

PH.D. THESIS IN
COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE

Extended Reality and Serious Games for Procedural and Cognitive Training in Healthcare

CANDIDATE

Alessandro Forgiarini

SUPERVISOR

Prof. Fabio Buttussi

Cycle XXXVIII — A.Y. 2024-2025

INSTITUTE CONTACTS

Dipartimento di Scienze Matematiche, Informatiche e Fisiche

Università degli Studi di Udine

Via delle Scienze, 206

33100 Udine — Italia

+39 0432 558400

<https://www.dmif.uniud.it/>

AUTHOR'S CONTACTS

Via delle Scienze, 206

33100 Udine — Italia

+39 0432 558413

forgiarini.alessandro@spes.uniud.it

Abstract

As the healthcare domain faces increasing demands for training and rehabilitation tools, Extended Reality (XR) and Serious Games (SGs) have emerged as promising solutions. This thesis investigates how design elements of XR systems and SGs affect knowledge and skill acquisition as well as user experience across three healthcare subdomains: procedural training for healthcare professionals, object-location spatial memory enhancement in healthy individuals, and executive function training in both healthy and clinical populations. The research proposes prototypes and describes user studies to evaluate both subjective user experience and objective performance metrics.

In procedural training, the thesis presents a study exploring the effects of real-time and post-performance feedback. Feedback, combined with refresher sessions, effectively improved the performance of the Basic Life Support with Defibrillator procedure. The thesis also presents qualitative and quantitative evaluations of a multi-user system for STEMI management. While performance improvements relative to a control group did not reach statistical significance, the results highlight a positive user experience.

Regarding object-location spatial memory, the thesis compared two versions of a SG: an immersive VR version for VR headsets and a non-immersive VR version for mobile devices. Results indicated that while immersive VR enhanced the sense of presence, the cognitive benefits were comparable to those of the non-immersive version. Furthermore, the thesis investigated the role of haptic interfaces in an object-location spatial memory task, showing that handheld controllers facilitated faster task completion than force-feedback gloves without altering memory performance.

In executive function training, the thesis proposes a SG with a sorting task to investigate the effects of adaptive difficulty and multi-user settings. The user study showed that the proposed heuristic-based difficulty adaptation did not result in a significant advantage over linear difficulty increase, and the multi-player setting did not demonstrate clear added value compared to the single-player. A variation of the SG based on the Go/No-Go task is presented and modelled as a Discrete-Time Markov Chain, indicating that the system could serve as a valuable instrument for clinicians to interpret patient performance and assess disease staging. In addition, the thesis presents an enhanced version of the SG that supports both immersive and non-immersive VR devices, along with an evaluation protocol for older patients with mild neurocognitive disorders.

Overall, the thesis showed positive effects of XR and SG on procedural and cognitive training in three different subdomains of healthcare. Most findings suggest that greater technological fidelity and some advanced features (e.g., immersive VR, haptic feedback, multi-player) do not necessarily lead to superior training outcomes, whereas other design elements (e.g., feedback, retraining) do. The thesis discusses expected and unexpected outcomes.

Contents

1	Introduction	1
1.1	Training Needs of the Healthcare Domain	2
1.1.1	Procedural Knowledge for Health Professionals	2
1.1.2	Cognitive Functions for Healthy People and Patients	3
1.2	Thesis Aims and Outline	6
2	Related Work	9
2.1	Level of Immersion	9
2.2	Use of Haptics	11
2.3	Feedback Modalities	13
2.4	Dynamic Difficulty Adjustment	14
2.5	Multi-user Setting	15
2.6	XR and SG for Training of Healthcare Professionals	17
2.7	XR and SG for Training Spatial Memory	18
2.8	XR and SG for Training Executive Functions	19
3	XR Procedural Training in Healthcare	21
3.1	Role of Feedback in XR Systems with Haptics for BLSD Training	21
3.1.1	Background	22
3.1.2	Description of the system and the simulation	23
3.1.3	Materials and Methods	25
3.1.4	Results	32
3.1.5	Discussion	42
3.1.6	Conclusion	44
3.2	Acceptance and Effectiveness of a Multi-user XR System for STEMI Training	45
3.2.1	Background	45
3.2.2	The Proposed VR System for STEMI Training	46
3.2.3	Qualitative evaluation of the STEMI training system	51
3.2.4	Quantitative evaluation of the STEMI training system	66
4	XR Training for Object Location Spatial Memory	79
4.1	Effectiveness of Display and Interaction Fidelity in a SG for OLSM Training	79
4.1.1	The proposed SG	80
4.1.2	User Study	84
4.1.3	Results	88

4.1.4	Discussion	90
4.1.5	Limitations	92
4.1.6	Conclusion and future work	93
4.2	Effects of Haptics in an XR System for Object-Location Spatial Memory Training	94
4.2.1	User Study	94
4.2.2	Results	98
4.2.3	Discussion	100
4.2.4	Limitations	102
4.2.5	Conclusion	102
5	XR Training for Executive Functions	105
5.1	Effects of Difficulty Adaptation and Multi-user Setting in an XR SG for Sorting Tasks	105
5.1.1	The Proposed Serious Game	106
5.1.2	User Study	112
5.1.3	Results	116
5.1.4	Discussion	119
5.1.5	Conclusions	122
5.2	Prediction of User Behavior in an XR SG for Inhibitory Control Training using Model Checking	122
5.2.1	Proposed Serious Game	123
5.2.2	Preliminaries on PRISM and Temporal Logics	125
5.2.3	Prism Model of the proposed SG	126
5.2.4	Probabilistic queries and results	130
5.2.5	Discussion and Conclusions	131
5.3	Acceptance and Efficacy of an XR SG for Inhibitory Control Training of Old Patients	132
5.3.1	The proposed SG	133
5.3.2	Methodology	136
5.3.3	Preliminary results	144
6	Conclusions and Future Work	155
6.1	XR Procedural Training for Nurses and Doctors	155
6.2	XR Training for Object Location Spatial Memory	156
6.3	XR Training for Executive Functions	157
6.4	Overall Conclusion	159

1

Introduction

In recent years, technologies capable of delivering eXtended Reality (XR) have gained increasing attention as immersive and interactive tools for professional education and training. XR acts as an umbrella term that includes Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (MR), which integrate real and virtual elements to different degrees. In the virtuality continuum proposed by Milgram and Kishino [221], AR enriches the physical world by overlaying digital content; MR blends real and virtual elements more dynamically, enabling their real-time coexistence and interaction; and VR immerses the user in a fully computer-generated Virtual Environment (VE). Consistently, Jerald defines VR as a computer-generated VE that users can experience and interact with as if it were real [150], where the sense of presence can be strong enough to create the perception of acting within the simulated world [339].

Some XR applications are designed as Serious Games (SGs), i.e., video games that use entertainment to further serious objectives [343]. The systematic review by Krath et al. [171] on the theoretical foundations of game-based learning, SGs, and gamification (i.e., the use of game elements in non-game contexts [84]) highlights several key mechanisms through which game elements can support learning and training: i) clear and relevant goals, as gamification can transparently illustrate objectives and their importance; ii) immediate feedback, allowing users to receive direct information about their actions; iii) positive reinforcement, rewarding users for performance and emphasizing the value of their achievements; and iv) adaptive content, which adjusts tasks and complexity according to the user's abilities and knowledge. SGs can also facilitate the "flow state", i.e., a mental state in which individuals are deeply engaged and intrinsically motivated to pursue the activity itself [76, 231].

By enabling realistic and repeatable simulation scenarios, XR technologies support training standardization, objective performance assessment, and large-scale dissemination. These features make XR particularly valuable in contexts where safety, cost, or logistical constraints limit the use of traditional training methods [155, 253].

XR and SGs have been widely adopted across multiple domains, including education, industrial training, workplace safety, and rehabilitation. For example, XR environments are increasingly used to support professional [217] and vocational [64, 78] training, from

health professions education [178] to broader occupational and industrial safety applications [258, 295]. In healthcare, XR has shown particular promise for training procedural knowledge and psychomotor skills, with systematic reviews indicating that VR-based training can improve learners' procedural performance, clinical skills, and attitudes compared with traditional approaches [153, 178].

Similarly, SGs have been deployed for education [308] and health-related purposes (e.g., promoting health among adolescents [13]), for teaching social skills to individuals with autism [299], medical training (e.g., training nursing students [223]) and rehabilitation ([238]), as therapeutic tools for individuals with mental health conditions [85], and stand-alone interventions as well as adjuncts to conventional therapy [110, 238].

1.1 Training Needs of the Healthcare Domain

In the healthcare domain, three primary groups of people who need training can be identified: professionals, healthy individuals, and patients. Health professionals, such as doctors and nurses, typically require procedural training that helps them acquire, update, and retain the procedures needed for safe and effective support for the population. Healthy individuals often engage in cognitive or physical training to maintain or enhance their abilities as part of preventive health strategies. Patients, for example, those with neurological conditions or age-related decline, need training and rehabilitation aimed at restoring or compensating for impaired cognitive or physical functions.

These different target users highlight the growing need for user-centered, adaptive, and scalable training solutions that can meet evolving clinical demands, support cognitive health, and ensure accessible rehabilitation, whether the target audience is healthcare professionals, healthy individuals, or patients with cognitive impairments.

1.1.1 Procedural Knowledge for Health Professionals

In healthcare education, procedural training for doctors and nurses focuses on acquiring and retaining the sequence of actions, decision points, and contextual conditions that underlie correct clinical performance, alongside the practical skills required to perform procedures (e.g., the correct execution of chest compressions in the Basic Life Support procedure). Developing robust procedural knowledge is essential to ensure that healthcare professionals can act consistently and effectively, particularly in uncertain or high-stress situations.

Continuing Medical Education (CME) is a globally recognized necessity for the medical profession [127]. It encompasses a broad spectrum of "educational activities which serve to maintain, develop, or increase the knowledge, skills, and professional performance and relationships that a physician uses to provide services for patients, the public, or the profession" [5]. CME represents any opportunity for clinicians to acquire new insights or refresh existing expertise [303]. These activities typically range from theoretical updates in diagnostic and therapeutic strategies to hands-on training for practical techniques and mixed events [7, 208, 303].

Evidence suggests that CME is most successful when it is treated as a truly continuous process rather than a series of isolated events [7]. Furthermore, its efficacy is

significantly enhanced by the use of interactive strategies and multiple exposures to educational content [208]. Beyond individual skill acquisition, CME is central to quality improvement [46]. By refining a practitioner’s knowledge, attitudes, and clinical behaviors, education can translate directly into more appropriate patient care and lead to better health outcomes [7, 106, 208]. Consequently, the World Health Organization [326] identifies the embedding of continuous learning within health systems as a key feature of healthcare quality.

However, traditional approaches, such as lectures, written protocols, or instructor-led simulations, often offer limited opportunities for active practice, personalized feedback, and realistic, context-based decision-making. In-person courses heavily depend on instructor availability and may lack the flexibility necessary to accommodate individual learning needs.

To address these pedagogical limitations, this thesis investigates the application of XR technologies to enhance the procedural knowledge acquisition specifically for emergency healthcare professionals. Among healthcare professionals, emergency physicians and nurses frequently face high-stress situations, and their training demands include remembering procedures, making rapid decisions, and coordinating in teams during high-stakes events such as cardiac arrest. Recent systematic reviews show that XR-based simulations are particularly promising in emergency settings, suggesting educational benefits in knowledge gain and retention as well as skills performance [1].

1.1.2 Cognitive Functions for Healthy People and Patients

Cognitive Functions (CF) encompass the mental abilities involved in daily living activities such as attention, executive functions, learning and memory, language, perceptual-motor, or social cognition [213]. Following the Diagnostic and Statistical Manual of Mental Disorders [213], a healthy individual presents cognitive functions within the normative range, allowing full independence in daily instrumental activities, while cognitive impairments are characterized by measurable declines in one or more cognitive domains, often accompanied by behavioral changes that interfere with everyday life. The degree of impairment distinguishes between mild Neurocognitive Disorder (mNCD), in which autonomy is preserved, and major Neurocognitive Disorder, in which independence is significantly compromised [213].

Because CFs are fundamental for maintaining autonomy and quality of life, preserving them through effective training is vital. Cognitive training aims to maintain or enhance these functions in healthy individuals and to mitigate decline or support rehabilitation in clinical populations. However, cognitive training programs can face challenges similar to those of procedural education, as professional-led sessions are costly, depend on specialist availability, and may offer limited adaptability or engagement. These constraints can result in low adherence, particularly when exercises are repetitive or detached from meaningful real-world contexts.

In this context, eHealth solutions seem promising in support to people aging with cognitive impairment but need further research [83], also addressing usability and participation [170].

The work in this thesis will focus on two CFs, Object Location Spatial Memory and Executive Functions, with a focus on Inhibitory Control. These skills are used in everyday life tasks and can deteriorate due to age or cognitive impairments. Therefore,

finding effective ways to enhance such skills is important for individuals with and without impairments.

Object Location Spatial Memory

Remembering where everyday items are placed (e.g., the keys in the purse, a pen on the desk, or a particular cup in the cupboard) and describing how objects are arranged relative to ourselves and to one another are tasks that rely on OLSM. This skill involves the ability to encode, store, and retrieve information about the location, orientation, and spatial relations of objects within an environment, across both small-scale settings (e.g., items on a desk) and larger spaces (e.g., rooms in a house) [225].

To perform OLSM tasks, individuals construct the “spatial image”, a short-lived internal representation in the working memory that can include multiple object locations, simple paths, and oriented elements in space [195]. This representation is modality-independent, meaning it can be formed and maintained through visual, auditory, or haptic cues, spatial language, or retrieval from long-term memory [195]. OLSM involves three mechanisms: (a) object processing – identifying the characteristics and meaning of the visual object, (b) spatial-location processing – determining where in the space the objects is, and (c) object-to-location binding, which creates a mapping between the two [252]. Two frames of reference for the spatial relations of objects are described in literature: egocentric self-to-object and allocentric object-to-object [9, 92, 283]. In the egocentric frame, object locations are encoded relative to the observer (e.g., “the shelf is behind me”). In the allocentric frame, spatial relations are encoded between objects within the environment (e.g., “the pen is left of the keyboard”), independent of the observer’s position.

OLSM tasks can be conducted in static environments or reaching spaces (i.e., they correspond to the portion of space within grasping distance and do not require participant displacement) or in navigational environments (i.e., involves the user performing either real or virtual displacements in the environments to reach the objects’ locations) [193].

In the educational context, research suggests that spatial thinking predicts achievement in STEM (i.e., Science, Technology, Engineering, and Mathematics). In particular, children with higher ability in visualizing spatial relationships in preschool have more advanced arithmetic skills in primary school [109, 291, 317, 318, 337].

Deficits in OLSM can arise from lesions in the hippocampus [160], infarcts in the right or left hemisphere of the brain [158], or more general memory deficiencies, like Alzheimer’s disease and dementia [159]. Indeed, impairments in spatial memory are considered one of the early indicators of Alzheimer’s disease [107]. Because such deficits severely affect independence and everyday functioning, spatial memory rehabilitation constitutes a meaningful target for cognitive interventions.

OLSM training is a cognitive intervention designed to target both working memory and spatial abilities. Although the primary goal is to enhance specific spatial skills, training has been shown to generalize, improving other spatial abilities that were not directly practiced [309]. Additionally, evidence suggests that OLSM training in rehabilitation settings can benefit broader cognitive functions, such as episodic memory and reasoning [340], as well as cognitive flexibility [220].

Executive Functions

Executive Functions (EFs) are described as a set of abilities that includes inhibitory control, working memory, cognitive flexibility, planning, reasoning, and problem solving [75]. In the model proposed by Diamond [88], it is reported how the three core EFs are Inhibitory Control, Working Memory, and Cognitive Flexibility.

- Working Memory (WM) comprises the systems necessary for retaining information while performing complex tasks such as reasoning, comprehension, and learning [24]. Notably, Miller [222] proposed the “magic number” 7 ± 2 of uncompressed chunks in WM. For example, when asked to memorize the sequence of ten total digits, a person tends to group the numbers into sets to remember them later, rather than memorizing the digits individually. The WM model consists of several subsystems; one of these is the visuospatial sketchpad, which is responsible for retaining visual material and element locations for short periods [23]. This subsystem is most effective within the first 600 milliseconds after a stimulus and decays over 10–20 seconds in the absence of active maintenance [248]. A common way to assess the visuospatial working memory is the Corsi Block-Tapping Task [73].
- Inhibitory Control (IC) refers to the ability to regulate one’s attention, behavior, thoughts, and emotions in order to override a strong internal tendency or an external distraction and instead act in a more appropriate, goal-directed way [88]. For example, IC allows individuals to suppress the impulse to attend to irrelevant stimuli while performing a task. This capacity plays an important role in everyday functioning and overall well-being. IC is commonly assessed through experimental paradigms such as the Stroop Task, Wisconsin Card Sorting Test, Go/No-Go Task, and Flanker Task [115].
- Cognitive Flexibility involves being able to shift attention or response sets, change perspective, and adapt to changing rules or environments in contrast to being rigid [88]. This ability builds upon the other two core EFs, as it generally requires inhibiting a previous perspective and loading a different rule into working memory [88]. For example, this ability supports behaviors such as revising a strategy when it no longer works, switching efficiently between multiple tasks, and considering alternative viewpoints in social interactions. Cognitive flexibility is typically assessed using paradigms that require set-shifting or task-switching, such as the Wisconsin Card Sorting Test [111].

EFs can decline due to multiple factors, including typical aging, neurocognitive disorders, stress, sleep deprivation, and certain psychiatric conditions [88, 213]. Given the central role of EFs in supporting everyday autonomy, their deterioration can substantially impact functional independence and well-being. Consequently, interventions aimed at strengthening or slowing age-related decline of EFs have become increasingly relevant. Evidence shows that EF performance can be enhanced through targeted cognitive training, physical exercise, and enriched, cognitively demanding activities, with improvements observed in both healthy individuals and populations at risk of cognitive impairment [88, 115].

1.2 Thesis Aims and Outline

Given the positive outcomes of integrating XR technologies [19, 178, 187] and SGs [3, 25, 49, 202] in healthcare training and rehabilitation, this thesis investigates how XR systems can support both procedural training for healthcare professionals and cognitive training for individuals with and without cognitive symptoms. In particular, it examines how different design elements influence users' experience and performance when acquiring and retaining procedures, as well as when performing specific cognitive tasks.

Aligned with the distinct needs of healthcare training, this thesis focuses on three main objectives targeting professionals, healthy individuals, and clinical populations:

1. Supporting the acquisition of procedural knowledge for healthcare professionals;
2. Maintaining and enhancing Object–Location Spatial Memory in healthy individuals;
3. Training executive functions in both healthy and clinical populations.

The cognitive training research follows a classical progression in which experiments are first conducted with healthy participants to establish feasibility, efficacy, and reference baselines. Once validated, these protocols are adapted and extended to clinical populations. In this thesis, the investigation of executive function training culminates in a dedicated study with patients with mNCD, developed through an international collaboration with the French research institute CoBTeK lab of the Université Cote d'Azur¹ and the Memory Center of the Nice University Hospital².

Chapter 2 presents the related work regarding the design elements investigated in this thesis, which are: level of immersion, use of haptics (e.g., haptic gloves), feedback modalities, dynamic difficulty adjustments, and multi-user setting. Furthermore, the chapter presents the use of XR and SGs in healthcare training by outlining current evidence on the effectiveness of XR technologies and SGs for procedural and cognitive training.

Chapter 3 focuses on procedural training for healthcare professionals, presenting three XR-based studies addressing different aspects of emergency procedural training. The first study explores feedback modalities in a Basic Life Support with Defibrillation (BLSD) training scenario, investigating the effects of real-time performance feedback and post-procedure feedback on learning outcomes and user experience. The second study and the third study, respectively, explore the qualitative and quantitative outcomes of a multi-user VR simulation for managing a ST-Elevation Myocardial Infarction (STEMI), involving three coordinated roles that simulate real emergency teamwork among medical professionals.

Chapter 4 explores XR-based cognitive training for healthy individuals, focusing on the maintenance and enhancement of OLSM. The first experiment compares immersive VR and non-immersive VR versions of a SG to assess the effects of display and interaction fidelity on performance and engagement. The second experiment investigates the effects of interaction fidelity in an immersive VR system, comparing standard VR controllers with haptic gloves that provide force feedback.

¹<https://www.cobtek.fr/english-website>

²<https://www.chu-nice.fr/>

Chapter 5 investigates SGs designed to train executive functions. The chapter first describes a study that examines the effects of adaptive difficulty and multi-user interaction in a sorting task, comparing adaptive and linear difficulty progression and single-player and multi-player modes. The chapter then introduces a theoretical model of a SG based on a Go/No-Go task, aimed at enabling rigorous assessment of user behavior across different cognitive health conditions. Finally, the chapter presents the design and preliminary results of a clinical protocol to evaluate the SG with older adults exhibiting mNCD.

Chapter 6 concludes the thesis by summarizing the main findings across procedural and cognitive subdomains of healthcare, and outlining directions for future research.

2

Related Work

This chapter examines the design elements of XR-based experiences explored in this thesis and reviews existing evidence regarding XR applications and SGs in healthcare training. The literature review specifically addresses procedural knowledge, spatial memory, and executive functions.

Design elements within XR and SGs can shape the user experience and impact performances, thereby influencing the effectiveness of training outcomes. This thesis focuses on five key dimensions: level of immersion, use of haptics (e.g., haptic gloves), feedback modalities, dynamic difficulty adjustment, and multi-user setting.

2.1 Level of Immersion

The distinction between immersive (iVR) and non-immersive (nVR) experiences constitutes a primary design element, defined by the degree to which users are perceptually isolated from the physical world and surrounded by the Virtual Environment (VE). Devices such as head-mounted displays (HMDs) or multi-projection environments (e.g., CAVE systems) can deliver iVR experiences that surround users with the VE and allow for real-time updates of the visual scene based on head and body movements [265]. In contrast, interaction with the VE in nVR setups typically occurs via flat displays and conventional input devices (e.g., a mouse or touchscreen). The level of immersion can significantly influence the user's sense of presence, the subjective feeling of "being there" in the virtual space [276].

Following the Cognitive Affective Model of Immersive Learning (CAMIL) proposed by Makransky and Petersen [205], the higher presence and agency that iVR systems can deliver may facilitate learning, interest, intrinsic motivation, and self-efficacy. However, high immersion acts as a double-edged design element as it may increase extraneous cognitive load depending on task complexity [205].

Meta-reviews indicate that iVR is particularly effective for perceptual-motor and spatial learning, supporting skills such as spatial reasoning, visual information processing, and hand-eye coordination [149]. Furthermore, iVR benefits the motivational and

affective domains, enhancing enjoyment and self-efficacy while reducing perceived exertion in physically demanding tasks [224]. Evidence suggests that these benefits have a greater impact in education and skills-focused training, notably in K–12 settings, science learning, and simulated practice, where iVR training may improve knowledge acquisition, skill development, and long-term retention [327].

Focusing on training for both physicians and nurses, iVR systems are frequently associated with higher learner satisfaction, engagement, and self-efficacy compared to nVR or traditional instructional approaches [54, 134, 190, 270]. In nursing education, iVR has demonstrated advantages in knowledge retention and skill performance [53, 134], with some reviews suggesting greater knowledge gains compared to nVR systems [218]. In medical education, head-mounted displays are perceived as salient and motivating [28, 63], and may reduce the learning curve for complex surgical procedures [305].

However, these advantages are not universal. Studies highlight that iVR offers no added benefit if the experience induces cybersickness or if overly intense immersion diverts attention from the learning objective [149]. Indeed, several works found no significant differences between iVR and nVR for acquiring spatial knowledge of object locations [42], gaining procedural knowledge in aviation safety [40], or promoting conceptual learning [227]. Similarly, the systematic review of Salatino et al. [275] on VR protocols for Unilateral Spatial Neglect (i.e., a debilitating disorder of spatial awareness) rehabilitation suggests that training effectiveness may depend more on simulation quality and task design than on immersion level alone.

Multiple studies report that objective learning outcomes (e.g., examination performance or procedural metrics) often do not significantly differ between iVR, desktop-based VR, MR, and traditional classroom instruction [89, 184, 190, 229, 270]. Similarly, engagement levels can be comparable across immersive and non-immersive VR modalities [181]. While virtual environments promote active learning through trial and error, they may fail to reproduce the psychological pressure and environmental complexity of real clinical emergencies, potentially limiting their effectiveness for high-stress training scenarios [306].

Highly immersive systems can also introduce practical limitations. Learners report technical difficulties, physical discomfort, and higher implementation costs compared to desktop or traditional solutions, which may hinder large-scale adoption [218, 274].

Evidence comparing different levels of immersion consistently indicates that benefits for OLSM depend more on task demands, environmental fidelity, and the availability of meaningful spatial cues than on immersion alone. MR appears particularly advantageous for recall tasks that benefit from anchoring virtual information to the physical environment. For instance, Hurter et al. [138] showed that MR outperformed two-dimensional displays and fully virtual conditions in recalling object locations and semantic details, despite comparable levels of reported presence in VR and MR. Their finding suggests that the coupling between virtual content and stable real-world reference frames plays a more critical role in memory recall than subjective immersion *per se*.

Within fully virtual environments, several studies have examined which aspects of immersive systems influence spatial memory acquisition and transfer. Xu et al. [330] investigated the transfer of spatial knowledge from a VE to the real world using different locomotion techniques and found no significant differences in placement errors, indicating that locomotion mode alone may not be a determining factor. Similarly, Hinterecker

et al. [128] reported that memory for object locations did not differ between horizontally and vertically aligned layouts when learning occurred from a single viewpoint, while also showing that judgments relied not only on object-to-object relations but also on object-to-environment relations. Complementing these results, Huang and Klippel [135] examined visual realism alongside individual differences, gaze, and movement, finding that high realism affords positive spatial learning opportunities even though no direct relationship with OLSM performance was observed.

Increasing immersion through head-mounted displays may support encoding by providing embodied orientation cues, particularly during learning phases. In the virtual memory palace task developed by Krokos et al. [174], iVR yielded higher recall accuracy than desktop-based nVR, with participants reporting an enhanced sense of spatial awareness attributable to head-tracked and body-based interaction. However, evidence for the benefits of immersion remains mixed. Kargut et al. [156] reported no clear consensus in the literature regarding the effectiveness of immersion for spatial learning and object selection. In their comparison of nVR and iVR setups for object location learning, more immersive conditions resulted in poorer selection performance and did not improve spatial learning outcomes. Despite this, most participants subjectively perceived immersive conditions as more effective for learning and expressed a preference for iVR. In two virtual OLSM tasks conducted inside an aircraft and outside the airport, Buttussi and Chittaro [42] found that although iVR improved distance estimation and perceived spatial presence, it did not outperform tablet-based nVR in object location retention. In line with these findings, the systematic review by Llana et al. [193] concluded that immersion alone is insufficient to support OLSM without rich and reliable environmental cues such as landmarks.

Comparisons between VEs and real-world learning further highlight the limitations of immersion in reproducing natural spatial experience. While both iVR and nVR can support spatial learning and transfer, Hejtmanek et al. [124] reported that real-world navigation resulted in superior performance, particularly during early stages of spatial acquisition. The authors suggested that the advantage may be attributed to the presence of intact visual detail, somatosensory feedback, and proprioceptive cues that were not fully replicated in the VE. When learning transfer over time is considered, AR shows complementary strengths. In a control-room spatial memory task Caluya et al. [47] reported that although VR supported stronger immediate recall, AR led to better delayed retention, likely due to the persistent alignment of learned spatial representations with real-world reference frames.

2.2 Use of Haptics

The second design element explored is the use of haptic feedback to increase the fidelity of immersive experiences by adding physically interactive elements.

In the context of XR and SGs, haptics involve simulating the sense of touch, which includes both cutaneous feedback (sensations on the skin surface, such as texture, temperature, and roughness) and kinesthetic feedback (forces felt in muscles and joints, including weight, inertia, and resistance) [237]. Technologies range from wearable haptic gloves to external force-feedback devices and instrumented mannequins. These additions can be designed to prevent users from unrealistically "passing through" virtual

objects (e.g., using an haptic glove to block the fingers when grabbing an object) and to be interacted with in order to increase the fidelity of the experience (e.g., by integrating sensors in real-world objects and then using the real-time data in the XR simulation).

