeons could verify their performance scores that were recorded by the simulator. Knowledge and skills were usually assessed prior to and post intervention. Surveys followed the Validation of Instructional Materials Motivation Survey [57] or the System Usability Survey for the measurement of participant perceptions on usefulness [40]. Participant self-assessment was also evaluated, in which authors asked students to assess their own performance, as self-written evaluation text or self-assessment questionnaire [2,52,53]. Other evaluation methods included using students’ drawings of the nerve head to assess diagnostic capabilities [36], saliva samples to elucidate stress levels during training [38], the use of dental computer tomography images to measure the differences between implant locations [45], and examination of the precision of acetabular placement [35].

Discussion
General Aspects
The results showed that HMDs are at least comparable to traditional methods of medical education and beneficial in terms of increasing students’ motivation for learning (see Multimedia Appendix 8 for an overview of benefits, shortcomings, and recommendations described within included studies). HMDs allow for repeated use of difficult training scenarios in an immersive and realistic environment, such as emergency procedures or rare complications during surgery. The studies found benefits and shortcomings of HMDs. Studies based their findings on improving and enhancing HMDs.

XR-based HMDs are currently dominantly used in high-income countries; only 2 [46,53] studies were conducted in LMICs. Nevertheless, HMDs may be particularly beneficial in a low-resource context to provide training for direly needed health care workers based on their versatile, mobile, and immersive nature. Bing et al [53] implemented and evaluated HMDs for surgical training in Zambia, which was found to be effective. The adoption of XR-based HMDs in other medical settings may most likely increase in the next years and may foster medical teaching and training especially in settings where there is a need for time-effective and cost-effective education [58]. Particularly, recent developments like the HMD Oculus Quest seem to be particularly promising for LMIC contexts. The potential for XR-based HMDs in other countries and settings, particularly in LMICs, need to be studied further. In addition, long-term effects of using HMDs on learners’ knowledge and skills in various medical education settings and the integration into medical curricula needs to be further researched. Further technical advancements are needed, and traditional methods of education are not to be ignored.
In the following sections, the Discussion is structured according to the research questions of this review: What is the effectiveness of using HMDs for medical education, specifically for knowledge and skills, and what are the strengths and weaknesses of HMDs in medical education?

**Effectiveness**

The studies in this review generally categorized an intervention as effective if the majority of the study population achieved higher scores in tests (pre-posttest, exercises) or participant observations as compared to traditional instructional approaches, such as books and analogue surgery or ultrasound procedures.

Not all studies reported effective outcomes for the use of HMDs in medical education. Some studies described the disadvantages of HMDs, such as motion sickness and nausea, technical problems, and stress. A systematic review underlined that these disadvantages may impede learning and training [9]. It is unclear if symptoms of motion sickness and nausea are related to beginners’ attempts to become familiar with the technology or if these persist long-term and may potentially impede learning or education. A study has found that women more often are faced with motion sickness using VR devices and that stabilization of the users’ body may alleviate symptoms [59]. In addition, effects may differ as AR devices combine real and virtual environments, which should reduce the experienced adverse health effects in VR applications, such as blurred vision, disorientation, and cybersickness. These adverse effects of individuals using HMDs may vary depending on the device. As, for example, AR devices combine real and virtual environments, they seem to mitigate negative health effects such as blurred vision, disorientation, and cybersickness [7].

It has to be taken into account that technology acceptance for XR-based HMDs may differ between individual learners, as inexperience may require more time and effort using HMDs for educational purposes.

In addition, realistic feedback was at times not implemented. For instance, “surgical errors that occurred in virtual training were not followed by complications such as simulated patient bleeding or variations in anatomy” [53]. Contextual factors such as sizes, sounds, and functionalities of instruments of virtual operating rooms need to be extremely precise and realistic. Otherwise, there is the potential of erroneous learning and training [8]. Nevertheless, with the continuous development of technologically sophisticated learning tools, more advancements become available that may supplement XR-based HMDs with audio and visual information [7].

**Skills**

The progression from knowledge to skills follows four levels of competences as per Miller’s Pyramid of Professional Competence [60]. Particularly, the see one, do one, teach one, and simulate one concept in line with Miller’s Pyramid underlines the potential of XR-based HMDs, which can implement this concept for medical teaching and skills training. Particularly, as students have limited time to learn with cadavers, if at all, they are required to supplement their anatomical knowledge through self-directed study. This material is frequently presented in the form of 2D supplementary resources such as lecture slides, textbooks, and flashcards [7]. HMDs, like other disruptive technologies, encourage students to be active and self-directed learners who set their own learning pace through hands-on experiences. Active learning has shown to lead to improved educational outcomes such as increased learning retention [61]. Through the use of XR-based HMDs, learners may gain a more realistic understanding of medical concepts [19]. Learners training with HMDs were shown to improve practical skills, reduce stress, and gain self-confidence for the actual surgery [32,32]. Furthermore, the high level of experience needed in the operating room can be easily and repeatedly gained with XR-based HMDs [8,16]. Our review highlighted possibilities of repeatedly practicing surgical operations such as hysterectomy, laparoscopy, or total hip arthroscopy via HMDs. In this review, HMDs were mostly used in the fields of surgery and anatomy with positive skill outcomes. The main benefits of studies have shown decreased surgical error rates, cost-effectiveness, and improved knowledge. The predominance of surgery and anatomy may be that both have a long-standing history of simulation training beginning in the late 1980s with simple simulation [62]. In ophthalmology, examining the virtual retina and gaze of a virtual patient increases practice time, as normally, practice is with the real pupils of a patient’s eye [31,36]. Resultantly, less patients need to participate in training sessions in ophthalmology [36]. A potential application that seems adequate for low-resource contexts could be HMD-based training for cataract surgery, as cataracts are still quite common in these settings [63].

During emergencies, physicians work under increased intensity while simultaneously working at reduced capacity due to increased stress. XR-based training scenarios can be effective in helping clinicians better understand the situations they need to be prepared for [8,64].

**Knowledge**

Some studies in this review pointed out that 3D models of human bodies significantly improved learning outcomes, with HMDs providing 360º views, as new structures and concepts are processed to enhance overall comprehension [65]. “Threshold concepts” are crucial to understand practical applications, such as with anatomy [65]. Without knowledge of anatomical structures, students will not be able to excel in surgery. The findings in this review suggest that XR-based HMDs set the hurdle lower for “threshold concepts” because students immerse themselves in physical structures, which increases understanding and later information retrieval during actual operations. In addition, using HMDs to view 3D brain structures appears to be more motivating and engaging than traditional methods such as books [8,39,51,66]. The lack of spatial understanding and imagination when reading traditional books might increase this phenomenon [39]. HMDs enable the understanding of complex organ structures [66]. There has been a significant increase in volume of medical information expected of students in modern times, and they are moving into technology-enhanced resources to increase learner engagement and authenticity [11].

Despite higher motivation and enjoyment for learning anatomy with HMDs, there was no significant difference in knowledge...
acquisition when compared to learning from textbooks [39,51]. This is similar to other reports [11], and similar to other XR applications, such as for patient treatment where studies that did examine efficacy demonstrated preliminary evidence of similar effectiveness or more effectiveness of XR interventions than their selected conventional therapy [67]. Ultimately, increased motivation to study and understand 3D structural complexities is an important aspect [51].

The duration of the various XR HMD study sessions has yet to be examined in detail in the numerous studies that have been done. Some may find it more pleasant to spend longer or shorter periods training with XR HMDs to gain the greatest possible knowledge [7]. The fact that there is no difference in learning outcomes when using XR-based HMDs is encouraging and validates the possibilities for incorporating these innovative technologies into medical education [11]. Likely, the most effective method to teach anatomy, for example, may be to combine multiple resources in addition to XR-based HMDs like plastic models, dissections, and learning software [7]. Particularly in LMICs, XR-based HMDs may be able to strengthen and scale interactive learning and training in anatomy and other surgical skills.

**Attitudes**

Attitudes are often neglected in medical education to the detriment of patients [68]. Communication and problem-solving skills as well as empathy are vital for patient care. Studies of this review showed that not only diagnostic and therapeutic skills can be trained, but also attitudes. For example, in the field of geriatrics, XR-based HMD training was shown to foster empathy for older adult patients [55]. The simulation was a scenario in which the user was an older person with an age-related disease such as Alzheimer disease. Being immersed in this surrounding provided a realistic point-of-view into older patients’ perceptions. In addition, physicians were successfully trained on communication skills while interacting with patients via VR-based avatars [69]. Through XR-based virtual patients, difficult conversations can be simulated [70]. In this way, communication and other soft skills may be trained more effectively and detailed for handling challenging patient situations. HMDs can be used for both diagnostic and therapeutic training as well as for communication skills and attitudes.

**Evaluation Methods**

Most included studies evaluated the effectiveness of HMDs (ie, through pre- and postknowledge tests, practical exercises, or self-assessments). The term effectiveness is used in many different contexts with varying definitions [71], consisting of several components: feasibility, cost, safety, and applicability to specific contexts [72]. A successful technological implementation and evaluation considers multiple levels: the individual learner, the learning environment, the context of the learning implementation, the technological environment, and the pedagogics involved in the learning implementation [73]. In studies, the criteria determining the effectiveness or ineffectiveness of an intervention was not based on a unified definition. Whereas some studies provided more information on HMD costs, others focused more on ways that XR-based HMDs can be used across various medical disciplines. Adhering to a standardized framework for implementation and evaluation of XR-based HMDs is key to generating good and comparable evidence that can guide digital health implementations. To this end, first endeavors seem to be on their way to improve methodological quality. The Virtual Reality Clinical Outcomes Research Experts (VR-CORE) international working group compiled a three-part framework for best practices in developing and testing VR treatments to improve patient outcomes. This framework targets the development of high quality, effective, and safe VR treatments [74]. Similarly, a framework would be needed to guide XR-based interventions for medical education. For the evaluation of XR-based interventions, the World Health Organization guide on monitoring and evaluating digital health interventions may already be used as a guide [58].

**Strengths and Limitations**

Study quality and bias were evaluated using three assessment tools (MERSQI, NOS-E, ACROBAT-NRSI). Most studies in this systematic review had relatively high quality scores for study design, sampling response rates, and types of data. This might be due to a large number of randomized controlled trials included in this review. This paper followed well-established methods of conducting and reporting systematic reviews [21,22].

The methodology of this review had limitations. Participants were restricted to medical professionals. In addition, solely English publications were included, which could have led to the exclusion of relevant articles in other languages. Only HMDs were included, even though other devices that use XR in medical education exist (eg, magic mirrors, large simulators, or serious games). Hence, the focus of this review was on mobile solutions in the field. Not all gray literature was screened, which could have also led to the exclusion of suitable papers. One further limitation is that we did not consider the costs of developing XR content, which is of particular relevance for XR-based applications in LMICs. Future studies should take into consideration effort and costs of developing XR-based content to provide a better basis for decision-making.

Although various studies have focused on specific countries or continents, this systematic review provided a global perspective. Types of health professionals were not limited to a specific medical discipline. The synthesis of findings has been reported narratively, as we included different study types.

**Conclusions and Recommendations**

The majority of studies included in this systematic review considered the XR-based intervention as at least noninferior to the traditional teaching methods. Most XR-based HMDs have been reported as an engaging and enjoyable tool for learners to improve their knowledge and skills. We approached this systematic review from a global perspective; however, we only found 1 study from a low-resource context using XR-based HMDs. Probably, this is an indicator that HMDs are not yet widely used in low-resource contexts, although this review shows that HMDs could provide a high quality element for medical education in LMICs. One positive aspect is that decreased access to cadavers cause high costs, which may be reduced by using XR-based HMDs for medical education [51].
Furthermore, HMDs offer the possibility of scalability and repeated practice, such as for anatomy, without adverse effects on the patient in various medical disciplines, especially in the field of surgery. The use of this technology may support the understanding of complex 3D structures, and the technology is a general training tool to prepare for the increase of technologies in the medical workspace. In addition, other disciplines such as pathology, ophthalmology, emergency medicine, gynecology, and dentistry reported effective outcomes with XR-based HMDs. XR-based HMDs can be seen as a valuable resource that has potential to strengthen medical education.

However, the deployment of HMDs in the medical setting requires further evaluation and technical advancement, for example, for HMD-based training tools. XR-based technology is still a rather novel technology, slowly unfolding its usefulness for medical education. Surgery and anatomy are quite prominent in HMD use, but it is unclear whether and how other medical disciplines may benefit, such as pediatrics. Currently, a framework or guideline for XR-based HMD interventions is lacking to guide implementations and evaluations, although initiatives like the VR-CORE international working group are working to close this gap. Further research is also needed for other requirements like financial aspects, implementation and training, technical feasibility, and reliability. Overall, based on the results of this review, XR-based HMDs seem to be a valuable component for medical education and can be recommended as an additional tool for teaching and learning particularly complex spatial structures. Based on our experience in low-income contexts, we also identified potential for the application of XR-based HMDs for medical education in LMICs, although more use cases and more research is needed.

Acknowledgments
We acknowledge financial support by the Open Access Publishing Fund of Ruprecht-Karls-Universität Heidelberg, Germany.

Authors’ Contributions
SB contributed toward the systematic review screening, conceptualization, and writing of the original draft and draft revision. LL contributed toward the systematic review screening and data synthesizing. TB and FN contributed toward the writing of the draft revision. CB contributed toward the writing of the original draft and visualization.

Conflicts of Interest
None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 Checklist.
[DOCX File , 32 KB - games_v9i3e29080_app1.docx ]

Multimedia Appendix 2
Search strings.
[XLSX File (Microsoft Excel File), 12 KB - games_v9i3e29080_app2.xlsx ]

Multimedia Appendix 3
Data extraction.
[DOCX File , 65 KB - games_v9i3e29080_app3.docx ]

Multimedia Appendix 4
Quality assessment.
[DOCX File , 80 KB - games_v9i3e29080_app4.docx ]

Multimedia Appendix 5
Three levels of knowledge.
[DOCX File , 60 KB - games_v9i3e29080_app5.docx ]

Multimedia Appendix 6
Details of number and type of study participants.
[DOCX File , 16 KB - games_v9i3e29080_app6.docx ]

Multimedia Appendix 7
Reported effectiveness.
[DOCX File , 19 KB - games_v9i3e29080_app7.docx ]
References


Abbreviations

ACROBAT-NRSI: A Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies
AR: augmented reality
HMD: head-mounted device
LMIC: low- and middle-income country
MERSQI: Medical Education Research Study Quality Instrument
MR: mixed reality
NOS-E: Newcastle-Ottawa Scale-Education
PICOS: population, intervention, comparison, outcome, study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analysis
SDG: Sustainable Development Goal
SWiM: Synthesis Without Meta-analysis
VR: virtual reality
VR-CORE: Virtual Reality Clinical Outcomes Research Experts
XR: extended reality

Edited by N Zary; submitted 25.03.21; peer-reviewed by C Moro, R Lundin; comments to author 16.04.21; revised version received 06.05.21; accepted 08.05.21; published 08.07.21.

Please cite as:
Barteit S, Lanfermann L, Bärnighausen T, Neuhann F, Beiersmann C. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. JMIR Serious Games 2021;9(3):e29080

URL: https://games.jmir.org/2021/3/e29080

doi:10.2196/29080

PMID: 34255668

©Sandra Barteit, Lucia Lanfermann, Till Bärnighausen, Florian Neuhann, Claudia Beiersmann. Originally published in JMIR Serious Games (https://games.jmir.org), 08.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Use of a Virtual Reality Simulator for Tendon Repair Training: Randomized Controlled Trial

Tsz-Ngai Mok¹, MD; Junyuan Chen¹, MD, PhD; Jinghua Pan², MD, PhD; Wai-Kit Ming³, MD, MPH, MMSc, PhD; Qiyu He⁴, MD, PhD; Tat-Hang Sin¹, MD; Jialin Deng¹, MD; Jieruo Li¹, MD, PhD; Zhengang Zha¹, MD, PhD

¹Institute of Orthopedics Diseases and Center for Joint Surgery and Sports Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China
²Department of General Surgery, The First Affiliated Hospital of Jinan University, Guangzhou, China
³Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou, China
⁴Pediatric Cardiac Surgery Centre, Fuwai Hospital, National Centre for Cardiovascular Diseases, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing, China

*these authors contributed equally

Corresponding Author:
Jieruo Li, MD, PhD
Institute of Orthopedics Diseases and Center for Joint Surgery and Sports Medicine
The First Affiliated Hospital of Jinan University
613 West Huangpu Avenue
Tianhe District
Guangzhou
China
Phone: 86 2038688563
Email: ilorugaie@163.com

Abstract

Background: Virtual reality (VR) simulators have become widespread tools for training medical students and residents in medical schools. Students using VR simulators are provided with a 3D human model to observe the details by using multiple senses and they can participate in an environment that is similar to reality.

Objective: The aim of this study was to promote a new approach consisting of a shared and independent study platform for medical orthopedic students, to compare traditional tendon repair training with VR simulation of tendon repair, and to evaluate future applications of VR simulation in the academic medical field.

Methods: In this study, 121 participants were randomly allocated to VR or control groups. The participants in the VR group studied the tendon repair technique via the VR simulator, while the control group followed traditional tendon repair teaching methods. The final assessment for the medical students involved performing tendon repair with the “Kessler tendon repair with 2 interrupted tendon repair knots” (KS) method and the “Bunnell tendon repair with figure 8 tendon repair” (BS) method on a synthetic model. The operative performance was evaluated using the global rating scale.

Results: Of the 121 participants, 117 participants finished the assessment and 4 participants were lost to follow-up. The overall performance (a total score of 35) of the VR group using the KS method and the BS method was significantly higher (P<.001) than that of the control group. Thus, participants who received VR simulator training had a significantly higher score on the global rating scale than those who received traditional tendon repair training (P<.001).