Force feedback gloves are hand gloves composed of sensors that can track the fingers' movements, apply vibrotactile feedback, and deliver kinesthetic feedback when grasping virtual objects. Force-feedback gloves, in particular, combine motion tracking, vibrotactile cues, and kinesthetic resistance to support realistic object manipulation. With the growing diffusion of consumer VR kits, several commercial companies have begun investing in haptic glove technologies [6]. Importantly, the addition of tactile other than visual information plays a critical role in forming efficient representations of near-body space, underscoring the relevance of haptics for spatial interaction and learning in XR environments [199].

Research consistently shows that haptic technologies enhance the fidelity and realism of immersive experiences by simulating physical constraints [261] and strengthening users' sense of presence [114, 173, 204, 226]. Specific studies showed how haptic feedback has been effectively applied in domains where realistic tactile interaction supports skill acquisition: in procedural tasks, force-feedback gloves have reduced execution time [173] and errors [232]; in entertainment and training-oriented gameplay, combining haptics with hand tracking and immersive displays improved usability and engagement compared to visual-only setups [226]. Comparative evaluations of glove configurations suggest trade-offs between task accuracy and perceived tactile quality [255]. Beyond performance, haptic devices can also increase user motivation, such as by enhancing perceived playfulness in home-based rehabilitation [320] or by enabling interactions that more closely mirror physical learning models in educational contexts [108].

Regarding training, haptic feedback have been reported to enhance the fidelity and realism of XR simulators and support training outcomes, particularly in surgical and other psychomotor domains [261]. Furthermore, integrating haptic technologies can lead to better task performance, retention, and transfer when compared to non-haptic systems [204]. Concluding, combining haptic cues with other sensory modalities (e.g., visual, auditory, and proprioceptive information) produces synergistic benefits for psychomotor skill acquisition and conceptual understanding, highlighting the role of haptics as a key design element for enhancing XR training effectiveness [114]. The integration of haptic feedback is widely regarded as critical for psychomotor skill acquisition in both nursing and medical training, as tactile sensation is essential for understanding tissue resistance, force application, and procedural accuracy [179]. Systematic reviews indicates that haptics can enhance procedural performance and accuracy [22, 305], and help bridge the theory–practice gap by enabling trainees to physically experience procedures rather than relying exclusively on visual cues [251].

For example, Gani et al. [104] showed that integrating haptic feedback into an immersive VR bone-drilling simulation resulted in significantly better performance compared to a non-haptic condition. Similarly, in CPR training, Delahaye et al. [82] reported that combining VR with physical mannequins substantially improved chest compression quality. However, findings are not uniformly positive. In a comparison between a MR system and traditional manikin-based simulation, Vega et al. [313] observed superior compression performance in the control group, potentially due to visual distortion or insufficient tactile fidelity in the MR system.

Limitations in haptic realism have also been highlighted in systematic reviews. Shorey and Ng [284] noted that early VR-based nursing simulations often failed to accurately replicate tactile sensations associated with procedures such as cannulation, reducing perceived realism. Overall, the widespread adoption of haptics remains constrained by high development costs and the need for specialized hardware, resulting in limited availability of such systems within standard curricula [22, 251].

The integration of haptics in XR systems can enhance users' spatial representations of virtual environments by simulating tactile interaction with virtual objects and surfaces. Through haptic feedback, users can perceive properties such as object shape, texture, and resistance, as well as environmental features, thereby enriching embodied interaction within the space where the task is performed.

The recent research of Siqueira Rodrigues et al. [287] in assessing the effectiveness of visuohaptic integration in promoting object retention across different display environment employed a Delayed Match-to-Sample (DMTS) task (i.e., a cognitive evaluation tool that involves presenting a sample stimulus for participant encoding, followed by a delay where the encoded stimulus disappears to prompt participants to retain its mental representation) and report that haptics can improve object retention without requiring fully immersive setups, offering insights for the design of interactive systems that assist professionals who rely on precise mental representations of digital objects.

2.3 Feedback Modalities

The third design element explores how feedback is delivered to the user.

XR systems can implement visual, auditory, and haptic feedback to facilitate learners to recognize procedural errors and enable timely self-correction, as well as evaluate their performance, identify knowledge gaps, and obtain correct knowledge [212]. For example, a system might simultaneously trigger a red glow, an error sound, and controller vibration when an incorrect button is pressed.

Feedback timing is generally categorized into two modalities: real-time feedback (rt-FB), which provides immediate information to promote faster and more accurate performance during the task [155, 259]; and post-performance feedback (pp-FB), which offers a summary after completion to foster reflection and long-term retention [97]. By using these mechanisms, XR serves as a powerful tool for delivering standardized, adaptive, and evidence-based training [285].

In the clinical context of many European countries, "Audit and Feedback" (A&F) strategies are frequently employed to align professional practice with normative benchmarks [147]. "Audit" refers to evaluating performance against explicit standards, while "Feedback" involves reporting these results back to professionals to illustrate their standing relative to peers or targets [144, 147]. A&F interventions can lead to improvements in the quality of care [144, 145] and the way feedback is provided may affect effectiveness [144]. The literature reports positive results also when the emergency setting is analyzed [186, 267].

Automating A&F process within XR systems allows for consistent, objective, and data-driven evaluation, thereby reducing instructor bias. For example, the systematic review conducted by Gugelmin-Almeida et al. [116] indicates that automated systems delivering real-time feedback in CPR training can significantly enhance skill acquisition

and procedural accuracy. Furthermore, the American Heart Association emphasizes that simulation feedback should be quantitative, reliable, and provided both during and after training to maximize learning outcomes [59, 60].

Meta-analytic evidence suggests that integrating performance feedback within XR environments can significantly enhance technical performance in healthcare training, with effects depending on feedback timing, modality, and learning objectives. Real-time and post-event feedback mechanisms are particularly effective for supporting psychomotor skill acquisition and adherence to clinical guidelines.

For instance, in CPR training the meta-analysis by Lyngby et al. [197] showed that real-time feedback significantly improves compression depth and rate, whereas post-event feedback primarily enhances compression depth and fraction. The authors conclude that combining real-time feedback for immediate correction with post-event feedback for reflective consolidation would produce the most significant performance improvements. These findings are supported by multiple controlled studies reporting higher proportions of guideline-compliant compressions when quantitative or visual feedback is provided compared to no-feedback conditions [21, 152, 281]. AR systems offering real-time visual overlays similarly improve CPR performance relative to no-feedback baselines, although no clear advantage over other feedback modalities has been established [61].

Beyond CPR, XR real-time feedback has shown benefits in other high-precision domains. In surgery, AR visualizations support improved accuracy and reduced errors during complex procedures (e.g., implant placement and neurosurgery) [305]. In MR environments, real-time feedback has been introduced to train non-technical skills, such as communication and empathy, among nursing students [166]. This approach has enabled novice learners to outperform more experienced peers in both cognitive and psychomotor assessments [168]. Systems that integrate multimodal feedback (e.g., audiovisual cues and physiological indicators, such as stress monitoring) provide actionable insights for both trainees and instructors [95, 334].

Despite these advantages, XR feedback is not consistently superior to conventional instructional methods. In CPR training, XR feedback improves performance relative to no-feedback conditions but does not outperform instructor-led or device-based feedback [61]. In addition, the effectiveness of visual feedback is context-dependent. For example, displaying performance indicators in HMDs provides limited added value when physical mannequins are already used [82], and some feedback elements increase learner motivation without measurably improving task performance [333].

In the context of spatial memory, feedback mechanisms in serious games serve to validate navigational decisions and reinforce spatial associations. For example, systems like "CityQuest" inherently provide outcome-based feedback where the successful location of landmarks validates the user's egocentric spatial processing [220].

2.4 Dynamic Difficulty Adjustment

The fourth design element explores the implementation of Dynamic Difficulty Adjustment (DDA) mechanisms. These systems are designed to maintain an optimal level of challenge by continuously monitoring user behavior and dynamically tuning task parameters to match the learner's current proficiency. DDA techniques seek to optimize

the user's engagement and performance throughout the training session.

In the context of XR and SGs, DDA strategies typically involve monitoring performance metrics (e.g., error rates, reaction times) or, in more advanced implementations, physiological states (e.g., heart rate variability) to infer user stress and engagement levels [250]. These data are utilized to adjust the parameters of the experience in real-time, such as altering the stimulus speed (e.g., by reducing the time interval for a reaction), modifying the required precision, or changing the number of distractors in the environment.

The systematic review of Mortazavi et al. [230] indicates that DDA can significantly influence user experience, with several studies reporting positive effects on enjoyment, flow, motivation, engagement, and immersion. Furthermore, Paraschos and Koulouriotis [242] observed in their literature review that adaptive and personalized games can enhance the experience for students and the rehabilitation process for patients. These benefits are often realized through the dynamic modeling of distinct user characteristics, including personal interests and exhibited emotions.

DDAs mechanisms in XR environments aim to maintain learners in a state of optimal challenge by continuously balancing task demands with individual competence. This approach is particularly relevant in healthcare training, where learners exhibit heterogeneous prior knowledge and tasks combine cognitive, psychomotor, and affective components.

An example of an area in which DDA mechanisms are supported is the emergency and disaster response training for paramedics. For example, in the VR triage system described by Vogt et al. [319], trainers can manually adjust parameters such as the number of victims, symptom complexity, environmental conditions, and sensory distractions based on trainee skill level. Although not automatically adaptive, this approach reflects a DDA-inspired progression strategy, and the system received high ratings for motivation and user experience.

Similarly, the fully immersive "First Responder" system proposed by Kman et al. [164] dynamically alters environmental chaos (e.g., lighting, smoke, debris, and ambient noise) as well as the number of patients and their injury severity. Novice trainees are exposed to controlled scenarios, while advanced learners encounter increasingly complex environments. Learner evaluations indicate that such progressive escalation enhances perceived realism and training effectiveness [322], supporting the relevance of DDA principles in high-stress XR simulations.

In the context of spatial memory, mechanisms for adaptive training can automatically adjust the complexity of spatial tasks—such as the number of landmarks to memorize or the intricacy of the maze—based on real-time performance tracking [288].

2.5 Multi-user Setting

The fifth design element concerns whether a task is performed individually or in a multi-user setting as they can introduce social dimensions that might influence engagement and learning outcomes.

The theoretical foundations for multi-user design choices draw from Social Learning Theory [26], suggesting that individuals learn effectively by observing and interacting with others. In XR and SG contexts, collaborative environments can enhance presence,

co-presence, and communication, strengthening users' sense of "being together" and enabling more naturalistic teamwork behaviors [290].

The implementation of multi-user settings in XR and SGs generally follows two main configurations: asymmetric and symmetric [38, 93]. Asymmetric interaction involves a single trainee in the VE interacting with others who play specific support roles, such as an instructor controlling a character. This setup is widely used in healthcare to train non-technical skills like communication, leadership, and crisis resource management [35]. Conversely, symmetric collaboration involves multiple learners with equal roles working together or competing. While this configuration exploits social interdependence to enhance motivation [151], it imposes coordination-related "transaction costs" that must be managed to prevent cognitive overload [163]. A systematic review conducted by van der Meer et al. [311] identified five key affordances of multi-user settings: enhanced engagement and motivation, support for remote collaboration, provision of interdisciplinary learning spaces, development of social skills, and compatibility with constructivist learning paradigms.

Research comparing individual and collaborative learning in XR and SG environments reveals distinct advantages for each approach. Individual scenarios often are better in efficiency and clarity with learners demonstrating faster task completion [280] and users rate single-user modes higher for perspicuity, efficiency, and dependability in specific tasks [102]. These individual settings tend to drive users toward specific performance goals, such as attaining the highest possible score [68]. Conversely, collaborative scenarios are linked to higher accuracy and significantly greater engagement of cognitive resources through distributed processing [280]. Multi-user dynamics can benefit knowledge construction, where activities like joint summarization prove more effective than individual efforts [246], and may induce players to adopt maximizing strategies for team roles [68]. However, these benefits can be offset by more complex shared environments and higher cognitive load as without immersive avatars or high social presence, collaborative scenarios may suffer from slightly lower usability and motivation [10].

Multi-user XR environments are increasingly recognized for supporting interprofessional education and the development of teamwork and communication skills. For nursing students, SGs and collaborative VR environments facilitate teamwork training [214, 251]. The study conducted by Yang [332] shows that multi-user VR significantly improves team coordination and system usability compared to single-user or control conditions in neonatal resuscitation scenarios. Furthermore, interprofessional courses combining medical and nursing students in VR have also demonstrated improvements in handover performance and learner confidence [233].

Schild et al. [277] introduced the EPICSAVE XR system, a multi-user training environment in which paramedic trainees collaborate while a trainer actively controls and adapts the emergency scenario. Their findings indicate that positive learning outcomes were closely linked to high levels of presence and interactivity. At the same time, the authors highlighted that usability limitations could have disrupted presence and increased cognitive load. These results suggest that the educational effectiveness of multi-user XR depends not only on shared virtual co-presence, but also on the quality of interaction design and the reliability of the underlying system.

Concluding, evidence regarding the impact of multi-user XR on complex collaborative competencies remains limited. Several reviews highlight a research gap concern-

ing interaction, teamwork, and decision-making in multi-user scenarios [268], and some learners report insufficient opportunities for authentic team dynamics and leadership development [306].

Collaborative XR environments introduce social dimensions to spatial learning, though the benefits regarding OLSM specifically are complex. Research utilizing a virtual shopping mall indicated that multi-user settings, whether cooperative or competitive, distinctly alter user navigation patterns and spatial knowledge acquisition compared to single-user modes [185].

2.6 XR and SG for Training of Healthcare Professionals

XR and SGs have emerged as tools in healthcare education, offering advantages to traditional training formats. A primary advantage of these technologies for healthcare professionals lies in their flexibility with respect to time and location. Unlike traditional role-play courses or mannequin-based simulations, XR-based training is not inherently location-dependent and allows for repeated practice, which is particularly valuable for mastering clinical skills [251, 284].

This adaptability also extends to safety. For instance, the work of Rezaei et al. [263] presented how VR-based training can reduce radiation exposure for cardiologists and nurses compared to traditional instruction, while simultaneously enhancing protocol retention [263]. XR simulations further enable safe practice in high-risk or restricted environments, such as isolation wards, without endangering patients or learners [336].

Meta-analytic evidence consistently indicates that, for nursing students, XR and SG interventions are effective in enhancing theoretical knowledge [134, 189, 191, 268, 286], practical skills [134, 189, 191], and learner satisfaction [134, 189, 191]. Across broader healthcare populations, including medical students and residents, immersive technologies similarly support knowledge acquisition and skill development [28, 270].

However, learning outcomes are frequently reported as comparable rather than superior to those achieved with conventional teaching methods [62, 190, 223, 270]. While benefits in cognitive and psychomotor domains are well supported, evidence for improvements in higher-order outcomes such as critical thinking and complex clinical decision-making remains mixed or limited [134, 191, 268].

VR allows trainees to learn in immersive and highly interactive environments that simulate real-world contexts (e.g., practicing procedures on virtual patients without posing risks to actual individuals), supporting experiential learning and learning by doing [207], in line with constructivist and situated learning theories [66]. In addition, VR could enhance learner motivation [136, 307], engagement [40, 41, 136, 270], and presence [40, 41] through high levels of immersion and interactivity, while also showing potential for improving post-intervention knowledge and skills compared to conventional approaches [178].

The scientific literature on education using VR for medical doctors [20, 28, 50, 86, 143, 219, 229, 269, 274, 302, 338] and nurses [131, 146, 251] has grown rapidly in recent years, underscoring the broader recognition of immersive technologies as valuable tools in health professional training [19, 187]. The COVID-19 pandemic further accelerated VR

adoption, particularly for distance learning, where it offered innovative ways to sustain educational continuity with generally positive outcomes for training medical doctors [20, 241, 274] and nurses [131, 181, 241]. In the context of cardiovascular medicine, VR is changing the training approach [18, 289, 312], with simulation-based training especially used in interventional cardiology [247].

XR-based training is an effective complement to traditional learning approaches [284, 286]. Surveys and studies indicate how VR training can enhance theoretical knowledge [57, 67, 134, 189, 190, 268, 284, 286, 335], practical skills [131, 134, 182, 190], and psychomotor performance [67, 189]. Positive outcomes have also been reported for procedure-specific training and motivation to learn [63, 284], learning satisfaction [63, 134], emergency response capacity [335], as well as skill retention [190, 263].

However, there is no clear evidence that VR training yields better results than traditional teaching methods for medical and nursing students [126], primarily due to differences in study design and the small number of participants involved [19, 57, 219, 229, 302]. In particular, two meta-analyses as well as a recent study found no statistically significant difference in nurses' procedural skills between learners using VR and those using conventional educational approaches [34, 54, 57]. Similarly, findings on the impact of VR training on critical thinking skills are inconsistent, as some meta-analyses report no significant effects [134, 190]. Finally, the evidence concerning self-efficacy is mixed, as several studies document improvements, while others show no meaningful change [63, 134, 146, 180, 189, 284]. Technological limitations and a lack of realism may have led to inconclusive findings, especially in early studies [284].

2.7 XR and SG for Training Spatial Memory

XR and SGs have been increasingly adopted to assess and train OLSM. The recent systematic review conducted by Llana et al. [193] about the use of XR in navigational OLSM highlights that, compared to other memory systems, OLSM has received less attention overall. Non-clinical OLSM research still concentrates mainly on healthy young adults. OLSM was primarily assessed using 2D or 3D computerized tasks, with high-semantic-value items for participants (e.g., everyday objects). The main focus on small-scale environments suggests a common trend toward investigating OLSM in simpler spatial contexts. [193].

XR and SG technologies can offer capabilities for standardization and rehabilitation. For example, in the medical context Maidenbaum et al. [203] compared the performance of epilepsy patients and healthy individuals in an OLSM task and showed that patients' data are comparable to the general population's data. Belger et al. [29] developed an iVR neuropsychological assessment tool for spatial memory that was positively rated in terms of usability, and Park [243] investigated the effects of XR based training for older adults with mNCD highlighting how the VR training yielded to statistically significant greater improvements in spatial cognition and episodic memory when compared to the control group.

Moreover, in previous sections, it is already noted that benefits for OLSM depend more on task demands, environmental fidelity, and meaningful spatial cues than on immersion alone. Hurter et al. [138] reported how MR settings demonstrate particular advantages for recall tasks through anchoring virtual information to stable real-world

reference frames, outperforming both 2D and fully virtual conditions. Head-mounted displays provide modest encoding support through embodied orientation cues, as evidenced in virtual memory palace tasks [174]; however, evidence remains mixed, with some studies showing limited performance and no improvement in spatial learning outcomes despite subjective user preferences for immersive conditions [42, 156].

Real-world navigation outperformed VE during early spatial acquisition due to intact visual detail, somatosensory feedback, and proprioceptive cues absent in VEs [124]. In a control-room spatial memory task Caluya et al. [47] reported that although VR supported stronger immediate recall, AR led to better delayed retention.

User studies investigating SGs for training OLSM have yielded positive outcomes across several spatial memory measures and executive functions [220]. Collaborative XR environments introduce social dimensions that distinctly alter navigation patterns and spatial knowledge acquisition compared to single-user modes [185].

2.8 XR and SG for Training Executive Functions

XR-based cognitive training offers an effective and engaging method to enhance cognitive functions. Preliminary evidence shows that immersive VR environments improve visual memory and sustain attention beyond training sessions, making them well-received tools for healthy elderly populations [298]. While most executive function interventions still use traditional paper-and-pencil tasks, recent pilots integrate SGs and exergames as innovative alternatives especially suitable for older adults or people with neurological impairments [87]. Targeting younger audiences, Homer et al. [130] proposed two versions of a SG (one focusing on speed, the other on accuracy) to train inhibitory control in adolescents. Their findings support the general claim that playing SGs can improve EF skills, particularly when the game is explicitly designed to target those skills. In addition, exergames show particular promise for cognitive-motor rehabilitation across neurological conditions [201].

SGs serve as valuable tools for both assessing and treating cognitive decline by capturing objective behavioral data to complement standard neuropsychological tests and enable early detection [264, 273]. Studies confirm the feasibility of SG-based cognitive assessments [304, 310], their efficacy in Alzheimer’s training [14, 148], and greater appeal to elderly users over conventional exercises [122, 215]. A systematic review by Zuo et al. found that digital SGs significantly improved cognition, daily functioning, and depression in adults with Alzheimer’s or MCI. However, results warrant caution due to limited randomized control trials, small samples, and low-quality evidence [342].

Long-term engagement poses a key challenge as SGs boost initial motivation, but sustaining interest is difficult [49]. The LightSword VR exergame, blending physical and cognitive tasks, improved inhibition in older adults over eight months, with effects lasting six months afterwards [49]. However, usability issues persist. For example, the “Brain-IT” exergame was acceptable for patients with mNCD but was limited by technical and hardware problems [206].

SGs and XR training for executive functions hold strong promise, but evidence on scalability, motivation, and sustained impact is limited. Findings from Hassandra et al. affirm good acceptance among older adults, yet efficacy validation remains essential [122].

3

XR Procedural Training in Healthcare

Building on the growing demand for Continuing Medical Education (CME) and the effectiveness of Audit and Feedback (A&F) strategies in XR, this chapter introduces two XR systems designed to optimize training for emergency healthcare professionals. These systems are the result of a multidisciplinary collaboration between the Human-Computer Interaction Lab at the University of Udine and the Azienda Sanitaria Universitaria Friuli Centrale (ASUFC).

The first system focuses on Basic Life Support and Defibrillation (BLSD). It features an iVR simulation integrated with a physical manikin, enabling real-time measurement and feedback on compression depth and frequency. The study evaluates how real-time and post-performance feedback influences the performance and user experience of trainees.

The second system is a multi-user application designed for training teams in the rescue of patients experiencing a STEMI (ST-Elevation Myocardial Infarction) heart attack. The system underwent two evaluations: a qualitative study capturing the subjective perceptions of emergency professionals, and a comparative study measuring training efficacy between a VR-trained group and a traditional control group.

3.1 Role of Feedback in XR Systems with Haptics for BLSD Training

This section presents a study investigating the impact of performance feedback on key metrics during BLSD training. This training is delivered via an XR system integrated with a physical manikin, which provides realistic haptic feedback to enhance the simulation's fidelity.

3.1.1 Background

BLSD is a life-saving technique applied to rescue an unresponsive person with abnormal or no breathing who is suspected to be in cardiac arrest. The BLSD algorithm consists of several steps: recognizing a cardiac arrest, alerting emergency medical services, running a high-quality cardiopulmonary resuscitation (CPR) and applying early defibrillation [236]. About one-third of the patients with out-of-hospital cardiac arrest enrolled in the large EuReCa Study ($n=37,054$) had a return of spontaneous circulation after bystanders performed CPR [112], but when health professionals performed resuscitation on in-hospital cardiac arrests, the goal was achieved in higher proportions, ranging from 36% to 79% [119, 256, 257, 278].

However, performing correct CPR requires procedural knowledge and manual skills that must be regularly practiced [31, 60] and many factors affect CPR skills retention among healthcare providers, such as educational level, experience, type and frequency of training, equipment and facilities, and clinical case review [154]. The most recent Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care published by the American Heart Association (AHA) [60], developed through a process of evidence review, strongly recommend booster training sessions to improve skills retention over time and provide a moderate recommendation for a spaced learning approach, while a mastery learning and deliberate practice model for resuscitation training is only weakly recommended [60].

Unfortunately, a recent review showed that instructors' assessment skills for chest compressions are suboptimal [165] and the 2022 International Liaison Committee on Resuscitation's Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations was not able to suggest the most appropriate training strategy to overcome this flaw, due to a lack of consistent scientific evidence [328]. In addition, extending traditional manikins with corrective feedback devices (such as a visual display of compression depth) or prompt devices (e.g., metronomes set to the correct compression frequency, pressure-sensing devices) is strongly recommended by AHA because it leads to skills improvement and retention [60], but may also help in reducing the variability of the instructor's subjective assessment. Furthermore, to improve training efficacy through increased realism, AHA moderately recommends using high-fidelity manikins (i.e., patient simulators reacting as if they were alive) and enriching traditional courses with in situ education (i.e., training is performed in the patient's actual setting) [60].

Building on the benefits of XR-based training in CPR, this study extends the XR system originally proposed by Buttussi et al. [43], which integrates immersive VR with a physical manikin. The system has been modified by adding defibrillation and updating the procedure to align with the European Resuscitation Council's COVID-19 guidelines [235] to evaluate the impact of real-time (rt-FB) and post-performance (pp-FB) feedback. To achieve this, a single-blinded controlled trial with doctors and nurses was conducted to determine if training with these feedback types improved efficacy compared to training without them (RQ1). Furthermore, the research measured knowledge retention after one month (RQ2), the effectiveness of a subsequent refresh training session (RQ3), and the participants' overall perceptions regarding the system's usability and value (RQ4).

3.1.2 Description of the system and the simulation

The XR system is composed of the following:

- a software that simulates a virtual emergency scenario (Fig. 3.1A);
- a workstation (CPU: i9-10900X, GPU: NVIDIA Quadro RTX 4000, RAM: 32GB) to run the software;
- a monitor (Dell P2720D) to allow the experimenter to see the progress of the simulation;
- a binocular VR headset (HTC Vive Pro Eye), sensing the head position and orientation, with integrated headphones (Fig. 3.1B);
- two wrist trackers (HTC Vive Tracker 3.0), sensing the position and orientation of the trainees' hands (Fig. 3.1C);
- a traditional BLSD manikin (Laerdal Resusci Anne), integrated with an Arduino-based pressure sensor; the manikin position in the real world is mapped in the VE with two trackers so that the trainees' interaction with the manikin in the real world is applied to the virtual patient in the VE (Fig. 3.1D).

The wrist trackers and the pressure sensor acquire information about trainees' actions and send them to the software.

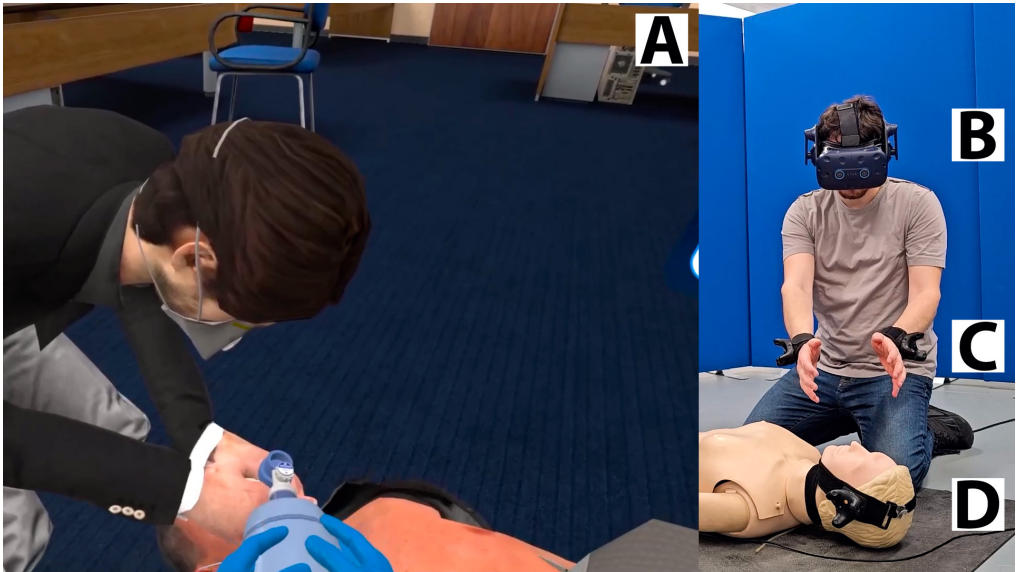


Figure 3.1: The software simulating the virtual scenario and a user wearing the devices to use it: A) the software as seen by the user, B) the VR headset, C) the wrist trackers, D) the manikin with trackers.

The virtual emergency scenario concerns a heart attack affecting an employee in his office. A threat to safety (personal computer with electrical problems) is also in the

office. The trainee is asked to help the victim following the BLSD sequence described in the European Resuscitation Council COVID-19 guidelines [235], as translated and integrated by the Italian Resuscitation Council [74].

The simulation involves a virtual colleague of the victim, who arrives on the scene after the 9th step of the sequence so that the trainee can send him to look for an AED. The virtual colleague also brings a reanimation balloon, places it on the victim's face, and keeps it in the correct position for ventilation.

The correct sequence of actions that need to be performed by the trainee, adapted from [74], is the following:

1. Wear gloves (self-protection)
2. Wear the FFP2 mask (self-protection)
3. Turn off the electricity (safety)
4. Shake the victim gently by the hips and ask: "Are you all right?"
5. Uncover the victim's chest
6. Check for thorax movements, coughing, and normal breathing for 10 seconds
7. Alert the emergency medical service through a mobile phone call
8. Put a surgical mask on the victim
9. Perform chest compressions
10. Send the virtual colleague to get an AED
11. Switch the AED on when it is brought on the scene
12. Attach the electrode pads to the victim's chest
13. Ensure no one is touching the victim
14. Push the shock button on the AED
15. Perform 30 chest compressions
16. Use the virtual reanimation balloon (2 insufflations)
17. Perform 30 chest compressions
18. Use the virtual reanimation balloon (2 insufflations, then the victim moves, coughs, and opens his eyes)
19. Check if the victim is conscious by looking for thorax movements, coughing, and normal breathing for 10 seconds

The XR system complies with the above-described sequence and can be run in two modes:

- Tutorial XR-BLSD, which gives trainees the chance to learn the sequence with the help of speech instructions as well as written text and illustrations in pop-up windows describing each step (Fig. 3.2D); in addition, semi-transparent head/hands (Fig. 3.2A) or directional arrows (Fig. 3.2B) suggest how/where to perform the step;
- Test XR-BLSD, in which trainees must repeat the same sequence without help, with the only exception of the semi-transparent head/hands (Fig. 3.2A) or directional arrows (Fig. 3.2B) if trainees were not able to proceed alone after a predefined time.

The XR system follows the A&F process. In the Audit phase, the headset and trackers acquire all the actions performed by the trainees, and the software compares them to the expected sequence (correct steps). Moreover, the system can acquire and calculate data about cardiac massage in terms of:

- compression depth (correct if ranging between 5 and 6 cm);
- compression frequency (correct if ranging between 100 and 120 compressions per minute);
- correct compressions considering depth, release, and hand position;
- too-weak compressions (less than 5 cm);
- too-deep compressions (more than 6 cm);
- compressions with incomplete release (depth at release higher than 1 cm);
- compressions with wrong hand position (more than 10 cm away from the center of the chest).

The Feedback phase provides real-time Feedback (rt-FB) and post-performance Feedback (pp-FB) implemented as follows:

- rt-FB, during the cardiac massage steps, showing a panel with the compression depth and frequency (Fig. 3.2C) supporting trainees in immediately correcting them to stay in the expected ranges;
- pp-FB, at the end of the simulation, displaying a debriefing panel (Fig. 3.2E) with trainees' performance at each step (right if performed correctly and within 5 seconds) as well as, for each of the three series of compressions, a graph showing the compression depth and frequency, and textual information about percentages of compressions that were correct, too weak, too deep, and with an incomplete release.

3.1.3 Materials and Methods

The study was conducted in the Virtual and Augmented Reality Laboratory of the Lab Village facility at the University of Udine. The protocol involved participants who agreed to train using the XR system on a voluntary basis by signing an informed consent module.

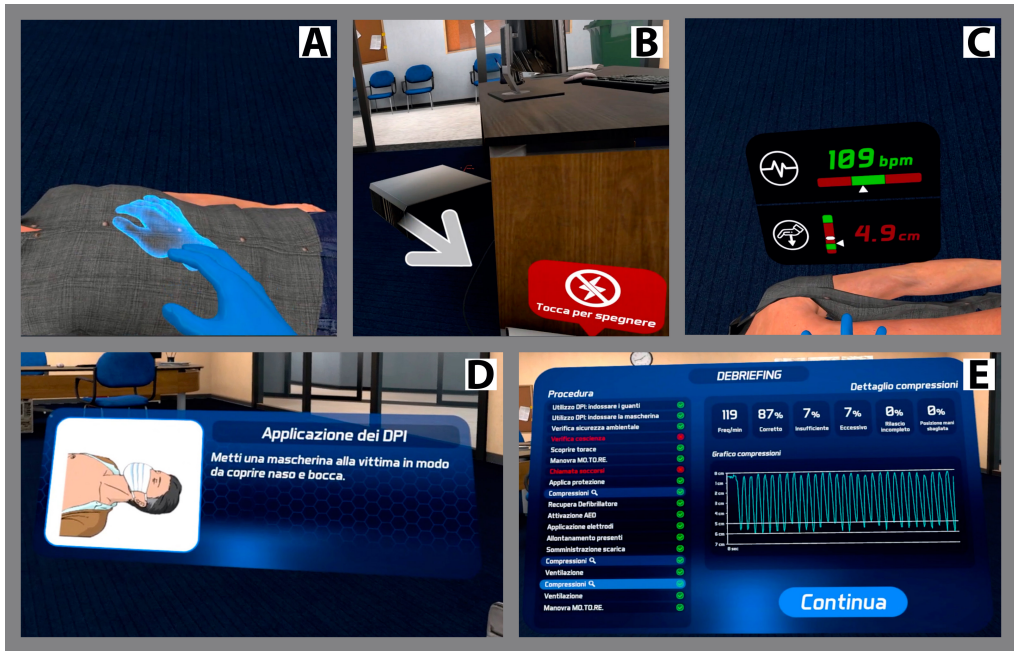


Figure 3.2: Interfaces displayed by the XR system: A) semi-transparent hands suggesting how to perform the step, B) directional arrows suggesting where to perform the step, C) panel with the compression depth and frequency, D) pop-up window with written text and illustrations describing each step, E) debriefing panel. All interfaces are localized in the participants' native language.

Design and Hypothesis

Our goal was to assess the effects of two independent variables in a two-part study.

In the first part, we organized participants into two groups, i.e., Group 1 (G1) and Group 2 (G2). Both groups underwent a first training session with the XR system, Session 1 (S1), where they first received initial training (Tutorial XR-BLS and Test XR-BLS), and then we assessed post-training performance (using Test XR-BLS). During the training, G1 received rt-FB and pp-FB from the system, while G2 did not. After the post-training performance, both groups received pp-FB. In this part, the independent variable (between-subjects) is *initial training feedback* and has two values: *present* if rt-FB and pp-FB were provided in the initial training phase with the XR system or *absent* if not.

In the second part, about a month after S1, all participants who agreed to continue underwent a second training session, Session 2 (S2). In S2, we first assessed retention performance (using Test XR-BLS), then all participants received the same refresh training (Tutorial XR-BLS and Test XR-BLS), and finally, we assessed post-refresh performance (using Test XR-BLS). In this part, the independent variable (within-subjects) is *time of measurement* and has three values: *post-training*, i.e., immediately after initial training; *retention*, i.e., about a month after initial training; or *post-refresh*,

i.e., immediately after refresh training.

We formulated the following hypotheses:

- *H1)* Participants who receive rt-FB and pp-FB in the initial training phase with the XR system (initial training feedback = present) should perform better than participants who do not receive it (initial training feedback = absent) because we conjecture that the positive effects of feedback found in traditional medical education [60, 249, 321] should extend to XR;
- *H2)* The performance of all participants should decrease from post-training to retention due to natural decay that occurs with the knowledge and skills that are not refreshed [17, 79];
- *H3)* Refresh training in XR should be effective in restoring knowledge and skills as it happens with traditional refresh training in medical education [12, 60, 152, 154].

Questionnaires

The study employed the following questionnaires:

- *Initial questionnaire.* It collects information about participants' age, gender, job role (doctor or nurse), working in the emergency area (yes or no), the number of resuscitations performed in the last 12 months, previous VR use (yes or no, and hours of use), experience as BLSD instructors (yes or no), months since last BLSD training, and participation in a workshop that included a brief presentation of the system (yes or no).
- *Usefulness, Satisfaction, and Ease of use (USE) questionnaire.* We used the translation of the USE questionnaire proposed in [196] and evaluated in [105]. It assesses an overall score and four subscales: usefulness (8 items), satisfaction (7 items), ease of learning (4 items), and ease of use (11 items) of the system, as perceived by participants, who are asked to rate their agreement with the statements through a seven-point Likert rating scale ranging from strongly disagree to strongly agree.
- *Willingness to use questionnaire.* It includes 9 items about participants' preferred setting (at work or home), purpose (personal or professional) of use of the system, and the preferred time lapse between training sessions (1 month, 6 months, 1 year). A blank box was available at the end of the questionnaire to collect any comments, suggestions, or advice concerning the system.

Data collected during XR sessions were recorded in “.json” files, while all questionnaire data were collected using Microsoft Forms ¹. All data were exported to “.csv” files for analysis.

Measures

To assess participants' knowledge about the BLSD procedure at post-training, retention, and post-refresh times of measurement, we measured CorrectSteps counting the number

¹<https://forms.office.com>

of correct steps autonomously performed by participants in the Test XR-BLSD mode without the help of the directional arrows or semi-transparent head/hands.

To assess participants' skills at the same measurement times, we considered the data acquired and calculated by the XR system (Section 3.1.2), and we derived the following measures:

- Frequency, i.e., the mean of compression frequency,
- FrequencyDifference, i.e., 0, if the mean compression frequency is in the correct range, mean compression frequency - 120, if the mean compression frequency is above 120, 100 - mean compression frequency, if the mean compression frequency is below 100,
- FrequencyInRange, i.e., a boolean value that indicates if the mean compression frequency is in the correct range,
- Depth, i.e., mean compression depth,
- DepthDifference, i.e., 0, if the mean compression depth is in the correct range, mean compression depth - 6, if the mean compression depth is above 6, 5 - mean compression depth, if the mean compression depth is below 5,
- DepthInRange, i.e., a boolean value that indicates if the mean compression depth is in the correct range,
- TooWeak, i.e., percentage of too-weak compressions,
- TooDeep, i.e., percentage of too-deep compressions,
- IncompleteRelease, i.e., percentage of compressions with incomplete release,
- WrongPosition, i.e., percentage of compressions with wrong hand position,
- CorrectCompressions, i.e., percentage of correct compressions.

The immersive XR simulation requires performing the chest compressions three times during each execution of the procedure. However, during the first sequence, the participant had to interrupt the compressions to send the virtual colleague to get an AED, and during the third sequence, tiredness could have affected the performance. Therefore, we decided to focus our attention on the data of the second sequence, which was the first that always required to perform 30 chest compressions.

Notably, the number of compressions with wrong hand position was 0 or close to 0 for all participants, in all executions of the procedure, so we decided to discard the WrongPosition measure from the results.

Considering subjective measures, we assessed the overall USE score and its four subscales, after training and after re-training with the XR system. We also assessed the 9 items about willingness to use the XR system at the end of S2.

In the first part of the study, to find possible differences between G1 and G2, we compared the measures of the two groups at post-training, considering data from all participants.

In the second part of the study, to assess possible knowledge and skill decay as well as the effect of refresh training, we compared the measures at the three times of measurement considering data from participants who attended both sessions.

Participant

For the first part of the study, we calculated the sample size considering a minimum mean difference of 2 correct steps between G1 and G2, i.e., about 10% of the 19 steps of the BLSD sequence, to find at least the effect reported by Kubal et al. [175]. Assuming a standard deviation of 3 and a ratio of 1 between G1 and G2, 36 participants per group is needed to obtain a power of 0.8 with a significance level of 0.05. Assuming a possible dropout of 10% during the first part of the study, we increased the sample size to 40 per group.

Participants were recruited among doctors and nurses working in public hospitals, departments, or healthcare districts of the Friuli Venezia Giulia Region. Inclusion criteria were a degree in Medicine or Nursing and participation in a residential BLSD training course within the previous 3 years. Having tried previous prototypes of the system was an exclusion criteria.

Specific advertising posters were affixed in all public health structures of Friuli Venezia Giulia Region, and a digital version of the poster was sent via e-mail to institutional recipients for dissemination. On the poster, a link to an online calendar was provided to allow willing health professionals to book participation in S1, choosing among the available dates and times. The booking process provided the information to reach the facility of the University of Udine where the sessions took place.

We assigned participants to G1 or G2, aiming at minimizing differences between groups in the results of the initial questionnaire. In particular, we aimed at having the same number of participants in each group and at minimizing the differences between the two groups in terms of:

1. Workshop participation
2. Being a BLSD instructor
3. Working in an emergency department
4. Professional role
5. Previous use of VR
6. VR hours of use, for those who previously used VR systems
7. Age
8. Gender
9. Number of resuscitations performed in the last 12 months
10. Months from the last retraining course

We devoted particular attention to having the same number of users per group for the first three characteristics because a participant who answered affirmatively to those questions would respectively: 1) have already seen the system (but without using it), 2) already have accurate BLSD knowledge, 3) be more likely to have performed the procedure on a patient during work time.

Between May and October 2022, we recruited 80 participants (55 doctors and 25 nurses), who were involved in the first XR-BLS training session in the same period. We assigned each participant to G1 (n=41) or G2 (n=39). Groups were balanced according to the variables collected through the initial questionnaire (Table 3.1).

Table 3.1: Participants characteristics (initial questionnaire). Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹Chi-square test; ²Wilcoxon rank sum test; ³Fisher's exact test.

	G1 n=41	G2 n=39	p-value
Gender			
male	19 (46.34%)	19 (48.72%)	1.000 ¹
female	22 (53.66%)	20 (51.28%)	
Age	33.30 $\pm$ 8.73 30.00 (28.00 - 35.00)	35.20 $\pm$ 9.81 32.00 (28.50 - 37.50)	0.322 ²
Job role			
nurse	13 (31.71%)	12 (30.77%)	1.000 ¹
doctor	28 (68.29%)	27 (69.23%)	
Working in the emergency area			
Resuscitations in the last 12 months	2.20 $\pm$ 4.74 0 (0 - 2.00)	2.49 $\pm$ 5.49 0 (0 - 2.00)	0.749 ²
Months since the last BLSD training	13.90 $\pm$ 13.30 11.00 (3.00 - 24.00)	14.80 $\pm$ 11.90 12.00 (6.50 - 21.00)	0.533 ²
BLSD instructor	8 (19.51%)	8 (20.51%)	1.000 ¹
Previous VR use	9 (21.95%)	7 (17.95%)	0.867 ¹
Hours of VR use (only users)	2.83 $\pm$ 3.14 2.00 (1.00 - 2.00)	3.04 $\pm$ 3.43 2.00 (1.00 - 3.50)	1.000 ²
Workshop participation	4 (9.76%)	4 (10.26%)	1.000 ³

Procedure

Both sessions lasted about an hour. In each one, participants were asked to train in BLSD using the XR system and to fill in two questionnaires (for a total of four), following the procedure explained below. All data were collected and recorded anonymously. To correctly match S1 and S2 results, a random code was assigned to each participant at the beginning of S1. The code was only kept by participants and was never associated with their identity.

Session 1 The experimenters led each participant to the laboratory, explained the aim of the research, and checked the fulfillment of the inclusion criteria. Participants provided written informed consent for participation and data collection. Then, the experimenters guided participants through S1, explaining each step of the procedure with the help of written instructions to guarantee the standardization of the procedure for each participant.

Each participant was asked to fill in the initial questionnaire, and answers were used to assign the participant to either G1 or G2. To smooth out possible differences in prior knowledge, we gave participants seven minutes to read a summary of the BLSD step-by-step instructions, consistent with the Italian Resuscitation Council Guidelines [74]. Then, the participants first experienced a training consisting in a Tutorial XR-BLSD and a subsequent Test XR-BLSD, both with rt-FB and pp-FB present or absent according to their group. At the end of Test XR-BLSD, the trainees were asked to complete the satisfaction, usability, and usefulness questionnaire. Finally, they were invited to experience a second Test XR-BLSD (post-training), in which no participant received rt-FB because this test was meant to assess performance without any help from the system. At the end of S1, however, both groups received pp-FB, so that they could all complete this session with an evaluation on their performance. A calendar link was provided to allow each trainee to book an appointment for S2, one month after S1.

Session 2 In S2, both groups followed the same procedure. The participants performed a Test XR-BLSD to evaluate the knowledge retained after one month (retention time), followed by a Tutorial XR-BLSD providing both rt-FB and pp-FB. The participants were then asked to complete the satisfaction, usability, and usefulness questionnaire. After filling in the questionnaire, the participants underwent another Test XR-BLSD (post-refresh), at the end of which both groups received pp-FB. Finally, the participants were asked to complete the willingness to use questionnaire.

3.1.4 Results

In descriptive analyses, we expressed continuous variables as mean $\pm$ standard deviation or median and first and third quartiles and categorical variables as absolute frequencies and percentages.

We checked for normality with Kolmogorov–Smirnov test.

To compare continuous variables between independent groups, we used Student's t-test for normally distributed variables and Wilcoxon–Mann–Whitney test otherwise. For categorical variables, we used the chi-square test when no more than 20% of cells

had expected frequencies <5 and no cell had an expected frequency <1 , and Fisher's exact test otherwise.

In the second part of the study, to compare dependent groups on repeated measurements, we used Friedman's test for not normally distributed interval variables and Cochran's Q test for dichotomous variables. To assess paired differences, we used Wilcoxon signed rank test.

Data analysis was performed using R software version 4.0.2 [101] and SAS© software version 9.4 [142].

RQ1: Effectiveness of rt-FB and pp-FB in an XR system for BLSD training

We compared the performance of G1 and G2 at the end of S1 to assess whether there were differences in the results due to the presence of rt-FB and pp-FB during training (Figure 3.3 and All participants columns in Table 3.2). CorrectSteps were similar between G1 and G2 (median 17.00 vs. 16.00, $p=0.472$). However, CorrectCompressions was much higher in G1 than in G2 (46.70% vs. 3.30%; $p=0.005$). Moreover, the DepthDifference was smaller in G1 than in G2 (0 vs 0.36; $p=0.018$) and, consistently, DepthInRange was higher (51.22% vs 30.77%), although the difference between groups, for this variable, did not reach statistical significance ($p=0.103$). No statistically significant difference was found for the other objective measures.

Table 3.2: Comparison between G1 and G2 at Post training (Session 1; all participants) and among Post training, Retention, and Post refresh (completers only). Variables are described as mean±standard deviation, median (first-third quartiles). ¹Wilcoxon-Mann-Whitney test; ²Wilcoxon signed rank sum test; ³Friedman’s test; ⁴Cochran’s Q test; ⁵McNemar test; ⁶Chi square test.

		All participants			Completers					
	Group	Post training G1=41 G2=39	G1 vs G2 p-value	Post training G1=28 G2=22	Retention G1=28 G2=22	Post training vs Retention	Post refresh G1=28 G2=22	Retention vs Post refresh	Post training vs Post refresh p-value	Trend Post training Retention Post refresh
CorrectSteps (out of 19)	G1	15.90±3.04 17.00(15.00-18.00)	0.472 ¹	15.50±3.29 17.00(13.80-18.00)	12.70 ± 3.16 13.00(11.00-15.00)	<0.001 ²	16.80±3.52 18.00(17.00-19.00)	<0.001 ²	0.015 ²	<0.001 ³
	G2	15.60±2.93 16.00(15.00-17.50)		15.50 ± 3.25 16.00(14.30-18.00)	13.60 ± 2.42 14.00(13.00-15.00)	0.008 ²	17.00±1.86 17.50(16.00-18.80)	<0.001 ²	0.006 ²	<0.001 ³
	G1+G2	15.70±2.97 16.00(15.00-18.00)		15.50±3.24 16.50(14.00-18.00)	13.10±2.87 13.50(12.00-15.00)	<0.001 ²	16.90±2.89 18 (16.30-19.00)	<0.001 ²	<0.001 ²	<0.001 ³
Frequency (cpr/min)	G1	109.00±13.40 106.00(110.00-113.00)	0.055 ¹	109.00±13.60 106.00(99.70-113.00)	108.00±10.40 106.00(100.00-111.00)	0.662 ²	107.00±13.30 106.00(102.00-109.00)	0.938 ²	0.973 ²	0.965 ³
	G2	113.00±20.00 109.00(105.00-122.00)		112.00±21.40 110.00(104.00-119.00)	106.00±14.80 105.00(101.00-112.00)	0.039 ²	109.00±7.85 109.00(103.00-113.00)	0.262 ²	0.545 ²	0.170 ³
	G1+G2	111.00±17.00 107.00(102.00-118.00)		110.00±17.40 107.00(102.00-115.00)	107.00±12.40 105.00(101.00-112.00)	0.251 ²	108.00±11.10 107.00(102.00-111.00)	0.385 ²	0.623 ²	0.375 ³
Frequency	G1	3.22±6.80 0 (0-3.76)	0.367 ¹	3.42±6.85 0 (0-4.79)	1.91±3.38 0 (0-2.53)	0.256 ²	2.47±9.04 0 (0-0.25)	0.485 ²	0.224 ²	0.471 ³
Difference (cpr/min)	G2	6.74±12.80 0 (0-6.63)		6.88±14.50 0 (0-4.58)	3.79±8.90 0 (0-1.43)	0.224 ²	0.78±1.88 0(0-0)	0.080 ²	0.014 ²	0.009 ³
	G1+G2	4.94±10.30 0 (0-5.12)		4.94±10.90 0 (0-4.93)	2.74±6.42 0 (0-2.06)	0.105 ²	1.73±6.87 0 (0-0)	0.089 ²	0.011 ²	0.016 ³
Frequency	G1	26 (63.41%)	0.681 ⁶	17 (60.71%)	17 (60.71%)	1.000 ⁵	21 (75.00%)	0.289 ⁵	0.427 ⁵	0.390 ⁴
In Range	G2	22 (56.41%)		13 (59.09%)	15 (68.18%)	0.724 ⁵	17 (77.27%)	0.617 ⁵	0.221 ⁵	0.264 ⁴
	G1+G2	48 (60.00%)		30 (60.00%)	32 (64.00%)	0.823 ⁵	38 (76.00%)	0.149 ⁵	0.118 ⁵	0.135 ⁴
	G1	4.89±1.10	0.246 ¹	4.77±1.16	4.69±1.23	1.000 ²	5.00±1.00	0.063 ²	0.099 ²	0.062 ³
Continued on next page										

Continued on next page

Table 3.2: Comparison between G1 and G2 at Post training (Session 1; all participants) and among Post training, Retention, and Post refresh (completers only). Variables are described as mean±standard deviation, median (first-third quartiles). ¹Wilcoxon-Mann-Whitney test; ²Wilcoxon signed rank sum test; ³Friedman’s test; ⁴Cochran’s Q test; ⁵McNemar test; ⁶Chi square test.

	All participants				Completers					
	Group	Post training G1=41 G2=39	G1 vs G2 p-value	Post training G1=28 G2=22	Retention G1=28 G2=22	Post training vs Retention p-value	Post refresh G1=28 G2=22	Retention vs Post refresh p-value	Post training vs Post refresh p-value	Trend Post training Retention Post refresh
TooDeep	G2	21.40±33.20 0 (0-46.60)		27.70±34.50 6.65(0-52.10)	32.30±39.40 8.35(0-76.70)	0.795 ²	32.10±31.20 20.00(4.98-55.00)	0.872 ²	0.286 ²	0.326 ³
	G1+G2	18.80±31.20 0 (0-22.50)		19.60±30.50 0 (0-35.00)	23.70±34.90 3.30(0-50.30)	0.568 ²	22.90±30.00 8.35(0-32.50)	0.577 ²	0.113 ²	0.053 ³
Incomplete Release	G1	8.12±19.70 0 (0-3.30)	0.306 ¹	8.21±21.00 0 (0-3.30)	7.26±17.00 0 (0-6.70)	1.000 ²	9.76±20.30 0 (0-10.00)	0.407 ²	0.637 ²	0.539 ³
	G2	12.70±26.90 0 (0-8.35)		17.60±33.10 1.65(0-12.50)	13.80±26.9 0 (0-8.33)	0.462 ²	9.55±21.20 0 (0-5.85)	0.332 ²	0.045 ²	0.180 ³
	G1+G2	10.30±23.50 0 (0-4.15)		12.40±27.10 0 (0-3.30)	10.10±21.9 0 (0-6.70)	0.673 ²	9.66±20.5 0 (0-6.70)	0.882 ²	0.356 ²	0.639 ³
	G1	39.20±32.50 46.70(3.30-66.70)	0.005 ¹	37.40±32.70 40.00(2.48-66.70)	27.30±31.20 16.70(0-45.80)	0.134 ²	42.10±30.50 45.00(15.00-67.50)	0.023 ²	0.509 ²	0.376 ³
Correct Compressions	G2	18.80±26.60 3.30(0-30.00)		27.60±30.90 15.00(0-44.40)	28.90±29.50 18.40(0-59.20)	0.542 ²	41.20±29.00 45.00(12.50-70.00)	0.236 ²	0.230 ²	0.539 ³
	G1+G2	29.20±31.30 16.70(0-54.20)		33.10±32.00 23.30(0-59.20)	28.00±30.20 16.70(0-51.70)	0.468 ²	41.70±29.50 45.00(11.70-70.00)	0.012 ²	0.188 ²	0.356 ³

Table 3.3: Comparison between completers and non completers, by group, in the initial questionnaire, post-training measures, and USE questionnaire scores. Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹Chi-square test; ²Wilcoxon rank sum test; ³Fisher's exact test.

		Completers (G1: n=28) (G2: n=22)	Non completers (G1: n=13) (G2: n=17)	Completers vs non completers p-value
INITIAL QUESTIONNAIRE				
Gender				
Male	G1	13 (46.43%)	6 (46.15%)	1.000 ¹
	G2	12 (54.54%)	7 (41.18%)	0.613 ¹
Female	G1	15 (53.57%)	7 (53.84%)	1.000 ¹
	G2	10 (45.46%)	10 (58.82%)	0.613 ¹
Age	G1	33.80 $\pm$ 9.64 30.00 (27.80-35.30)	32.20 $\pm$ 6.54 30.00 (29.00-32.00)	0.877 ²
	G2	35.50 $\pm$ 10.50 31.00 (28.00-38.30)	34.80 $\pm$ 9.18 33.00 (30.00-35.00)	0.776 ²
Job role				
Nurse	G1	8 (28.57%)	5 (38.46%)	0.720 ¹
	G2	4 (18.18%)	8 (47.06%)	0.082 ³
Doctor	G1	20 (71.43%)	8 (61.54%)	0.720 ¹
	G2	18 (81.82%)	9 (52.94%)	0.082 ¹
Working in the emergency area	G1	9 (32.14%)	6 (46.15%)	0.604 ¹
	G2	5 (22.73%)	10 (58.82%)	0.049 ¹
Resuscitations in the last 12 months	G1	1.04 $\pm$ 2.13 0 (0-2.00)	4.69 $\pm$ 7.40 0 (0-5.00)	0.148 ²
	G2	0.91 $\pm$ 3.26 0 (0-0)	4.53 $\pm$ 7.05 1.00 (0-4.00)	0.010 ²
Months since last BLSD training	G1	15.90 $\pm$ 14.90 11.50 (4.50-24.00)	9.54 $\pm$ 7.71 7.00 (2.00-15.00)	0.285 ²
	G2	16.40 $\pm$ 12.50 12.00 (10.30-22.50)	12.90 $\pm$ 11.10 11.00 (5.00-15.00)	0.340 ²
BLSD instructor	G1	6 (21.43%)	2 (15.39%)	1.000 ¹
	G2	2 (9.09%)	6 (35.29%)	0.059 ¹
Previous VR use	G1	6 (21.43%)	3 (23.08%)	1.000 ¹
	G2	5 (22.73%)	2 (11.77%)	0.438 ¹
Hours of VR use (only users)	G1	0.73 $\pm$ 2.17 0 (0-0)	0.39 $\pm$ 0.77 0 (0-0)	0.923 ²
	G2	0.86 $\pm$ 2.34 0 (0-0)	0.13 $\pm$ 0.49 0 (0-0)	0.341 ²
Workshop participation	G1	4 (14.29%)	0 (0%)	0.288 ¹
	G2	2 (9.09%)	2 (11.77%)	1.000 ¹
POST-TRAINING MEASURES				
CorrectSteps (out of 19)	G1	15.50 $\pm$ 3.29, 17.00 (13.80-18.00)	16.70 $\pm$ 2.29, 17.00 (16.00-18.00)	0.295 ²
	G2	15.50 $\pm$ 3.25 16.00 (14.30-18.00)	15.80 $\pm$ 2.54 16.00 (15.00-17.00)	0.932 ²
Frequency (cpr/min)	G1	109.00 $\pm$ 13.60 106.00 (99.70-113.00)	108.00 $\pm$ 13.50 104.00 (101.00-112.00)	0.945 ²
	G2	112.00 $\pm$ 21.40 110.00 (104.00-119.00)	114.00 $\pm$ 18.70 108.00 (106.00-124.00)	0.585 ²
FrequencyDifference (cpr/min)	G1	3.42 $\pm$ 6.85 0 (0-4.79)	2.80 $\pm$ 6.94 0 (0-1.79)	0.559 ²
	G2	6.88 $\pm$ 14.50 0 (0-4.58)	6.57 $\pm$ 10.60 0 (0-9.35)	0.708 ²
FrequencyInRange	G1	17 (60.71%)	9 (69.23%)	0.858 ¹
	G2	13 (59.09%)	9 (52.94%)	0.953 ¹

Continued on next page

Table 3.3: Comparison between completers and non completers, by group, in the initial questionnaire, post-training measures, and USE questionnaire scores. Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹Chi-square test; ²Wilcoxon rank sum test; ³Fisher's exact test.