Conclusions: Our study shows that compared with the traditional tendon repair method, the VR simulator for learning tendon suturing resulted in a significant improvement of the medical students in the time in motion, flow of operation, and knowledge of the procedure. Therefore, VR simulator development in the future would most likely be beneficial for medical education and clinical practice.

Trial Registration: Chinese Clinical Trial Registry ChiCTR2100046648; http://www.chictr.org.cn/hvshowproject.aspx?id=90180

(JMIR Serious Games 2021;9(3):e27544) doi:10.2196/27544

KEYWORDS
virtual reality simulators; tendon suture; medical education
Introduction

The incidence of tendon rupture has been increasing owing to the increasing number of people participating in recreational and competitive sports [1]. Various factors such as the intensity of exercise, overuse, genetic predisposition, and aging can cause tendon rupture [2-4]. Ruptured tendons can have delayed recovery and a high frequency of recurrence [5,6]. Tendon repair is one of the most commonly used techniques in the orthopedics field [7]. Accuracy and proficiency in the suture technique is a fundamental surgical skill that will directly influence the result of the performed operation [8]. Medical students should master both basic and practical suturing concepts to perform an outstanding operation [9]. The standardized goal for flexor tendon repair is to obtain sufficient tensile strength and to facilitate a better recovery [10]. In addition, the repair will allow early mobilization, prevent adhesion formation [11], stimulate tendon healing [12], and improve clinical outcomes. Therefore, tendon repair is an essential skill in the orthopedics field [13,14]. Tendon repair and suture techniques are part of the orthopedic course during clinical training. In traditional training, students only have limited time to practice owing to cost limitations [15,16]. During clinical practice, a real patient is involved at every step in the process. Because of the lack of practical training, medical students need to be guided through a long learning curve to become independent in clinical practice [17].

In recent years, the use of virtual reality (VR) simulators has become widespread in medical school, and VR simulators are promoted for both medical students and resident training [18]. VR simulators for orthopedics provide a holistic learning application, which produces a close-up surgical training experience. Students are provided with a 3D human model to directly observe human details from the point of view of multiple senses, including vision and hearing, and they can participate in an environment relatively close to reality. The use of virtual practical teaching content expands and enriches teaching quality and content [19]. In addition, the use of a VR simulator is a skill that requires time and practice to master. Skills involved in using a simulator include adjusting the master controls, endo wrists, camera navigation, and the requirement for students to picture the VR simulator environment in reality [20]. Recent studies have applied VR simulators in surgical training [21-23], but the aim of this study was to promote a new approach consisting of a shared and independent study platform for medical student education. This study compared traditional tendon repair training with VR simulation of tendon repair and evaluated future applications of VR simulation in the academic medical field.

Methods

Study Design

This study is a parallel-design randomized controlled trial comparing VR and control groups. This study was approved by the ethics committee of the First Affiliated Hospital of Jinan University and registered in the Chinese Clinical Trial Registry (Registry: ChiCTR2100046648). Information was collected from all participants after obtaining written informed consent in accordance with the Declaration of Helsinki. All participants were required to complete the final assessment, which was performing tendon repair on synthetic models with 2 different knots, that is, the “Kessler tendon repair with 2 interrupted tendon repair knots” and the “Bunnell tendon repair with figure 8 tendon repair” (KS and BS methods, respectively). The CONSORT checklist was used for this trial.

Participants

Senior medical students were the eligible participants in this study. They were required to complete the following fundamental courses before entering the randomized control trial: (1) human anatomy, (2) physiology, (3) biochemistry, (4) pathology, (5) pathophysiology, (6) diagnostics, (7) internal medicine, (8) orthopedics, (9) surgical probation, and (10) other professional basic clinical courses. This study excluded any participant who did not meet the above requirements. Written informed consent with a clearly stated study plan was given to all participants. The purpose of this trial was explained to the participants. After informed consent had been signed, we asked the medical students to perform tendon repair on synthetic simulations. A baseline score was given by an orthopedic specialist. Other baseline information, including gender, age, and grade point average, was collected from the medical school database.

Allocation

All participants provided written, informed, and oral independently witnessed consent to participate in the research study. A random allocation sequence was generated using a random number table. A sequence was used to allocate the groups of participants to the VR and control groups. For the examination, the students performed tendon repair on a synthetic model. All participants were randomly assigned to one of the two groups. Participants in the VR group (n=61) learned the technique of tendon repair through the VR simulator method, whereas the control group (n=60) used the traditional tendon repair teaching method. The examiners were well-trained surgeons and unaffiliated with the medical school; they evaluated and assigned a score to each final product immediately without knowing the allocation list in a nonbiased manner during preintervention and postintervention assessments. In order to ensure the rigor of the examination, we included a short training period during which the examiners were given a detailed explanation of the exam process, including the study plan, marking assignments, and how to mark the scores. The examiners were required to attend a training session before the examination to ensure consistency in marking behavior. At the end of the training, examiners also did a marking exercise to scrutinize examiners’ marking behavior. During the examination, medical students were asked not to tell the examiner which group they were assigned to (Figure 1).
Interventions

The control group participants were required to participate in complete 8 hours of lectures and a 6-hour practical class in medical school for 2 weeks. The participants learned about traumatic orthopedic theory and the fundamentals of tendon repair during the lectures. They practiced tendon repair on synthetic models under the professor’s guidance. In the practice class, students were given a PowerPoint presentation, which provided illustrations, photographs, and step-by-step instructions. They were instructed to review the training material for 1 hour. The VR group participants were required to take the same course as the control group, except for the guided PowerPoint review part. Instead, they practiced with VR simulators (including the VR version and the personal computer [PC] version) for 1 hour in class. The medical students practiced under guidance with detailed instructions. The VR simulator focuses on every participant’s performance while performing tendon repair. The operation in the VR simulator is divided into practice and examination modes. Corresponding notes for each step during the practice mode were provided; however, no notes were provided for the examination mode (Multimedia Appendix 1). The students were required to finish all the required learning in the practice mode before entering the examination mode for assessment. For the VR training section, while half of the students were practicing in the VR simulator, the rest of the students were practicing on the PC version in the training center. These students shifted the training modes after 30 minutes of VR training (see Table 1). All trainings were performed within the classes, and both groups had exactly the same opportunity for practice time.

Table 1. Illustration of the trainings undertaken by each group.

<table>
<thead>
<tr>
<th>Trainings</th>
<th>Virtual reality group</th>
<th>Control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lectures (total 8 h)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Practical class (total 6 h)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Guided PowerPoint Review (within the practical class)</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Virtual reality + personal computer practice (within the practical class)</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Virtual reality + personal computer assessment (within the practical class)</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
The VR Platform

The VR simulator used in this study was created by Jinan University and the Department of Orthopedic Surgery and Sports Medicine Center [24]. All VR simulators were classified as HTC Vive VR [25], and the software was SteamVR (JinKe Lu) [26]. The VR simulator method used in this study included an independent study section in which each participant was required to study all the theories on the VR simulator website. This website is open to the public after registration on the website. Two versions are available (ie, PC and VR simulator versions). Students were able to learn the method of tendon repair by controlling the keyboard and mouse in the PC version. Although the PC version is on a computer, a 360° scenario to enhance the study environment is still included. The VR simulator version requires the student to have the HTC Vive VR controller to practice repairing a tendon. Both versions were open to students for practice according to their needs. In addition, the practical study section of tendon repair is performed using the VR simulator with individual steps for each procedure (comprising 7 steps in total). The website provides a section in which students and teachers can communicate and share ideas with each other. This program allows the students to learn in-depth while enhancing their learning through the communication section of the VR simulator. This web-based discussion can overcome the barriers of time and distance. In addition, we measured the effectiveness of this platform, including the PC and VR versions. Because the whole training package was considered, the VR group students were divided into VR and PC groups (Figure 2).
Performance and Assessment

Both the control and VR groups participated in the research study for 14 days. The results were calculated using the global rating scale. Seven dimensions were incorporated into the tool. The global rating scale shows different aspects of the operative performance. This technology was compared with the traditional teaching method by using the global rating scale for several aspects: (1) repair with respect to tissue, (2) time in motion, (3) instrument handling, (4) tendon repair skill, (5) flow of operation, (6) knowledge of procedure and final suture, and (7) qualitative and objective assessment of all tendon repair performances [27,28]. Each column was scored on a 5-point scale. Explicit descriptors were designed to guide the examiner when evaluating the students’ performance. Each item was scored from 1 (poor performance) to 5 (good performance). Higher scores on the 35-point global rating scale in the final assessment indicated that the participants using the specific method had better performance with respect to the tendon suturing technique [27]. The global rating scale is widely used in the evaluation of surgical behavior, including objective and subjective criteria [29]. It has been used for various types of surgical evaluations such as arthroscopic surgery [30], endoscopic surgery [31], pediatric surgery [32], and orthopedic surgery [33]. Additionally, it has been validated for use with
VR systems because it measures nontechnical cognitive skills such as decision making and judgment [34].

**Statistical Analysis**

Data were analyzed using the SPSS 23.0 (IBM Corp) software package [35]. The baseline information, including age and grade point average, was analyzed using the independent two-tailed t test for parametric data [36]. Differences in the objective and semiojective measurements between the 2 groups were analyzed using the Mann-Whitney U test for nonparametric data [37]. The level of agreement between the semiojective assessments made by the 2 experts was estimated by the Cohen k coefficient. P values less than .05 were considered significant [38].

**Results**

Between August 1, 2019, and August 12, 2020, 121 potential participants were assessed for study participation in the Medical School of Jinan University. Four participants from the control group dropped out of the program for personal reasons. All participants were required to undergo a final assessment on synthetic models, and the overall score sheet was used to calculate the results. This study analyzed all participants by using the global rating scale described above. The global rating scale baseline is shown for assessing tendon repair differences in the control and VR groups (Table 2). A comparison of the participants in both groups according to age, gender, grade point average, and pretest evaluation revealed no educationally relevant or significant differences. The follow-up ended on September 30, 2020.

Posttraining scores on the global rating scale were used to assess tendon repair by the two groups. Table 3 shows a comparison of the global rating scale scores between the KS and BS methods.

With respect to tissue, no significant difference was found between the VR and control groups using the KS method (P=.22) and the BS method (P=.21). Participants in the VR group showed higher scores than the control group for time in motion (P<.001) for the KS method and the BS method, thereby indicating that the VR group produced better results for time in motion. Regarding instrument handling, no significant difference was found between the VR and control groups for either the KS or the BS method (KS and BS methods P=.31 and .16, respectively). With respect to suture skill, the VR group performed better than the control group by using the BS method than the KS method (P<.001). In the flow of operation, the VR group performed better than the control group with the KS and BS methods (P<.001). With respect to procedure knowledge, the VR group performed better than the control group when performing the KS method (P<.001) and the BS method (P<.001). The KS method (P=.048) and the BS method (P<.001) yielded significant results in the final product. Thus, the overall performance of the VR group was significantly better (P<.001) than that of the control group with both KS and BS methods (P<.001).

### Table 2. Baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Control group (n=56)</th>
<th>Virtual reality group (n=61)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>23.07 (0.97)</td>
<td>22.93 (1.01)</td>
<td>.46</td>
</tr>
<tr>
<td>Gender (male), n (%)</td>
<td>24 (43)</td>
<td>31 (51)</td>
<td>.39</td>
</tr>
<tr>
<td>Grade point average, mean (SD)</td>
<td>3.02 (0.54)</td>
<td>3.19 (0.49)</td>
<td>.66</td>
</tr>
<tr>
<td>Kessler tendon repair with 2 interrupted tendon repair knots, median (IQR)</td>
<td>8.00 (7-9)</td>
<td>8.00 (7-9)</td>
<td>.13</td>
</tr>
<tr>
<td>Bunnell tendon repair with figure 8 tendon repair, median (IQR)</td>
<td>8.00 (7-9)</td>
<td>8.00 (7-9)</td>
<td>.25</td>
</tr>
</tbody>
</table>
Table 3. Posttraining scores on the global rating scale to assess tendon repair.*

<table>
<thead>
<tr>
<th>Repair method, aspects considered, global rating scale score (1: poor, 5: good)</th>
<th>Control group (n=56)</th>
<th>Virtual reality group (n=61)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kessler tendon repair with 2 interrupted tendon repair knots</strong></td>
<td></td>
<td></td>
<td>.22</td>
</tr>
<tr>
<td>With respect to tissue, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>8 (13)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 (5)</td>
<td>41 (67)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>52 (93)</td>
<td>10 (16)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1 (2)</td>
<td>2 (3)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Time in motion, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1</td>
<td>2 (4)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14 (25)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>23 (41)</td>
<td>20 (33)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>17 (30)</td>
<td>40 (66)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Instrument handling, n (%)</td>
<td></td>
<td></td>
<td>.31</td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8 (14)</td>
<td>13 (21)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>42 (75)</td>
<td>43 (71)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>6 (11)</td>
<td>5 (8)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>Suture skill, n (%)</td>
<td></td>
<td></td>
<td>.05</td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11 (20)</td>
<td>6 (10)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>36 (64)</td>
<td>38 (62)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>9 (16)</td>
<td>14 (23)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>3 (5)</td>
<td></td>
</tr>
<tr>
<td>Flow of operation, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19 (34)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33 (59)</td>
<td>11 (18)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4 (7)</td>
<td>49 (80)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>Knowledge of procedure, n (%)</td>
<td></td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>15 (27)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>41 (73)</td>
<td>1 (2)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>42 (69)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>17 (28)</td>
<td></td>
</tr>
<tr>
<td>Final product, n (%)</td>
<td></td>
<td></td>
<td>.048</td>
</tr>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 (5)</td>
<td>12 (20)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>49 (88)</td>
<td>24 (39)</td>
<td></td>
</tr>
<tr>
<td>Repair method, aspects considered, global rating scale score (1: poor, 5: good)</td>
<td>Control group (n=56)</td>
<td>Virtual reality group (n=61)</td>
<td>P value</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>4</td>
<td>3 (5)</td>
<td>22 (36)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>1 (2)</td>
<td>3 (5)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Overall performance, median (range)

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 (18-22)</td>
<td>24 (21-29)</td>
<td>0.001</td>
<td></td>
</tr>
</tbody>
</table>

**Bunnell tendon repair with figure 8 tendon repair**

**With respect to tissue, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.21</td>
</tr>
<tr>
<td>2</td>
<td>15 (27)</td>
<td>23 (38)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>41 (73)</td>
<td>38 (62)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Time in motion, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>15 (27)</td>
<td>5 (8)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>41 (73)</td>
<td>32 (53)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>23 (74)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Instrument handling, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.16</td>
</tr>
<tr>
<td>2</td>
<td>8 (14)</td>
<td>15 (25)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>32 (57)</td>
<td>33 (54)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>16 (29)</td>
<td>13 (21)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Suture skill, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>17 (30)</td>
<td>10 (16)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>39 (70)</td>
<td>30 (49)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>21 (34)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Flow of operation, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>27 (48)</td>
<td>1 (2)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>25 (45)</td>
<td>26 (43)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>0 (0)</td>
<td>28 (46)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>4 (7)</td>
<td>6 (10)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Knowledge of procedure, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
<tr>
<td>2</td>
<td>21 (38)</td>
<td>2 (3)</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>21 (38)</td>
<td>11 (18)</td>
<td>0.001</td>
</tr>
<tr>
<td>4</td>
<td>14 (25)</td>
<td>47 (77)</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>0.001</td>
</tr>
</tbody>
</table>

**Final product, n (%)**

<table>
<thead>
<tr>
<th></th>
<th>Control group</th>
<th>Virtual reality group</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Several surgical trainings involve composite models, VR simulation, cognitive task analysis, and cadaver specimens. VR-simulated studies for surgical practice, such as laparoscopic, cardiovascular, and arthroscopic surgeries, can be found in the literature [39-41]. In the field of orthopedics, researchers mainly focus on sophisticated surgical procedures, for example, (thoracic) pedicle screw placement and insertion (lateral mass screw placement) [42-44], (percutaneous) vertebroplasty [45], knee arthroscopy [38], and hip arthroplasty [46]. These procedures are performed by surgeons who are specialized in orthopedics. In medical schools, for every general practice physician, standard tendon repair is of prime importance in fundamental surgical skills. Thus, it is necessary to develop tendon repair training [47]. To our knowledge, this is the first study to adopt VR simulation for tendon repair training. Adopting VR simulation in regular curricula is challenging owing to the limited efficacy of VR as a learning tool [48]. To clarify the effectiveness, we demonstrated that the VR simulator was an effective tool in the acquisition of tendon repair in our blinded randomized trial. Modern VR simulations have a common disadvantage, that is, high cost. Clarke [48] reported that individual simulators cost up to 6-figure sums. Our platform removed the cost for the students and was open to the public to maximize cost-effectiveness. To consider whether the VR simulator is an educational tool, the results have to be statistically significant with positive feedback. If the surgical performance of the VR group participants did not improve, the VR simulator was not considered as part of the regular training.

A VR simulator plays a major role in the medical field. The lack of medical practice and uneven distribution of medical resources in various regions has resulted in a decrease in clinical practice opportunities for medical students. In addition to tendon repair studies, the VR simulator can perform simulated surgery. Future surgeons can practice with a VR simulator until they are comfortable performing the operation. In addition, experienced surgeons can also study the simulation aspect of the VR simulator to learn and explore new surgical techniques or to discover other surgical options. The application of virtual technology in medicine, medical education, and clinical treatment will have a major impact on the medical system.