		Completers (G1: n=28) (G2: n=22)	Non completers (G1: n=13) (G2: n=17)	Completers vs non completers p-value
Depth (cm)	G1	4.77 $\pm$ 1.16 5.14 (4.43-5.44)	5.17 $\pm$ 0.92 5.10 (5.05-5.75)	0.298 ²
	G2	5.02 $\pm$ 1.08 5.33 (4.22-5.78)	3.63 $\pm$ 1.54 3.43 (2.54-4.67)	0.006 ²
DepthDifference (cm)	G1	0.54 $\pm$ 0.94 0.06 (0-0.57)	0.29 $\pm$ 0.61 0 (0-0.16)	0.313 ²
	G2	0.48 $\pm$ 0.71 0 (0-0.78)	1.58 $\pm$ 1.23 1.57 (0.33-2.46)	0.002 ²
DepthInRange	G1	13 (46.43%)	8 (61.54%)	0.506 ¹
	G2	10 (45.46%)	2 (11.77%)	0.037 ³
TooWeak	G1	47.70 $\pm$ 39.90 35.00 (9.18-87.50)	33.60 $\pm$ 36.20 30.00 (0-46.70)	0.249 ²
	G2	40.30 $\pm$ 43.60 13.50 (4.15-95.00)	79.20 $\pm$ 38.30 100 (73.30-100)	0.018 ²
TooDeep	G1	13.20 $\pm$ 25.80 0 (0-12.50)	22.80 $\pm$ 36.00 0 (0-20.00)	0.306 ²
	G2	27.70 $\pm$ 34.50 6.65 (0-52.10)	13.30 $\pm$ 30.50 0 (0-0)	0.052 ²
IncompleteRelease	G1	8.21 $\pm$ 21.00 0 (0-3.30)	7.95 $\pm$ 17.30 0 (0-3.30)	0.666 ²
	G2	17.60 $\pm$ 33.10 1.65 (0-12.50)	6.27 $\pm$ 14.40 0 (0-6.70)	0.604 ²
CorrectCompressions	G1	37.40 $\pm$ 32.70 40.00 (2.50-66.70)	43.10 $\pm$ 33.10 53.30 (10.00-70.00)	0.682 ²
	G2	27.60 $\pm$ 30.90 15.00 (0-44.40)	7.46 $\pm$ 13.70 0 (0-6.70)	0.023 ²
USE QUESTIONNAIRE				
Usefulness score	G1	5.63 $\pm$ 1.07 5.75 (4.94-6.44)	5.48 $\pm$ 1.29 5.63 (4.63-6.63)	0.855 ²
	G2	5.22 $\pm$ 1.19 5.56 (4.63-6.00)	5.57 $\pm$ 1.22 6.00 (4.88-6.38)	0.228 ²
Ease of use score	G1	5.82 $\pm$ 0.98 6.18 (5.27-6.64)	5.54 $\pm$ 1.01 5.36 (5.00-6.36)	0.369 ²
	G2	5.55 $\pm$ 1.04 5.73 (5.30-6.32)	6.01 $\pm$ 0.80 6.36 (5.45-6.45)	0.100 ²
Ease of learning score	G1	6.08 $\pm$ 1.07 6.50 (5.44-6.75)	6.19 $\pm$ 0.82 6.50 (5.75-7.00)	0.943 ²
	G2	5.76 $\pm$ 1.06 5.88 (5.31-6.69)	6.43 $\pm$ 0.62 6.50 (6.00-7.00)	0.036 ²
Satisfaction score	G1	6.04 $\pm$ 0.75 6.07 (5.54-6.61)	6.10 $\pm$ 1.02 6.57 (5.43-7.00)	0.499 ²
	G2	5.73 $\pm$ 1.14 6.00 (5.32-6.54)	5.79 $\pm$ 1.11 6.00 (5.14-6.71)	0.910 ²
Overall score	G1	5.85 $\pm$ 0.88 6.02 (5.25-6.54)	5.74 $\pm$ 0.97 6.07 (5.20-6.43)	0.695 ²
	G2	5.53 $\pm$ 1.00 5.63 (5.18-6.21)	5.90 $\pm$ 0.86 6.30 (5.20-6.50)	0.202 ²



Figure 3.3: Comparisons of the measures between G1 and G2 at the end of S1. Capped vertical bars indicate the standard error for non-dichotomous measures. The horizontal grey bar in the “Frequency (cpr/min)” and “Depth (cm)” graphs indicate the correct ranges. The * and *** signs indicate differences with p-values < 0.05 and < 0.005 respectively.

RQ2: Knowledge retention after one month

Between June and November 2022, 50 of 80 recruited participants attended S2 (time elapsed since S1: median 35 days; first and third quartile 32 and 42 days, respectively): 28 out of 41 in G1 and 22 out of 39 in G2, with similar dropouts (31.71% and 43.59%, respectively; chi-square $p=0.386$). However, the comparison between completers and non-completers in each group showed a differential dropout in G1 and G2 (Table 3.3). In fact, while there were no statistically significant differences between completers and non-completers in G1, non-completers from G2 were more likely than completers to work in the emergency area (58.82% vs 22.73%; $p=0.049$) and performed more resuscitations in the previous year (median 1.00 vs 0; $p=0.010$), but had poorer results in terms of Depth (median 3.43 cm vs 5.33; $p=0.006$), DepthDifference (median 1.57 cm vs 0; $p=0.002$), DepthInRange (11.77% vs 45.46%; $p=0.037$), TooWeak (100% vs 13.50%; $p=0.018$), and CorrectCompressions (0% vs 15.00%; $p=0.023$).

Due to differential dropout, we cannot attribute differences between groups in S2 to receiving rt-FB and pp-FB or not during training in S1, so we did not compare groups in S2; on the other hand, since training was different, G1 and G2 cannot be simply considered as a whole. Thus, in the following sections, we decided to report both pooled and separate results for each group.

In S2, we measured performance during the first XR-BLSD and after refresh training.

The comparison between the results of S1 and those of the first XR-BLSD of S2 (Figure 3.4 and Completers columns in Table 3.2) showed a natural decay in CorrectSteps when G1 and G2 were analyzed separately (G1 median: 17.00 vs 13.00; $p<0.001$; G2 median: 16.00 vs 14.00; $p=0.008$) and pooled (median: 16.50 vs 13.50; $p<0.001$). A decrease in Frequency between post-training and retention was only found in G2 (110.00 vs 105.00; $p=0.039$), while a decay in CorrectCompressions was only found in G1 and was not statistically significant (40.00% vs 16.70%; $p=0.134$). No statistically significant difference was found for the other objective measures.

RQ3: Effectiveness of a refresh training session with rt-FB and pp-FB

Friedman's and Cochran's tests analyses revealed statistically significant differences between the three measurements over time for CorrectSteps (G1: $p<0.001$; G2: $p<0.001$; pooled: $p<0.001$), FrequencyDifference (G2: $p=0.009$; pooled: $p=0.016$), Depth (pooled: $p=0.024$), DepthDifference (G1: $p=0.042$; pooled: $p=0.003$), DepthInRange (G2: $p=0.020$, pooled: $p=0.020$).

The comparison between the retention and post-refresh measurements performed during S2 (Figure 3.4 and Completers columns in Table 3.2) showed that refresh training was effective in improving CorrectSteps (G1 median: 13.00 vs 18.00, $p<0.001$; G2 median: 14.00 vs 17.50, $p<0.001$; pooled median: 13.50 vs 18.00, $p<0.001$), Depth (pooled: 5.29 vs 5.43, $p=0.037$), DepthDifference (G1: 0.31 vs 0, $p=0.007$; G2: 0.98 vs 0, $p=0.033$; pooled: 0.13 vs 0, $p<0.001$), DepthInRange (G2: 40.91% vs 72.73%, $p=0.045$; pooled: 40.00% vs 60.00%, $p=0.024$). Despite Friedman's tests were not significant, we also found a noticeable difference in TooWeak (G1: 68.30 vs 33.40, $p=0.041$; pooled: 28.40 vs 11.70, $p=0.021$) and CorrectCompressions (G1: 16.70% vs 45.00%, $p=0.023$; pooled: 16.70% vs 45.00%, $p=0.012$).

Moreover, the comparison between the post-training and post-refresh measurements

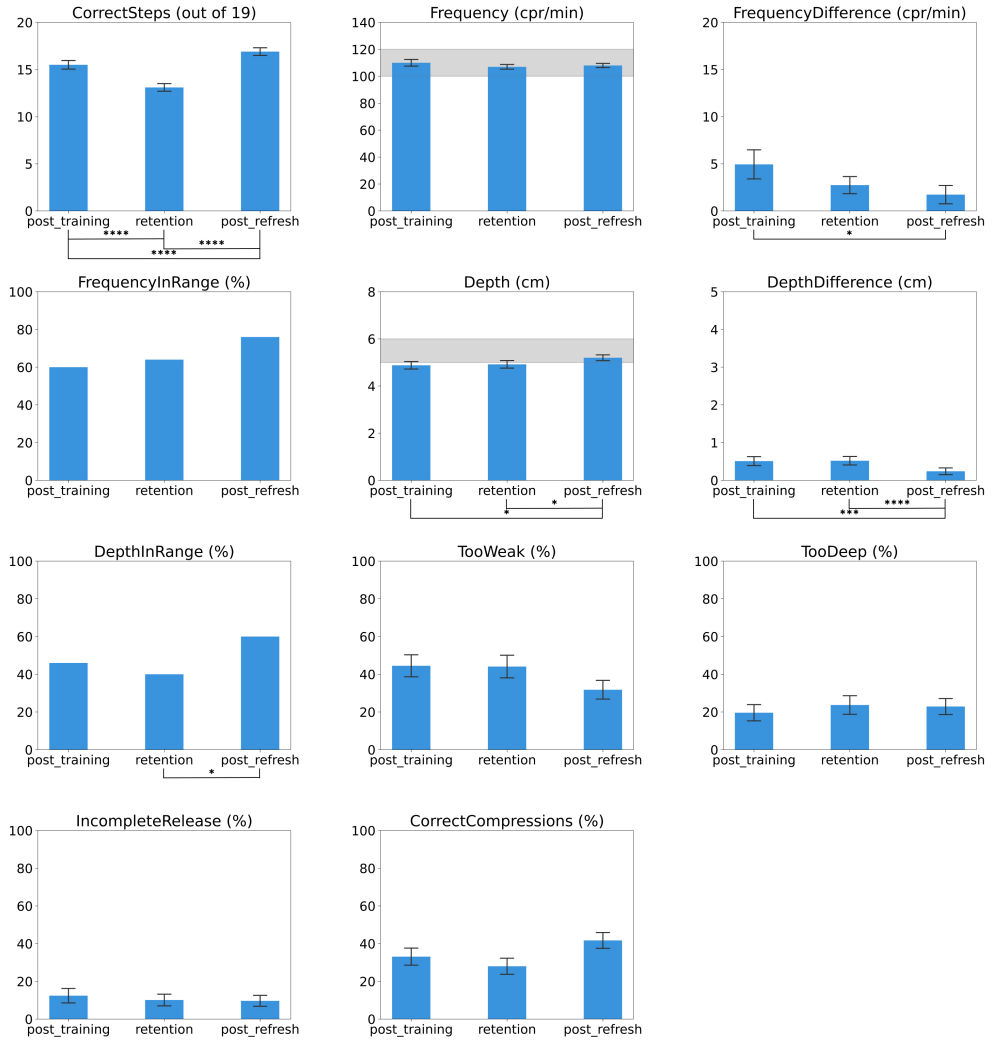


Figure 3.4: Comparisons of the measures for all completers between post_training, retention and post_refresh. Capped vertical bars indicate the standard error for non-dichotomous measures. The horizontal grey bar in the “Frequency (cpr/min)” and “Depth (cm)” graphs indicate the correct ranges. The *, ***, and **** signs indicate differences with p-values < 0.05 , < 0.005 , and ≤ 0.001 respectively.

(Figure 3.4 and table 3.2) showed that the refresh training with the XR system further improved knowledge and skills with respect to the training in S1 in terms of CorrectSteps (G1 median: 17.00 vs 18.00, $p=0.015$; G2 median: 16.00 vs 17.50, $p=0.006$; pooled median: 16.50 vs 18.00, $p<0.001$), FrequencyDifference (G2: 0 vs 0, $p=0.014$, pooled: 0 vs 0, $p=0.011$), Depth (pooled median: 5.26 vs 5.43, $p=0.023$), and DepthDifference (G1: 0.06 vs 0, $p=0.037$; pooled: 0.07 vs 0, $p=0.004$). Despite the Friedman's tests were not significant, we found a noticeable difference in DepthDifference also in G2 (0.07 vs 0, $p=0.044$), as well as a difference in TooWeak (pooled: 26.70 vs 11.70, $p=0.048$) and IncompleteRelease (G2: 1.65 vs 0, $p=0.045$).

RQ4: Participants' perception of the XR system

At S1, users described the XR system as useful (median 5.75), easy to use (6.05), easy to learn (6.50) and satisfying (6.00). The median overall USE score was 6.08. There were no statistically significant differences between G1 and G2 for the overall USE and its subscales. Even better results were reported by completers at S2 when compared to S1 in terms of overall score (G2: 5.98 vs 5.63, $p=0.022$; pooled: 6.00 vs 5.82, $p=0.005$), Usefulness score (G2: 6.00 vs 5.56, $p=0.004$; pooled: 6.06 vs 5.69, $p=0.001$), Ease of Use score (pooled: 6.00 vs 5.95, $p=0.02$), Ease of Learning score (G1: 6.63 vs 6.50, $p=0.023$; pooled: 6.25 vs 6.13, $p=0.017$). Only the increase in Satisfaction score was not statistically significant. All data about each subscale of USE for G1, G2, and pooled at S1 and S2 are available in Table 3.3.

In S2, most participants assigned the highest score (median 7.00) to the willingness to use for professional reasons (86.00% agree or strongly agree) and personal improvement (92.00% agree or strongly agree), at home (82.00% agree or strongly agree) and in their workplace (86.00% agree or strongly agree). Most participants would use it again in six months (76.00% strongly agree, 14.00% agree) and would gladly use similar XR applications to improve other knowledge/skills (76.00% strongly agree, 20.00% agree). All data about each item of the questionnaire for G1, G2, and pooled are available in Table 3.3.

3.1.5 Discussion

To assess whether the combination of rt-FB and pp-FB is effective for BLSD training with an XR system (RQ1), we compared two groups: one that received both types of feedback and the other that did not receive these types of feedback during initial training. Participants who received rt-FB and pp-FB during initial training performed compressions significantly better during assessment, consistently with the effect of rt-FB reported by Augusto et al. [21]. The statistically significant positive effect of rt-FB and pp-FB was confirmed also for DepthDifference, which was smaller in G1 than in G2, a result that is in line with found by Augusto et al. [21]. Also FrequencyInRange was higher when rt-FB and pp-FB were present during initial training, although the difference was not statistically significant. A similar but statistically significant result was reported by a study on healthcare professionals receiving rt-FB [21]. On the contrary, the two groups showed a similar number of correct steps. This can be partly explained because, in our XR system, both rt-FB and pp-FB provide information on compression depth and frequency, and the former allows the trainees to adapt their ac-

tions according to the measurements displayed while they are performed. Instead, only pp-FB conveys information on steps' correctness, and such information can be used to improve procedure performance in the following sessions. Since pp-FB, by its nature, is provided when the simulation is over, it might be overlooked if trainees do not take their time to examine the summary assessment. Nevertheless, we can accept our H1 hypothesis because participants who received rt-FB and pp-FB in the initial training phase performed better compressions than participants who did not receive it.

At S2, 37.5% of participants dropped out. Although a similar dropout occurred and completers were still balanced between G1 and G2 for the variables collected through the initial questionnaire, a difference between participants who dropped out in the two groups was still a concern [30]. In fact, the comparison between participants attending only S1 and completers revealed that, while in G1 the differences between completers and non-completers were not statistically significant, the worst performers dropped out in G2: G2's non-completers performed worse than G2's completers as for Depth, DepthDifference, DepthInRange, TooWeak, and CorrectCompressions. Surprisingly, in the same group, non-completers were more likely to work in the emergency area than completers and consistently had performed more resuscitations in the previous 12 months. It may be hypothesized that those performing worse in G2 could have decided not to participate in S2 because of the lack of feedback, which prevented them from ascertaining their need for improvement. This could have been further underrated by those working in the emergency area, who were more likely to abandon the study. Moreover, not having their performance measured and fed back could have caused those in G2 to lose interest in the XR system, so only the more motivated participants could have remained in the study.

To assess RQ2, we measured performance decay one month after the initial training and observed similar results in both groups. As expected, completers showed a statistically significant decay in the number of correct steps, consistently with the results reported in a study conducted on medical students [228], and confirming our H2 hypothesis. Performance also worsened for Depth and Frequency, as well as for CorrectCompressions (from 23.30% to 16.70%), although differences did not reach statistical significance or were statistically significant only in one group, possibly due to the small sample size and the short time lapse between the two measurements. Even in a 3-month retention study conducted on nursing students, correct compressions did not differ between post-training and 3 months later, but the performance was much higher [16] than in our study. Moreover, a study on 6-month retention among health sciences students, training in the VE using a mannequin providing rt-FB, showed a decay at correct hand positioning, median compression depth, correct compressions, and correct compression frequency [51].

Considering RQ3, we assessed whether a refresh training session with rt-FB and pp-FB is effective in restoring knowledge and skills after the decay. Refresh training was effective in terms of number of correct steps, Depth, DepthDifference, and DepthInRange, confirming our H3 hypothesis. Notably, these measures improved even when compared to post-training: correct steps, FrequencyDifference, Depth, and DepthDifference. A study using a mannequin with a device providing pp-FB on a screen showed an improvement in the mean compression depth and frequency after retraining [152].

Considering subjective measures (RQ4), we did not find statistically significant dif-

ferences in overall score and subscales of USE between the two groups. Indeed, scores were positive in both groups, showing that the proposed XR system was perceived as useful, easy to learn, easy to use, and satisfactory, regardless of the presence of rt-FB and pp-FB. It is however interesting to note that in an item of the Usefulness subscale, we found a statistically significant difference favoring G1: the system with rt-FB and pp-FB feedback was perceived as more helping to be more effective. It is also interesting to note that in S2, the system was perceived as more useful, easy to learn, and easy to use than in S1, leading also to a statistically significant difference in the overall USE score. This can be likely due to two reasons: i) during S2 both groups received rt-FB and pp-FB, and the introduction of feedback could have motivated G2 participants to give higher ratings; ii) all participants could have become more familiar with the system and appreciated its usefulness also as a frequent retraining tool. The latter motivation is supported by the very positive scores in the Willingness to use questionnaire. The large majority of participants said they were willing to use the system in their workplace or at home for both professional reasons and personal improvement. In particular, 90% of them agreed on using it every six months, and almost all participants were open to using VR to improve other knowledge/skills. This result is of particular importance, given that knowledge and skills naturally decay if they are not refreshed [17, 79].

A limitation of the study is the lack of a baseline BLS D performance measurement, which would enable us to completely exclude that differences among groups are due to pre-experimental different skills. However, groups were balanced for many subject characteristics, including age, which may influence both the performance [12] and the effect of feedback [209], and gender, a known predictor of CPR performance [12]); thus, a selection bias can reasonably be excluded. Another limitation is that the Willingness To Use questionnaire was not validated, and the USE questionnaire was not validated in its Italian version.

3.1.6 Conclusion

We evaluated an immersive XR system for BLS D training conducting a study on doctors and nurses. To the best of our knowledge, this is the first study to combine rt-FB and pp-FB in the context of an immersive XR system to teach BLS D. The study also observed the effects of training after one month and the effects of a refresh training session with the XR system, as well as participants' subjective opinions.

We showed that rt-FB and pp-FB as well as refresh training are useful in improving and maintaining BLS D performance among nurses and doctors. Since, even after retraining, the median percentage of correct compressions was only 45.00%, with one quarter of participants performing 11.70% or less correct compressions, systems that allow for frequent retraining and provide feedback, as the one we described, are fundamental to maintain CPR skills. For this reason, a future version of the system could invite trainees to repeat the simulation until a satisfactory performance level is reached, according to objective and quantitative measurements. The positive willingness to use the XR system reported by participants who completed both sessions is promising to support its adoption as a regular retraining tool.

Another possible future work could consist in nesting pp-FB inside a replay of the simulation so that the trainee could possibly be more engaged and not skip pp-FB. In such replay, the whole performance should be described with a positive approach,

highlighting correct behaviors and giving advice about how to correct inaccurate or ineffective actions.

Additional hints for future work came from the participants' feedback in the final box of the last questionnaire: a participant suggested to add the count of the compressions as an additional type of rt-FB, another to enable speech interaction with the virtual characters, one to allow performing the first three steps (self-protection and safety) in any order, two to offer the simulation also to lay people, and three to use similar systems also for training in other scenarios or procedures (e.g., advanced life support).

3.2 Acceptance and Effectiveness of a Multi-user XR System for STEMI Training

This section presents a multi-user VR application designed for training teams in the rescue of patients experiencing a STEMI (ST-Elevation Myocardial Infarction) heart attack. The system underwent two evaluations: a qualitative study capturing the subjective perceptions of emergency professionals, and a comparative study measuring training efficacy between a VR-trained group and a traditional control group.

3.2.1 Background

STEMI is a life-threatening acute coronary syndrome caused by complete thrombotic occlusion of a coronary artery branch. It is characterized by a distinct electrocardiographic pattern and represents a true time-dependent emergency [71]. Timely initiation of appropriate reperfusion therapies (e.g., primary percutaneous coronary intervention, stenting, or thrombolysis) is essential to improve prognosis and limit irreversible myocardial damage [162]. Early diagnosis is therefore a critical determinant of outcomes, and current European Society of Cardiology guidelines recommend that, when an acute coronary syndrome is suspected, an ECG is performed and interpreted by an experienced clinician within 10 minutes of first medical contact [44].

To address such time-critical emergencies, healthcare authorities establish action plans and procedures and implement targeted training programs. For example, the Friuli Venezia Giulia region in Italy published a document in 2019 and updated it in 2022² that describes the procedures to adopt for rescuing, diagnosing, providing clinical assistance, and providing cardiologic rehabilitation of patients with STEMI.

Currently, nurses and doctors in emergency medical services usually follow Advanced Life Support (ALS) courses, consisting of theoretical and practical content and a final exam [94]. These courses provide specific skills to deal with cardiac arrest and peri-arrest situations, manage resuscitated patients until their transfer to an intensive care setting, and communicate effectively with the other rescue team members and the patient's family [94]. However, specific training courses to teach how to deal with STEMI patients are rare and based on role-playing simulations where different team members pose as STEMI patients, family members, or healthcare providers. Unfortunately, training experiences conducted by instructors suffer from high variability, issues in replicability,

²https://www.regione.fvg.it/rafvig/export/sites/default/RAFVG/salute-sociale/sistema-sociale-sanitario/FOGLIA53/allegati/11022019_PDTA_STEMI-FEB.pdf

and subjective evaluation [120, 157, 165, 262, 293, 328, 329].

As shown in Section 2.6, VR has been found to be a valuable tool in medical and nursing education, offering opportunities for standardized training and objective evaluation [72, 169, 253, 338], and its usage in the broader healthcare field is expanding [2, 35, 254].

Notably, in the context of pre-hospital emergency care training for nurses, current VR applications are focused primarily on simulating minor accident scenes and disaster response scenarios [184]. Consequently, further research is needed in other time-critical areas, one of which is the STEMI pre-hospital management. To the best of our knowledge, no VR system has been yet proposed to train nurses and doctors in the STEMI pre-hospital management.

This section presents a VR training system used to reproduce scenarios representing patients with out-of-hospital STEMI heart attacks. The system engages two trainees simultaneously: an emergency nurse who manages the patient in the VE, and a case manager (CM) doctor in the real world. The CM can either be a cardiologist working in a coronary care unit or an emergency room doctor: they evaluate the clinical information and electrocardiogram received by the nurse, managing the patient's pathway from diagnosis to treatment. The proposed VR system uses an A&F strategy [144]. This aims to encourage healthcare professionals to improve clinical practice by correcting behaviors not meeting standards.

3.2.2 The Proposed VR System for STEMI Training

The proposed system simulates the rescue operation of a STEMI patient in several scenarios characterized by different anamnestic data and vital parameters. The system comprises three applications used by one instructor and two trainees simultaneously:

- The course instructor uses a tablet-based Instructor Application to control different aspects of the simulation in real time (e.g., changing vital parameters).
- The case manager doctor uses a tablet-based Case manager Application that provides a phone call communication with the nurse, and displays the data sent by the nurse during the simulation (e.g., patient's electrocardiogram).
- The nurse wears a VR headset (Meta Quest 2)³ in which a Nurse Application (NA) displays the VEs.

Instructor Application

The Instructor Application manages the entire simulation. In addition to serving as a server for communication between the various devices, its functionalities include:

- Loading premade scenarios.
- Real-time modification of the patient's vital parameters during the simulation.
- Managing the loading of the different scenes.

³<https://www.meta.com/it/en/quest/products/quest-2/>

- Storing the actions performed by the nurse during the simulation.

Before the simulation starts and during its execution, the instructor can change the patient's vital parameters by using the graphical interface in the application, as shown in Fig. 3.5. The interface is divided into four parts: i) the left side contains the log of every action the nurse had performed in the simulation; ii) the center part contains the anamnestic data of the patient for the simulation; iii) the right side contains the current state of the patient that can be updated in real-time by the trainer; iv) the bottom part let the instructor load the different VEs.

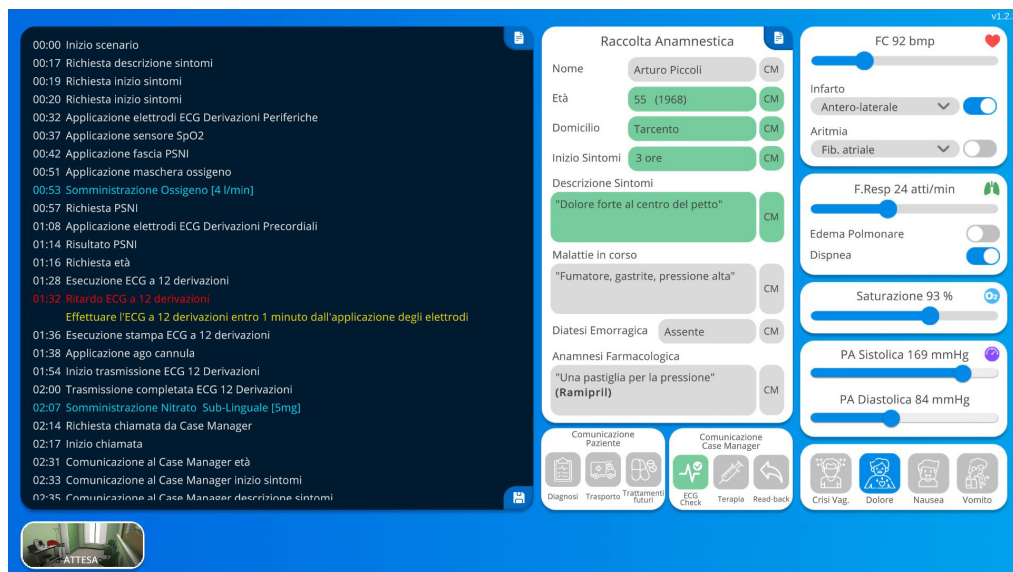


Figure 3.5: The user interface of the Instructor Application, localized in the language of course instructors.

The anamnestic data of the patient comprise name, age, domicile, time of the beginning of the symptoms, description of symptoms, current illnesses, presence of bleeding diathesis, and which medications have already been taken by the patients. The patient's vital parameters are heart rate, presence and type of infarction, presence and type of arrhythmia, respiratory rate, presence of pulmonary edema or dyspnea, blood saturation, blood and systolic pressure, presence of vagal seizure, presence of pain, and presence of nausea and vomiting.