During the COVID-19 pandemic, the use of technology has become a popular topic in the medical education field. Tendon repair is a procedure that requires senior professional surgeons; therefore, medical students and junior doctors may not have sufficient practice to be able to perform suturing independently. A possible solution to this problem is that junior doctors practice using the VR simulator, thereby becoming more familiar with the procedure and more confident when performing it. The VR simulator can maximize a medical student’s efficiency with respect to mastering this technique. It has been proven that the VR simulator in a simulation laboratory rather than in an operating room is a better practice method than the traditional classroom study in terms of the flexibility of location [20]. Medical students or residents can perform tendon repair via the VR simulator before performing a formal operation. Using the VR simulator serves the purpose of shortening the operating time, reducing operation errors, and alleviating patients’ postprocedural pain [49]. However, the expense of textbooks or teaching assistance when using the traditional method has no significant comparison with the investment in equipment for VR simulators.

The VR simulator can provide a realistic surgical scenario, thus allowing students to train for a particular skill continuously or to master any unfamiliar procedure. The findings of our study show that students learning via the VR simulator had significantly better scores than those learning via the traditional method with respect to the tendon repair technique \((P<.001)\). This finding may indicate that students using a VR simulator will be able to follow the whole operation more carefully and master the knowledge of the procedure in the future. While we were developing the VR simulator, we tried to mimic reality based on mathematical models by simulating a surgical setting, instruments, training objects, and interactions. We aimed to create an ideal simulator that is realistic in multiple dimensions such as simulation physics, optical properties, and haptic feedback [50]. The system allows students to apply their knowledge and practical skills in a realistic and tactile environment. Another advantage of the VR simulator method is that students can independently perform the surgery, hence increasing the student’s ability to master the technique fully. A VR simulator allows a student to practice at his/her convenience with multiple repetitions, regardless of the availability of cadavers or human trials [51]. Using the VR simulator, the student can maximize his/her surgical efficiency at his/her own learning time even after the scheduled class, which will serve to increase the study’s performance and study result.

A VR simulator can reduce the high costs of conducting animal or human trials [52]. More realistic training usually involves training on animals, but this method is expensive and not available in many medical schools. A VR simulator would...

---

### Discussion

#### Principal Results

Several surgical trainings involve composite models, VR simulation, cognitive task analysis, and cadaver specimens. VR-simulated studies for surgical practice, such as laparoscopic, cardiovascular, and arthroscopic surgeries, can be found in the literature [39-41]. In the field of orthopedics, researchers mainly focus on sophisticated surgical procedures, for example, (thoracic) pedicle screw placement and insertion (lateral mass screw placement) [42-44], (percutaneous) vertebroplasty [45], knee arthroscopy [38], and hip arthroplasty [46]. These procedures are performed by surgeons who are specialized in orthopedics. In medical schools, for every general practice physician, standard tendon repair is of prime importance in fundamental surgical skills. Thus, it is necessary to develop tendon repair training [47]. To our knowledge, this is the first study to adopt VR simulation for tendon repair training. Adopting VR simulation in regular curricula is challenging owing to the limited efficacy of VR as a learning tool [48]. To clarify the effectiveness, we demonstrated that the VR simulator was an effective tool in the acquisition of tendon repair in our blinded randomized trial. Modern VR simulations have a common disadvantage, that is, high cost. Clarke [48] reported that individual simulators cost up to 6-figure sums. Our platform removed the cost for the students and was open to the public to maximize cost-effectiveness. To consider whether the VR simulator is an educational tool, the results have to be statistically significant with positive feedback. If the surgical performance of the VR group participants did not improve, the VR simulator was not considered as part of the regular training.

A VR simulator plays a major role in the medical field. The lack of medical practice and uneven distribution of medical resources in various regions has resulted in a decrease in clinical practice opportunities for medical students. In addition to tendon repair studies, the VR simulator can perform simulated surgery. Future surgeons can practice with a VR simulator until they are comfortable performing the operation. In addition, experienced surgeons can also study the simulation aspect of the VR simulator to learn and explore new surgical techniques or to discover other surgical options. The application of virtual technology in medicine, medical education, and clinical treatment will have a major impact on the medical system.

During the COVID-19 pandemic, the use of technology has become a popular topic in the medical education field. Tendon repair is a procedure that requires senior professional surgeons; therefore, medical students and junior doctors may not have sufficient practice to be able to perform suturing independently. A possible solution to this problem is that junior doctors practice using the VR simulator, thereby becoming more familiar with the procedure and more confident when performing it. The VR simulator can maximize a medical student’s efficiency with respect to mastering this technique. It has been proven that the VR simulator in a simulation laboratory rather than in an operating room is a better practice method than the traditional classroom study in terms of the flexibility of location [20]. Medical students or residents can perform tendon repair via the VR simulator before performing a formal operation. Using the VR simulator serves the purpose of shortening the operating time, reducing operation errors, and alleviating patients’ postprocedural pain [49]. However, the expense of textbooks or teaching assistance when using the traditional method has no significant comparison with the investment in equipment for VR simulators.

The VR simulator can provide a realistic surgical scenario, thus allowing students to train for a particular skill continuously or to master any unfamiliar procedure. The findings of our study show that students learning via the VR simulator had significantly better scores than those learning via the traditional method with respect to the tendon repair technique \((P<.001)\). This finding may indicate that students using a VR simulator will be able to follow the whole operation more carefully and master the knowledge of the procedure in the future. While we were developing the VR simulator, we tried to mimic reality based on mathematical models by simulating a surgical setting, instruments, training objects, and interactions. We aimed to create an ideal simulator that is realistic in multiple dimensions such as simulation physics, optical properties, and haptic feedback [50]. The system allows students to apply their knowledge and practical skills in a realistic and tactile environment. Another advantage of the VR simulator method is that students can independently perform the surgery, hence increasing the student’s ability to master the technique fully. A VR simulator allows a student to practice at his/her convenience with multiple repetitions, regardless of the availability of cadavers or human trials [51]. Using the VR simulator, the student can maximize his/her surgical efficiency at his/her own learning time even after the scheduled class, which will serve to increase the study’s performance and study result.

A VR simulator can reduce the high costs of conducting animal or human trials [52]. More realistic training usually involves training on animals, but this method is expensive and not available in many medical schools. A VR simulator would...
provide many benefits in this situation. VR simulators allow students to repeat the exercise several times—an action that can ensure that students are totally familiar with all important concepts before surgery. In addition, students are able to familiarize themselves with the technique before surgery, which also guarantees safety for the patient. Studies show that surgeons using VR and other simulation methods for practicing surgical techniques can reduce operative time or possible errors, thereby increasing their confidence and decreasing uncertainty in the procedure outcomes [53]. In addition, most surveyed residents and directors believe that a surgical simulation is a useful tool for complementing traditional forms of training on animals, cadavers, or synthetic models [51]. Although the texture of operation cannot be simulated exactly on a VR simulator, using animals, cadavers, or synthetic models with a VR simulator could also allow students to practice their techniques with more understanding, hence reducing the cost and availability issue of animals, cadavers, and synthetic models.

Surgical training is different from other courses in medical school because students need to be exposed to nonstandard or beyond standard levels of teaching to help them understand how much they need to learn or improve. Studying with different methods can help students have a clear understanding of this curriculum. Surgical education departments have already purchased some effective training equipment that can help medical students achieve consistent and measurably high level of performance. Therefore, combining traditional and alternative teaching methods such as composite models and cognitive task analysis is necessary for the future [54]. Al-Nammari et al [55] found that the average duration of orthopedic education in medical schools in the United Kingdom was approximately 2.65 weeks; however, medical schools may not provide their students with enough scheduled course time to master musculoskeletal medicine. One of the solutions to this problem is that students combine VR simulation with traditional methods, which will save time and maximize students’ efforts in musculoskeletal medicine. If the student is interested in this field, this study also shows that students with interest in orthopedics had a statistically better understanding of musculoskeletal medicine than their counterparts. In a broader view, more elite students entering the field of musculoskeletal medicine will increase the possibility of new discoveries in musculoskeletal medicine.

VR simulators have good reusability. Once a VR simulator is established, the study will not be limited to that particular location or the number of students. The VR software can be freely downloaded by students from the internet. Therefore, medical students can study in any location and are not limited to the classroom or training room. Rather than taking turns practicing on subjects, students are able to practice together using the VR simulator. The 3 important advantages of students using VR simulators concurrently are as follows: (1) VR simulators allow students to perform the experiment together, (2) VR simulators allow the exchange of ideas or knowledge of a topic at the same level, and (3) VR simulators ensure that everyone is on schedule for the course.

VR simulators will be part of the future technology in the medical field, and medical students should be exposed to this technology as soon as possible to familiarize themselves with this method. In addition, feedback from students using a VR simulator can help improve the simulator, and further developments may be performed. The application of virtual technology in the field of medicine, medical education, and clinical treatment will have a major impact on health care and the development of the medical field and on the training of young doctors. Both software and hardware in the VR simulator can be flexibly configured according to the needs of teaching. In addition to tendon repair, VR simulators provide a set of virtual instruments that can be developed into different functions. The scope of virtual orthopedics is very wide and includes the fields of endoscopic protection, radiosurgery, microsurgery, and remote surgery. Moreover, a VR simulator is currently used in neurosurgery, eye, heart, plastic, and abdominal surgeries, and many other areas. Therefore, the widespread use of a VR simulator indicates that studying or performing surgery via a VR simulator is an effective method for completing this training task.

This study is the first to show that a VR simulator can also be applicable in medical schools and that it is a very effective method for medical students to enhance their learning through multiple repetitions by participating in a 360° VR environment. In addition, a VR simulator should be used in sections not only for complex surgeries but also in important presurgical sections, for example, tendon suture. Studies have shown that medical students enjoy early exposure to clinical practice and perceive that it is valuable for future studies [56,57]. As mentioned earlier, multiple medical professions have attempted to utilize VR simulators for their medical training. However, most professions cannot overcome a critical issue—the high cost. This high cost also results in only using the compatible device of the VR simulator to minimize costs. The VR simulator program was uploaded on the internet, and all medical students could choose the versions according to their own needs and the available equipment. The PC version can be downloaded on the internet at no cost. This inexpensive, remote, detailed, sustainable, and effective platform should be used in medical schools for all medical students. A successful education system should be accessible to all students [38]. This VR simulator with high accessibility allows different students to have multiple practice opportunities and to participate in in-depth discussions in the specific communication section. Designing this program was costly; however, with more students using this program, the cost-effectiveness of this system can maximized.

Despite the benefits of using VR simulators to train medical students, there are some challenges for medical schools to adopt this new approach in the future. For designing the program, the development corporation or the program writers who are not medical professionals have limited concepts or knowledge of the orthopedic procedure. This may lead to simulation results of nondetailed procedures, and medical students may miss the surgical details when practicing surgery in VR simulators. To overcome this obstacle, we had extensive discussions and we shared opinions back and forth between the medical school and the development team during the early period of establishing the VR simulator. It took at least a year to achieve an acceptable VR simulator for medical students. The other problem of VR simulators is the lack of validated score measures. Therefore,
the final assessment was evaluated on synthetic models by using the global rating scale. Surgical training is a comprehensive process; therefore, medical students and surgeons may take years to master a surgical technique. The VR simulator is part of the comprehensive training but not the only training technique. Our findings show that the VR simulator is an effective method among the traditional teaching methods. We hope that future studies can focus on an effective VR simulator measurement scale or comparative evaluations between simulators.

Limitations
The follow-up period in this study can only reflect the short-term effect of the VR simulator, which was a limitation of this study. The long-term effects on orthopedic specialists who practice on VR simulators could take years to evaluate.

Conclusion
Compared to the traditional tendon repair method, the VR simulator for learning tendon repair significantly improves medical students’ time in motion, suture skill, the flow of operation, and knowledge of the surgical procedure. Our paper sets an example for future VR simulator development for medical curricula.

Authors’ Contributions
TNM, JP, and WKM designed the study and wrote the manuscript. WKM and QH reviewed the risk of bias of the studies and the manuscript. JC and JL extracted the data from the studies. THS and JD interpreted the results. JL and ZZ supervised the entire study. The authors read and approved the final manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Virtual reality simulator showing every step of the suturing process.
[DOCX File, 16 KB - games_v9i3e27544_app1.docx ]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1295 KB - games_v9i3e27544_app2.pdf ]

References


24. Orthopedic trauma emergency integrated diagnosis and treatment simulation course. ilab-x. 2019. URL: http://www.ilab-x.com/details?id=2934&isV=true [accessed 2021-06-23]

25. SteamVR. URL: https://store.steampowered.com/steamvr/ [accessed 2020-09-01]

26. SteamVR. URL: https://store.steampowered.com/steamvr/ [accessed 2020-09-01]


31. Mok et alJMIR SERIOUS GAMES 2021 | vol. 9 | iss. 3 | e27544 | p.118https://games.jmir.org/2021/3/e27544


Abbreviations

BS: Bunnell tendon repair with figure 8 tendon repair
KS: Kessler tendon repair with 2 interrupted tendon repair knots
PC: personal computer
VR: virtual reality
Medical Students’ Perceptions of Play and Learning: Qualitative Study With Focus Groups and Thematic Analysis

A E J Van Gaalen¹, MD; A D C Jaarsma², DVM, Prof Dr; J R Georgiadis¹, PhD
¹Department of Biomedical Sciences of Cells and Systems, Section Anatomy & Medical Physiology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands
²Center for Education Development and Research in Health Professions, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
A E J Van Gaalen, MD
Department of Biomedical Sciences of Cells and Systems, Section Anatomy & Medical Physiology
University Medical Center Groningen
University of Groningen
Antonius Deusinglaan 1
Groningen, 9713AV
Netherlands
Phone: 31 640994891
Email: a.e.j.vangaalen@umcg.nl

Abstract

Background: In times where distance learning is becoming the norm, game-based learning (GBL) is increasingly applied to health profession education. Yet, decisions for if, when, how, and for whom GBL should be designed cannot be made on a solid empirical basis. Though the act of play seems to be intertwined with GBL, it is generally ignored in the current scientific literature.

Objective: The objective of our study was to explore students’ perceptions of play in leisure time and of GBL as part of a mechanistic, bottom-up approach towards evidence-informed design and implementation of GBL in health profession education.

Methods: We conducted 6 focus group discussions with medical and dentistry students, which were analyzed using thematic analysis.

Results: A total of 58 students participated. We identified 4 major themes based on the students’ perception of play in leisure time and on the combination of play and learning. Our results indicate that, while play preferences were highly various in our health profession student cohort, pleasure was the common ground reported for playing. Crucially, play and the serious act of learning seemed paradoxical, indicating that the value and meaning of play are strongly context-dependent for students.

Conclusions: Four key points can be constructed from our study. First, students play for pleasure. Perceptions of pleasure vary considerably among students. Second, students consider play as inefficient. Inefficiency will only be justified when it increases learning. Third, play should be balanced with the serious and only be used for difficult or tedious courses. Fourth, GBL activities should not be made compulsory for students. We provide practical implications and directions for future research.

(JMIR Serious Games 2021;9(3):e25637) doi:10.2196/25637

KEYWORDS
gamification; serious games; game-based learning; medical education; computers; new technology; focus group; play; qualitative

Introduction

In times where distance learning is becoming the norm, game-based learning (GBL) is increasingly applied to health profession education [1,2]. Yet, recent reviews in this field indicate that GBL research is still in its infancy and that robust study designs based on sound theoretical foundations or supporting scientific evidence are scarce [3-5]. Although certain trends in GBL use can be observed, there seems to be little theoretical support to clarify the effects of GBL on academic learning [4]. Most studies report on the use of GBL applications that are tailored to specific local settings [4]. Therefore, decisions about implementing game-like interventions — if, when, how, and for whom — cannot be made on a solid empirical basis. This increases the likelihood of suboptimal and even counterproductive educational design. In this study, we took a user-driven approach in an attempt to unravel key
processes that could explain why and how GBL does or does not work in health profession education, therewith advancing the science and practice in this field.

GBL has been applied based on the idea that play and learning are closely intertwined [6,7]. Intriguingly, studies investigating GBL in academic settings do not seem to explicitly scrutinize, measure, or mention play [3-5,8,9]. GBL studies mainly focus on outcome measures such as learning outcome, motivation, and likability of GBL [4] but seem to dismiss play itself. This situation seems to persist despite significant scientific interest in the fundamentals of play-learning interaction [10-13]. Generally, empirical evidence supports claims that play positively influences the learning of problem solving [14-16], creativity [17,18], and self-regulatory skills [19]. Most knowledge on this play-learning interaction stems from early experimental animal research [20,21] or research on developmental psychology in children (eg, [14,22]). By comparison, play in the realm of adult learning has received very little attention, especially with regard to GBL.

The studies that came closest to research on play-learning interactions in educational contexts for adult learning were quantitative in nature and primarily used questionnaires aimed at examining experiences with already existing games or to inform the design of a specific game [23-27]. However, these studies did not investigate how play can be elicited in students or, more specifically, which type of play can benefit or support student learning, in which situations, or under which conditions. Furthermore, because of the specific study set-up or study intent, participants in such studies may have directed their answers to a specific game or game design, which does not allow for generalization of the results to other contexts or game designs. Next to quantitative studies, qualitative approaches have been employed in order to describe adult playfulness [28] or inform game design [29-33]. Findings of these studies gave insight into self-perceived reasons for adult engagement in play such as stress relief, challenge, and friendship. However, whether these needs for play in adults can also be met in combination with learning was not explored in these studies. Thus, the links between play, academic learning, and GBL remain a blind spot in the literature.

Provided that there are meaningful play-learning interactions in GBL, even when the nature of that interaction is unknown, we need to understand how to elicit play in students. But what exactly is play, how do we define play, and how do we relate play to GBL? There is no univocal answer to any of these questions, since there is considerable disagreement in the scientific literature as to what constitutes play and games [13,34]. Interestingly, and perhaps as a logical consequence of this disagreement, there is strong consensus that playfulness is an individual predisposition [35] and that the liking of play is dependent on personal opinions, characteristics [36-38], and context [34]. Some propositions have been made by play scholars to classify different expressions of play and distinguish play from other behaviors such as exploration [13,39]. Probably most interesting in this regard is the distinction between paidia (free, spontaneous, expressive, creative forms of play) and ludus (rule-bound play) [40]. These heuristics can be very valuable for the theoretical conceptualization of GBL, because GBL design seems to relate much more to rule-bound “ludic” play [4,26,40] than to free, creative “paidic” play. Furthermore, the strong individual character of play that has been established in the literature seems to require qualitative research approaches to understand students’ perceptions of play and academic learning, especially in relation to GBL.

In the present study, we took inspiration from play research as a first step towards a mechanistic analysis of GBL effects. We employed the qualitative method of open focus group discussions to help us gain deeper insight into medical and dentistry students’ perceptions of play and learning by exploring their ideas, interpretations, feelings, and actions [41] as well as favorable circumstances or limitations for engaging in GBL activities. Although, at this point, we do not have scientific reasons to assume that the range of opinions on play would vary significantly across students as a function of the academic level or discipline they are enrolled in, we chose to focus on medical and dental students for 2 main reasons. First, our main teaching experience as well as our research interest in GBL lie within the context of health profession education. Second, if play preferences are indeed highly individual and contextual, this would also apply to students enrolled in a particular program or discipline. In summary, in this focus group study, we explored health professions students’ perceptions of what constitutes play, play-learning interaction, and if, how, and when GBL material should be designed and implemented in health professions education to foster their learning.

Methods

Context

We conducted this study at the University of Groningen, University Medical Center Groningen, the Netherlands, between March 2019 and April 2019. The 6-year undergraduate medical curriculum of the University of Groningen is comprised of a 3-year Bachelor’s phase and a 3-year Master’s phase. The Bachelor’s program includes 2 Dutch-taught and 2 English-taught tracks, called learning communities. The program is problem-based and patient- and student-centred, with a focus on tutor groups, practicals, and e-learning rather than lectures. The students are expected to be proactive, and they are encouraged to develop self-regulated and self-directed learning skills to pursue lifelong learning. The 3-year Master’s program includes 2.5 years of clinical rotations (1 year of junior clerkships, 1 year of senior clerkships, 0.5 year of elective clerkship), and 0.5 year master thesis.

The 6-year undergraduate dentistry curriculum of the University of Groningen, likewise, is comprised of a 3-year Bachelor’s phase and 3-year Master’s phase and has a patient-centred approach. Compared to medicine, the dentistry Bachelor’s phase has a stronger focus on lectures and practicals and is taught in Dutch only. The Master’s phase consists of 1 year of mainly skills labs and practicals, while the final 2 years mainly consist of clinical rotations and a master thesis. Both medicine and dentistry students use e-learning, and teachers sometimes apply GBL, but there is no considerable nor structural implementation of GBL in either curriculum.
Participants and Ethical Considerations

We used convenience sampling and invited all medical and dental students to participate in our study via an online announcement on the virtual learning environment of the University of Groningen, which is also used as a communication platform and is visible to all students (N=1600). We explained that the purpose of the focus group study was to gain more insight into students’ perceptions of play due to increasing interest in GBL. It was communicated that students not interested in games would also be able to participate in this study. We did not set specific exclusion criteria.

Ethical approval was obtained via the Netherlands Association for Medical Education (January 11, 2019). Prior to the start of each focus group session, the participants signed an informed consent form and completed a brief demographic questionnaire. They were informed that there were no correct nor incorrect answers and that we were interested in all ideas and perceptions. The discussions were structured around a short break. Before the break, discussions aimed at exploring playful behavior in leisure time. After the break, discussions continued and focused on participants’ ideas and perceptions of the play-learning interaction and how GBL could be implemented in the curricula to foster their learning. We used a topic list with open-ended questions (Textbox 1) and encouraged further discussion. The first 4 sessions were moderated by 1 of the authors (AJ). An observer (Ob1) was seated outside the group and took detailed field notes of group dynamics, atmosphere, and nonverbal communications. The last 2 sessions were moderated by the observer of the first 4 sessions and, consequently, a different observer (Ob2) was used. To create an open and social atmosphere, pizza and soft drinks were served.

After 4 sessions, our sample provided sufficient information power to address the aims of this study [43]. The information we had gathered from these focus groups was used to fine-tune and add some questions to the topic list for the final 2 focus group sessions (Textbox 2). Since no new information was obtained in these 2 sessions, we decided not to schedule any further sessions [42].

Focus Group Sessions

The focus group sessions followed the guidelines as described by Krueger et al [42] as well as the Association for Medical Education in Europe (AMEE) guideline on using focus groups [41]. Initially, 6 focus groups sessions (4 Dutch and 2 English sessions, with a maximum capacity of 13 students per session) were planned, each lasting 1.5-2 hours.

With the consent of the students, all meetings were audiotaped for later transcription and analysis. It was explained that there were no correct nor incorrect answers and that we were interested in all ideas and perceptions. The discussions were structured around a short break. Before the break, discussions aimed at exploring playful behavior in leisure time. After the break, discussions continued and focused on participants’ ideas and perceptions of the play-learning interaction and how GBL could be implemented in the curricula to foster their learning. We used a topic list with open-ended questions (Textbox 1) and encouraged further discussion. The first 4 sessions were moderated by 1 of the authors (AJ). An observer (Ob1) was seated outside the group and took detailed field notes of group dynamics, atmosphere, and nonverbal communications. The last 2 sessions were moderated by the observer of the first 4 sessions and, consequently, a different observer (Ob2) was used. To create an open and social atmosphere, pizza and soft drinks were served.

After 4 sessions, our sample provided sufficient information power to address the aims of this study [43]. The information we had gathered from these focus groups was used to fine-tune and add some questions to the topic list for the final 2 focus group sessions (Textbox 2). Since no new information was obtained in these 2 sessions, we decided not to schedule any further sessions [42].
Textbox 2. Additional questions

- What does a game make ‘addictive’?
- Does anyone not like to play games?
- When do you prefer to play?
- Do you ever play drinking games? Why do you think that is attractive?
- Which type of play elements do you believe would work in education?

Data Analysis

All audiotapes were transcribed verbatim and anonymized before analysis. Atlas.ti (version: 8.4) was used as software to help us manage and analyze the data [44]. The method of thematic analysis was used to evaluate the data [42]. We used the most widely adopted approach for thematic analysis [45] outlined by Braun and Clark [46] and consisting of 6 steps: (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, (6) producing the manuscript. Notably, this method of analysis is recursive, meaning that each subsequent step in the analysis might have prompted us to circle back to earlier steps in light of newly emerged themes or data [45]. The detailed observers’ field notes facilitated additional exploration of themes when needed throughout the entire process.

First, coders (AvG and Ob1) familiarized themselves with the data by examining and re-examining the transcripts and audiotapes. Second, initial codes were generated (Avg and Ob1) to organize the data on potential items of interest [45]. One focus group discussion was coded (AvG and Ob1); thereafter, the coders discussed and defined a coding framework for the remaining dataset while denoting possible patterns or discrepancies between the codes (Table 1) [46]. All disagreements between coders were resolved via discussion between the coders. Open coding was used to ensure flexibility to incorporate themes outside our questioning route or initial coding table (Table 1) [47]. Third, the identified codes from all focus groups were discussed with the entire team in order to construct themes. We inductively [41] and iteratively constructed themes by comparing, analyzing, combining, and mapping codes [45]. Fourth, the team (iteratively) reviewed the identified themes to examine whether they were sufficiently common and coherent, but also whether they were sufficiently distinct from each other to justify separation [45,46,48]. Fifth, we ensured that the denominators of our themes were adequately clear and descriptive. Finally, we wrote the manuscript, which proved to be a continuation of the iterative interpretation and analytic process of thematic analysis [49].

Table 1. Initial coding framework.

<table>
<thead>
<tr>
<th>Preliminary codes</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>Luck and unpredictability, ownership, meaning</td>
</tr>
<tr>
<td>Escapism</td>
<td>Fantasy, immersion, escapism, relaxation</td>
</tr>
<tr>
<td>Social</td>
<td>Being together, helping each other</td>
</tr>
<tr>
<td>Strategy</td>
<td>Strategy</td>
</tr>
<tr>
<td>Mechanics</td>
<td>Duration, variation</td>
</tr>
<tr>
<td>Achievement</td>
<td>Challenge, winning, losing, competition, revenge, provocation</td>
</tr>
<tr>
<td>Devotion</td>
<td>Dark play, eagerness</td>
</tr>
<tr>
<td>Exploration</td>
<td>Storytelling, learning new things, curiosity</td>
</tr>
<tr>
<td>Applicability for learning</td>
<td>Difficulty subject, boring, paradox with leisure</td>
</tr>
</tbody>
</table>

Reflections

Our research team consisted of researchers with various backgrounds, supporting a critical examination and interpretation of the data from multiple perspectives. During the team discussions, we deliberately addressed all these perspectives, while allowing each team member to make an equal contribution.

AvG has a medical degree, is appointed as a lecturer (ie, anatomist), has a research interest in the motivational pull of play and games, and develops GBL strategies. JG is an associate professor of anatomy with a research interest in affective neuroscience and motivational forces in education. AJ is a full professor of Health Professions Education and Research with ample experience in qualitative research. Ob1 is a master student in Dentistry, assists multiple (clinical) educators in developing e-learning, and helped perform this study as part of her graduation assignment.

AvG and JG did not join the focus group sessions because they might know the participating students; AJ did not know any of them. Ob1 knew some participants in 2 out of 4 focus group sessions she observed; however, these participants did not consider this to be a problem, and they felt free to speak their minds. When Ob1 acted as a moderator, she did not know any of the participants.
Results

A total of 58 participants volunteered to participate (41 women and 17 men; mean age 22.8 years, range 18-31 years). This sample was comprised of 30 Bachelor students in Medicine, 8 Master students in Medicine, 2 Bachelor students in Dentistry, and 18 Master students in Dentistry. Of the participants, 51 reported to be of Dutch nationality and 7 of a nationality other than Dutch (Brazilian, French, Israeli, Italian, Saudi Arabian, Romanian, South African). The number of participants joining each focus group ranged between 7 and 13 students. One focus group session only included dentistry students (n=13), 1 session was predominantly attended by dentistry students (6 of 7 were dentistry students), 1 session was predominantly attended by medicine students (11 of 12 were medical students), and 3 sessions (including the 2 English sessions) were only attended by medical students. We found no distinct differences between the opinion on play or GBL between dentistry and medical students. The detailed field notes yielded no additional results for analysis.

We chose to present our findings based on the structure of the focus group discussions: first, students’ perceptions of play in leisure time, then their perceptions of GBL, and finally, the interaction between play and learning. In the following sections, quotations are used to illustrate the findings. Identified themes are in bold and italic, and identified sub-themes are bold.

Perceptions of Play in Leisure Time

At the start of each focus group session, students discussed their favorite games in leisure time. A great diversity offavorite game genres was mentioned by the students, for example, puzzle/jigsaw games, shooting games, strategy games, sport games, and adventure games. As one student stated: “I think there’s no game that everyone likes to play…”

All students liked to play, but the amount of play in leisure time ranged considerably from only once a year to daily. Pleasure seemed to be the common ground as to why students engaged in games, in all their diversity.

Pleasure

Whether the games were solitary (eg, patience, jigsaw, or shooter games) or multiplayer or collaborative (eg, Monopoly, settlers of Catan, or FIFA), students felt that games should be fun. However, ways to achieve a pleasurable experience from play varied considerably across students.

For instance, fun could come from the joy of winning (“I really like winning.”), from the feeling of supremacy and achieving something (“You are special. You have something that others don’t have.”), or from the delight of getting a good story out of a game (“I’ve always seen video games as ‘my book kind of thing’. I don’t read a lot of books, so I get my stories from games.”)

A striking aspect that was highlighted in the discussions was that not only the pleasure experience itself (eg, the experience of a victory) but also the sense of pleasurable anticipation motivated students to continue playing: “I continue playing until I win the final match.”

Students indicated that pleasure should not be easily obtainable: “it has to be a challenge to win.”

Reward uncertainty seemed to modulate the impact of the pleasure experience, such that uncertain wins were associated with greater pleasure than certain wins. Upon analyzing the students’ statements, it became clear that they experienced greater joy after a difficult win, compared to easy wins:

… father always wins [at Scrabble], that’s not the worst. … it also gives more satisfaction if you beat him.

What I like, is when you really make a brilliant move, so someone else just doesn’t win, but you do.

However, the degree of reward uncertainty seemed to have an optimum. Students said that if the reward seemed out of reach and their chances of winning were little to none, or even close to a certain loss, all anticipatory tension was gone. When students no longer had fun or prospect of pleasure, they felt less motivated to continue playing:

When you keep losing, you’re done with it [the game]

It [Monopoly] takes too long. You’re like “let’s just stop, do we really have to finish [the game]?“

The final major part of the pleasure experienced in play that was brought up in the discussions was social pleasure. Students tended to play games in groups of close friends or family or with new people (met in pubs, societies, or a digital world), which helped them gain or strengthen the sense of collectiveness and sociability:

It [playing a game with each other] makes you feel like you are in your own world.

You can talk about the game and about everyday life, which offers opportunities for discussion.

… then you just want it [the game] to last a long time, because you have such a good time with each other.

Students mentioned that play more easily creates a bond, a sense of social togetherness, which in turn can also be enriched through play; “… it makes you feel connected.”

However, the sociability of play could also backfire when players with competitive spirits who could not win (sore losers) ruined the game:

I’m very fanatic. If I lose, I’ll also be grumpy for an hour. A lot of people also don’t like to play a game with me.

The sociability of play could also backfire when players who disliked strong competition were disappointed because play was merely reduced to competitiveness and the desire to win:

If they are all very fanatic, it (the game) doesn’t matter that much to me anymore.

Perceptions of Game-Based Learning

In order to keep the balance between play and the serious act of learning, participants brought up and discussed their perceived requirements and the relevance for implementing GBL.
**Perceived Requirements**

Despite a possible unpleasant confrontation with the serious world (see the Paradox section), competition was believed to enhance learning:

> The more you compete, the more you will learn by yourself, because you want to improve.

However, in order to keep competition playful and in balance with the more serious part of learning, students felt that players’ identities should remain confidential in GBL using competition (for example by choosing a nickname) or that players should be grouped into (collaborative) teams:

> They [the other students] are allowed to see the [game] results, but then anonymously.

> When you’re losing [a teamplay game], you’re not losing alone. So that’s also nice.

Students’ perceptions of meaningful GBL design generally stayed close to their learning task at hand (ie, the learning task itself could be easily recognized). Students particularly referred to disciplines they found difficult or tedious, such as anatomy, physiology, cell biology, immunology, or statistics:

> … if courses are really tedious and dry, it [GBL] shows you that it’s [the course content] is useful, and if you play it right, the [new] knowledge sticks.

> … if you can find a game to make people understand physiology, you’re a genius!

Furthermore, students often mentioned game versions of their future workplace (based on The Sims game), which gave them opportunities to learn playfully by building their own practice:

> I used to play The Sims a lot and really liked to build.

> … Wouldn’t it be great to build your own dentistry practice in a Sims kind of way!? Designing your practice, doing treatment, making money to go to courses in which you can learn new treatments, through which you can make even more money so you can improve your practice, can get more staff etc. …

**Relevance**

Students’ opinions on the need for GBL were divided:

> I don’t necessarily want to play a game every time I go to class.

> … yes, I think we need it [GBL].

Although there was some debate about how frequently GBL should be used, the consensus seemed to be that GBL could support learning. At times, the medical education continuum was experience as long-lasting, hectic, and stressful:

> It [dentistry] is really a study for the long haul.

> We are all really stressed, and everyone’s stressing each other out, like: “Have you passed the exam/the test?” or “Did you hear? He hasn’t passed it [the exam/the test]!”

Subsequently, students mentioned that adding playful fun to learning might help relieve stress in stressful times:

> Why not make it a little more light-hearted? Just to relieve the tension every now and then.

Nevertheless, students felt that there has to be a balance between the playful and the serious, which has to be respected:

> If we turn aspects of the 6-year learning process into play, it feels as if the/all seriousness has been lost.

The extent to which the serious and the playful should be balanced depended very much on personal preferences. Therefore, an approach tailored to students’ needs would be the best fit according to the students:

> Make it an extra activity, because playing a game just doesn’t work for some people…

> I think it’s also important to keep in mind that everybody is different …

**Perceptions of the Interaction Between Play and Learning**

The combination of play and the serious act of learning seemed paradoxical to students (Figure 1).
Paradox

Students overwhelmingly indicated that GBL should not be made compulsory. Compulsory play sounded for them like a “contradictio in terminis”: Play would become serious, which cannot be play:

If you are forced to play games, it would be like school.
If it [a GBL activity] becomes compulsory, then I don’t like it anymore.

Students considered play to be a leisure-time activity to temporarily escape the serious demands of daily life:

It’s really relaxation, just something completely different [play], which has nothing to do with anything else.

Integrating play and the serious act of learning into education, therefore, seemed to be a paradox:

I think it’s strange that you can be enjoying gameplay in your private life for fun and relaxation, but apparently, if you frame it as “education,” it suddenly becomes too much.

Indeed, although students were apt to think of play as pleasure during leisure time, it was difficult for them to link play and pleasure to academic learning:

I find it difficult to see it [learning/GBL] as a game because it’s all so serious. Something is depending on it. And when I think of playing games, I think “Ah, cozy, fun!”
If I have to get together with everyone for a joint activity [GBL], then I think “No, I just have my own way of studying. And if I deviate from that, then I get really upset.”

Students believed that adding play [to education] would reduce their learning efficiency:

You probably have to “camouflage” the learning [part], which will probably require more study time.
I think (educational) games just have to be short and efficient…
If something [GBL] really takes a lot of time, then people are inclined to think, as always: “I just quickly read this [book] instead of wasting my time on a game.”

In addition to reduced learning efficiency, the paradox of play and learning was attributed to a mismatch in identity in play as compared to reality:

That’s the funny thing with games; you can pretend to be different than you normally are.