As the instructor changes the patient's status, it is transmitted to the simulation running in the VR headset, and the virtual patient changes according to the updated values of the vital parameters. Fig. 3.9 shows an example of changes in the virtual patient's status.

The Instructor Application is also responsible for storing all the actions performed by the nurse, changes in the patient's vital parameters by the instructor, details on calls made between the case manager doctor and nurse (timestamp and the subject of request events, call initiation and termination), and communication events between nurse, patient, and case manager doctor indicated by the instructor.

During scenario simulation, feedback messages are displayed in the control app corresponding to errors the nurse performs during the rescue. From the analysis of the generated log file, the application generates a final report that identifies whether specific actions necessary for the proper execution of a rescue intervention were performed correctly.

Case Manager Application

The case manager is the medic responsible for the management of the patients. They can talk with the nurse via a phone call and receive the data gathered by the nurse in the simulation. The data concerns vital information about the patient as the 12-derivations ECG track, blood pressure, and heart rate. Figure 3.6 shows the interface of the Case Manager application. It shows the patient’s data mentioned above and two buttons to handle the call with the nurse (answer the phone call and mute microphone).

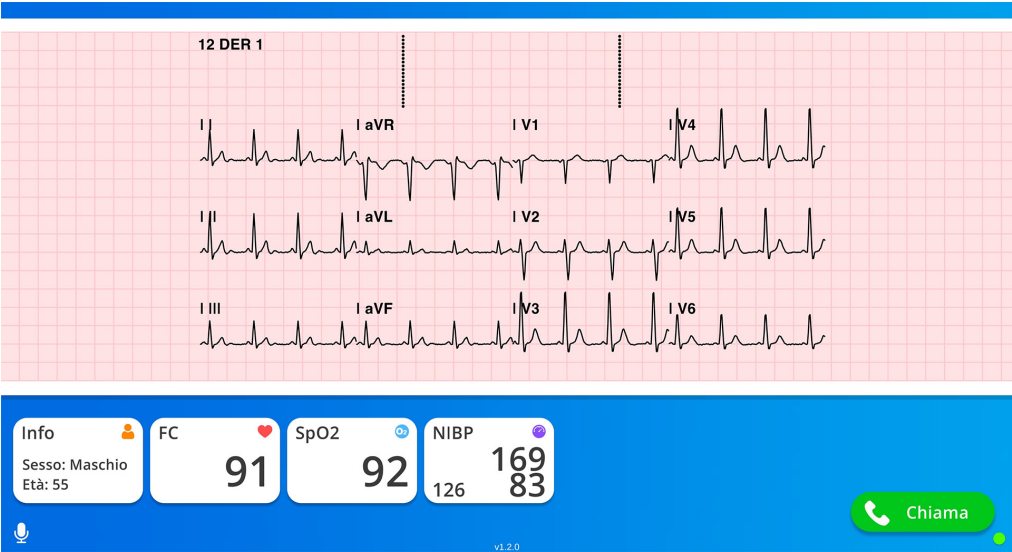


Figure 3.6: The user interface of the Instructor Application, localized in the language of course instructors.

Nurse Application

The Nurse Application (NA) has three Virtual Environments (VE) and supports different interactions with the objects in the VE and the patient.

VEs in the NA The three VEs (Fig. 3.7) represent the sequence of real-world environments encountered during a STEMI rescue: i) the crew room, ii) the ambulance, and iii) the patient’s house.

- *Crew Room* In this VE (Fig. 3.7a), the nurse can become familiar with the VR headset and hand tracking. The VE includes a table with the medical equipment

to be used in the training scenarios (monitor with defibrillator, medication bag, smartphone, oxygen tank). In this scenario, the trainee does not use yet the equipment.

- *Ambulance* In this VE (Fig. 3.7b), the nurse receives briefing information while reaching the emergency scene.
- *Patient's house* In this VE (Fig. 3.7c), the nurse acts on the patient. Emergency equipment is placed near the nurse to minimize the need for walking around.



Figure 3.7: The user interface of the Instructor Application, localized in the language of course instructors.

Interactions The interactions (Fig. 3.8) supported by the Nurse Application exploit both head and hands tracking. The simulation was designed to allow for a natural interaction with the VE. The nurse can interact with the patient and the objects in the living room VE in four different ways:

- *Movement.* The nurse can move freely by walking directly to the desired location. We minimized the need for movement, ensuring that a simple rotation allows to reach objects and interact with them.
- *Proximity.* Some objects are automatically activated when the nurse's hand is close (e.g., when it approaches the pouches at the side of the monitor, they display their contents as in Fig. 3.8a) to simplify interaction with objects, allowing the nurse to focus on reasoning and learning about the medical procedures.
- *Grasping.* Objects can be grasped and moved by approaching them with the hand and performing the pinch gesture. Regardless of the object's size, when the fingers come in contact, the object is grasped and can be moved to the desired position. When the nurse moves the two fingers apart, the object is released. Some objects are automatically released if they are moved close to specific interaction points (e.g., the face mask in Fig 3.8b automatically sticks to the patient's face).
- *Press.* The monitor buttons can be pressed using the forefinger, as in the real world. The button lights up when the finger is sufficiently close (Fig 3.8c), indicating that continuing the movement will complete the interaction.

Interactions with Objects The nurse can use all the tools typically used during a rescue:

- *Monitor with defibrillator (LIFEPAK 15)*⁴ to monitor the patient's vital parameters. Several sensors (SpO2, band for noninvasive pressure measurement, peripheral and precordial leads and electrodes for electrocardiogram measurement, plates for external electrostimulation) are available in two side pouches attached to the monitor and must be applied to the patient to display his vital parameters. To allow for proper functionality during the rescue, the relevant monitor's functions are reproduced during the simulation: manual pressure measurement and setting automatic pressure measurement in preset time intervals; calculation of the 12-lead electrocardiogram and printing of the paper report; sending the electrocardiogram to the CM; administration of resuscitation discharges; and activation of external pacing. Fig. 3.8a and 3.8c show examples of interaction with the monitor.
- *Oxygen cylinder and mask.* A mask connected to an oxygen cylinder can be attached to the patient's mouth to administer oxygen (Fig. 3.8b), regulating its flow through an interface appearing at the cylinder's regulating valve.
- *Medication bag.* Pills, intravenous drugs, and intravenous bags can be administered to the patient during the simulation. All medications are contained in the medication bag (Fig. 3.8d) and can be accessed through an interface that opens as the user's hand approaches the bag.
- *Smartphone.* This device manages the communications between the nurse and the CM. When the nurse's hand approaches the smartphone in the scene, an interface appears to make, terminate, or answer a call (Fig. 3.8e).



Figure 3.8: The user interface of the Instructor Application, localized in the language of course instructors.

⁴<https://www.stryker.com/us/en/emergency-care/products/lifepak-15.html>

Interactions with the Patient The nurse can interact with the patient by applying sensors, oxygen masks, and administering medications. During the simulation, the instructor impersonates the patient, responding directly to questions the nurse asks.

The patient's representation in the VE adapts according to the vital parameters set by the instructor via the IA. The patient's state is visualized through facial animation, body animation, and skin characteristics, that change according to vital parameters (Fig. 3.9):

- When events that are painful to the patient occur, his face takes on increasingly pronounced suffering expressions. In case of fainting or deep sedation, eyelids are lowered.
- To give more direct visual feedback, the animations the patient performs change according to his consciousness, respiratory rate, and pain.
- The parameters representing the patient's skin are changed according to vital parameters, to simulate decreased perfusion or sweating.

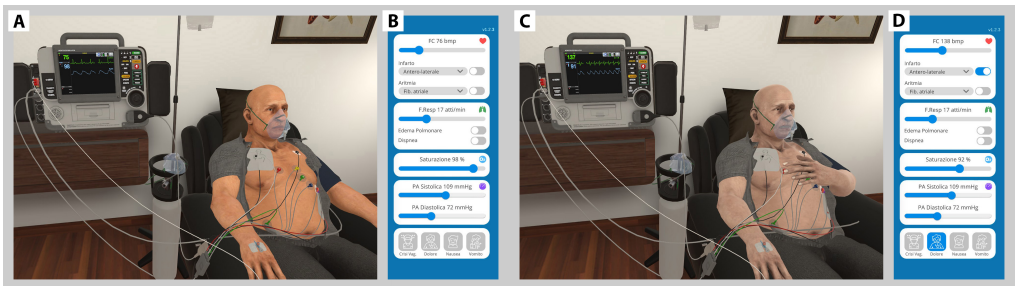


Figure 3.9: The user interface of the Instructor Application, localized in the language of course instructors.

3.2.3 Qualitative evaluation of the STEMI training system

The study's goal was to evaluate the perceptions and opinions of nurses and doctors attending the CME blended learning course with the VR system to train participants in rescuing STEMI patients.

Methods

To evaluate the subjective perceptions and opinions of the learners about the VR system, we conducted an interview consisting of a sequence of structured open-ended questions (Fig. 3.10). Questions were intentionally general and allowed to collect information from the point of view of both nurses and doctors. Data were analyzed by employing thematic analysis, as described by Braun and Clarke [37]. The study was approved by the Institutional Review Board of the Department of Mathematics, Computer Science, and Physics of the University of Udine.

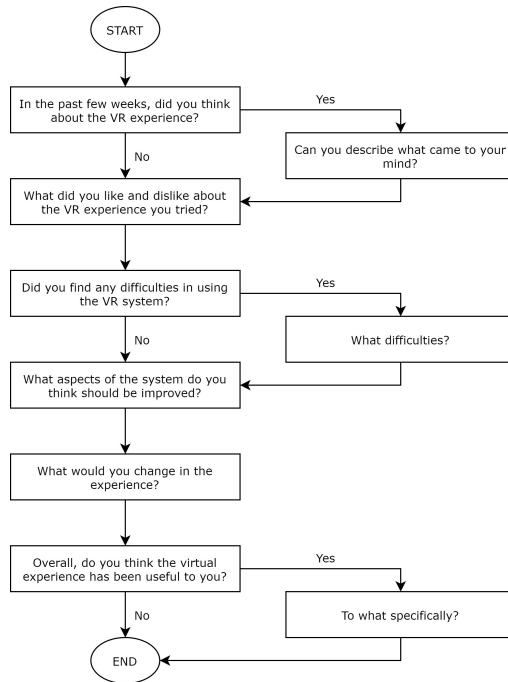


Figure 3.10: Flow diagram of questions in the interview. The interview was conducted in the language of participants. Sentences have been translated here for readers' convenience.

Use in the course The iVR system was integrated into the educational protocol of an institutional Continuing Medical Education (CME) accredited training course, resulting in a Blended Learning Course (BLC). The course was conducted in a total of six editions from October 2022 to January 2023 at the Azienda Sanitaria Universitaria Friuli Centrale in Northeastern Italy.

Each edition consisted of a four-hour session, beginning with a theoretical introduction and followed by practical simulations of STEMI patient rescue using the iVR system. During the simulations, the cardiologist instructor operated the Instructor Application, role-playing the patient and dynamically adjusting the scenario according to the learner's actions (e.g., if the scenario involved nausea symptoms and the nurse omitted the prescribed drug, the virtual patient began vomiting after a set time), without offering feedback on the correctness of the actions performed during the simulation.

Trainee doctors assumed the role of Case Manager doctors through the Case Manager Application, coordinating the in-hospital response. Since there were fewer Case Manager doctors than nurses, each doctor used the CA for one or more scenarios. Trainee nurses engaged with the Nurse Application on the VR headset in one scenario. To replicate the spatial separation typical of real-world prehospital emergencies, the doctor and nurse completed the simulation from different rooms, communicating exclusively through the immersive VR system.

Each edition of the BLC included the same seven scenarios to illustrate the variety of medical situations that may arise in STEMI rescue. Each simulated scenario lasted about 10-15 minutes. While one doctor–nurse pair actively engaged with the iVR system, the other participants observed the session projected on a wall. The projected image combined the nurse's headset view and the patient's physiological parameters as seen by the instructor. This participatory course approach was meant to allow nurses and doctors to gain the most from the experience.

After each simulation, the cardiologist instructor conducted a debriefing session to review the actions performed by the nurse in the Nurse Application, explaining both correct and incorrect steps, and encouraged discussion among participants.

At the end of the course, all participants were invited, via a dedicated form, to indicate their availability for follow-up contact by phone or email to participate in future user studies. When all six editions of the course were concluded, the report containing all the errors made during the simulation was sent to the learners.

Participants Among the 48 emergency nurses and 15 CM doctors who attended one of these editions, 17 participants volunteered to be interviewed about their experience with the VR system and the blended learning course. Twelve nurses and five doctors (9M, 8F) among the course participants volunteered to be interviewed. All but one also provided information about their age, years of experience in the role and in the emergency area, and previous use of VR (Table 3.4).

Table 3.4: Participants’ gender, age range, professional role, years of experience in the role and in the emergency area, and previous use of VR.

Participant	Gender	Age range	Professional role	Years of experience in the role	Years of experience in emergency area	Previous use of VR
D1	Male	35-44	Doctor	10-14	10-14	No
D2	Male	35-44	Doctor	10-14	10-14	Yes
D3	Female	35-44	Doctor	5-9	5-9	No
D4	Female	55+	Doctor	25-29	20-24	Yes
D5	Male	35-44	Doctor	10-14	<5	No
N1	Female	55+	Nurse	25-29	5-9	No
N2	Female	25-34	Nurse	5-9	5-9	No
N3	Male	55+	Nurse	30-34	30-34	No
N4	Male	35-44	Nurse	10-14	5-9	No
N5	Male	25-34	Nurse	5-9	5-9	No
N6	Male	25-34	Nurse	5-9	<5	Yes
N7	Male	45-54	Nurse	25-29	25-29	Yes
N8	Female	25-34	Nurse	10-14	5-9	Yes
N9	Female	25-34	Nurse	5-9	<5	Yes
N10	Female	45-54	Nurse	25-29	25-29	Yes
N11	Female	45-54	Nurse	25-29	15-19	No
N12	Male	NA	Nurse	NA	NA	NA

Procedure At the end of the CME course, we asked participants if they would like to volunteer for a subsequent interview about the experience. The volunteers left their e-mail and phone number to be contacted and book an interview time slot. After the interviewer defined the available time slots and assigned an online meeting link for each of them, two other team members contacted the volunteers and sent them the privacy policy and the consent module. After receiving the signed consent, the team members recontacted each volunteer, arranged a time slot and then confirmed it via e-mail, also including an anonymous code and the instructions to blind their identity to the interviewer (i.e., join the meeting without using a personal account, use the anonymous code and not the real name, do not use a webcam, possibly apply a filter to change the voice). Participants were also informed that the interview would be audio recorded and then deleted after transcription. At the defined time slot, the interviewer met the participant online and started the audio recording. After briefly introducing the interview and inviting the participant to freely express opinions, taking the time needed to reason about them, the interviewer asked the openended questions following the flow in Fig. 3.10. When needed, the interviewer also asked some additional questions (e.g., what, when, how?) to clarify participants answers.

Each audio recording was saved with the participant's anonymous code and then verbatim transcribed. Two members of the research team, with different background (one in medicine and one in computer science), performed the thematic analysis by repeatedly listening to the interviews, checking the transcripts back against the original audio recording, actively reading the transcripts many times, generating initial codes using a data-driven approach to include all potentially interesting content, collating them into potential themes and subthemes, and finally reviewing and refining them to create a codebook [37, 81]. The final themes and subthemes were chosen to describe the entire blended learning course (ReflectOnExperience, PositiveAspect, Critique) and the VR system (Realism, UsageDifficulty), and to acquire useful information for improvement (Proposal, Opinions). After agreeing on the codebook with the supervision of a third team member (a senior medical doctor), each of the coders independently assigned sub-themes to chunks of data extracted from the interviews using the Taguette software [260]. Then, the two coders discussed on the disagreements. In case of assignments for which no agreement was found, the third team member was involved to break the tie.

Results

The coders identified seven themes and 31 sub-themes in the interviews (Table 3.5). After the independent assignment of the codes, Cohen's Kappa inter-coder reliability was 0.76 (95% CI 0.71-0.82). After discussion, the two coders reached an agreement on all but six assignments, which were decided by the third team member.

Table 3.5: Identified themes and sub-themes with their description and participants who talked about them.

Theme	Sub-theme	Description	Participants
ReflectOn Experience	Yes	In the following weeks, the user reflected on the experience in virtual reality and / or talked about it with colleagues	D1; D2; D3; D4; N1; N2; N4; N5; N6; N7; N8; N10; N11; N12
ReflectOn Experience	No	In the following weeks, the user did not reflect on the virtual reality experience	D5; N3; N9
PositiveAspect	Positive Experience	In general, participation in the course was perceived to be positive and enjoyable and/or valuable and/or the participant would not change anything of the experience	D1; D2; D3; D4; D5; N1; N2; N3; N4; N5; N6; N7; N8; N9; N10; N11; N12
PositiveAspect	Engaging Learning	The experience is engaging and enables to memorize notions as effectively learned	D2; D5; N2; N12
PositiveAspect	Innovative	The course uses innovative technology as compared to traditional courses	D1; D2; D3; N1; N2
PositiveAspect	Technology Potential	The technology used is beneficial for this educational aim	D2; D3; N2; N3; N6; N7; N8; N9; N11; N12
PositiveAspect	Retry	The user would have liked to try the system several times, perhaps with different scenarios	N3; N8
PositiveAspect	Future Participation	The user would like more sessions of the course to be held	D5; N5; N11
PositiveAspect	Realism	The experience during the course is like what happens during an actual rescue operation	D2; D3; D4; D5; N1; N2; N3; N4; N5; N8; N9; N10; N12
PositiveAspect	Challenge	The user liked to challenge themselves and put their knowledge to the test	D2; D5; N2; N10; N11

Table 3.5: Identified themes and sub-themes with their description and participants who talked about them.

Theme	Sub-theme	Description	Participants
PositiveAspect	Increased Confidence	The course helped to increase the confidence level in one's knowledge	D2; N4; N6; N11
PositiveAspect	SafeEnvironment	VR allows trainees to practice risky situations in a safe environment	N3; N6; N7
PositiveAspect	ShareOpinions	The course can be a good opportunity to exchange views with others	D3; D4; N2; N4; N5; N6; N7; N10
PositiveAspect	Highlighting Problems	The course is a good opportunity to highlight issues encountered during operations on the field	D2; D3; D4; N11
PositiveAspect	Instructors Expertise	Course instructors are trained and experienced	D4; N4
PositiveAspect	NoIssues	The user found no negative aspects in the experience	D1; D3; D4; D5; N5; N6
Critique	Discomfort	The presence of other people makes the learner feel uncomfortable and under pressure	D4; N1; N7; N10; N11
Critique	FamiliarTopics	The topics in the course are already known, nothing new is provided	D4; N1; N2; N7; N8; N9; N12
Realism	Over Simplification	The system simplifies / speeds up some procedures that are not simple or that in the real world take a certain amount of time to perform	N1; N7; N8; N9; N10
UsageDifficulty	MissingActions	It is not possible to perform within the system all the actions that would be performed during a rescue	D2; N7
UsageDifficulty	Patient	The user stated that the patient is unrealistic, too still, and was not able to notice pallor or sweat	N7; N8; N12
UsageDifficulty	Sensorial	Only sight and hearing are used; stimulation of the other senses is lacking	D2; N7; N10; N12

Table 3.5: Identified themes and sub-themes with their description and participants who talked about them.

Theme	Sub-theme	Description	Participants
UsageDifficulty	BecomeFamiliar	The user had difficulty since it was their first time using the system and they did not have a chance to become familiar with it before the test	D3; D4; D5; N1; N2; N3; N4; N7; N8; N9; N10; N11; N12
UsageDifficulty	Interaction	Need to use unnatural gestures to interact with various elements	D2; D3; D4; N1; N2; N3; N7; N8; N9; N11
UsageDifficulty	Cybersickness	The user reported experiencing headache or nausea or dizziness or mentioned that colleagues suffered from it	N2; N4; N9; N10; N12
UsageDifficulty	None	No difficulties were identified in using the system	D1; D3; D4; D5; N1; N4; N5; N6; N8; N12
Proposals	Environments	The user suggests increasing the inter-scenario variability to avoid making them too repetitive	D2; N3; N8; N10; N12
Proposals	Uncommon Environments	The user suggests using this technology for creating unusual scenarios, which do not happen frequently	D2; N1; N3; N7; N11
Proposals	Realism	The user wishes that the system was more like reality	D2; N1; N3; N6; N7; N9; N10
Opinion	UsefulForNovices	Positive experience for new staff rather than people who have been working for years	D4; N1; N7; N9; N12
Opinion	UsefulForExperts	For experienced staff it can be a good knowledge check, to repeat protocols	D2; D4; N2; N5; N6; N8

Here we report in more depth the results of the qualitative analysis by including sample interview extracts. In the extracts, the parts in square brackets are words or phrases added to clarify the sentence.

- *ReflectOnExperience_Yes*. Fourteen people reflected on the virtual reality experience and/or talked about it with colleagues in the following weeks. For example:
N4 - I've thought about it.
- *ReflectOnExperience_No*. Three people did not reflect on the virtual reality experience in the following weeks. For example:
N9 - No.
- *PositiveAspect_PositiveExperience*. Seventeen people perceived participation in the course to be positive and enjoyable and/or valuable and/or would not change anything of the experience. For example:
D3 - We can simply say I liked it.
- *PositiveAspect_EngagingLearning*. Four people praised the application stating that the experience engages those who experience it and enables them to memorize the notions effectively learned. For example:
N2 - [This mode of explaining a topic] is very active learning.
- *PositiveAspect_Innovative*. Five people praised the innovative technology of the course compared to the traditional material used in other lessons. For example:
D3 - [I appreciated] the quality of the experience itself, the technique, and my experience. Moreover, the fact that I had not tried [something like that] before left me pleasantly surprised.
- *PositiveAspect_TechnologyPotential*. Ten people told us how the technology used is beneficial for this educational aim. For example:
N2 - I think it gives you many extra things in specific courses. Because you really drop into the actual situation, and so you get yourself involved.
- *PositiveAspect_Retry*. Two people would have liked to try the system several times, perhaps with different scenarios. For example:
N8 - I would maybe try a few more times. We each tried once. Possibly in the idea of having more time, it would be nice to do several different scenarios and try several times.
- *PositiveAspect_FutureParticipation*. Three people would like more sessions of the course to be held. For example:
N5 - The only thing I hope is that it is a course that is done every year because it fits, as I said before, relating to the various figures, in my opinion, is essential, as far as our work is concerned.

- *PositiveAspect_Realism*. Thirteen people stated how the experience during the course is like what happens during an actual rescue operation. For example:

N5 – It very much reflects reality. Of course, you emphasize a little bit more, in the sense that when you talk to the cardiologist, you lose a little bit more time. However, it's a course that allows you to relate to the patient as you relate to them in the real world. And the treatments you do are the real ones, whether it's the ECG, administering the medications, or contacting the cardiologist, which is an important thing anyway.
- *PositiveAspect_Challenge*. Five people appreciated challenging themselves by putting their knowledge to the test. For example:

D2 – Well, let's say [that the application also helps] to understand the difficulties in interfacing with practitioners who are not your colleagues. I'm used to being a case manager for my nurses who, in short, I know off the top of my head, so I know what I [expect] when someone tells me something, and [I understand] what they mean. When I give directives, I know what is received. Doing it instead with people I don't know is definitely something that, let's say, is very helpful.
- *PositiveAspect_IncreasedConfidence*. Four people praised the course for helping them to increase the confidence level of their knowledge. For example:

N4 – In my opinion, it gives you more confidence; afterward, you will know what to do in the real world [when you are in similar situations].
- *PositiveAspect_SafeEnvironment*. Three people praised the safety of the virtual environment, stating that it is possible to practice handling risky situations in a protected environment. For example:

N3 – This [mode of learning] is the middle ground. It allows you to apply practical theory while being in a safe scenario.
- *PositiveAspect_ShareOpinions*. Eight people told us how the course can be a good opportunity to exchange views with others. For example:

N7 – I liked the fact that we still got together with different operators from the same company. We have become a big company now, so this is a time to meet [with other colleagues].
- *PositiveAspect_HighlightingProblems*. Four people praised the course as an opportunity to highlight issues encountered during operations on the territory. For example:

N11 – There is a big gap between those who work on the ground and those who work in the hospital. They don't understand our situation, so they don't help us when requests are made.
- *PositiveAspect_InstructorsExpertise*. Two people expressed how the course instructors were well-trained and experienced. For example:

N4 – The interlocutors, that is those who empathized with the part, were also realistic, so it was a very good experience.

- *PositiveAspect_NoIssues*. Six people did not report any negative aspects of the experience. For example:

D1 – In fact, there is not one thing that I did not like.

- *Critique_Discomfort*. Five people report how the presence of other people makes them feel uncomfortable and under pressure. For example:

D4 – There were still many other people, and somehow [I had] the feeling of being a little bit under examination.

- *Critique_FamiliarTopics*. Seven people report how the topics in the course are already known, and so nothing new is provided for them. For example:

N7 – So, honestly, that course was not helpful to me in the sense that all the topics that were being talked about I have already consolidated over time.

- *Realism_OverSimplification*. Five people report how the system simplifies or speeds up some procedures that are not simple or that, in the real world, take a certain amount of time to be performed. For example:

N10 – Instead, the difficulty and the time to do those tasks in the real world are not like that. That is, the fact that I touched the wires and brought them to the patient was instantaneous. In a real situation, it takes me several seconds in which more questions can be processed, and I take more time to think about what I need to do. It was too fast.

- *UsageDifficulty_MissingActions*. Two people complained about the lack of the possibility to perform within the system all the actions that would normally be performed during a rescue. For example:

D2 – It certainly wasn't entirely realistic, having to be limited to two or three actions that need to be codified at some point.

- *UsageDifficulty_Patient*. Three people stated that the patient is unrealistic, too still, and there is no noticeable pallor or sweat. For example:

N8 – The operator was giving some somewhat nonspecific information that then, in the mannequin within the virtual reality, you could not see. The person was standing there, so there was a moment of inconsistency.

- *UsageDifficulty_Sensorial*. Four people reported how only sight and hearing are used, leading to a lacking of stimulation of the other senses. For example:

D2 – The learner perceives not only by hearing and with sight but with all the senses, including touch and smell.

- *UsageDifficulty_BecomeFamiliar*. Thirteen people had difficulties with the system since it was their first time using it, and they did not have a chance to become familiar with it before the test. For example:

N2 – You have to get a little bit of practice with this kind of technology.

- *UsageDifficulty_Interaction*. Ten people had difficulties interacting with the system because of the need to use unnatural gestures to interact with various elements. For example:

N8 – I found myself not performing things that I have performed, as if my intervention was not detected by the instrument.

- *UsageDifficulty_Cybersickness*. Five people reported experiencing headaches, nausea, dizziness, or colleagues suffering from it. For example:

N2 – To some people who were with me at the course, the system made them a little dizzy.

- *UsageDifficulty_None*. Ten people reported no difficulties when using the system. For example:

D5 – The tablet that was being provided and the communications were both [good] at the level of technology and the degree of communication; I think they were absolutely congruent with the situation [we were facing].

- *Proposals_Environments*. Five people suggest increasing the inter-scenario variability to avoid making them too repetitive. For example:

D2 – Maybe trivially change the structure of the room or the patient lying down rather than finding them in bed or on the stairs. [These situations] in practice often change the practitioner's approach.

- *Proposals_UncommonEnvironments*. Five people suggest using this technology to create unusual scenarios which do not happen frequently. For example:

D2 – In the sense that along the lines of this course, I think so many other kinds of pathways can be done, right? I'm thinking of the stroke pathway rather than a trauma pathway, the shock patient pathway, that is, in those in which you have to add the figure of the central doctor.

- *Proposals_Realism*. Seven people wish that the system was more similar to reality. For example:

N9 – So, the monitor definitely maybe needs to be fixed a little bit in the sense that it's not really very accurate. When you go to press some of the times, it takes one thing for another. And the backpack definitely, the medication backpack, is not really with the proper dosages. In my opinion, the medication dosages should be adjusted a little bit.

- *Opinion_UsefulForNovices*. Five people think that the experience is more positive for new staff rather than for people who have been working for years. For example:

N1 – I would have done them very gladly at the beginning of my career.

- *Opinion_UsefulForExperts*. Six people think the system could be a good knowledge check to repeat protocols for experienced staff. For example:

N2 – You have to [already] know the things. It's not like you learn them there. I mean, it could be a suitable method of verifying learning and not so much for

education because the things that we learned, I mean, in my opinion, [within the system] we verified what we know or not during the course, we didn't learn new things.

Discussion

The methodological paper by Pedram et al. [245] recommends that VR systems should “provide motivation through enjoyment, satisfaction, engagement, attention and challenge” and in other studies a positive experience was reported in terms of enjoyment [103, 274], satisfaction [52, 161, 307], attentiveness [274, 307], motivation [307], and acceptability and feasibility [86]. The results of our study confirm those of the literature in the context of STEMI: all participants appreciated the course and the use of the VR system, and 14 out of 17 said they thought about the experience in the weeks following the course (Reflect on Experience theme).

All participants contributed to at least 3 sub-themes of the broad Positive Aspect theme and six did not find any negative aspects; the VR system was considered engaging (n=4), as reported in other studies [86, 132, 272, 307], innovative (n=5) and promising for training courses (n=10). For example, N2 reported “I think it gives you many extra things in specific courses. Because you really drop into the actual situation, and so you get yourself involved” and D3 declared “[I appreciated] the quality of the experience itself, the technique, and my experience. Moreover, the fact that I had not tried [something like that] before left me pleasantly surprised”.

In addition, some participants would have liked to try more simulations (n=2) or more course sessions (n=3), as reported in other studies where the willingness to use a VR system again was described [18, 132].

The VR system was also described as an opportunity to test knowledge and skills in a safe environment (n=3), as positively described in previous work [18, 270]. Some trainees described the blended learning course as an opportunity to discuss and exchange opinions (n=8) and to point out issues (n=4); for example, N11 reported: “There's a big gap between those who work on the ground and those who work in the hospital. They don't understand our situation, so they don't help us when requests are made”. The presence of experienced instructors was also appreciated (n=2), consistently with other studies, where a supervisor and peer interaction were positively considered [272, 307].

Moreover, most participants reported that the VR experience was highly similar to what happens during actual emergency interventions (n=13), as in other studies [274]. For example, N5 reported: “It very much reflects reality. Of course, you emphasize a little bit more, in the sense that when you talk to the cardiologist, you lose a bit more time. However, it's a course that allows you to relate to the patient as you relate to them in the real world. And the treatments you do are the real ones, whether it's the ECG, administering the medications, or contacting the cardiologist, which is an important thing anyway” and it is interesting to note that realism was appreciated both by doctors and nurses, although only nurses lived the VR experience using the VR headset, while all participants watched it on the projected screen. However, some participants complained about the procedures' oversimplification (n=5) and missing actions (n=2) and suggested to improve realism (n=7), as in Hood et al. [132]. For example, under the Realism theme, N10 reported: “Instead, the difficulty and the time to do those tasks in the real world are not like that. That is, the fact that I touched the wires and brought them to the

patient was instantaneous. In a real situation, it takes me several seconds in which more questions can be processed, and I take more time to think about what I need to do. It was too fast.”

Participants also showed their interest (Proposal and Opinion themes) suggesting to modify scenarios to increase their variability ($n=5$; e.g. D2: “Maybe trivially change the structure of the room or the patient lying down or finding them in bed or on the stairs”), like requested in another study [132], and to describe infrequent situations ($n=5$). Five participants reported it could be beneficial to naive health personnel (N1: “I would have done them very gladly at the beginning of my career”), while six reported that experienced personnel could benefit from it as a refresher training.

Nine participants contributed to the Critique theme, seven of them describing the blended learning course as not useful for them because they already knew the topics and five reporting discomfort and embarrassment due to the presence of other people (e.g. D4: “There were still many people, and somehow [I had] the feeling of being a little bit under examination”). On the other hand, four found the course beneficial to improve self-confidence and five said that they liked the challenge. For example, D2 reported: “Well, let’s say [the system also helps] to understand the difficulties in interfacing with practitioners who are not your colleagues. I’m used to being a case manager for my nurses who, in short, I know off the top of my head, so I know what I [expect] when someone tells me something, and [I understand] what they mean. When I give directives, I know what is received. Doing it instead with people I don’t know is definitely something that, let’s say, is very helpful”. Challenge was indeed one of the aims suggested by Pedram et al. [245] and the trainee confidence following the VR training was already described as an advantage in previous studies [86, 274].

The Usage Difficulty theme revealed that ten did not mention any but a few participants thought the virtual patient was too static, finding it difficult to appreciate clinical signs such as paleness or sweating ($n=3$), or felt the interaction limited by the only use of sight and hearing ($n=4$). Moreover, thirteen reported the need to get familiar with the system and ten to learn specific interaction gestures (e.g. N8: “I found myself not performing things that I have performed, as if my intervention was not detected by the instrument”). This happened despite all participants were invited to watch a video tutorial before attending the course. These results are consistent with those described by other authors [132, 241, 272] and explain the recommendation for VR knowledge or practice before approaching VR systems for medical education [272].

Finally, three participants suffered dizziness or nausea; this was expected, since cybersickness represents a known drawback of VR systems [272, 323].

There are different positive aspects to notice about the system and the course. We implemented a VR system that supports all control requirements recommended by Pedram et al. [245]: i) including “gesture recognition”, ii) providing “motion sensitivity of controls”, iii) meeting “minimum position tracking thresholds”, iv) enabling “kinesthetic manipulation of virtual objects of interest”, v) providing “efficient movement controls”, vi) providing “control element granularity commensurate with the virtual layout needed”. Moreover, it provides “effective feedback to the user” [245], both during the simulation (when the patient reacts to the user’s actions) and at the end, through a description of the steps and their correctness. Participants in Hood et al. [132] appreciated feedback on timing and decision making and asked for more specific feedback, such as more in-

formation on mistakes. In our study, further feedback was provided by the instructor, a senior cardiologist complementing VR simulation.

As recommended by Pedram et al. [245], visual and auditory cues (e.g., monitor's display contents and sounds) are also provided, while haptic cues are currently missing. As for the course, training was performed in small groups (from 6 to 13 participants), and this is described as a facilitator aspect [272]. Moreover, in contrast to another study, where the roles of distinct health workers were not differentiated, resulting in the disappointment of participants [132], nurses and doctors played their own roles in our system and their collaboration was accurately reproduced, as also reported by some participants. This partnership is essential in the STEMI patient rescue, since the emergency nurse on the field and the doctor in the hospital need to work together to save the patient's life.

Costs are traditionally mentioned among the limitations of VR systems [241, 307], of which users are also aware [272]. However, we specifically implemented our system to run on the low-cost Meta Quest 2 VR headset. Moreover, cost-effectiveness may improve if economic costs are analyzed in the long-term, and counterbalanced by potential risks and medical errors that could be avoided thanks to a training conducted in a safe environment [18].

Conclusions

This is the first proposal and study of a VR system to train nurses and doctors in the pre-hospital management of STEMI patients. The system was embedded in a CME course, making it a blended learning course. We performed a thematic analysis of the answers of course participants who volunteered to be interviewed. This possibly led to a self-selection of those more interested and satisfied and to a consequent overestimation of the blended learning course and VR system. An additional limitation was the small sample size, even though representative of the highly specialized healthcare personnel who deal with prehospital STEMI emergencies. In addition, the interviews did not take place immediately after the course and this allowed participants to reflect on the course, but it could also have caused memory fading and detail loss.

The analysis showed that all participants perceived the VR system and the blended learning course as a positive experience and provided interesting feedback on positive aspects and possible areas of improvement.

Following the advice, we worked on additional scenarios to increase variety. Moreover, we plan to improve the VR system as an evaluation tool and use it to compare the performance of participants of the course with the performance of nurses and doctors who had not yet attended the course, to quantitatively measure the effectiveness of the proposed form of VR-based medical education.

As the main implication for practice, this study highlighted that a VR system integrated in a blended learning course with a strong collaborative approach is much appreciated by all participants and could be inspiring when designing novel medical education programs for nurses and doctors. Interestingly, the training course also received a positive reaction at the regional level and was included in the current Training Plan of the Friuli Venezia Giulia Region, thus allowing nurses and doctors to receive a consistent education, smoothing possible geographical differences.

The work presented in this section was published in the *Cyberpsychology, Behavior, and Social Networking* journal [99].

3.2.4 Quantitative evaluation of the STEMI training system

The previous section presented the study based on the thematic analysis of how participants felt subjectively the VR system reporting how the training with this technology was engaging, realistic, and well-received. This section describes the evaluation investigating the educational efficacy of integrating our immersive VR system into the BLC by comparing a group of participants who followed the BLC with a control group.

Methodology

The goal of this study is to investigate the efficacy of the BLC held between October 2022 and January 2023. The analysis focuses specifically on nurses, as they were the primary users of the immersive VR system employed in the BLC. Case manager physicians were excluded from this analysis because, unlike nurses, they did not directly utilize the immersive VR system. To evaluate the impact of the BLC on procedural knowledge acquisition, the study employed a between-subjects design, comparing nurses who participated in the BLC (G1) with a control group of nurses who did not (G2). All participants in both G1 and G2 received traditional training, consisting of the Advanced Life Support or Advanced Cardiac Life Support course. Thus, the VR experience represented an additional component integrated into the traditional training for G1.

Following the literature suggesting that VR training is an effective complement to traditional learning approaches [284, 286] and the positive subjective outcomes reported in our previous study [99], our hypothesis is that G1 would report greater procedural knowledge than G2.

Simulation Setup The iVR system was employed for evaluation using the 7 clinical scenarios originally designed for the BLC, which encompass the combinations of anamnestic data and vital parameters addressed during the course. For the purposes of this study, in order to reduce the heterogeneity among the simulation tested by the participants, the scenarios presenting a higher level of difficulty due to additional complications were simplified by excluding those complications. In addition, four new scenario variations were specifically developed and used exclusively within this user study. Overall, the use of 11 scenario variations allowed for a more comprehensive and balanced assessment, aiming to have an aggregate measure of the group's competence across a broad yet controlled range of clinical conditions.

The Nurse Application was installed on a Meta Quest 2 headset, and the Instructor Application and Case Manager Application were installed on two Lenovo Tab P11 Pro tablets.

Each simulation involved an instructor (using the Instructor Application) who also impersonated the patient, a medical doctor (using the Case Manager Application) who impersonated the case manager physician, and another experimenter acting as a nurse of the regional operating center who briefed the participant before the interaction with the patient with standardized information on the patient's location (e.g., city), triage code,

age, and symptoms reported during the emergency call. Although 5 experimenters alternated in these roles, the combinations were balanced across groups to minimize potential confounding variables. Additionally, the same software developer was always present to manage the system setup and address any technical issues that arose. No further operators or observers were involved to minimize distractions and maintain experimental control.

Measures Nurses' procedural knowledge of STEMI management was evaluated using a checklist based on their performance during the VR simulation. This checklist comprises 33 distinct procedural steps, each corresponding to a recommended action in pre-hospital STEMI management according to the guidelines developed in the region where the study was conducted. Certain steps involve specific operational parameters (e.g., acquiring data within a defined time frame following sensor placement). In such cases, the parameter-dependent component was considered as a separate parameter-component step. The procedure, which combines distinct steps and parameter-component steps, comprises a total of 40 steps (Table 3.6).

The measures to assess the effectiveness of the BLC are divided into step-level and nurse-level:

- At the step-level, the aim is to identify which procedural steps were performed correctly or incorrectly most frequently. The analysis evaluated performance on each of the 40 checklist items individually across the entire participant group, thereby assessing adherence to specific steps of the procedure by counting how many nurses per group performed correctly a specific step.
- At the nurse level, a score is assigned at each step that reflects the clinical importance (Table 3.6). The analysis aggregated each participant's performance across the steps of the checklist. The steps were grouped into the six clinical areas detailed in Table 3.7. The total score for each of the six areas was calculated by summing up the scores of its constituent steps.

The metrics for the step and nurse level are summarized in Table 3.8.

Table 3.6: Checklist of the steps the nurse has to perform in the simulation with score assigned that reflects the clinical importance. Steps with “p” suffix refers to the parameter component of the related step.

Step	Category	Action	Score
S1	Demographics	Asked for patient's name and surname	0.5
S2	"	Asked for patient's age	0.5
S3	Symptoms	Asked patient to describe symptoms	0.5
S4	"	Asked for time of symptom onset	0.5
S5	Pathological history	Asked about bleeding diathesis	0.5
S6	"	Asked about current illnesses	0.5
S7	Pharmacological history	Asked about ongoing treatments	1.0
S8	Blood pressure	BP sensor applied	0.5
S8p	"	BP sensor applied within 2 minutes	0.5
S9	SpO ₂	SpO ₂ sensor applied	0.5
S9p	"	SpO ₂ sensor applied within 2 minutes	0.5

Table 3.6: Checklist of the steps the nurse has to perform in the simulation with score assigned that reflects the clinical importance. Steps with “p” suffix refers to the parameter component of the related step.

Step	Category	Action	Score
S10	Venous access	Venous line applied	0.5
S10p	"	Venous line applied within 5 minutes	0.5
S11	ECG – Peripheral leads	Electrodes applied	0.5
S11p	"	Electrodes applied within 2 minutes	0.5
S12	ECG – Precordial leads	Electrodes applied	0.5
S12p	"	Electrodes applied within 2 minutes	0.5
S13	ECG – Recording	ECG print performed	0.5
S13p	"	ECG print within 1 minute from electrode placement	0.5
S14	ECG – Transmission	ECG sent to case manager	0.5
S14p	"	ECG sent within 1 minute	0.5
S15	Call	Nurse called CM after sending ECG	1.0
S16	Data transfer	Communicated patient’s name and surname	0.5
S17	"	Communicated patient’s age	0.5
S18	"	Communicated symptom description	0.5
S19	"	Communicated time of symptom onset	0.5
S20	"	Communicated bleeding tendency	0.5
S21	"	Communicated current illnesses	0.5
S22	"	Communicated ongoing treatments	1.0
S23	MANOB protocol	Correct administration of morphine	1.0
S24	"	Correct administration of acetylsalicylic acid	1.0
S25	"	Correct administration of sublingual nitroglycerin	1.0
S26	"	Correct administration of oxygen	1.0
S27	"	Correct administration of beta-blocker	1.0
S28	Other therapy	Correct intravenous administration of heparin	1.0
S29	"	Correct administration of ticagrelor	1.0
S30	Communication	Informed patient of diagnosis	0.5
S31	"	Informed patient of destination hospital	0.5
S32	"	Informed patient of future treatments	0.5
S33	Defibrillation	Pads applied after STEMI identification on ECG	1.0

Table 3.7: Overview of Assessment Areas and total score obtainable by each nurse during the simulation.

Category	Number of items	Focus Area	Composing Steps	Area Score
Area 1	7	Patient’s medical history	Steps from S1 to S7	4
Area 2	15	Acquisition of medical data	Steps from S8 to S15 considering parameter-dependent component	8
Area 3	7	Nurse - Case Manager communication	Steps from S16 to S22	4
Area 4	7	Ancillary therapies	Steps from S23 to S29	7
Area 5	3	Nurse - Patient communication	Steps from S30 to S32	1.5
Area 6	1	Defibrillation	Step S33	1
Total	40	-	-	25.5

Table 3.8: Summary of Performance Metrics at the step and nurse level.

	Metric	Description
For each step	Nurse_completed	Number of nurses who completed the specific procedure step.
For each step with a parameter	Nurse_correct_parameters	Number of nurses who met the required parameters in the parameter-specific step.
For each nurse	Total_score	Total score obtained by the nurse
For each nurse and each area	Total_score_in_area	Total score obtained by the nurse in the specific area

Each checklist item was marked as either 'performed' or 'not performed'. Verbal items were assessed by the instructor using a tablet to track each step of the procedure. Other items were automatically tracked by the system. The list of items performed was then verified by a second medical doctor supported by the software developer using system logs, and a third medical doctor resolved any remaining discrepancies.

Participants The sample size calculation was based on the total performance score as the primary outcome, with a maximum achievable score of 25.5. Power analysis assumed a comparison between two groups of equal size, aiming to detect a minimum difference of 5 points in the mean total score. A standard deviation of 4, an alpha level of 0.05, and a statistical power of 0.8 were applied. Based on these parameters, a minimum of 11 participants was required per group.

The evaluation involved 22 nurses (G1: N = 11; G2: N = 11). All participants were volunteers. Nurses in G1 were those who had previously indicated their availability for follow-up studies after the BLC, while nurses in G2 were recruited through nurse coordinators and personal contact. An initial questionnaire collected demographic data, professional background, and prior training. Table 3.9 presents the questions asked in the questionnaire along with corresponding results.

Table 3.9: Demographic characteristics, professional background, prior training, and self-rated knowledge of STEMI procedures of participants. Data are reported as absolute values with percentages relative to each group, and self-rated knowledge is presented as Mean / SD.

	G1	G2	Total
ALS certification			
No	2 (18.2%)	4 (36.4%)	6 (27.3%)
Yes	9 (81.8%)	7 (63.6%)	16 (72.7%)
ACLS certification			
No	8 (72.7%)	2 (18.2%)	10 (45.5%)
Yes	3 (27.3%)	9 (81.8%)	12 (54.5%)
Sex			
F	4 (36.4%)	3 (27.3%)	7 (31.8%)
M	7 (63.6%)	8 (72.7%)	15 (68.2%)
Age range			
25-34	5 (45.5%)	3 (27.3%)	8 (36.4%)
35-44	3 (27.3%)	6 (54.5%)	9 (40.9%)
45-54	2 (18.2%)	2 (18.2%)	4 (18.2%)
55+	1 (9.1%)	0 (0%)	1 (4.5%)

Table 3.9: Demographic characteristics, professional background, prior training, and self-rated knowledge of STEMI procedures of participants. Data are reported as absolute values with percentages relative to each group, and self-rated knowledge is presented as Mean / SD.

	G1	G2	Total
Years of work experience in emergency and urgent care (hospital or prehospital)			
0-4	3 (27.3%)	3 (27.3%)	6 (27.3%)
5-9	4 (36.4%)	2 (18.2%)	6 (27.3%)
10-14	0 (0%)	1 (9.1%)	1 (4.5%)
15-19	1 (9.1%)	1 (9.1%)	2 (9.1%)
20-24	2 (18.2%)	4 (36.4%)	6 (27.3%)
25-29	1 (9.1%)	0 (0%)	1 (4.5%)
Work in a Hub or a Spoke center			
Hub	5 (45.5%)	1 (9.1%)	6 (27.3%)
Spoke	6 (54.5%)	10 (90.9%)	16 (72.7%)
Years of work experience in current professional role			
0-4	1 (9.1%)	1 (9.1%)	2 (9.1%)
5-9	4 (36.4%)	4 (36.4%)	8 (36.4%)
10-14	2 (18.2%)	1 (9.1%)	3 (13.6%)
15-19	1 (9.1%)	1 (9.1%)	2 (9.1%)
20-24	1 (9.1%)	3 (27.3%)	4 (18.2%)
25-29	2 (18.2%)	0 (0%)	2 (9.1%)
30-34	0 (0%)	1 (9.1%)	1 (4.5%)
Self-rated knowledge of the PDTA for STEMI patients (0 = no knowledge, 10 = complete knowledge)	7.09 / 0.83	6.91 / 1.14	7.00 / 0.98

Procedure The evaluation took place from March to April 2023. A researcher conducted an initial telephone contact to confirm participant availability, provide a concise overview of the study protocol and simulation procedures, and address preliminary questions. The researcher then sent the study information sheet and written consent form to participants by email for prior review.

On the day of the user study, written informed consent was obtained in person by the experimenter, who collected the signed forms and assigned each participant an anonymous identification code to be used for questionnaires and simulations, ensuring confidentiality. After providing consent, the participant filled up the initial questionnaire and then attended a standardized briefing session. The instructor and the software developer guided them through an identical series of exercises to build familiarity with the immersive VR system and develop the manual skills needed for interactions with the simulation tools.

Each nurse completed one simulation scenario lasting approximately 10 minutes. Overall, the 11 scenarios were each performed twice, once by a G1 participant and once by a G2 participant, so that every nurse in each group faced a different case.

Results

All data have been statistically analyzed with R version 4.3.0 [301]. Given the relatively small sample size, comparisons are performed using the Mann–Whitney test for continuous variables and the Fisher’s exact test for categorical variables.

Metrics at step-level Considering the metrics at step level for completed steps (Table 3.10 and Fig 3.11) and parameter dependent steps (Table 3.11 and Fig 3.12), the analysis reported no statistically significant differences among all the metrics between the two groups.

Table 3.10: Number of nurses who completed the step of the procedure. ¹Fisher's exact test; ²Chi-square test.

		Overall (N=22)	G1 (N=11)	G2 (N=11)	p-value
S1	Not Completed	13 (59.1%)	5 (45.5%)	8 (72.7%)	> 0.05 ¹
	Completed	9 (40.9%)	6 (54.5%)	3 (27.3%)	
S2	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S3	Not Completed	5 (22.7%)	2 (18.2%)	3 (27.3%)	> 0.05 ¹
	Completed	17 (77.3%)	9 (81.8%)	8 (72.7%)	
S4	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S5	Not Completed	20 (90.9%)	9 (81.8%)	11 (100%)	> 0.05 ¹
	Completed	2 (9.1%)	2 (18.2%)	0 (0%)	
S6	Not Completed	12 (54.5%)	5 (45.5%)	7 (63.6%)	> 0.05 ¹
	Completed	10 (45.5%)	6 (54.5%)	4 (36.4%)	
S7	Not Completed	11 (50.0%)	5 (45.5%)	6 (54.5%)	> 0.05 ²
	Completed	11 (50.0%)	6 (54.5%)	5 (45.5%)	
S8	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S9	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S10	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S11	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S12	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S13	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S14	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S15	Not Completed	3 (13.6%)	1 (9.1%)	2 (18.2%)	> 0.05 ¹
	Completed	19 (86.4%)	10 (90.9%)	9 (81.8%)	
S16	Not Completed	21 (95.5%)	11 (100%)	10 (90.9%)	> 0.05 ¹
	Completed	1 (4.5%)	0 (0%)	1 (9.1%)	
S17	Not Completed	4 (18.2%)	1 (9.1%)	3 (27.3%)	> 0.05 ¹
	Completed	18 (81.8%)	10 (90.9%)	8 (72.7%)	
S18	Not Completed	5 (22.7%)	2 (18.2%)	3 (27.3%)	> 0.05 ¹
	Completed	17 (77.3%)	9 (81.8%)	8 (72.7%)	
S19	Not Completed	1 (4.5%)	0 (0%)	1 (9.1%)	> 0.05 ¹
	Completed	21 (95.5%)	11 (100%)	10 (90.9%)	
S20	Not Completed	20 (90.9%)	9 (81.8%)	11 (100%)	> 0.05 ¹
	Completed	2 (9.1%)	2 (18.2%)	0 (0%)	
S21	Not Completed	13 (59.1%)	6 (54.5%)	7 (63.6%)	> 0.05 ¹
	Completed	9 (40.9%)	5 (45.5%)	4 (36.4%)	
S22	Not Completed	14 (63.6%)	6 (54.5%)	8 (72.7%)	> 0.05 ¹
	Completed	8 (36.4%)	5 (45.5%)	3 (27.3%)	
S23	Not Completed	12 (54.5%)	6 (54.5%)	6 (54.5%)	> 0.05 ²

Table 3.10: Number of nurses who completed the step of the procedure. ¹Fisher's exact test; ²Chi-square test.

		Overall (N=22)	G1 (N=11)	G2 (N=11)	p-value
	Completed	10 (45.5%)	5 (45.5%)	5 (45.5%)	
S24	Not Completed	9 (40.9%)	4 (36.4%)	5 (45.5%)	> 0.05 ¹
	Completed	13 (59.1%)	7 (63.6%)	6 (54.5%)	
S25	Not Completed	3 (13.6%)	1 (9.1%)	2 (18.2%)	> 0.05 ¹
	Completed	19 (86.4%)	10 (90.9%)	9 (81.8%)	
S26	Not Completed	12 (54.5%)	7 (63.6%)	5 (45.5%)	> 0.05 ¹
	Completed	10 (45.5%)	4 (36.4%)	6 (54.5%)	
S27	Not Completed	10 (45.5%)	5 (45.5%)	5 (45.5%)	> 0.05 ²
	Completed	12 (54.5%)	6 (54.5%)	6 (54.5%)	
S28	Not Completed	13 (59.1%)	6 (54.5%)	7 (63.6%)	> 0.05 ¹
	Completed	9 (40.9%)	5 (45.5%)	4 (36.4%)	
S29	Not Completed	11 (50.0%)	5 (45.5%)	6 (54.5%)	> 0.05 ²
	Completed	11 (50.0%)	6 (54.5%)	5 (45.5%)	
S30	Not Completed	6 (27.3%)	2 (18.2%)	4 (36.4%)	> 0.05 ¹
	Completed	16 (72.7%)	9 (81.8%)	7 (63.6%)	
S31	Not Completed	6 (27.3%)	4 (36.4%)	2 (18.2%)	> 0.05 ¹
	Completed	16 (72.7%)	7 (63.6%)	9 (81.8%)	
S32	Not Completed	20 (90.9%)	10 (90.9%)	10 (90.9%)	> 0.05 ¹
	Completed	2 (9.1%)	1 (9.1%)	1 (9.1%)	
S33	Not Completed	18 (81.8%)	8 (72.7%)	10 (90.9%)	> 0.05 ¹
	Completed	4 (18.2%)	3 (27.3%)	1 (9.1%)	

Table 3.11: Number of nurses who met the required parameter in the parameter specific step. ¹Fisher's exact test.

		Overall (N=22)	G1 (N=11)	G2 (N=11)	p-value
S8p	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S9p	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S10p	Not Completed	1 (4.5%)	0 (0%)	1 (9.1%)	> 0.05 ¹
	Completed	21 (95.5%)	11 (100%)	10 (90.9%)	
S11p	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S12p	Not Completed	0 (0%)	0 (0%)	0 (0%)	/
	Completed	22 (100%)	11 (100%)	11 (100%)	
S13p	Not Completed	8 (36.4%)	3 (27.3%)	5 (45.5%)	> 0.05 ¹
	Completed	14 (63.6%)	8 (72.7%)	6 (54.5%)	
S14p	Not Completed	1 (4.5%)	0 (0%)	1 (9.1%)	> 0.05 ¹
	Completed	21 (95.5%)	11 (100%)	10 (90.9%)	

Metrics at nurse-level Considering the metrics at nurse level (Table 3.12 and Figure 3.13), the analysis reported no statistically significant differences among all the metrics between the two groups.

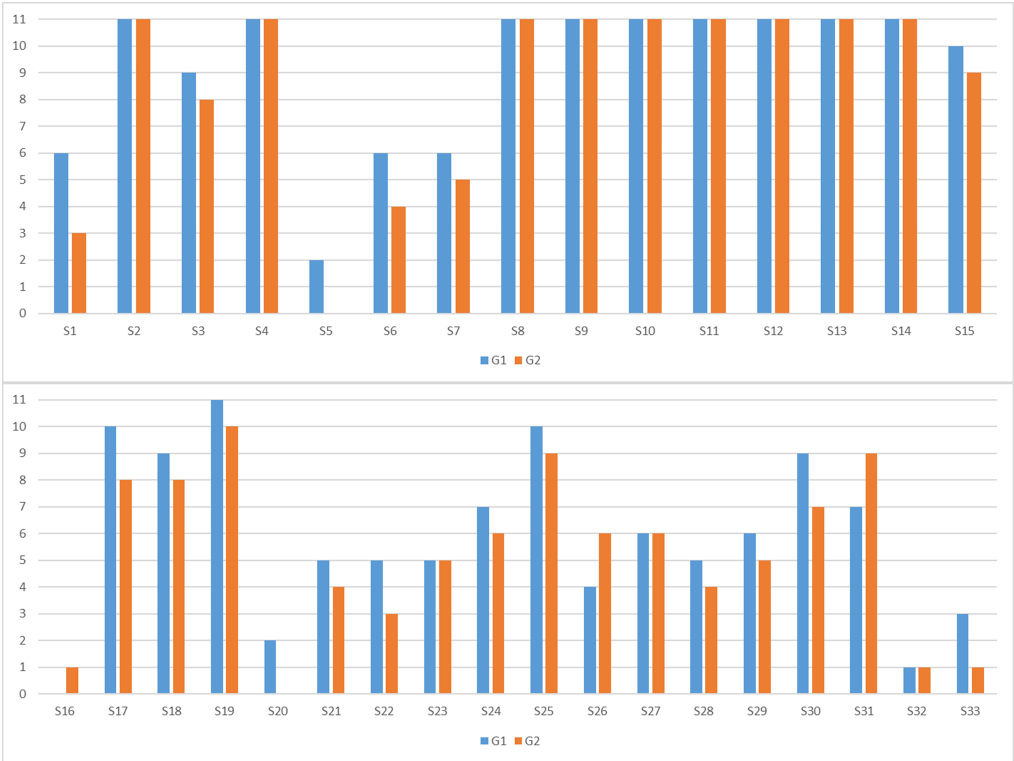


Figure 3.11: Number of nurses who completed the specific step in groups G1 (blue) and G2 (orange). The y-axis indicates the total number of nurses per group.