Losing a game in the imaginary play environment was never seen as fun but considered trivial nonetheless; from a gameplay point of view, the game was over, and the ending was (most of time) appreciated: “It’s just a game.”

However, students felt that their playful imaginary identity would be lost in learning. Losing a game in a learning activity or environment was considered to possibly lead to unpleasant and stressful confrontations with the real world:

… because it’s a game, just a one-time thing. And here, even if it’s just a Kahoot, and in general, sometimes you just have a group of questions you really don’t know anything about. But you can take it personally, even though you don’t have to. And think “I’m not a good student, but I want to become a good doctor” and “they’re all going to be better doctors than me” and “my resume is not good enough.”
In such cases, students particularly mentioned that competition had influenced this unfavorable confrontation with the serious:

*I think you don’t want to show [your peers] that you’re not able to do something [well], and if you do it [GBL] in the form of a competition, that there’s always someone better than you. You have the feeling that you’re less good at it.*

Or, students were concerned that competitive behavior in games would become prominent in their education as well:

*I’m already chasing all the credits [in the curriculum] and I feel like it [competitive GBL] would make me too competitive, too reward focused.*

**Discussion**

In our study, we took inspiration from play research as a first step towards a mechanistic analysis of GBL effects. On the basis of open focus group discussions, we explored how medical and dental students perceived play in leisure time as well as in the context of academic learning, GBL in particular. The student samples were representative of the student population in terms of age, intellectual level (university students), and academic interest (medicine and dentistry). All students reported that they liked to play in leisure time. However, analysis of the transcripts showed that they had very different ideas about how pleasure could be achieved through play. Although we intentionally did not refer to a specific definition or conceptualization of play during the focus group discussions, students naturally discussed play in the context of digital, card, and board games. At the evaluative level, we observed a strong tendency towards rule-bound “ludic” play and only a weak tendency towards free, creative “paidic” forms of play [34,40,50-52]. An important observation from our analysis pertains to the context-dependency of the reported playfulness. Students openly and enthusiastically discussed play in leisure time, but when they were asked to discuss play in the context of GBL and academic learning, they began to carefully formulate their perceptions of play. They became cautious and began to change their perceptions of play, and many even became sceptical or disapproving. It seems that the outcomes of our focus group study not only allowed us to confirm some of the key principles of adult play (eg, challenge and sociability) [34,50,51,53] but also enabled us to generalize these to the context of health profession education. Moreover, we were able to identify key elements to consider in deciding if, how, and when to adopt GBL in health profession education to, possibly, foster learning.

Pleasure was a central theme in our open focus group discussions. This is not surprising because play, in its most fundamental expression, is seen as one of the primary positive emotions common to all mammals. Interestingly, students’ perceptions of what made play pleasurable varied considerably and involved not only positive affect (eg, fun, sociability) but also affective states that can be taken in a more negative way (eg, the urge to win). This variation persisted across participants and focus group sessions, even though the focus group composition was similar regarding demographic characteristics. This finding is consistent with previous literature stating that play preference, inclination to play, and the meaning of play are associated with many variables such as culture [54-56], personality [57-59], gender [60-62], and play frequency [63,64].

Attesting to the variable nature of play liking is that even negative affect, such as feelings of guilt or antisocial behaviors like sadism and violence [65-67], can be pursued in play and games and might be considered pleasurable in certain contexts [68-71]. It thus seems clear that in humans, any analysis of the interaction between pleasure and play must also exceed the level of primary emotion.

In our analysis of the transcripts, we adopted a multilevel conceptualization of pleasure [72], where we consider pleasure as more than just the joy of playing, which resonates findings from other fields such as developmental psychology, psychoanalysis, and neuroscience [18,72-74]. Pleasure research showed that various positive and negative behaviors and incentives can activate the same pleasure system in the brain and that pleasure is contextual and mainly dependent on individual experiences [75,76]. Pleasure can also refer to mood states (eg, happiness [77], a feeling of content), which can be maintained by perseverance [78,79], even at the cost of momentary negative affect [80,81]. Human play has also been associated with interest [82-84], surprise [85-87], and arousal [88-90]. In our focus group discussions, some students mentioned that the anticipatory joy of possibly winning as a main reason why playing was fun for them. Students also discussed the pleasure of uncertainty in this respect. On one hand, they perceived the pleasure of uncertainty (imagining winning the game) as more motivating than the pure pleasure of certainly winning (“difficult wins over easy wins”). On the other hand, they perceived a high probability of not being able to achieve the desired outcome (certain loss) as demotivating and in fact, as a reason to end the game, which is in line with the literature [91]. Indeed, is it well established that reward probability has a powerful effect on the anticipatory state of pleasure; the greater the reward uncertainty, the greater the motivating effect will be, but only if there is (at least) some probability in receiving that reward [91,92].

Another frequently discussed element of play pleasure was sociability. Many students believed that playing together was way more fun than playing alone. This is in line with findings of research in animals other than humans, in which social play is characterized as a high-level reward [93,94], more pleasurable than other forms of social interaction [95-97], and, intriguingly, at times even more pleasurable than food [98,99]. Also in human studies, it has empirically been shown that a prominent characteristic of social play is its high reward value [98,100]. However, some students appeared to be hesitant of social play; they perceived that pleasure gained from social play was sensitive to any dominance hierarchy within the player group.

A main observation from the focus group sessions was that students’ enthusiasm about play dampened when the context of the discussion shifted from leisure time to academic learning and GBL. Students mentioned many instances where play could be beneficial for their learning or even for their personal well-being. However, students also mentioned that GBL felt, at times, like a paradox and that play cannot be implemented in every course. Perhaps, this might be attributed to a shift from
intrinsic motivation (free play in leisure time) and extrinsic motivation (when play becomes task-based and therefore, possibly, less fun). Another important aspect of this particular discussion was that students perceived the implementation of play as the opposite of efficient learning. This criticism had to do with the underlying belief that academic learning is serious and that the opposite of seriousness is play. Students indicated that the act of academic learning should be efficient but saw play as inefficient. Nevertheless, they seemed to justify the inefficiency of play when it could increase the efficiency of learning. They saw the greatest benefits for somewhat tedious, difficult subject matter. As a corollary, this may also imply that if students judge a particular learning activity as too playful, they will criticize it as inefficient and rather prefer to avoid participating. This is in line with the moderate enjoyment hypothesis theorizing that the general relationship between entertainment and learning is an inverse u-shape [101]. According to this hypothesis, entertainment (and the resulting pleasure) only facilitates learning up to a certain point, the peak of the inverted u-shape. After that point, learning will decrease, possibly due to distraction (leading to inefficiency) of entertainment [88]. Interestingly, but paradoxically enough, pleasure associated with play may be perceived as hindering the achievement of higher (academic) goals, which, in turn, is also part of the pursuit of happiness and well-being by providing long-term pleasures [75].

Considering play as inefficient corresponds with the literature on this subject. For instance, Suits [51] stated that all play involves sacrifice of efficiency; there are always easier ways to obtain goals than through play. In golf, for example, there are far more efficient ways to get a small round object into the ground than with the swing of a golf club, but the voluntary acceptance of game rules permits the player to do so [51]. He and many others argued that, without the voluntary acceptance of these rules (with an inherent loss of efficiency), play will be lost [34,50-52]. Voluntarily accepting rules in favor of less efficient means also resonates with our finding. Students stated that play implemented as a learning tool (ie, GBL) should not become a compulsory activity for students. This is in line with work from developmental psychologists [102,103], research on motivational theories applied to play [104], and views from play scholars [34,50,51].

Students’ opinions about competition in play varied considerably, depending on the context. When students considered competition in the context of leisure time play, their focus was on the (prospective) joy of winning, and they could also interpret winning and losing as trivial outcomes of play. Nevertheless, competition was believed to improve oneself. Others, however, did not play for competence, but rather for the sake of sociability (pertaining to the “relatedness” need). Also differing from the relatedness need was the finding of students who were interested in a single player game that draws them into the storyline. Autonomy on the other hand was very much agreed upon; play should not be compulsory. Building on the self-determination theory, Nicholson’s RECIPE for meaningful gamification is a design theory that describes 6 elements (Reflection, Engagement, Information, Choice, Exposition, and Play) in order to attain intrinsically motivated usage [109]. This theory is in line with many of our findings. For instance, GBL/play should be free (Play element), should be a choice (Choice element), should not deviate too much from the real-world setting (otherwise it might be deemed inefficient; Information and Exposition element), should be challenging and socially engaging (Engagement element), and should have a narrative (Exposition element).

Practical and Research Implications

In research on GBL, design choices have rarely been made explicit, and most studies use the same type of design [4]. However, perceptions of play are highly individual, contextual, and variable, so one-design-fits-all approaches do not seem to work well for GBL research, according to our results. A thorough understanding of specific students’ perceptions within a culture or university might therefore play a pivotal role in utilizing the full potential of GBL. In the future, researchers and educators should map students’ play preferences before implementing GBL. Such information is essential for both the design of effective GBL activities and transfer of existing research into educational practice. As a first step in designing GBL and engaging in evidence-based decision making, teachers need to compare the play preferences of research participants in the study design with the play preferences of their own target group. This information also helps researchers understand and clarify what type of design works, for whom, and in what situation or circumstances. Thanks to the manageable and flexible possibilities of digital media, such an approach will bring tailor-made education a step closer.

Educators who want to implement GBL should aim at balancing the interaction between play and learning, harmonizing the right amount of play with the serious to increase efficiency of learning. What educators first should determine when they engage in teaching difficult subjects is whether there is a real need for play by identifying problems in students’ learning attitudes or learning behaviors [4]. Currently, determining the right amount of play and how learning efficiency can be improved seem to depend on intuition and personal perceptions.
rather than evidence-based decision and is therefore an area for future research.

One of our main findings was that participants strongly expected, or found, pleasure in play. The pleasure in games is strongly related to playing time [90]. Longer periods of time spent at GBL might indicate increased repetition of the learning material which, in turn, will lead to improved learning outcomes and retention [4,110-113]. Educators who want to design GBL could, therefore, adopt different positive motivational forces of pleasure as a method to guide their design. Using pleasure as a motivational force also opens up exciting new ways for research. Negative motivational forces (eg, violence) are also often observed in games [65,66], but their roles have rarely been investigated in the context of GBL [3,4,9,114].

We identified sociability as a major incentive for medical and dental students to play. Consequently, social play might be an interesting design option for GBL material. Strikingly, social play is underrepresented in GBL research, since most studies adopt a single player approach [4,9,26,114,115]. Although students felt that competition could enhance their learning, educators should be careful with implementing competitive elements in GBL, because these may also cause undesired effects such as increased stress. Playing in teams or in anonymity may be more appropriate options for such scenarios.

Strengths and Limitations

A strength of our focus group study may be that it generated a rich understanding of students’ perceptions, experiences, and beliefs with respect to play and the interaction between play and learning. An experienced moderator guided the focus group sessions and stimulated in-depth discussions; we thoroughly explored students’ perceptions; 2 independent researchers identified codes; and the whole team, including the moderator, discussed and reflected on themes.

As in any focus group study, the identified themes unavoidably bear some relation to the original impetus for asking the questions and designing the interview guide. We tried to counteract this by actively encouraging input from all students during the sessions, even if it deviated from the original topic list.

In focus group studies, researchers sometimes meet with their participants to verify the generated themes. Although we did end every focus group discussion with a summary, to check whether our summary was appropriate to how the participants experienced it, we did not opt to meet with our participants, which potentially could have altered our outcome.

The gender ratio in our sample was imbalanced in favor of female students (70% were women). Although this ratio represents the Dutch medical student population [116], some countries might have gender ratios more balanced towards men (eg, medical schools in the United States have around 50% of their students as men [117]). Literature argues that the liking of play differs between genders [118,119]. Therefore, there is a possibility that our findings are more pertinent to female students. Yet, our aim was not to provide a generalization, or find a consensus, during the focus group. Even in our coding, we aimed to include varied opinions on play and GBL. However, although we tried to counteract an imbalance towards female students, results might be localized to the students of Dutch medical schools.

The meaning of play is associated with many variables such as culture [54-56]. Since we used convenience sampling, the participants in this study predominantly had a similar ethnic background (European/Caucasian) and possibly also a similar socioeconomic background [120]. Therefore, cultural and regional differences might have affected the results of our study. In a different setting, the same methodology might yield different results.

Finally, our findings on the play-learning interaction reflect students’ perceptions. Perceptions, however, do not always reflect actual behavior, meaning that students do not always do what they say they do. For example, in a study on using leader boards to increase the use of laparoscopic simulators, the majority of the surgical residents mentioned that they were not motivated by leader boards. However, the results showed that the time students in the competition group had spent on the simulation was higher than in the control group [121]. Additionally, because students in our sample had little experience with GBL, they might not have fully understood all the possibilities of GBL and, therefore, might have provided limited answers. Nonetheless, we believe that our findings offer important insights for future research to examine which design and GBL situation hold the highest promise for learning.

Conclusion

With this focus group study, we aimed to explore students’ perceptions of play and the play-learning interaction. We explored what they considered to be play and how they believed it could interact with their learning. Four key points emerge from our study. First, students play for pleasure. Perceptions of pleasure vary considerably among students. Second, students consider play as inefficient. Inefficiency will only be justified when it increases learning. Third, play should be balanced with the serious and only used for difficult or tedious courses. Fourth, GBL activities should not be made compulsory for students, since there is a discrepancy between the serious of compulsory activities and the free nature of play.

Acknowledgments

The authors would like to thank the students for participating in this study. We also thank Tineke Bouwkamp-Timmer for her valuable comments on the manuscript and checking the English. We thank Irena Middeljans (Ob1) for her persistence and accuracy in transcribing the manuscripts as well as creativity in finding codes. We thank Nora Spraakman (Ob2) for her helpfulness in observing the last 2 focus group discussions.

https://games.jmir.org/2021/3/e25637
Conflicts of Interest
None declared.

References


34. Eberle S. The Elements of Play Toward a Philosophy and a Definition of Play. Journal of Play 2014;6(2):214-233 [FREE Full text]


Abbreviations

AMEE: Association for Medical Education in Europe
GBL: game-based learning
Association of Extensive Video Gaming and Cognitive Function Changes in Brain-Imaging Studies of Pro Gamers and Individuals With Gaming Disorder: Systematic Literature Review

Eunhye Choi¹, BSc, MSc; Suk-Ho Shin², PhD; Jeh-Kwang Ryu³, PhD; Kyu-In Jung¹, MD, PhD; Yerin Hyun¹, MA; Jiyea Kim¹, MA; Min-Hyeon Park¹, MD, PhD

¹Department of Psychiatry, Eunpyeong St. Mary’s Hospital, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea
²Dr. Shin’s Child and Adolescent Psychiatry Clinic, Seoul, Republic of Korea
³Department of Physical Education, College of Education, Dongguk University, Seoul, Republic of Korea

Corresponding Author:
Min-Hyeon Park, MD, PhD
Department of Psychiatry
Eunpyeong St. Mary’s Hospital, College of Medicine
The Catholic University of Korea
1021, Tongil-ro, Eunpyeong-gu
Seoul, 03312
Republic of Korea
Phone: 82 2 2030 2720
Email: neominnie00@daum.net

Abstract

Background: The World Health Organization announced the inclusion of gaming disorder (GD) in the International Classification of Diseases, 11th Revision, despite some concerns. However, video gaming has been associated with the enhancement of cognitive function. Moreover, despite comparable extensive video gaming, pro gamers have not shown any of the negative symptoms that individuals with GD have reported. It is important to understand the association between extensive video gaming and alterations in brain regions more objectively.

Objective: This study aimed to systematically explore the association between extensive video gaming and changes in cognitive function by focusing on pro gamers and individuals with GD.

Methods: Studies about pro gamers and individuals with GD were searched for in the PubMed and Web of Science databases using relevant search terms, for example, “pro-gamers” and “(Internet) gaming disorder.” While studies for pro gamers were searched for without date restrictions, only studies published since 2013 about individuals with GD were included in search results. Article selection was conducted by following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Results: By following the PRISMA guidelines, 1903 records with unique titles were identified. Through the screening process of titles and abstracts, 86 full-text articles were accessed to determine their eligibility. A total of 18 studies were included in this systematic review. Among the included 18 studies, six studies included pro gamers as participants, one study included both pro gamers and individuals with GD, and 11 studies included individuals with GD. Pro gamers showed structural and functional alterations in brain regions (eg, the left cingulate cortex, the insula subregions, and the prefrontal regions). Cognitive function (eg, attention and sensorimotor function) and cognitive control improved in pro gamers. Individuals with GD showed structural and functional alterations in brain regions (eg, the striatum, the orbitofrontal cortex, and the amygdala) that were associated with impaired cognitive control and higher levels of craving video game playing. They also showed increased cortical thickness in the middle temporal cortex, which indicated the acquisition of better skills. Moreover, it was suggested that various factors (eg, gaming expertise, duration or severity of GD, and level of self-control) seemed to modulate the association of extensive video game playing with changes in cognitive function.

Conclusions: Although a limited number of studies were identified that included pro gamers and/or individuals who reported showing symptoms of GD for more than 1 year, this review contributed to the objective understanding of the association between extensive video game playing and changes in cognitive function. Conducting studies with a longitudinal design or with various comparison groups in the future would be helpful in deepening the understanding of this association.
Introduction

Background

Video game playing has become one of the most popular leisure activities [1]. With the growing popularity of video game playing, a minority of individuals have been reported to play video games in problematic ways, resulting in negative consequences (eg, withdrawal from socializing and death) [2-4]. By focusing on problematic video gaming in a minority of individuals, the World Health Organization (WHO) recently announced that the International Classification of Diseases, 11th Revision (ICD-11) included gaming disorder (GD) as a syndrome [5,6]. GD refers to the persistent engagement on the internet in playing games despite the psychological distress and the interference with daily activities for more than 12 months [7]. However, there are concerns about the inclusion of GD as one of diseases in the ICD-11 [6,8], in that the objective evidence that showed there were harmful effects of GD was not sufficient (ie, little research examined causality and the persistence of symptoms) [9].