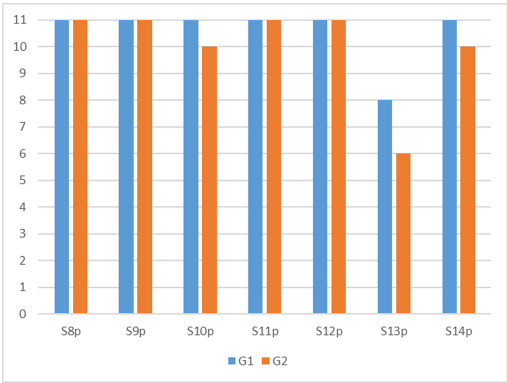


Figure 3.12: Number of nurses who completed the specific step meeting the required parameter in groups G1 (blue) and G2 (orange). The y-axis indicates the total number of nurses per group.

Table 3.12: Mean, Standard deviation, and Wilcoxon-Mann-Whitney test results for the metrics at nurse level.

	Overall		G1		G2		Wilcoxon	
	M	SD	M	SD	M	SD	z	p
Total Score	16.70	2.46	17.50	2.16	15.90	2.60	1.22	0.22
Area 1 Score	2.36	0.71	2.59	0.58	2.14	0.78	1.32	0.19
Area 2 Score	7.64	0.58	7.77	0.34	7.50	0.74	0.88	0.38
Area 3 Score	1.91	0.70	2.14	0.60	1.68	0.75	1.49	0.14
Area 4 Score	3.82	1.50	3.91	1.51	3.73	1.56	0.23	0.81
Area 5 Score	0.77	0.43	0.77	0.47	0.77	0.41	0.11	0.92
Area 6 Score	0.18	0.40	0.27	0.47	0.09	0.30	1.03	0.30



Figure 3.13: Mean scores for nurse level metrics in groups G1 (blue) and G2 (orange). Error bars represent $\pm$ SE. The y-axis indicates the full obtainable score range for the scale (minimum to maximum).

Discussion

This study aimed to evaluate the efficacy of the BLC incorporating the iVR system to enhance nurses' pre-hospital procedural knowledge of STEMI management.

Notably, nearly all nurses in both groups successfully completed the steps and parameters related to medical data acquisition via sensors (Figure 3.11, S8–S14, and Figure 3.12). In contrast, fewer than half of the participants in either group performed the steps involving the collection (Figure 3.11, S5–S7) and communication (Figure 3.11, S20–S22) of pharmacological and pathological history. Furthermore, almost none of the participants completed the item regarding communicating name and surname to the case manager doctor (S16) and future treatments to the patient (S32). This performance gap may be attributed to the nature of real-world clinical practice. On one hand, nurses typically operate within a team, thus participants may have struggled to execute all procedural steps individually without the support of colleagues they usually rely on. On the other hand, nurses may have prioritized the assessment of current symptoms (S3 and S4) and the communication of this information to case managers doctors (S18 and S19), as this reflects standard protocols for pre-hospital emergency care.

The findings provide inconclusive support for the hypothesis; while G1 participants achieved higher mean scores than G2 across several metrics, these differences did not reach statistical significance. Consequently, the hypothesis that G1 would have greater procedural knowledge could not be confirmed. This is particularly noteworthy because our previous qualitative analysis [99] indicated that participants found the iVR system immersive and motivating. However, this positive subjective perception did not translate into a statistically significant improvement in performance compared to G2. These results align with literature suggesting that integrating VR into traditional healthcare courses does not always yield superior knowledge gains compared to conventional approaches [134, 190]. The following discussion explores the potential factors that may have limited the efficacy of VR integration.

Statistical power and participants knowledge Regarding the statistical power, our power analysis was set so that the study had sufficient power to detect a large effect size (a minimum difference of 5 points in the mean total score). The observed results of measurements at nurse-level (Table 3.12), report how G1 achieved higher mean scores than G2 without statistical significance, suggesting that the true effect size is smaller than anticipated and would require a larger sample to be detected. Furthermore, an additional analysis of the initial question about self-perceived knowledge about reported in Table 3.9 revealed no statistically significant differences (Wilcoxon test, $z=0.17$, $p > 0.05$) between G1 and G2. The lack of statistically significant differences may indicate that nurses in both groups had similar self-perceived knowledge of the pre-hospital management procedure for STEMI. Therefore, it is plausible that while G1 acquired its knowledge through our structured BLC, G2 may have gained comparable knowledge from other sources, such as field experience or other educational courses, thus narrowing the measurable gap between the groups.

VR influence on steps Another potential explanation for the lack of significant differences may lie in the specific implementation of VR during the course. The system immersed nurses within a VE where they interacted with a virtual patient managed by the instructor, thereby mirroring the instructional design of the traditional course. While the iVR system enables simple interactions with medical equipment (e.g., simulating and simplifying practical tasks such as electrode placement or IV line insertion), the increased

immersion provided by the headset may not have enhanced the complex clinical decision-making required. In this scenario, nurses must still recall and execute a specific sequence of steps in response to the evolving clinical situation.

Our results aligns with existing literature suggesting that evidence for VR-based training of higher-order outcomes, such as critical thinking, remains mixed or limited [134, 191, 268].

Observation role might have mitigated results and learn by doing Another factor contributing to the lack of statistical significance may be the structure of the VR simulation activities. Throughout the course, seven scenarios were presented in each edition. However, G1 nurses engaged in active practice with the iVR system in only one scenario, while observing the remaining six. This limited hands-on engagement may have reduced the potential impact of the immersive VR training, as participants had fewer opportunities for direct active learning compared to their time spent as observers.

As reported in studies [33, 117] and a scoping review by Rogers et al. [266] in the broader field of simulation-based education, observational learning can produce similar outcomes to active participation under certain conditions. For instance, Ham et al. [117] found no statistically significant differences in knowledge or clinical reasoning between observers and active participants in a study on nursing care for hip arthroplasty. In our study, observers were actively engaged through instructor-led group discussions focused on the errors made during the VR simulations.

Therefore, the combination of limited iVR practice and the observation of many scenarios as in the traditional courses may not have provided enough differentiation to outperform traditional learning or field experience. However, since G1's scores were higher overall, it is possible that multiple hands-on sessions with the VR system may be required to produce more conclusive effects, as Woon et al. [325] suggest in their meta-analysis.

Feedback A further explanation for the lack of significant differences between G1 and G2 may relate to the feedback modality used in the iVR system and the timing of feedback delivery. In the BLC, feedback regarding the correctness of actions was provided at two specific points: first, by the instructor during the post-simulation discussion phase, and second, upon the conclusion of all six editions of the BLC, coinciding with the invitation for G1 participants to join this user study.

Providing feedback at the end of a simulation aligns with traditional simulation-based learning, where post-simulation debriefing is a common practice in nursing education [96]. While instructor-led debriefing allows for tailored feedback based on a learner's preparation level, it introduces subjectivity. Instructors must constantly balance the need to address clinical mistakes with the challenge of preserving the students' sense of psychological safety [183]. Furthermore, instructors in this study prioritized procedural errors based on clinical relevance, focusing on major mistakes while placing less emphasis on minor deviations. Consequently, participants in both G1 and G2 likely received similar feedback regarding the most critical steps of STEMI management, while less clinically relevant actions received limited attention. This selective instructional focus may have attenuated potential differences in procedural knowledge acquisition between the groups regarding the specific management of STEMI.

A primary advantage of XR-based systems is the potential to implement A&F mechanisms. Such automation would allow the iVR system to objectively identify when an action is incorrect or deviates from established protocols. This could ensure that feedback is delivered in a consistent and standardized way, either in real-time during the simulation or during a structured post-session debriefing.

Furthermore, the provision of personalized debriefing was subject to considerable delay. This personalized debriefing provided the specific mistakes made by participants within the iVR system. Because these reports were sent only after all six editions of the BLC were completed, some learners received feedback one week after their session, while others waited up to three months. This delay likely hindered the reinforcement of procedural knowledge.

Therefore, the absence of real-time feedback, coupled with the delayed provision of personalized reports and the subjectivity of instructor-dependent debriefing, likely resulted in a learning experience for G1 that was comparable to G2, attenuating the potential added value of the iVR component. Our findings align with current literature suggesting that iVR learning outcomes are frequently comparable, rather than superior, to those achieved through conventional teaching methods [62, 190, 223, 270].

Knowledge Decay and Refresh The timing of the post-training assessment may have influenced the observed results. In this study, evaluations were conducted between March and April 2023, whereas the course took place from October 2022 to January 2023. Consequently, participants were reassessed between two and six months after completing the BLC. The literature indicates that procedural skills with a substantial motor component can deteriorate markedly within weeks or months [300], while knowledge related to advanced life support generally declines after approximately six months [331].

The checklist comprises 33 distinct procedural steps based on guidelines developed in the region where the study was conducted. The first part of the checklist (items S1 to S22) involves gathering patient information and communicating it to the hospital case manager. This is standard practice in other pre-hospital protocols (e.g., in managing the n-stemi protocol ⁵) thus, nurses working in the field perform these steps regularly. In contrast, the second part (S23 to S33) involves tasks that are performed less frequently. Therefore, the absence of significant differences between G1 and G2 may also reflect a decay of procedural knowledge over time in G1 rather than a lack of initial learning. If this interpretation holds, these findings underscore the importance of regular refresher training, as recommended by CME guidelines, to mitigate knowledge loss and support long-term retention.

Conclusions

This study investigated the efficacy of integrating an iVR system into the educational protocol of an institutionale CME training course for the pre-hospital management of STEMI. While the results did not report statistically significant differences between

⁵https://www.regione.fvg.it/rafvig/export/sites/default/RAFVG/salute-sociale/sistema-sociale-sanitario/FOGLIA53/allegati/11022019_PDPA_NSTEMI-FEB.pdf

nurses who followed the BLC course and the control group, a consistent positive trend was observed in the mean scores of G1.

The positive subjective impressions reported by the participants in the qualitative analysis, alongside the upward trend in performance metrics, suggest that the value of iVR extends beyond immediate score improvements. In the context of medical education, iVR have been reported as a powerful support tool that enhances the learning experience. By providing a safe and engaging environment for high-stakes rehearsal, iVR may increase motivation, potentially encouraging nurses to engage more frequently in voluntary refresher training.

Future research should investigate training designs that explore iVR's motivational potential through repeated hands-on practice and shorter assessment intervals to better capture long-term retention and clinical readiness in time-critical care.

4

XR Training for Object Location Spatial Memory

Building on the need for cognitive training aimed at preserving and enhancing mental functions, this chapter focuses on the use of XR technologies to support Object–Location Spatial Memory (OLSM) in healthy individuals. Through two studies involving XR technologies, the chapter aims to understand how different design choices can affect training efficacy and user experience.

The first system investigates display fidelity by comparing two versions of the same serious game (headset and smartphone), evaluating how the degree of immersion influences spatial memory performance in a real-world task, as well as inside the SG, engagement, and perceived effort.

The second system examines user experience and performance during a spatial task by comparing the use of standard VR controllers with haptic gloves capable of delivering force feedback.

4.1 Effectiveness of Display and Interaction Fidelity in a SG for OLSM Training

Given the positive results in literature of iVR, nVR, and SGs for different kinds of training, and current research on the use of VR in training in spatial abilities is promising 2.7, and that research in OLSM training is limited [193], this section presents a serious game designed to train OLSM, developed for both VR headsets and mobile devices. The study evaluates user perception regarding immersion, usability, and player experience, as well as its efficacy in improving performance in a real-world OLSM task.



Figure 4.1: Setting of the VE (left) and of the real environment (right) where the user study was conducted.

4.1.1 The proposed SG

The SG to train OLSM is designed for both nVR and iVR devices, specifically mobile smartphones and VR headsets. The SG was developed with Unity 2020.3 LTS¹ and the 3D models were made with Blender².

Game VE

The SG is set in a VE (Fig. 4.1 on the left) that replicates a room from the University of Udine building (Fig. 4.1 on the right). The objects that the players interact with are common items (i.e., a white bottle, a pile of poker fiches, a blue book, a chessboard, an orange videotape cover, a black shoe box, a red bowl, and a red cube) without labels or distinctive markings to prevent the player evoking specific memories or associations. Players can place objects in 12 designated positions represented by fading black boxes, as shown in the left part of Fig. 4.1. Each of the three tables in the room has four positions above it.

Players can place objects using the quick inventory and the shelf inventory (Fig 4.2 on the right). The first, with a light-blue background, temporarily stores objects removed from the room for quick reuse and displays a numbered label indicating the quantity in the inventory. The second, with a yellow background and grid layout, provides an unlimited supply of objects. Figure 4.3 illustrates both inventories in the iVR and nVR versions.

OLSM training task and implementation in the SG

The OLSM training task we designed consists of three phases: memorization, rearrangement, and restoration.

1. Memorization: A set number of virtual objects are placed randomly in the room based on the level's difficulty. Some objects may appear multiple times (e.g., two tapes, one white bottle, one chessboard). To proceed, the player, without any

¹<https://unity.com/>

²<https://www.blender.org/>

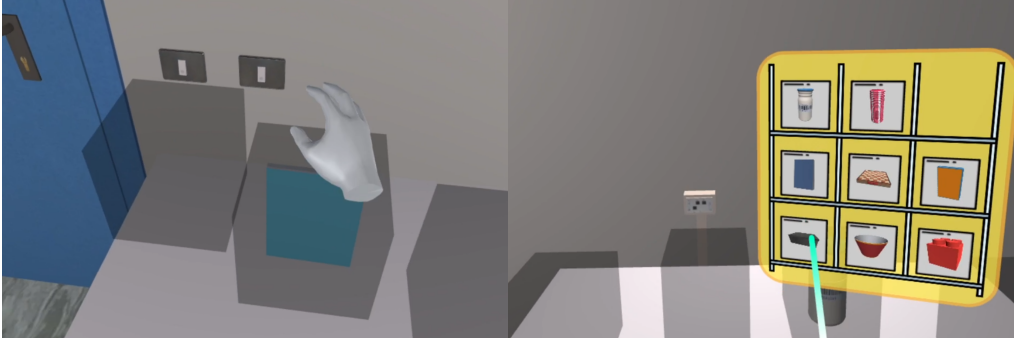


Figure 4.2: User playing with the iVR version of the SG during the memorization phase (on the left) and selecting an object from the shelf inventory during the restoration phase (on the right).

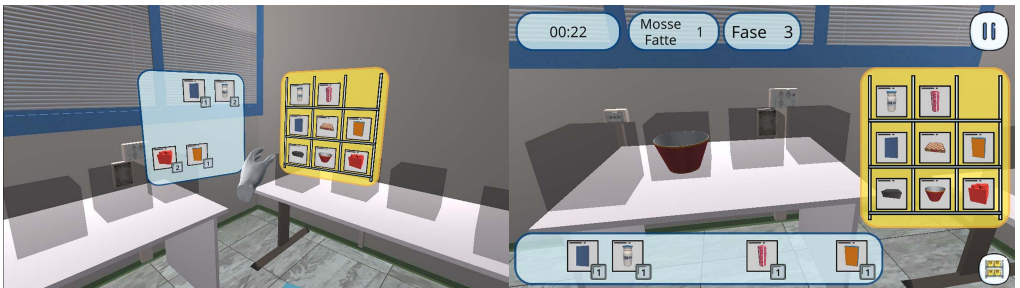


Figure 4.3: User interfaces of the Serious Game showing the quick inventory and the shelf in the iVR (left) and nVR (right) versions.

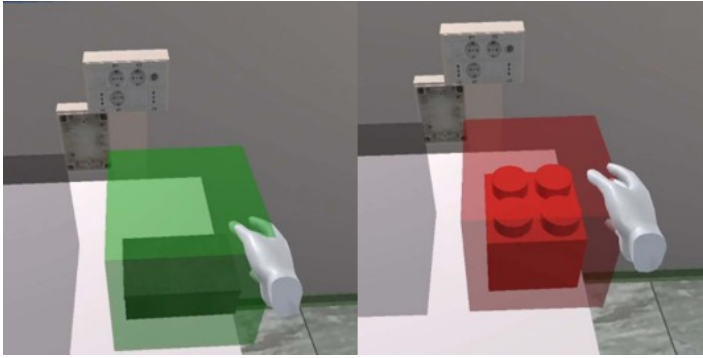


Figure 4.4: Visual feedback on the locations given to the player when the object placed is correct (on the left) or wrong (on the right).

time limit, must interact with each object to ensure they have observed all the objects they need to remember (as in Fig 4.2 on the left).

2. **Rearrangement:** The room undergoes up to four possible changes: (i) adding a new object, (ii) moving an object, (iii) removing an object and placing it in the quick inventory, or (iv) removing an object without adding it to the quick inventory. To prevent reliance on short-term visual memory, all objects are temporarily hidden. The waiting time is set to five seconds, allowing the visuospatial sketch pad to fade its maximum efficiency so the player has to create the spatial configuration of the objects over the tables in their working memory. The time of five seconds is also set aiming to maintain engagement without loss of concentration.
3. **Restoration:** The player must restore the room’s original arrangement from memory. Objects can be moved, retrieved from the quick inventory, or taken from the shelf (as in Fig 4.2 on the right). The system provides immediate visual feedback during the restoration phase to indicate whether an object has been correctly placed. When an object is positioned, the black transparent box at that location changes color to green for a correct placement (Fig 4.4 on the left) and red for an incorrect one (Fig 4.4 on the right), allowing players to instantly recognize and adjust their actions accordingly. If players fail to complete the task within the time limit, they lose the level.

SG levels progression

The SG is designed to allow the level designer to adjust difficulty by modifying parameters for each level, which are the total number of objects to remember, and how many to add, move, remove, or move to the quick inventory. In its current configuration, the first four levels introduce game mechanics, while subsequent levels gradually increase in difficulty so players can play at their current skill level, possibly contributing to reaching the flow state [231]. The “target time” value estimates the level’s difficulty by calculating the time a player has to complete the restoration phase to earn the highest trophy

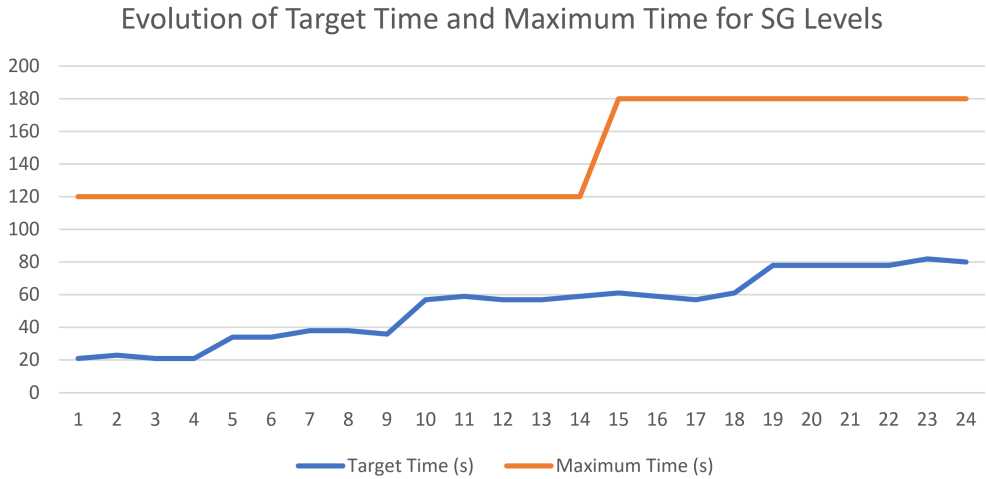


Figure 4.5: Player's target time in the SG to obtain the best trophy for the different levels (bottom blue line) and maximum time to complete the level (top orange line). The X axis indicates the levels, the Y axis indicates the time in seconds.

successfully. This value is based on all level parameters, assuming two seconds per required move and an increase relative to the total number of objects to remember. The "maximum time" is calculated by rounding the "target time" up to the next full minute, then adding one additional minute to allow extra time. As a result, the average player is expected to complete the restoration phase of each level within the range between the target time to earn the highest trophy and the maximum time before failure, while the more skilled players can complete the restoration phase closer to the target time or even faster. Figure 4.5 shows how the target time and the maximum time change across the 24 levels of the SG.

To encourage progression in the SG, players earn a trophy upon completing a level, the quality of which (bronze, silver, and gold) depends on completion time and the number of errors (i.e., fewer mistakes and more remaining time result in a better trophy).

Interactions with the SG in iVR and nVR

In the iVR version (Figure 4.3, on the left), players move physically in the VE to reach the different locations. User interface elements appear as floating posters, and interaction involves pointing a ray from the player's hand and pressing the controller's trigger button. To remove an object, players grab it while pressing the grip button and move it outside the designated area. To add an object, players open the quick inventory (left controller button) or shelf (right controller button), then either point the ray at the desired object and press the trigger button to spawn it in the air or grab it directly by pressing the grip button. Finally, they place the object by moving it to the desired location and releasing the grip button.

In the nVR version (Figure 4.3, on the right), players can see the VE by moving the



Figure 4.6: A user playing the SG developed in the iVR version (on the left) and nVR version (on the right).

smartphone, exploiting its gyroscope, and pinching the screen to zoom. They interact with the user interface via touchscreen, while UI elements overlay the environment. To remove an object, players tap it, adding it to the quick inventory. To place an object, they select it from the inventory and either tap the desired location or use drag-and-drop from the inventory or shelf.

Figure 4.6 shows a player playing with both versions of the SG.

4.1.2 User Study

The user study aims to examine how the iVR and nVR versions of the SG impact player experience and OLSM training. We designed a between-subjects study, with two groups: the iVR group uses a VR headset, and the nVR group plays on a mobile device.

We formulated the following hypotheses:

H1) We expect that the iVR group will feel a higher sense of presence than the nVR group as in the related work [41, 45, 140, 192, 194, 239, 240].

H2) We expect that both groups will perform better in a OLSM task after being trained with the SG, considering the positive effects of VR and SGs on training in general [56], on cognitive training [49] and on training in spatial abilities [124].

Regarding users' ability improvement to perform the OLSM task in the real world after the training, the study takes an exploratory approach. According to the framework proposed by [205], a higher sense of presence can lead to greater satisfaction, which in turn may enhance learning outcomes. However, contrasting findings from [8] suggest that increased spatial presence may actually impair memory recall. Moreover, some previous studies have reported better training outcomes in iVR conditions [42, 316], others have found nVR systems to be more effective for delivering procedural knowledge [325], and others found no significant differences in training outcomes attributable to the level of immersion [40, 275].

Finally, the study is explorative considering usability, players' experience, and progress

in the SG: on one hand, the nVR group may find the SG more usable, have a smoother experience, and progress faster in the SG because of the use of a familiar device³; on the other hand, the iVR group might benefit from the natural interaction with the tracked controllers and the higher sense of presence (if H1 is confirmed).

Materials

We conducted the user study in the room of the University of Udine that was presented in Fig. 4.1 on the right. This room is dedicated to user studies, with access restricted to authorized personnel only and windows covered to prevent participants from being distracted by outside stimuli. The nVR version has been tested with the Zte Axon 7 Android smartphone. The iVR version has been tested with the HTC Vive Pro with the application running on a desktop computer with Intel i7-3820 processor and Nvidia GTX 1070 graphics card.

The performance of the player within the SG has been recorded with a logger included in the SG that saves four kinds of data: i) the level's current configuration, ii) every action of the player inside the application (e.g., objects added and removed during the gameplay, iii) the user results at the end of the level, and iv) the hardware characteristics of the device the SG is running on.

Measures

Sense of presence We measured sense of presence with the Igroup Presence Questionnaire [279]. It includes a specific item for the sense of "being there" IPQ-G1P, three different subscales (spatial presence IPQ-SPP, involvement IPQ-INVP, and realness of the simulation IPQ-REALP), and a global value of presence IPQ-TOTP calculated by averaging all the items, after reverting the scale of negative items. These scales range from 0 to 6, where 6 indicates the greatest sense of presence.

Usability The usability of the SG was evaluated with the System Usability Scale (SUS) [39] questionnaire, which includes 10 items and gives results on a scale 0-100.

Player Experience We measured player experience using the Player Experience Inventory [4], which includes 30 items assessing experience on two levels. The first, Functional Consequences (PXFUN), captures immediate reactions to game design elements such as audiovisual appeal and ease of control. The second, Psychosocial Consequences (PXPSY), reflects emotional and social responses as a second-order effect of those game design choices. Both PXFUN and PXPSY are scored on a scale from 1 to 7, with 7 indicating the most positive result.

Progress within the SG We measured the progress in the SG with the logger inside the application. The MLW value indicates the maximum level the players have won in their gameplay.

³https://www.radicati.com/wp/wp-content/uploads/2021/Mobile_Statistics_Report,_2021-2025_Executive_Summary.pdf

Performance in an OLSM task in the real world We designed a real-world task that was similar to the one in the SG. We attached tape on the tables in the real room (Figure 4.1, on the right) to indicate the 12 locations of the objects, and the participants could use two boxes, one in the center of the room as the quick inventory and another at the side of the room as the shelf with all the objects. The selected objects for the real-world task were: a box of paper clips, an envelope, a blue folder, a green folder, a box, and a remote controller.

The task is structured in three phases that correspond to those in the SG:

1. **Memorization.** Participants have two minutes to remember the position of the seven items placed above the crosses marked with tape over the three tables in the room. They are allowed to move freely around that part of the room but are not permitted to move the objects.
2. **Rearrangement.** The experimenter performs the following room modifications: relocating an object, introducing a new object, placing an object in the quick inventory box, and completely removing an object from the tables. To prevent participants from observing these changes, the experimenter brings them outside the experiment room. The modifications take approximately 30 seconds, and bringing the participant in and out of the room adds about another 30 seconds.
3. **Restoration.** Participants are informed that some objects have been added, removed, and moved compared to the original setup. Without any time limit, they have to restore the object configuration based on their memory and indicate completion by saying “I’m done.” The experimenter then starts a stopwatch to measure the restoration time and, upon task completion, stops the timer and takes a photo of the final arrangement to assess errors later.

After completing the restoration phase, the experimenter removes all the objects from above the tables. The participant performs the task two times, one before and one after the gameplay session. Figure 4.7 shows the objects configuration of the first task (top row) and second task (bottom row) before (left column) and after (right column) the rearrangement.

The number of errors and the time elapsed during the restoration phase are the measures used to track the user performance in the task.

Participants

We involved a sample of 51 students, all from the computer science bachelor degree programs at the University of Udine. After excluding those involved in pilot tests ($N=4$) and those who had technical problems during the experience ($N=2$), the data of 45 participants (43M, 2F) were used for the analysis. Participants’ age ranged from 20 to 26 ($M=21.44$, $SD=1.47$). We asked participants how many hours they play with videogames (VGH) on average every week ($M=6.56$, $SD=6.70$). We also asked participants if they had previously used VR headsets and, if they had, for how many hours (VRH): 29 had not used them, while 16 participants had used them ($M=1.75$, $SD=1.26$). Participants were assigned to the two groups (iVR: 22 participants; nVR: 23 participants) in such a way that the groups were similar in terms of gender (both the iVR and the nVR group



Figure 4.7: Object configurations for the first (A, B) and second (C, D) real-world OLSM tasks. Configurations A and C show the initial setups to be memorized by the participant, while B and D show the room arrangements after the experimenter’s modifications.

had one female participant), age (iVR: $M=21.36$, $SD=1.26$; nVR: $M=21.52$, $SD=1.68$), frequency of video game use (iVR: $M=7.05$, $SD=6.96$; nVR: $M=6.09$, $SD=6.57$), and previous experience with VR headsets (iVR: $M=1.79$, $SD=1.47$; nVR: $M=1.71$, $SD=1.16$). Each of the variables for which we aimed at minimizing between-group differences was submitted to a t-test that confirmed no statistically significant difference between the two groups.