After the announcement by the WHO, GD was reported in the media to result in structural alterations in brain regions based on the results of a cross-sectional study that compared the brain structures of individuals with GD to those of healthy controls; this emphasized the necessity of treatment for GD [10]. Although the tendency of GD was found to be negatively associated with the volume of gray matter (GM) in the prefrontal brain regions that are involved in cognitive control and sensorimotor functioning [11], it is difficult to confirm the causality of the association in the cross-sectional studies. Moreover, unlike the focus on GD, playing video games was positively associated with cognitive function [12-15]. Video game players, compared to non–video game players, showed more integrated white matter (WM) in motor and visual pathways [14] and higher levels of activation in the frontoparietal brain regions to detect visual stimulus despite the comparable cognitive performance [15]. Playing video games for a longer duration was also associated with thicker cortices in the brain regions for attention, navigation, visuomotor function, and the resolution of ambiguity (eg, the left frontal eye field, the left dorsolateral prefrontal cortex [DLPFC], and the bilateral entorhinal cortex) [12,13]. That is, while the association of GD with alterations in brain regions was more focused, playing video games was also associated with cognitive enhancement.

Furthermore, there are individuals who play video games extensively for more than 10 hours a day without reporting disrupted lifestyles (eg, a disrupted sleep-wake cycle) [16]; these individuals are called pro gamers. They refer to a group of people who belong to a team through a contract and who make economic profits by taking part in e-sports competitions [17]. The mean age of pro gamers in major leagues was reported to be 22 years [18]. Although statistics for their mean age of retirement were not available, more than half of pro gamers reported that their retirement was dependent on their judgment of their performances in competitions [17]. Since cognitive-motor speed (ie, the speed at which the cognitive process initiates actions) was found to start to decline at 24 years in a sample of StarCraft II players who were aged between 16 and 44 years old, regardless of their expertise level [19], pro gamers were assumed to retire at 25 to 27 years of age. Taken together, playing video games for a longer duration did not result in the development of GD, and only a minority of people reported the development of GD [6,7]. Unlike negative opinions in the media toward video games and playing video games, playing them was associated with cognitive enhancement (eg, Zhang et al [14] and Richlan et al [15]). Video games were also suggested as a potential tool for clinical intervention for individuals with mental disorders (eg, Alzheimer disease) [20]. Thus, it is necessary to explore the association between video game playing and alterations in brain regions in a more objective manner. Moreover, despite the comparable amount of video game experience between individuals with GD and pro gamers who play video games extensively without any symptoms of GD (eg, higher impulsivity for gaming) [16], more studies have been conducted that focused on individuals with GD, and more review studies about GD have been conducted (eg, Leeman and Potenza [21] and Wei et al [22]). Since pro gamers, in addition to individuals with GD, are a population of interest for investigating the association between extensive video gaming and changes in cognitive function, reviewing studies that recruited pro gamers would deepen the understanding and effects of playing video games extensively.

Objective

This systematic review aimed to explore the association between extensive video game playing and changes in cognitive function. That is, this study reviewed brain-imaging studies that included pro gamers and/or individuals with GD.

Methods

Search Strategy

Literature searches were conducted in two databases: PubMed and Web of Science. Studies about pro gamers were searched for with the following search terms, without a restriction on the date: “pro-gamers,” “pro video game players,” “action video game experts,” “video gaming experts,” and “long-term video game players.” Studies about individuals with GD were searched for with the following search terms, with date restrictions:“(Internet) gaming disorder,”“(Internet) gaming addiction,” and“(online) video game addiction.” As many studies about GD have been conducted, only studies that were published since 2013 were included in the results of the literature search.
Study Selection

Overview

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23] were followed in this study. Firstly, duplicates from the search results from the two databases were removed using EndNote X9 (Clarivate Analytics). After the removal of duplicates, the titles and abstracts of the remaining articles were screened to determine if they were eligible for full-text assessment. Secondly, full-text articles were carefully reviewed to determine their eligibility for this review based on selection criteria.

Inclusion Criteria

Inclusion criteria for articles in this review were as follows: (1) original research articles published in English, (2) the use of brain-imaging techniques, and (3) the recruitment of pro gamers and/or individuals with GD. Definitions of pro gamers and GD in this study are as follows:

- Pro gamers are defined as individuals who (1) belong to e-sports teams, (2) are highly experienced video game players without reporting any problematic daily lifestyle behaviors, and/or (3) have won video game playing competitions.
- GD refers to persistent engagement in video game playing despite psychological distress and interference with daily activities [24] for more than 1 year [7].

That is, articles about pro gamers were selected for inclusion when recruited participants met at least one description of pro gamers above. For example, a study that recruited video game experts, who were recognized as top-ranking players [25], was included because the second description of pro gamers was met. The types of video games (e.g., StarCraft and League of Legends) were not restricted for the selection. Articles were also selected for the review when recruited participants were confirmed to be diagnosed with GD for more than 12 months or when the stated mean duration of GD symptoms in participants was more than 1 year. In the process of the selection of articles about individuals with GD, the age of the individuals with GD was not restricted. This was because adults, in addition to adolescents, have also shown GD symptoms, despite the report that the prevalence of GD has increased especially in adolescents [26] whose cognitive control is developing [27] with different developmental trajectories of the limbic system and prefrontal cortex (PFC) regions [28].

Exclusion Criteria

Articles were excluded when full texts were not available and when participants in studies did not meet any of the descriptions of pro gamers defined above. Articles were also excluded when they did not confirm that individuals with GD showed GD symptoms for more than 1 year or when they did not present the information of the mean duration of GD.

Data Extraction

The following data were extracted from the selected articles: information about the study (i.e., study design, participants, duration of GD, and brain-imaging techniques used) and the brain regions that were associated with extensive video game playing.

Results

Literature Overview

The searches of PubMed and Web of Science resulted in the identification of 2571 records. After the removal of duplicates, 1903 studies with unique titles were obtained for the screening of titles and abstracts. A total of 1761 studies were removed after the screening of titles, and another 56 studies were removed after the screening of abstracts. After excluding 1817 records, 86 full-text articles were comprehensively reviewed in order to assess their eligibility for inclusion in this review. After conducting assessments based on the inclusion criteria, 68 articles that did not meet the inclusion criteria or did not have full-text access were excluded (Figure 1). Thus, this review included 18 articles. The results of the extracted data from the selected articles are presented in three subsections: (1) information about the study, (2) alterations in brain regions in pro gamers, and (3) alterations in brain regions in individuals with GD.
Information About the Study

Study Design

As seen in Table 1 [16,20,25,29-43], 16 studies had a cross-sectional design. One study [16] had a correlational design and one study [25] had a longitudinal design.
Table 1. Summary of 18 studies included in this review.

<table>
<thead>
<tr>
<th>Authors; study design</th>
<th>Participant information</th>
<th>Brain-imaging technique</th>
<th>Alterations in brain regions associated with extensive video gaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyun et al [16]; correlational</td>
<td>Pro gamers: N=23; all males; mean age 19.8 (SD 1.7) years</td>
<td>Magnetic resonance imaging (MRI)</td>
<td>Cortical thickness in the right superior frontal gyrus, the right superior parietal gyrus, and the right precentral gyrus</td>
</tr>
<tr>
<td>Tanaka et al [29]; cross-sectional</td>
<td>1. Pro gamers: N=17; all males; mean age 24.1 (SD 2.9) years 2. Age- and educational level–matched control group: N=33; all males; weekly gaming duration was less than 2 hours</td>
<td>Structural MRI</td>
<td>Gray matter (GM) volume in the right posterior parietal cortex</td>
</tr>
<tr>
<td>Gong et al [30]; cross-sectional</td>
<td>1. Pro gamers: N=27; mean age 23.26 (SD 0.4) years 2. Amateur players: N=30; mean age 22.3 (SD 0.38) years; did not habitually engage in video game playing, and video gaming experience was less than 1 year</td>
<td>MRI</td>
<td>Functional connectivity (FC) and GM volume in the insular subregions</td>
</tr>
<tr>
<td>Gong et al [31]; cross-sectional</td>
<td>1. Pro gamers: N=23; all males; mean age 23.3 (SD 4.3) years 2. Amateur players: N=22; all males; mean age 22.3 (SD 3.46) years; video gaming experience was less than 1 year</td>
<td>Resting-state functional MRI (fMRI)</td>
<td>FC within and between the salience network (SN) and the central executive network (CEN)</td>
</tr>
<tr>
<td>Gong et al [20]; cross-sectional</td>
<td>1. Pro gamers: N=28; all males; mean age 24.6 (SD 1.4) years 2. Amateurs players: N=30; all males; mean age 24.3 (SD 1.8) years</td>
<td>Diffusion tensor imaging (DTI)</td>
<td>White matter (WM) networks in the prefrontal network, the limbic system, and the sensorimotor network</td>
</tr>
<tr>
<td>Gong et al [25]; longitudinal</td>
<td>1. Pro gamers: N=20; all males; mean age 21.42 (SD 1.64) years 2. Amateur players: N=20; all males; mean age 22.25 (SD 1.65) years; gaming experience was less than 1.5 years</td>
<td>Resting-state fMRI at the beginning and end of the study</td>
<td>Amplitude of low-frequency fluctuation (ALFF) in the brain regions of the default mode network (DMN), the CEN, and the SN</td>
</tr>
<tr>
<td>Han et al [32]; cross-sectional</td>
<td>1. Individuals with gaming disorder (GD): N=20; all males; mean age 20.9 (SD 2) years; mean duration of gaming disorder 4.9 (SD 0.9) years 2. Pro gamers: N=17; all males; mean age 20.8 (SD 1.5) years 3. Age- and educational level–matched healthy controls (HCs): N=18; all males; mean age 20.9 (SD 2.1) years</td>
<td>MRI</td>
<td>GM volume in cingulate gyrus, thalamus, and occipitotemporal areas</td>
</tr>
<tr>
<td>Ko et al [33]; cross-sectional</td>
<td>1. Participants with GD: N=15; mean age 24.67 (SD 3.11) years; mean education duration 15.47 (SD 1.56) years 2. Remission subjects from GD: N=15; mean age 24.8 (SD 2.68) years; mean education duration 15.87 (SD 1.41) years 3. HC: N=15; mean age 24.47 (SD 2.83) years; mean education duration 16 (SD 1.13) years</td>
<td>Task-based fMRI (the presentation of neutral vs online game-related screenshots)</td>
<td>The bilateral dorsolateral prefrontal cortex (DLPFC), the precuneus, the left parahippocampus, the posterior cingulate, the right anterior cingulate, and the left superior parietal lobe</td>
</tr>
<tr>
<td>Yuan et al [34]; cross-sectional</td>
<td>1. Adolescents with GD: N=18; n=12 males; mean age 19.4 (SD 3.1) years 2. Age- and gender-matched HCs: N=18; daily gaming duration was less than 2 hours</td>
<td>MRI</td>
<td>Cortical thickness in the left lateral orbitofrontal cortex, the insula cortex, lingual gyrus, the right postcentral gyrus, the entorhinal cortex, the inferior parietal cortex, the left precentral cortex, the middle temporal cortices, the precuneus, the middle frontal cortex, and the inferior temporal cortices</td>
</tr>
<tr>
<td>Authors; study design</td>
<td>Participant information</td>
<td>Brain-imaging technique</td>
<td>Alterations in brain regions associated with extensive video gaming</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Yuan et al [35]; cross-sectional  | 1. Participants with GD: N=18; n=12 males; mean age 19.4 (SD 3.1) years  
2. Age- and gender-matched HCs: N=18; n=12 males; mean age 19.5 (SD 2.8) years; daily gaming duration was less than 2 hours | Resting-state fMRI | ALFF in brain regions, including major regions of the DMN |
| Ko et al [36]; cross-sectional  | 1. Male subjects with GD: N=26; mean age 24.58 (SD 3.23) years  
2. Male HCs: N=23; mean age 24.35 (SD 2.12) years | Task-state fMRI (go/no-go task) | Activation in the frontostriatal network |
| Chen et al [37]; cross-sectional  | 1. Individuals with GD: N=15; all males; mean age 24.67 (SD 3.12) years  
2. HCs: N=25; all males; mean age 24.47 (SD 2.83) years | Task-state fMRI (go/no-go task) | Activation in the right supplementary motor area (SMA) or pre-SMA |
| Ko et al [38]; cross-sectional  | 1. Male adults with GD: N=30; mean age 23.57 (SD 2.5) years  
2. Age- and educational level–matched HCs: N=30; mean age 24.23 (SD 2.47) years | Resting-state fMRI | 1. GM density in bilateral amygdala  
2. FC of amygdala with the left DLPFC, the orbitofrontal lobe (OFL), and the contralateral insula |
| Cai et al [39]; cross-sectional  | 1. Individuals with GD: N=27; n=23 males; mean age 17.9 (SD 0.9) years  
2. Age-, gender-, and educational level–matched HCs: N=30; n=22 males; mean age 18.3 (SD 1.6) years | MRI | The volume in the striatum |
| Chen et al [40]; cross-sectional  | 1. Individuals with GD: N=28; all males; mean age 23.64 (SD 2.54) years  
2. Age- and educational level–matched HCs: N=28; all males; mean age 24.14 (SD 2.53) years | Resting-state fMRI | FC between the left insula and the left DLPFC and OFL, and between interhemispheric insula |
| Jin et al [41]; cross-sectional  | 1. College students with GD: N=25; n=16 males; mean age 19.12 (SD 1.05) years  
2. Age- and gender-matched HCs: N=21; n=14 males; mean age 18.76 (SD 1.81) years | Resting-state fMRI | 1. GM volume in the prefrontal regions and the right SMA  
2. FC of prefrontal regions with temporal and occipital regions, and between several regions, including the bilateral caudate, the thalamus, the putamen, the insular cortex, and the right SMA |
| Yuan et al [42]; cross-sectional  | 1. Individuals with GD: N=43; n=32 males; age range 16-22 years; primary internet activity was to play League of Legends  
2. Age- and gender-matched HCs: N=44; n=34 males; age range 15-23 years | Resting-state fMRI | 1. Volume in the striatum  
2. Resting-state FC within the dorsal and ventral striatum networks |
| Zhai et al [43]; cross-sectional  | 1. Subjects with GD: N=16; n=11 males; mean age 19.1 (SD 1.3) years; primary internet activity was to play League of Legends  
2. Age- and gender-matched HCs: N=16; n=11 males; mean age 18.4 (SD 1.9) years | DTI | The global and local efficiency of WM networks |

**Participants**

Among the included 18 articles, six articles included pro gamers (Table 1). While pro gamers were compared with amateur video game players who had less gaming experience in five articles [20,25,29-31], one article [16] included pro gamers with varied gaming expertise. One article [32] included both pro gamers and individuals with GD. A total of 11 articles [33-43] included individuals with GD (Table 1). While one article [33] recruited a remission group in addition to individuals with GD and healthy controls, 10 articles recruited individuals with GD and healthy controls. Moreover, six articles [34,35,39,41-43] included female participants.

**Brain-Imaging Techniques**

Two studies used diffusion tensor imaging and six studies used magnetic resonance imaging to investigate the alterations in brain regions associated with extensive video game playing.
Seven studies used resting-state functional magnetic resonance imaging (fMRI) and three studies used task-state fMRI. Tasks that were used in the studies included the presentation of gaming cues [33] and tasks to measure cognitive control [36,37] (Table 1).

Alterations in Brain Regions in Pro Gamers

As seen in Table 1, pro gamers showed structural alterations in brain regions that were different from those in amateur players and individuals with GD. Pro gamers showed higher GM volume in the left cingulate gyrus [32], the right posterior parietal cortex (PPC) [29], and insula subregions, including the left long insular gyrus and central insular sulcus [30], compared to amateur players. The GM volume in the left cingulate gyrus was also higher in pro gamers than in individuals with GD [32]. Moreover, pro gamers showed decreased GM volume in some brain regions, including in the left middle occipital gyrus and the right inferior temporal gyrus, compared to healthy controls [32]. The GM volume in the left thalamus was lower in pro gamers than in individuals with GD [32]. Moreover, the pro gamers with longer career lengths were shown to have a thicker cortex in the right superior frontal gyrus, right superior parietal gyrus, and right precentral gyrus, and the pro gamers who won more in the competitions were shown to have a thicker PFC [16].

Pro gamers also showed functional alterations in brain regions. They showed increased amplitude of low-frequency fluctuation (ALFF) in brain regions of the default mode network (DMN), the central executive network (CEN), and the salience network (SN) compared to amateur players [25]. That is, pro gamers showed increased ALFF in the posterior cingulate cortex (PCC), the right angular gyrus, the right DLPFC, the anterior cingulated cortex (ACC), and the right anterior insula [25]. The WM network in the prefrontal regions, the limbic system, and the sensorimotor network was also more integrated in pro gamers compared to amateur players [20]. Moreover, pro gamers, compared to amateur players, showed more functionally connected networks, not only between anterior and posterior insula subregions [30] but also within and between the SN and CEN [31].

Alterations in Brain Regions in Individuals With GD

Individuals with GD showed structural alterations compared to healthy controls and pro gamers (Table 1). Compared to both pro gamers and healthy controls, individuals with GD showed increased GM volume in the left thalamus [32]. Compared to healthy controls, individuals with GD showed increased volume in the striatum (ie, right caudate and right nucleus accumbens [NAc]) [39,42]; they also had a thicker cortex in the left precentral cortex, the middle temporal cortices, the precuneus, the middle frontal cortex, and the inferior temporal cortices [34]. Moreover, individuals with GD showed decreased GM volume in some brain regions compared to healthy controls. Individuals with GD were found to have decreased GM volume in the amygdala [38], the temporal-occipital cortex (ie, the right middle occipital gyrus and the left inferior occipital gyrus) [32], the prefrontal regions (ie, the bilateral DLPFC, the orbitofrontal cortex [OFC], and the ACC), and the right supplementary motor area (SMA) [41]. The cortical thickness in the left lateral OFC, the insula cortex, the lingual gyrus, the right postcentral gyrus, the entorhinal cortex, and the inferior parietal cortex was also found to be decreased in individuals with GD compared to healthy controls [34].

Individuals with GD showed functional changes in brain regions in the resting state compared to healthy controls. The short path length in individuals with GD was found to be increased [43]. Individuals with GD, compared to healthy controls, also showed increased ALFF in the left medial OFC, the left precuneus, the left SMA, the right parahippocampal gyrus, and the bilateral middle cingulate cortex [35]. Moreover, individuals with GD showed increases in the resting-state functional connectivity (FC), not only between the bilateral amygdala and the contralateral insula [38] but also between the bilateral insula [40].

Individuals with GD showed decreased resting-state activation or FC in some brain regions. The global and local efficiency of WM networks was found to be reduced in individuals with GD [43]. The FC between prefrontal regions (ie, ACC, OFC, and DLPFC) and temporal and occipital regions (ie, pallidum, thalamus, caudate, and putamen) also decreased in individuals with GD compared to healthy controls [41]. Moreover, individuals with GD, compared to healthy controls, showed both reduced FC in the dorsal and ventral striatum networks (ie, FC between the right caudate and the DLPFC, and FC between the right NAc and the OFC) [42] and decreased FC of the left insula with the left DLPFC and the orbitofrontal lobe (OFL) [40]. The FC of the bilateral amygdala with the left DLPFC and the FC of the right amygdala with the OFL were found to be decreased in individuals with GD [38].

Additionally, studies that used task-based fMRI showed that individuals with GD showed increased activation in the frontostriatal network (ie, bilateral OFL, ACC, left putamen, right DLPFC, and middle temporal lobe) [36] but decreased activation in the right SMA or pre-SMA [37] compared to healthy controls in the task that required inhibition. Furthermore, when gaming cues were presented, individuals with GD showed higher activation in the bilateral DLPFC, the precuneus, the left parahippocampus, the posterior cingulate, the right anterior cingulate, and the left superior parietal lobe compared to healthy controls [33]. Individuals with GD also showed higher activation in the right DLPFC, the left parahippocampus, and the left middle temporal gyrus in response to gaming cues compared to the remission group of participants [33].

Discussion

Overview

This review aimed to explore the association between extensive video game playing and changes in cognitive functions by focusing on pro gamers and individuals with GD. That is, this study systematically reviewed the brain-imaging studies that included pro gamers and/or individuals with GD. By following PRISMA guidelines, 18 studies were included in this review. Based on selected studies, it was found that pro gamers and individuals with GD showed different structural and functional
alterations in brain regions, despite the comparable level of gaming engagement.

**Primary Results of the Studies Including Pro Gamers**

Results showed both increased and decreased GM volume in brain regions of pro gamers. Pro gamers, compared to both healthy controls and individuals with GD, showed a thicker cortex in the left cingulate cortex, which is involved in the maintenance of attention and control over executive functioning [32]. Pro gamers also showed structural enhancement in brain regions that are involved in visual working memory, attention, and sensorimotor function (eg, the right PPC and insular subregions) compared to amateur video game players [29,30]. The increased GM volume in the right PPC was positively associated with better visual working memory performance in pro gamers [29], and the increased GM volume in insular subregions was suggested to contribute to functional integration within insular regions [30]. However, pro gamers showed decreased GM volume in occipitotemporal regions for visual processing (eg, the left middle occipital gyrus) compared to amateur players, and they showed decreased GM volume in the left thalamus, which is involved in reward processing, compared to individuals with GD [32]. These structural alterations in pro gamers suggested that pro gamers did not show impaired reward processing but showed enhanced cognitive function (eg, cognitive control and visual working memory), along with reduced cortical thickness in brain regions that are involved in the processing of visual stimuli. Moreover, since pro gamers—who reported longer video game experience or higher rates of winning in competitions—were found to show a thicker cortex in frontal regions for cognitive flexibility (eg, the right superior frontal gyrus) [16], it was suggested that gaming expertise seemed to modulate the association between extensive video game experience and cognitive enhancement in pro gamers.

Results also showed functional enhancement in brain regions of pro gamers. Pro gamers showed more functional integration between anterior and posterior insular subregions [30] and within and between the SN and the CEN [31]. That is, the attention and sensorimotor functions were more coordinated in pro gamers [30], and they showed improvement in the ability to process information [31]. The plasticity of WM networks in brain regions for sensorimotor function and cognitive control (eg, the sensorimotor network and the prefrontal network) was also more enhanced in pro gamers compared to amateur players [20]. Pro gamers were found to integrate information more efficiently by showing nodal and global enhancement in WM networks [20]. Moreover, it was found that activation in the DMN (eg, PCC), the CEN (eg, the right DLPCS), and the SN (eg, the ACC), which was higher in pro gamers than in amateur players at the beginning of the study, decreased after the pro gamers were asked not to play video games for 1 year [25]. The results of that longitudinal study suggested that extensive video game playing seemed to enhance the development of brain regions [25]. Consistent with structural alterations in pro gamers, when compared to amateur players, pro gamers showed functional enhancements within and between the brain regions that are involved in attention, visual processing, sensorimotor function, and cognitive control.

**Primary Results of the Studies Including Individuals With GD**

Individuals with GD were found to show structural alterations in frontostriatal regions. While the cortical thickness of the brain regions that were associated with executive function and decision making (eg, the DLPFC, the OFC, and the amygdala) decreased in those with GD [34,38,41], the volumes of the brain regions for reward processing (eg, the striatum) increased in individuals with GD as compared to healthy controls [39,42]. Individuals with GD also showed increased volumes in the brain region that is involved in the expectation of rewards (ie, the left thalamus) compared to pro gamers [32]. The structural alterations in brain regions of individuals with GD suggested that they showed impairment in executive functioning and higher levels of craving to play video games [34]. In particular, the increased volume in the striatum was found to be associated with impairment of cognitive control [39]. However, individuals with GD showed increased volume in the middle temporal cortex, which is involved in the acquisition of skills, compared to healthy controls [34]. That is, despite impaired cognitive control in individuals with GD, their higher level of video game playing experience, compared to that of healthy controls, was associated with improved skills.

Individuals with GD were found to show not only structural alterations but also functional alterations compared to healthy controls. Individuals with GD showed reduced levels of integration within WM networks [43]. The increased short path length [43] and increased resting-state FC within and between the bilateral insula and the amygdala [38,40] in individuals with GD were found to be associated with a higher level of impulsivity. An increased resting-state activation in certain brain regions (eg, the left medial OFC) [35] was also associated with impairment of cognitive control in individuals with GD. Moreover, the CEN (eg, DLPFC) and the reward circuits (eg, amygdala) in individuals with GD, compared to healthy controls, were found to be less functionally integrated in the resting state [38,40]. That is, the FC within the frontostriatal networks, which are involved in the processing of motivation and cognitive control, was found to decrease in individuals with GD during the resting state [41,42]. It was suggested that they showed both higher levels of impulsivity and impaired cognitive control.

Consistent with the resting-state functional alterations in individuals with GD, the task-state fMRI studies showed impairment of cognitive control in individuals with GD. When the execution of cognitive control was required, individuals with GD showed increased activation in the frontostriatal networks, unlike healthy controls who showed increased activation only in the DLPFC [36]. That is, not only frontal regions but also striatal regions were activated in individuals with GD for response inhibition. However, activation in the brain region that is involved in executing proper behavior (eg, pre-SMA) decreased in individuals with GD in the cognitive control task [37]. Moreover, when the gaming cue was presented, individuals with GD showed increased activation in brain regions that are involved in the processing of affective or salient stimuli and craving (eg, the bilateral DLPFC, the posterior cingulate, and the anterior cingulate) compared to healthy controls [33]. The remission group, who showed lower
levels of craving video game playing than individuals with GD, also showed higher activation in the brain region for visual attention (ie, the superior parietal lobe), though the difference in activation between the remission group and healthy controls was not significant [33].

Furthermore, there was a difference in structural and functional alterations in brain regions among individuals with GD. Individuals who reported more severe levels of GD showed an increased volume in the NAc, which is involved in the processing of rewards [39]. Individuals who reported having GD for a longer duration showed not only an increased volume in the left precentral gyrus and precuneus but also a decreased volume in the lingual gyrus [34]. They also showed more abnormal resting-state activation in the left medial OFC and the left precuneus [35]. Moreover, the level of self-control in individuals with GD was negatively associated with activation of the bilateral caudate nucleus [36]. These findings suggest that it was plausible that the severity or duration of GD and impaired self-control mediated the association between extensive video game playing and changes in cognitive function.

Taken together, pro gamers and individuals with GD showed different structural and functional alterations in brain regions despite comparable extensive engagement in video game playing. Pro gamers showed enhancement in cognitive function (eg, attention and visuomotor function) and better cognitive control. Unlike pro gamers, individuals with GD showed impairment in cognitive control and higher levels of craving video game playing. They also showed improvement in the acquisition of skills. Moreover, factors that seemed to modulate the association of extensive video game playing with changes in cognitive function (ie, gaming expertise, duration or severity of GD, and level of self-control) were identified. That is, although individuals with GD showed impaired executive functioning, extensive video game playing was associated with enhancement in cognitive function in not only pro gamers but also in those with GD.

Limitations
There were three limitations in this review. The first limitation was the limited number of studies. Only a few studies were identified that included pro gamers. Most studies that included individuals with GD did not consider or state the duration of GD despite its importance. According to the WHO’s announcement, the diagnosis of GD should be based on the report of symptoms for more than 12 months [7]; only a subgroup of individuals were found to show persistent symptoms of GD over 12 months [44]. That is, more studies should be conducted that include pro gamers and individuals who had reported GD for more than 1 year. The second limitation was the design of the studies. Although one study had a longitudinal design, most included studies had cross-sectional designs. More brain-imaging studies with longitudinal designs should be conducted, as these would be helpful in tracking alterations in brain regions and in understanding causality in the association between extensive video game playing and changes in cognitive function. The last limitation was the comparison group. Most included studies recruited participants who did not habitually play video games as the comparison group. There was a group of highly engaged video game players who were not pro gamers and did not show any symptoms of GD [45]; therefore, comparing alterations in brain regions between individuals with GD and highly engaged video game players or pro gamers would be helpful to deepen the understanding of the effect of video game playing on structural and functional alterations in brain regions and to identify the mediating factors of their association. Further studies should be conducted while considering these limitations.

Conclusions
Pro gamers and individuals with GD showed differences in structural and functional alterations in certain brain regions. While pro gamers showed enhancement in cognitive functions (eg, cognitive control), individuals with GD showed impaired cognitive control despite the acquisition of better skills compared to non–video game players. Mediating factors (eg, the duration of GD) were found to be associated with different alterations of brain regions in pro gamers and individuals with GD. That is, it was suggested that various factors seemed to modulate the association of extensive video game playing with changes in cognitive function. However, a limited number of brain-imaging studies included pro gamers and/or individuals who reported symptoms of GD for more than 1 year. Thus, more studies that include pro gamers and/or individuals with GD, as well as more diverse comparison groups, and ones that longitudinally track alterations in brain regions should be conducted in the future.
None declared.

References


Abbreviations

ACC: anterior cingulate cortex
AI: artificial intelligence
ALFF: amplitude of low-frequency fluctuation
CEN: central executive network
DLPFC: dorsolateral prefrontal cortex
DMN: default mode network
FC: functional connectivity
fMRI: functional magnetic resonance imaging
GD: gaming disorder
GM: gray matter
ICD-11: International Classification of Diseases, 11th Revision
NAc: nucleus accumbens
OFC: orbitofrontal cortex
OFL: orbitofrontal lobe
PCC: posterior cingulate cortex
PFC: prefrontal cortex
PHR: personal health record
PPC: posterior parietal cortex
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SMA: supplementary motor area
SN: salience network
WHO: World Health Organization
WM: white matter

©Eunhye Choi, Suk-Ho Shin, Jeh-Kwang Ryu, Kyu-In Jung, Yerin Hyun, Jiyea Kim, Min-Hyeon Park. Originally published in JMIR Serious Games (https://games.jmir.org), 09.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.
Using Visual Guides to Reduce Virtual Reality Sickness in First-Person Shooter Games: Correlation Analysis

Kwang-Ho Seok¹, PhD; YeolHo Kim¹, MS; Wookho Son², PhD; Yoon Sang Kim¹, PhD

¹BioComputing Lab, Institute for Bio-engineering Application Technology, School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan-si, Republic of Korea
²Electronics & Telecommunications Research Institute, Daejeon, Republic of Korea

Corresponding Author:
Yoon Sang Kim, PhD
BioComputing Lab, Institute for Bio-engineering Application Technology, School of Computer Science and Engineering
Korea University of Technology and Education
1600, Chungjeol-ro, Byeongcheon-myeon, Dongnam-gu
Cheonan-si, 31253
Republic of Korea
Phone: 82 0415601496
Email: yoonsang@koreatech.ac.kr

Abstract

Background: The virtual reality (VR) content market is rapidly growing due to an increased supply of VR devices such as head-mounted displays (HMDs), whereas VR sickness (reported to occur while experiencing VR) remains an unsolved problem. The most widely used method of reducing VR sickness is the use of a rest frame that stabilizes the user's viewpoint by providing fixed visual stimuli in VR content (including video). However, the earth-fixed grid and natural independent visual background that are widely used as rest frames cannot maintain VR fidelity, as they reduce the immersion and the presence of the user. A visual guide is a visual element (eg, a crosshair of first-person shooter [FPS]) that induces a user's gaze movement within the VR content while maintaining VR fidelity, whereas there are no studies on the correlation of visual guide with VR sickness.

Objective: This study aimed to analyze the correlation between VR sickness and crosshair, which is widely used as a visual guide in FPS games.

Methods: Eight experimental scenarios were designed and evaluated, including having the visual guide on/off, the game controller on/off, and varying the size and position of the visual guide to determine the effect of visual guide on VR sickness.

Results: The results showed that VR sickness significantly decreased when visual guide was applied in an FPS game. In addition, VR sickness was lower when the visual guide was adjusted to 30% of the aspect ratio and positioned in the head-tracking direction.

Conclusions: The experimental results of this study indicate that the visual guide can achieve VR sickness reduction while maintaining user presence and immersion in the virtual environment. In other words, the use of a visual guide is expected to solve the existing limitation of distributing various types of content due to VR sickness.

(JMIR Serious Games 2021;9(3):e18020) doi:10.2196/18020

KEYWORDS

virtual reality; motion sickness; VR sickness; visual guide; VR fidelity

Introduction

Recently, virtual reality (VR) content based on head-mounted display (HMD) has been expanded to various industrial fields such as sports, medical care, education, and social network. Moreover, such content has been used in video games. However, most users have experienced “VR sickness” while using HMD-based VR content. VR sickness has symptoms similar to motion sickness, including nausea, oculomotor discomfort, and disorientation caused while experiencing an HMD-based VR [1]. To investigate the causes of VR sickness, various theories are being studied in a cognitive science approach. The popular sensory conflict theory [1] proposes that VR sickness is induced by an inconsistency between the visual and the vestibular or proprioceptive senses. In particular, the vection that occurs during the VR experience is the biggest cause of sensory conflict [2-4]. Additionally, the postural instability theory posits that VR sickness is caused by changes in human balance [5]. Furthermore, VR sickness may also be induced by various
individual characteristics, such as age, gender, prior user experience, concentration, medical history, mental rotation, perceptual style, and dominant eye [6].

Until now, various studies have only been partially successful in attempting to reduce HMD-induced VR sickness by focusing on the device and the content. A typical method focusing on the device is the optimization of the delay time caused by head movement tracking response, rendering, image transmission, and display response speed [7]. Studies have also been conducted to reduce VR sickness in device elements such as resolution, frame rate, viewing angle, binocular parallax, and flicker fusion frequency [8-13]. Recent studies have evaluated methods focusing on content. These methods investigated the use of dynamic blurring with retinal tracking [14], optical flow reduction of peripheral vision [15], field of view control [16], and viewpoint snapping [17].

However, there is a disadvantage that these dynamic blurring methods limit the user's experience. Therefore, in other studies, VR sickness has been reduced by adding fixed or dynamic visual stimuli regardless of the motion of objects in the content [18-20]. This visual stimulus includes a rest frame that serves as a reference frame designed to induce a user's effective spatial perception. In particular, VR sickness was reduced by applying a virtual human nose as a rest frame to the content [21]. However, this artificial rest frame could not maintain VR fidelity, as it reduced user presence and immersion and provided a strong sense of heterogeneity to the user. However, this artificial rest frame could not maintain VR fidelity, as it reduced user presence and immersion and provided a strong sense of heterogeneity to the user. VR sickness has further been reduced using the earth-fixed grid or the natural independent visual background (IVB) with rest frames, whereas it did not increase the presence and immersion of the user [22,23]. Particularly, VR sickness in first-person shooter (FPS) games was reduced when cockpits were added to the IVB, while the VR fidelity of the users was disturbed [24].

To overcome the problem of IVB, we discuss another visual element applied to VR FPS games called the visual guide, which refers to a visual element that induces a user's gaze movement within VR content. To our knowledge, there is no study investigating the effect of visual guide on the reduction of VR sickness while maintaining VR fidelity in VR FPS games. Therefore, in this study, we investigated how VR sickness reduction is affected by visual guide in FPS games. To do this, we performed experiments on a VR FPS game consisting of eight scenarios, including visual guide on/off, game controller on/off, and varying the size and position of the visual guide.

Methods

Visual Guide Design in VR FPS Games

In this section, we describe the visual guide used in this study. In VR FPS games, crosshairs, character path indicators, and maps are used as visual guides to provide situational awareness for the user to spatially determine their position in a virtual environment. In this study, we used crosshair as a visual guide to induce a user's gaze movement while maintaining VR fidelity in a space-environment FPS game. Crosshair is a 2D image composed of color, shape, line thickness, depth, size, and position elements. Figure 1 shows an example of crosshair used in the VR FPS game. The brilliant color, complex shape, bold line, and depth of a visual guide can reduce the presence and immersion of the user by increasing the visual stimulation [18]. Therefore, these elements are fixed, and they are designed as a white, circle, 1.0 px, and 0, respectively by referring to commercial VR FPS games. Table 1 shows various crosshair in commercial VR FPS games.

Figure 1. Example of crosshair used in the virtual reality first-person shooter game. HMD: head-mounted display; SSQ: Simulator Sickness Questionnaire.
Table 1. Various crosshair in commercial virtual reality (VR) first-person shooter (FPS) games.

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Crosshair image</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Super Stardust Ultra VR</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>2</td>
<td>End Space VR</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>3</td>
<td>Elite: Dangerous</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>4</td>
<td>Gunjack</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>5</td>
<td>Sublevel Zero</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>6</td>
<td>VR Galaxy Wars</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>7</td>
<td>Fusion Wars</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>8</td>
<td>EVE: Valkyrie</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>9</td>
<td>Rez Infinite</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
<tr>
<td>10</td>
<td>EVERSPACE</td>
<td><img src="image" alt="Crosshair" /></td>
</tr>
</tbody>
</table>

The size and position of the visual guide is considered a variable for VR sickness measurements. Therefore, we designed it as a variable. First, the size was designed to be 0%, 10%, 30%, and 50% of the size of the aspect ratio (with 0% implying that there is no visual guide). Second, the position was synchronized to “none,” “head-tracking direction with HMD (H),” “movement direction with game controller (G),” and “head-tracking direction with HMD and movement direction with game controller (H & G).” That is, if the position is “none,” then there is no visual guide. If the position is “H,” then the position of the visual guide is synchronized with the head-tracking direction by the HMD operation. If the position is “G,” then the position of the visual guide is synchronized in the direction of forwarding movement, pitch, yaw, and roll rotation by the game controller operation. Finally, if the position is “H & G,” then the position of the visual guide is synchronized with the head-tracking direction by the HMD operation and the direction of forwarding movement, pitch, yaw, and roll by the game controller operation.

Experiments

In this section, we describe the experimental environment and methods used to examine the effect of visual guide on VR sickness. The experimental environment consisted of the participant, experiment device, and experimental content.

We used a paired t-test method to analyze the results. This requires the experimental results to approximate the normal distribution. Therefore, the minimum sample size was set to 30 to satisfy a central limit theorem [25]. Eventually, 32 individuals (male: 23, female: 9) participated in the experiment.

All the participants were in their twenties without any HMD VR experience, and none had a medical history of hearing or balancing disorder. We received an institutional review board approval for testing of VR sickness by KOREATECH.

Before performing the experiment, each participant was administered a preliminary questionnaire (M1) comprising the Simulator sickness Questionnaire (SSQ) items, such as measurement, reliability, validity, score interpretation, etc [26,27], to evaluate their current motion sickness state. After the SSQ (M1) was completed, the participant wore the HMD, and the experiment was initiated.

The experimental equipment used was a head-tracking interface device HTC VIVE HMD and an Xbox One Wireless Controller. With these devices, the participants played the VR FPS game and answered the SSQ. The game controller has several functionalities including forward movement; rotation based on pitch, yaw and roll axes; missile launch targeting; and selection of answers in the SSQ. Experimental content is a VR FPS game in space environment because it easily causes VR sickness due to multiaxis movement. The user uses the HMD and the game controller operations to trace the enemy target and score it with the missile.

The experimental protocol consisted of preliminary experiment (M2), eight experimental scenarios (S1–S8), SSQ, and rest videos. Figure 2 shows the experimental protocol for VR sickness measurement. In the preliminary experiment (M2), the participants were taught how to operate the HMD and the game controller. The eight experimental scenarios were designed to measure VR sickness. If a participant is exposed to VR content for a long time, the VR sickness level will potentially increase. [28] Therefore, each scenario was composed of content of 60 s for the safety of the participants. These eight experimental
scenarios were designed to have the visual guide on/off, the game controller on/off, and to vary sizes and positions of the visual guide to determine its effect on VR sickness. Table 2 shows the features of these eight scenarios.

In addition, the experimental scenarios were randomly placed and used to ensure reliability. For 240 s after each experimental scenario (including M2) ended, participants entered SSQ and watched the rest video to relax VR sickness. All participants were equally exposed to this rest video. When the experiment was completed, the SSQ data for the eight scenarios were automatically saved. From the VR sickness measurement experiment, preexperimental SSQ data (M1) and postexperimental SSQ data (S1-S8) were collected for each participant.

**Figure 2.** Experimental protocol for virtual reality sickness measurement. HMD: head-mounted display; SSQ: Simulator Sickness Questionnaire.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Feature</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Game controller</strong></td>
<td></td>
<td>Off</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
<td>On</td>
</tr>
<tr>
<td><strong>Visual guide size (%)</strong></td>
<td></td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>30</td>
<td>0</td>
</tr>
<tr>
<td><strong>Visual guide position</strong></td>
<td></td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>H&lt;sup&gt;a&lt;/sup&gt;</td>
<td>H</td>
<td>H</td>
<td>G&lt;sup&gt;b&lt;/sup&gt;</td>
<td>H &amp; G&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>H: head-tracking direction with head-mounted display.
<sup>b</sup>G: movement direction with game controller.
<sup>c</sup>H & G: head-tracking direction with head-mounted display and movement direction with game controller.

Using these experimental protocols, we conducted two experiments to examine the effect of the visual guide on VR sickness. Experiment I investigated the effect of visual guide on/off on VR sickness. In the first step of experiment I, we investigated the effect of VR sickness when the visual guide is on/off with the game controller off. Thus, the SSQ data of S1 and S2 were used for the comparison of VR sickness. Figure 3 shows the S1 and S2 used in the first step of experiment I.

In the second step of experiment I, we investigated the effect of the visual guide on/off with the game controller on. Thus, the SSQ data of S3 and S5 were used for the comparison of VR sickness. Figure 4 shows the S3 and S5 used in the second step of experiment I.

Experiment II investigated the effects of the visual guide’s size and position on VR sickness. In the first step of experiment II, we investigated the effect of the visual guide’s size with the game controller on. Thus, the SSQ data of S4, S5, and S6 and the SSQ data of S3 were used for the comparison of VR sickness. Figure 5 shows the S3, S4, S5, and S6 used in the first step of experiment II.

In the second step of experiment II, we investigated the effect of the visual guide’s position with the game controller on. Thus, the SSQ data of S5, S7, and S8 using a visual guide with different positions at 30% of the size of the aspect ratio and the SSQ data of S3 were used for comparison of VR sickness. Figure 6 shows the shows S3, S5, S7, and S8 used in the second step of experiment II.
Figure 3. S1 and S2 used in the first step of experiment I. VG: visual guide.

Figure 4. S3 and S5 used in the second step of experiment I. VG: visual guide.
Results

Overview
First, we compared the SSQ preexperiment (M1) and postexperiment (S1-S8) data to verify that the experimental protocol for VR sickness measurement was well designed. From the paired $t$ test results, a significant difference was observed in VR sickness between M1 and S1-S8. Significantly higher nausea and disorientation symptoms were noted in S1-S8 than in M1. Hence, we confirmed that the participants had experienced nausea and disorientation. In addition, oculomotor discomfort in S1-S8 did not significantly increase because many participants had eye fatigue from M1. From the different SSQ values before and after VR exposure, the increase in the values of nausea and disorientation symptoms of the participants was confirmed. It was concluded that the scenarios for the VR sickness experiment were appropriately produced.

Experiment I: Effects of Visual Guide On/Off
In the first step of experiment I, the results of the paired $t$ test showed no significant difference in VR sickness with respect to visual guide on/off when game controller was off; there was a negligible difference in nausea, oculomotor discomfort, disorientation, and total score values. In other words, there was no effect of visual guide with game controller off (watching video) on VR sickness. Table 3 shows the results of the paired $t$ test of visual guide on/off with game controller off. Figure 7 shows the results of the first step of experiment I (ie, difference in VR sickness with respect to visual guide on/off with game controller off).
In the second step of Experiment I, the paired t test showed a significant difference in VR sickness with visual guide on/off and game controller on: nausea and total score values were significantly decreased, whereas oculomotor discomfort and disorientation values had no significant differences. Table 4 shows the results of the paired t test of visual guide on/off with the game controller on. Figure 8 shows the results of the second step of Experiment I (difference in VR sickness of visual guide on/off with game controller on).

Table 3. Results of the paired t test of the visual guide on/off with game controller off.

<table>
<thead>
<tr>
<th>Scenario (state)</th>
<th>Nausea</th>
<th>Oculomotor discomfort</th>
<th>Disorientation</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>t (df)</td>
<td>P value</td>
<td>Score</td>
<td>t (df)</td>
</tr>
<tr>
<td>S1 (visual guide off)</td>
<td>8.65</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>N/A</td>
<td>17.84</td>
</tr>
<tr>
<td>S2 (visual guide on)</td>
<td>8.65</td>
<td>0 (31)</td>
<td>&gt; .99</td>
<td>13.50</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.

Figure 7. Results of the first step of experiment I (difference in virtual reality sickness with respect to visual guide on/off with game controller off). SSQ: Simulator Sickness Questionnaire; VG: visual guide.

Table 4. Results of the paired t test of the visual guide on/off with game controller on.

<table>
<thead>
<tr>
<th>Scenario (state)</th>
<th>Nausea</th>
<th>Oculomotor discomfort</th>
<th>Disorientation</th>
<th>Total score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>t (df)</td>
<td>P value</td>
<td>Score</td>
<td>t (df)</td>
</tr>
<tr>
<td>S3 (visual guide off)</td>
<td>9.24</td>
<td>N/A&lt;sup&gt;b&lt;/sup&gt;</td>
<td>N/A</td>
<td>15.87</td>
</tr>
<tr>
<td>S5 (visual guide on)</td>
<td>5.66</td>
<td>2.547 (31)</td>
<td>.02&lt;sup&gt;b&lt;/sup&gt;</td>
<td>13.03</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.
<sup>b</sup>Italicized values indicate statistical significance.
**Figure 8.** Results of the second step of experiment I (difference in virtual reality sickness of visual guide on/off with game controller on). SSQ: Simulator Sickness Questionnaire; VG: visual guide.

**Experiment II: Effect of Visual Guide Size and Position**

In the first step of Experiment II, the paired *t* test showed that there was a significant difference in VR sickness with the size of visual guide used in S5: nausea and total score values were significantly decreased, whereas oculomotor discomfort and disorientation values had no significant differences. However, there was no significant difference in VR sickness with the size of visual guide used in S4 and S6. Table 5 shows the results of the paired *t* test of the visual guide size with the game controller on. As shown in Figure 9, VR sickness was lower than that observed with other sizes when the visual guide was 30% of the size of the aspect ratio. Figure 9 shows the results of the first step of Experiment II (ie, difference in VR sickness with respect to the visual guide size when the game controller was on).

In the second step of Experiment II, the paired *t* test showed that there was a significant difference in VR sickness with the position of visual guide used in S5: nausea and total score values were significantly decreased, whereas oculomotor discomfort and disorientation values had no significant differences. However, there was no significant difference in VR sickness with the position of the visual guide used in S7 and S8. Table 6 shows the results of the paired *t* test of the visual guide position with the game controller on. As shown in Figure 10, VR sickness was lower than that observed with other positions when visual guide was in the head-tracking direction. Figure 10 shows the results of the second step of Experiment II (ie, difference in VR sickness with respect to visual guide position when the game controller is on).

**Table 5.** Results of the paired *t* test of visual guide size with game controller on.

<table>
<thead>
<tr>
<th>Scenario (size)</th>
<th>Simulator Sickness Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nausea</td>
</tr>
<tr>
<td></td>
<td>Score</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>S3 (0%)</td>
<td>N/A</td>
</tr>
<tr>
<td>S4 (10%)</td>
<td>7.45</td>
</tr>
<tr>
<td>S5 (30%)</td>
<td>5.66</td>
</tr>
<tr>
<td>S6 (50%)</td>
<td>7.16</td>
</tr>
</tbody>
</table>

*a* N/A: not applicable.

*Italicized values indicate statistical significance.*
Figure 9. Results of the first step of experiment II (difference in virtual reality sickness with respect to visual guide size when game controller was on). SSQ: Simulator Sickness Questionnaire; VG: visual guide.

Table 6. Results of the paired \( t \) test of the visual guide position with game controller on.

<table>
<thead>
<tr>
<th>Scenario (position)</th>
<th>Simulator Sickness Questionnaire</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nausea</td>
</tr>
<tr>
<td></td>
<td>Score</td>
</tr>
<tr>
<td>S3 (none)</td>
<td>9.24</td>
</tr>
<tr>
<td>S5 (H)</td>
<td>5.66</td>
</tr>
<tr>
<td>S7 (G)</td>
<td>9.24</td>
</tr>
<tr>
<td>S8 (H &amp; G)</td>
<td>7.75</td>
</tr>
</tbody>
</table>

aN/A: not applicable.
bItalicized values indicate statistical significance.

Figure 10. Results of the second step of Experiment II (difference in virtual reality sickness with respect to visual guide position when game controller is on). SSQ: Simulator Sickness Questionnaire; VG: visual guide.
Discussion

Principal Findings

In this study, we analyzed the correlation between VR sickness and crosshairs, which is widely used as a visual guide in an FPS game. To do this, eight scenarios were designed: visual guide on/off, game controller on/off, and various sizes and positions of visual guide. Experiments were performed using a protocol that consisted of the abovementioned eight scenarios, SSQ, and rest video. Results of Experiment I showed no reduction of VR sickness of the visual guide when the game controller was not operated, whereas there was a reduction when the game controller was operated. While the user operated the HMD and the game controller in the game, the user's gaze movement was synchronized with the motion of the visual guide, thereby reducing VR sickness. The results suggest that visual guide can reduce VR sickness caused by sensory conflicts between content and users when manipulating content. It can be interpreted that visual guide should be used effectively to reduce VR sickness because most VR content requires a game controller. Results of experiment II confirmed a difference in VR sickness with respect to the size and position of the visual guide with game controller operation. VR sickness according to sizes 10% and 50% and positions G and H & G of the visual guide was not significant, whereas VR sickness according to size 30% and position H of the visual guide was significant. Moreover, VR sickness was lower when the visual guide was 30% of the size of the aspect ratio and positioned in the head-tracking direction compared with other sizes and positions. The various sizes of the visual guide reduced nausea, oculomotor discomfort, disorientation symptom, and total score when compared to scenarios when a visual guide was not used. In particular, the visual guide at 30% the size of the aspect ratio further reduced these symptoms to a larger extent than other sizes. The visual guide at 50% of the size of the aspect ratio reduced the disorientation symptom, showing that disorientation was minimized due to the high synergy with the gaze movement when the visual guide size increased. However, if the size of the visual guide increases, the VR fidelity cannot be maintained because it lowers the immersion and presence of the user. The visual guide positioned in the head-tracking direction reduced the symptoms of nausea, oculomotor discomfort, disorientation, and total score more than other positions. As mentioned above, crosshair was the visual guide function to reduce VR sickness, as well as maintain VR fidelity in FPS games.

When recruiting participants, we tried to keep the number of men and women the same. As a result, the number of male (n=23) participants exceeded the number of female (n=9) participants recruited. This may have affected the results of the experiments, as it is reported that men are stronger than women with regard to motion sickness [29]. Therefore, the effects of the proposed method in a same-gender ratio environment should be included in further study.

Conclusions

In this study, we used an experimental protocol consisting of scenarios such as visual guide on/off, game controller on/off, and various sizes and positions of the visual guide to analyze the correlation between VR sickness and crosshair that is widely used as a visual guide in FPS games. VR sickness was found to be significantly correlated with visual guide on/off, and the use of a visual guide was very effective in reducing VR sickness when using a game controller. VR sickness was reduced by synchronizing the user's gaze movement to the motion of the visual guide while operating the HMD and game controller within the game. In addition, VR sickness reduced when the visual guide was 30% of the size of the aspect ratio and positioned in the head-tracking direction. From these findings, it is confirmed that using a visual guide can be an effective method to reduce VR sickness. The experimental results of this study indicate that the visual guide can achieve VR sickness reduction while maintaining user presence and immersion in the virtual environment. In other words, the use of visual guide is expected to solve the existing difficulty in disseminating various VR content due to VR sickness.

Acknowledgments

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (Ministry of Science and ICT [MSIT]) (No. 2017-0-00289-001). This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2020R1F1A1076114).

In addition, this study was partially supported by the Sabbatical Year Research Program of KOREATECH in 2019.

Conflicts of Interest

None declared.

References


**Abbreviations**

- **FPS**: first-person shooter
- **G**: movement direction with game controller
- **H**: head-tracking direction with head-mounted display
- **H & G**: head-tracking direction with head-mounted display and movement direction with game controller
- **HMD**: head-mounted display
- **IITP**: Institute for Information & Communications Technology Promotion
- **IVB**: independent visual background
- **MSIT**: Ministry of Science and ICT
- **SSQ**: Simulator Sickness Questionnaire
- **VR**: virtual reality

©Kwang-Ho Seok, YeolHo Kim, Wookho Son, Yoon Sang Kim. Originally published in JMIR Serious Games (https://games.jmir.org), 15.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on https://games.jmir.org, as well as this copyright and license information must be included.