Procedure

The experimenter told the participants we were testing a SG that aimed at training memory. The experimenter also informed participants that VR users might suffer from nausea or headache in rare cases, and they could refrain from continuing the experiment at any time and for any reason. The experimenter then explained that he would ask participants to fill out some questionnaires in an anonymized form. After they gave written informed consent, the experimenter asked participants to turn off or silence their smartphone to minimize distractions during the experiment and then asked them to complete an initial questionnaire concerning the information described in Section 4.1.2.

Then, after the preparation of the room, the experimenter explained and asked participants to perform the first OLSM task in the real world as described in Section 4.1.2.

After removing all the objects from the tables, the experimenter asked participants to move in the middle of the tables and provided them with a one-minute explanation on how to interact with the technology (headset for the iVR version and smartphone for the nVR one). Regardless of the version, participants played with the SG for 15 minutes (as the user in Figure 4.6). For the iVR version of the SG, the experimenter positioned himself where he could monitor participants for their safety without interfering with their experience. For the nVR version, the experimenter stood behind a column in an alcove of the room, out of the participants' view, to avoid distracting them.

Then, the experimenter asked participants to fill out the IPQ, SUS, and PXI questionnaires. Finally, participants performed the second OLSM task in the real world as described in Section 4.1.2.

4.1.3 Results

All data were statistically analyzed with SPSS software⁴, version 28. Results concerning the sense of presence, usability, player experience, and progress within the SG were analyzed with the t-test. Results concerning the performance in the OLSM task were analyzed with a mixed-design ANOVA where the version of the SG is the between-subjects variable, and the instant of measurement (pre or post training with the SG) is the within-subjects variable.

Sense of Presence

Regarding presence (Table 4.1), the analysis revealed statistically significant differences between the groups for several measures. The general sense of being there (IPQ-G1P, $p <$

⁴<https://www.ibm.com/spss>

Table 4.1: Mean, standard deviation, t-test results, and effect size for presence (IPQ scales), usability (SUS), player experience (PXFUN and PXPSY), and progress in the SG (MLW).

	iVR		nVR		t-Test			Effect size
Measures	M	SD	M	SD	t	df	p	Cohen's d
IPQ-G1P	4.59	1.33	3.09	1.16	4.04	43	< .001	1.20
IPQ-SPP	4.35	0.76	3.18	1.08	4.16	43	< .001	1.24
IPQ-INVP	4.11	1.21	3.39	1.36	1.88	43	> .05	0.56
IPQ-REALP	2.92	0.66	2.68	0.81	1.07	43	> .05	0.32
IPQ-TOTP	3.89	0.67	3.09	0.85	3.47	43	< .005	1.04
SUS	75.80	12.01	79.13	7.93	-1.10	43	> .05	0.33
PXFUN	5.50	0.75	5.46	0.76	0.17	43	> .05	0.05
PXPSY	4.84	1.02	4.26	1.12	1.83	43	> .05	0.55
MLW	11.73	1.86	14.57	2.54	-4.27	43	< .001	1.28

0.001) and the spatial presence subscale (IPQ-SPP, $p < 0.001$) both showed statistically significant differences, as did the overall presence score (IPQ-TOTP, < 0.005), with higher means in the iVR group compared to the nVR group and a large effect size for all of them. Instead, no statistically significant differences were found for involvement (IPQ-INVP, $p > 0.05$) or experienced realism (IPQ-REALP, $p > 0.05$).

Usability

Regarding usability (Table 4.1), the analysis showed no statistically significant difference between groups in the overall System Usability Scale (SUS, $p > 0.05$).

Player Experience

Regarding player experience (Table 4.1), the analysis showed no statistically significant differences between groups in both functional consequences (PXFUN, $p > 0.05$) and psychosocial consequences (PXPSY, $p > 0.05$) scores.

Progress in SG

Regarding the progress in the SG (Table 4.1), the analysis revealed that the difference in the MLW variable was statistically significant ($p < 0.001$), with higher means in the nVR group compared to the iVR group and a large effect size.

OLSM task in the real world

Figure 4.8 shows the number of errors and the time elapsed in the pre and post (on x axis) OLSM tasks in the real world and the version of the SG (separate columns). The analysis showed a main effect of the instant of measurement for the number of errors ($f(1,43)=6.51$, $p < 0.05$, $\eta_p^2 = 0.13$) and the time elapsed ($f(1,43) = 21.03$, $p < 0.001$, $\eta_p^2 = 0.33$), both higher before the training with the SG, with a medium effect size for the number of errors and a large effect size for the time elapsed. Instead, there was

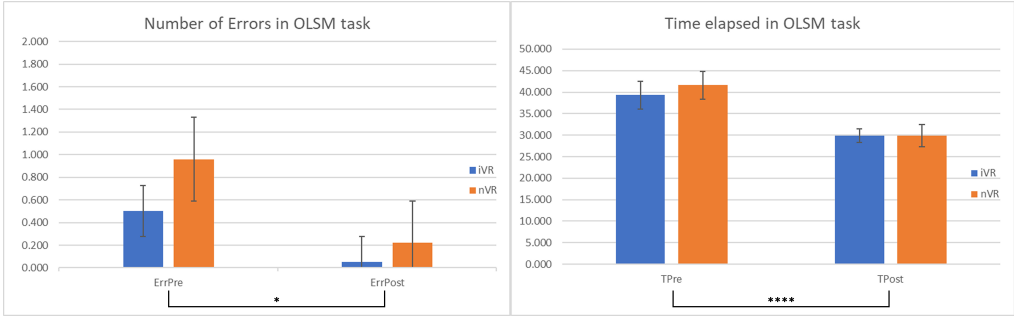


Figure 4.8: Number of errors (left part) and time elapsed (right part) in the pre and post (on x axis) OLSM tasks in the real world and version of the SG (separate columns). Capped vertical bars show standard error. The * and **** signs indicate differences with p-values respectively < 0.05 and < 0.001

no main effect of the version of the SG either for the number of errors ($f(1,43) = 1.90$, $p > 0.05$, $\eta_p^2 = 0.04$) or for the time elapsed ($f(1,43) = 0.15$, $p > 0.05$, $\eta_p^2 < 0.01$). There was also no interaction between the independent variables either for the number of errors ($f(1,43) = 0.37$, $p > 0.05$, $\eta_p^2 = 0.01$) or for the time elapsed ($f(1,43) = 0.23$, $p > 0.05$, $\eta_p^2 = 0.01$).

4.1.4 Discussion

The aim of this study was to explore how the iVR and nVR versions of the SG affect users' sense of presence, usability of the SG, player experience, and progress within the SG, and if the training with the SG could enhance users' ability to perform OLSM tasks in the real world.

As hypothesized (H1), participants who played the SG in the iVR version reported a significantly higher sense of presence than those in the nVR version. This was reflected in: the general sense of “being there” (G1P), the spatial presence subscale (IPQ-SPP), and the overall presence score (IPQ-TOTP). These results confirm our hypothesis, align with previous findings in the literature [41, 45, 140, 192, 194, 239, 240], and extend them to a SG designed for OLSM training. Notably, according to the interpretation of IPQ presence scores of [216], the average IPQ-TOTP for the iVR version is satisfactory, while the average IPQ-TOTP for the nVR version is unacceptable. No statistically significant differences were observed for involvement (IPQ-INVP), although the effect size was medium. Moreover, the average IPQ-INVP for the iVR version is satisfactory, while it is unsatisfactory for the nVR version. Statistical significance was not reached also for experience realism (IPQ-REALP), but for this subscale this is likely because both versions used the same 3D models. Notably, for both versions, the average IPQ-REALP scores were unsatisfactory according to interpretation of [216]. This can be due to the absence of detailed textures (e.g., with labels or distinctive markings) that was intended to prevent the player evoking specific memories or associations as described in Section 4.1.1.

Regarding usability, both versions of the SG received high SUS scores, with no

statistically significant difference between the two versions, and a small effect size. Since the minimum acceptable SUS score for usability is 70, and both versions scored above 75 (iVR=75.79, nVR=79.13), participants found both versions of the SG to have good usability [27]. The lack of differences between iVR and nVR aligns with the results of [42, 192, 240], suggesting that, despite most of the participants lacked prior experience with VR headsets, the natural interaction offered by the use of tracked controllers in the iVR version may have compensated the familiarity with the smartphones.

Regarding player experience at the functional consequences level (PXFUN), the mean scores (iVR=5.50, nVR=5.46) indicate that players gave a positive evaluation to both versions of the SG, with no statistically significant differences between groups, and a negligible effect size. This high similarity in the PXFUN scores is likely due to the fact that the core game content remains the same, and the differences lie only in the interaction methods and participants rated the usability good for both versions of the SG, as described above. For psychosocial consequences (PXPSY), while no statistically significant difference was found, the mean scores (iVR=4.84, nVR=4.26) suggest that players in the iVR group reported a more favorable emotional response, with a medium effect size. A possible explanation for the medium effect size for the PXPSY score is the novelty effect, since, for many participants, it was the first time they used an immersive VR headset, and this may have influenced their overall perception of the experience.

Regarding progress within the SG, a statistically significant difference was found in the maximum level reached. On average, the nVR group completed three more levels than the iVR group. This result is different from the one of [240], who did not find differences between the final level reached by their participants. A possible explanation for our result is that participants in the iVR group might have spent part of the gameplay time familiarizing themselves with the VR headset and the tracked controller (especially those using a VR headset for the first time) rather than immediately focusing on progressing through the levels.

Considering the users' ability improvement in performing the OLSM task in the real world, the mixed-design ANOVA confirmed hypothesis H2. Specifically, both the number of errors and the time elapsed significantly decreased between the pre-test and the post-test, supporting the effectiveness of the SG for OLSM training and extending to this kind of training the positive effects of VR SGs from the literature [56]. Although participants in the nVR group completed more levels, thus training in more complex OLSM tasks, their improvement in the real-world OLSM task was comparable to that of the iVR group. Statistical analysis found no significant difference in effectiveness between the two SG versions (no main effect of the SG version and no interaction), and both groups showed similar benefits. In the literature, some studies have reported superior performance in iVR conditions when training specific abilities [42, 316], while others have found nVR systems to be more effective for delivering procedural knowledge [325]. Meta-analyses have also concluded that no single VR setup is universally superior across training domains [133]. These findings suggest that the benefits of higher immersion may be task-dependent rather than universally applicable. This perspective is supported by [275], who reported that the success of VR protocols in rehabilitating Unilateral Spatial Neglect, a spatial awareness disorder, depended more on other aspects than the level of immersion alone.

In particular, our design of the OLSM task mainly relies on an allocentric strategy,

as the positions of the objects are anchored to fixed landmarks in the environment, such as tables (e.g., the blue book is to the right of the table under the window). While both versions of the SG support the allocentric object-to-object frame of reference, the iVR version potentially supports also the egocentric self-to-object frame, as users can remain stationary while exploring the environment by simply turning their heads, thus they can more easily define a consistent egocentric frame of reference (e.g., “in front of me,” “to my left,” “to my right”). Instead, in the nVR version, users must physically turn around to explore the VE and observe the positions of the different objects, which makes it more difficult to rely on egocentric self-to-object references. Possibly, in a OLSM task relying more on the egocentric frame, the iVR version might lead to better performance.

Another possible explanation for the lack of statistically significant differences between the iVR and the nVR versions on the improvement in the real-world OLSM task can be that the possibly positive effect of completing a higher number of levels in the nVR version was compensated by the greater sense of presence experienced in the iVR version, which could have enhance training outcomes as suggested by [205].

Overall, the study provides evidence that the proposed SG can effectively train users' spatial abilities, with both the iVR and nVR versions contributing to an improvement in an OLSM task. Interestingly, the increased sense of presence of the iVR version did not result in significantly better performance in the real-world OLSM task compared to the nVR version. Therefore, while iVR can offer a more immersive experience, the comparable outcomes suggest that even the non-immersive version of the SG can effectively support the training of spatial abilities. These results suggest that future research should explore the development of mobile-based OLSM SGs, which, due to the widespread availability and familiarity of smartphones, have the potential to benefit a broader population compared to the number of individuals who own an iVR headset at home. This approach was already successfully explored in the context of aviation safety, where [40] showed that an nVR and an iVR version of the same SG led to comparable knowledge gain and retention, and then released a more rich nVR version of the SG called “Prepare For Impact” on mobile app stores, reaching several million users [65].

4.1.5 Limitations

Some limitations of our study should be acknowledged. First, the sample of involved participants was not diverse, as they were mostly young individuals with a background in computer science, and only two female participants were included. This limits the generalizability of the findings to broader populations. Future studies should involve a more heterogeneous sample to assess whether the observed trends hold across different age groups, genders, as well as educational and technological backgrounds.

Second, the lack of a statistically significant difference in the real-world OLSM task may be due to the task not being sufficiently challenging to reveal differences in training effectiveness between the two SG versions. The low error rates in the post-test (iVR=0.05; nVR=0.22) suggest that the task may have been too easy, leading to a ceiling effect. Therefore, future research should consider OLSM tasks with increased complexity to assess whether, in this case, a difference between the iVR and nVR versions could be noted.

Considering the two limitations together, the lack of statistically significant differences in OLSM task performance may have been influenced by the sample used in the

study. The same real-world OLSM task might yield different outcomes if it were conducted with individuals affected by medical conditions that impair spatial abilities. In such cases, the considered task might be considered more challenging by the users and lead to higher times and errors. To support the training of individuals with spatial impairments, the SG's adjustable level parameters (described in Section 4.1.1) could be used to tailor the difficulty, making the SG more suitable for enhancing OLSM skills in that specific population.

A possible additional limitation that could have reduced the advantages of the iVR version of the SG is the use of controllers instead of bare-hand or glove-based hand tracking. A more natural interaction might have led to shorter familiarization times and a higher number of completed levels with the iVR version of the SG, possibly resulting in a performance improvement in the OLSM task.

4.1.6 Conclusion and future work

Considering the positive results of VR and SGs in training found in the literature and the limited research on VR and SGs for OLSM training, we proposed a SG for OLSM training, also comparing a iVR and nVR version and evaluating their efficacy with a user study.

Both versions were found usable, and participants positively perceived the experience, but the iVR group had a greater sense of presence compared to the nVR group, while the progress in the SG was greater for the nVR group. OLSM results show that playing the SG leads to a statistically significant improvement in the performance of a real-world OLSM task (fewer errors and less time elapsed after playing), regardless of the SG version used.

The task designed in this study primarily engages an allocentric object-to-object frame of reference, emphasizing relationships independent of the user's position. It would, therefore, be interesting in future research to compare the two SG versions using a task that relies more on an egocentric self-to-object reference frame. Such a comparison could help determine whether the iVR version provides advantages compared to the nVR version for that kind of task.

Since our study involved young people with a computer science background, we plan to extend our results to people of different ages and technological competencies in future studies. In particular, we would like to exploit the SG to train older adults with impairments related to spatial memory.

It will also be interesting to study the long-term effects of the SG by training OLSM with it over time, e.g., by deploying the nVR version on mobile app stores and the iVR version on stores available from VR headsets. Finally, following the results of [118], it would be interesting to add cues and other stimuli near the objects in the room and then compare the results with and without those stimuli to eventually support mnemonic strategies for remembering object locations in a rehabilitation context.

4.2 Effects of Haptics in an XR System for Object-Location Spatial Memory Training

This section details an experiment in which users perform an OLSM task using either handheld controllers or force-feedback gloves, investigating how different haptic devices influence performance and metrics of presence, usability, and fatigue.

Given the positive results of using iVR in OLSM tasks and the enhanced experiences users perceive with FGs, this work will investigate if and how FGs could improve OLSM performance in a specific task in iVR.

4.2.1 User Study

This user study investigates whether the hardware used to interact in the iVR can influence OLSM. More precisely, two interaction methods with virtual objects are compared: 6-DOF tracked handheld controllers (HCs) and 6-DOF tracked force-feedback gloves (FGs).

We designed a task to train OLSM that consists of remembering different objects' spatial locations. The task is composed of four steps:

1. Memorize the locations of different objects over a table.
2. Move the objects one by one to a box over another table.
3. When all objects are moved inside the box, wait for the objects to be shuffled.
4. Move the objects from the box to the starting table in the memorized locations.

The task is repeated three times with an increasing level of difficulty. The difficulties were chosen according to Miller [222] magic number 7 ± 2 of uncompressed chunks in working memory[211]. The first level (3 items) has few objects to remember, so the users can perform the task while familiarizing themselves with the system. The difficulty of the second level (5 items) was chosen as the lower bound of the magic number, and the third level (10 items) as the upper bound with an additional object to give a higher challenge.

Design and Hypotheses

We conducted a within-group user study to evaluate the HCs and FGs interaction methods. Half of the users tried the HCs first and the FGs next, while the other half did the opposite. All the users attempted the three different levels of the application two times, one for each interaction method.

The within-group design was chosen to balance the possible differences in the OLSM ability of the participants, since each user provides data for both conditions, and to improve the statistical efficiency, as pointed out by Greenwald [113]. The main problem of this design is that the subjects are susceptible to the possible effect of practice, sensitization, and carryover [113]. In this user study, however, these effects are limited for these reasons: i) the task is focused mainly on WM, and the task does not involve LTM; ii) the objects to remember are not significative, and thus they should not stimulate

LTM; iii) between the two trials, there is a break from performing the task where the participants have to fill out the questionnaires and, therefore, it should empty the WM; iv) the groups are counter-balanced in the order they tried the two conditions.

Besides evaluating the performance of the task, we considered other aspects of the experience that may have an impact on training: i) presence may influence the user's focus on the OLSM task rather than being distracted by external distractions in the real environment; ii) the system's usability could lead to a high task success rate, as pointed out by [167], and thus it may lead to more successful training of the OLSM; iii) muscle fatigue could lead to a deterioration in cognitive performance [324], thus to an inefficient effect of the OLSM training.

We made seven hypotheses: four regarding the performance of the OLSM tasks and three for the overall experience.

- H1) We expected that users would remember more object locations with the FGs than HCs throughout the levels because, as described in Section 1, the FGs could provide haptic inputs about the shape of the objects. This FG capability can contribute to updating users' spatial image of the environment [195], enhancing the OLSM.
- H2) We expected that the proportion of correctly placed objects after the shuffle would be higher in the first two levels than in the third one, because the number of objects in the third level is higher than the upper bound of Miller [222] magic number, while it is lower in the first two levels.
- H3) We expected that users would complete the task in less time with the FGs than HCs because they can use their hands like in the real world, thus using the same abilities to move objects acquired during their lifetime.
- H4) We expected that the average time elapsed to handle an object would be higher in the first level than in the last two, because the first time the users try a new interaction method they would need to familiarize with it.
- H5) We expected that users would feel a greater sense of presence with the FGs because of the more natural interaction with the objects, in line with the results seen in [226].
- H6) We expected that users would find the application more usable with the FGs because of the more natural interaction with the objects, in line with the results seen in [226].
- H7) We expected that users would feel more fatigue in the fingers and the wrist with the FGs because of the weight of the glove, as suggested by [255].

Materials

Figure 4.9 shows the VE used for the OLSM task. The VE is composed of an empty room to limit distractions, the two tables necessary for the task, and the instructions in the users' native language. There are several slots based on the number of objects to remember over the table on which objects are located at step 1 of the task. The objects

used for the OLSM task are 3D shapes of different solids: a cube, an ovoid, a capsule, a sphere, and a cylinder. The application has three levels of difficulty as described above.

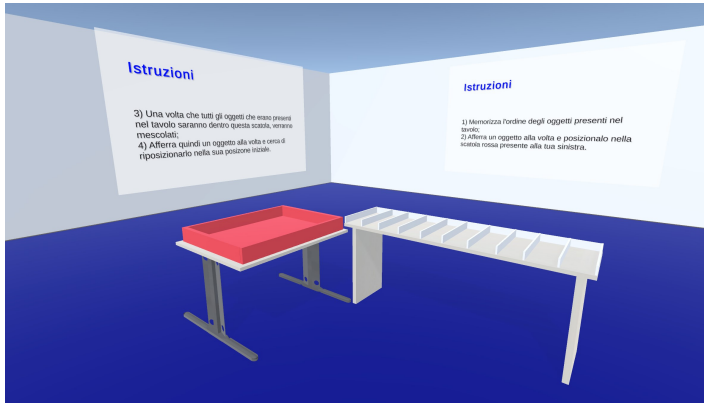


Figure 4.9: Virtual Environment used for the task composed by an empty room, the two tables, and the instruction in the users' native language.

The HCs employed for the task were the HTC Vive Pro Controllers ⁵. Users could press or release the “Grip” or “Trigger” buttons to grab or drop objects. Instead, the FGs worn by users were the SenseGlove DK1 gloves ⁶. Users could grab and release objects like in real life by closing or opening their hands near a target object. Figure 4.10 presents an example of the user interacting with the HC and the FG. In both conditions, users could see the VE using a commercial VR headset (HTC Vive Pro).



Figure 4.10: The two interaction methods used to handle virtual objects. On the left side, the user has the HC (HTC Vive Pro Controller) in his hand, and on the right side, he is using the FG (Senseglove DK1).

The application was developed with Unity 2020.3 LTS ⁷ and the Open XR framework

⁵<https://www.vive.com/us/accessory/controller/>

⁶<https://www.senseglove.com/>

⁷<https://unity.com/>

⁸. A logger in the application saved in a file the total number of correctly placed objects after the shuffle and the time elapsed to complete the task. The simulation was run on a PC equipped with a 3.30 GHz Intel i7-5820K processor, 16 GB RAM, and a Nvidia GTX 980 graphic card.

Participants

The evaluation involved a sample of 24 participants (17M, 7F). Participants were volunteers that received no compensation and were recruited through personal contact. Age ranged from 20 to 28 ($M = 22.71$, $SD = 2.01$). We asked participants if they had previously used any HMDs and, in case of a positive answer, for how many hours. Answers of the 9 participants who had previously used an HMD ranged from 0.3 to 20 hours ($M = 3.59$, $SD = 6.31$). We also asked all participants how many hours per week they use video games. Answers ranged from 0 to 18 hours ($M = 4.44$, $SD = 5.34$), with 7 of them not playing at all and 4 of them playing at least 10 hours.

Measures

Performance of the OLSM task The performance was evaluated with the data gathered by the application logger. More precisely, we considered two measures: i) the proportion of correctly placed objects (PO) after the shuffle (correctly placed objects over total objects in that level); ii) the average time elapsed (TE) to handle an object during the OLSM task at each level (total time spent for the task without the waiting time over total objects in that level).

Sense of presence We measured sense of presence by administering participants the Igroup Presence Questionnaire [279]. It includes a general item for the sense of “being there” (IPQ-G1P), and 13 other items organized in three different subscales (spatial presence IPQ-SPP, involvement IPQ-INVP, and experienced realism IPQ-REALP). The global value of presence IPQ-TOTP is calculated by averaging answers from all items. Global value and subscales can range from 0 to 6, where 6 indicates the greatest sense of presence.

Usability The system usability was evaluated with the well-known SUS questionnaire [39], which includes 10 items and gives the result on a scale from 0 to 100.

Fatigue The fatigue was measured with the Device Assessment Questionnaire (DAQ) [90], which includes 13 items concerning the required force, smoothness, mental and physical effort, the accuracy and speed of the movements, the fatigue in different parts of the body, the general comfort, and the ease of use. Each item can be rated on a 5-point scale, ranging from 1 to 5. For items 1, 3, 4, 6, the most desirable value is at the central point (3). For items 2, 12, 13 the most desirable value is the highest endpoint (5). For items 5, 7, 8, 9, 10, 11 the most desirable value is the lowest endpoint (1).

⁸<https://www.khronos.org/OpenXR/>

Procedure

The experimenter told participants we were testing different interaction methods in an iVR experience. He also informed participants that VR users could suffer from nausea or headache in rare cases, and they could refrain from continuing the experiment at any time and for any reason. The experimenter then explained that he would ask participants to fill out some questionnaires in an anonymized form. After they gave informed consent, participants completed an initial questionnaire concerning the information described in Section 4.2.1.

Then, the experimenter asked participants to go to the middle of the room, and he helped them wear the HMD and the HCs or the FGs. As described in Section 4.2.1, half of the participants tried the HCs first and the FGs next, while the other half did the opposite. For the HCs, the experimenter explained the “trigger” and “grip” buttons to interact with for grabbing the objects. Instead, for the FGs, the experimenter asked the participants to move their fingers to calibrate the gloves and told them that they can grab and drop the virtual objects like in the physical world.

Then the experimenter explained the task participants had to perform in the iVR experience as described in Section 4.2.1. The experimenter told the participants to notify him when they had finished the task so he could make the application proceed to the next level.

After completing all levels for the first interaction method, the experimenter helped the participants to remove the devices and invited them to fill out the three questionnaires described in Section 4.2.1: IPQ, SUS, and DAQ. Then participants tried the second interaction method and filled the questionnaires a second time.

4.2.2 Results

All data have been statistically analyzed with SPSS, version 28⁹. Results concerning performance in the OLSM task were analyzed using 3x2 repeated measures ANOVA, since each participant tried all the three levels with both interaction methods. When Mauchly’s test indicated that the assumption of sphericity was violated, degrees of freedom were corrected using Greenhouse-Geisser. When we found a main effect of level, we proceeded with Bonferroni pairwise comparisons. When we found an interaction between the two independent variables, we analyzed simple effects, considering the effects of level separately for each of the two interaction methods, and the effects of interaction method separately for each level. Results concerning measures of presence and usability were analyzed using paired t-Tests. Results concerning the items of DAQ questionnaire were analyzed using Wilcoxon Signed-Ranks Tests.

Performance

Figure 4.11 shows the results of the performance measures for each level (on x axis) and condition (separate lines).

Considering PO (Figure 4.11a), we found no main effect of interaction method, $F(1, 23) = 0.00$, $p > 0.05$, $\eta_p^2 = 0.00$, a main effect of level, $F(2, 46) = 61.41$, $p < 0.001$, $\eta_p^2 = 0.73$, and no interaction, $F(2, 46) = 0.68$, $p > 0.05$, $\eta_p^2 = 0.03$. Post-hoc analyses

⁹<https://www.ibm.com/products/spss>

showed that PO was higher in the first two levels than in the third, $p < 0.001$, and no statistically significant difference between the first and the second level, $p > 0.05$.

Considering TE (Figure 4.11b), we found a main effect of interaction method, $F(1, 23) = 74.15$, $p < 0.001$, $\eta_p^2 = 0.76$, a main effect of level, $F(1.59, 36.55) = 28.79$, $p < 0.001$, $\eta_p^2 = 0.56$, and an interaction between the two independent variables on TE, $F(1.24, 28.62) = 5.91$, $p < 0.05$, $\eta_p^2 = 0.20$. TE was higher with FG than HC, $p < 0.001$. Post-hoc analyses showed that TE was higher in the first than in second and third level, $p < 0.001$, and that TE was lower in the second than in the third level, $p < 0.05$. The analysis of simple effects showed that TE was higher with FG than HC at all levels, $p < 0.001$. When interaction method was FG, TE was higher in the first than in the last two levels, $p < 0.001$, and no statistically significant difference was found between the second and the third level. When interaction method was HC, TE was higher in the first than in the second level, $p < 0.005$, TE was lower in the second than in the third level, $p < 0.05$, and no statistically significant difference was found between the first and the third level.

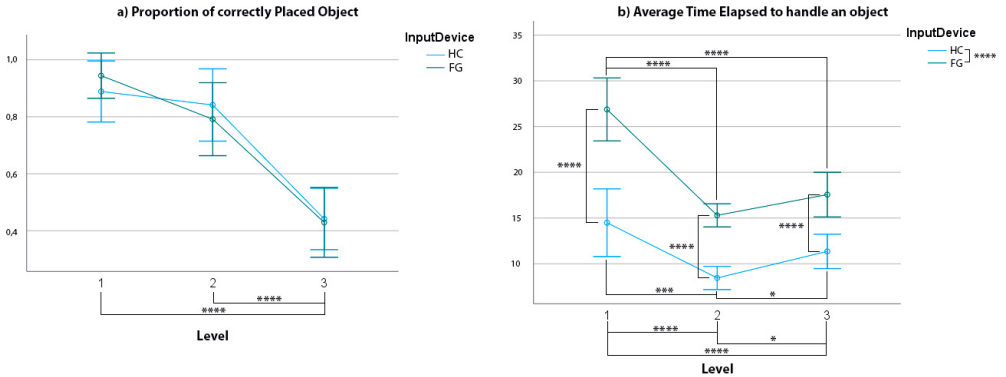


Figure 4.11: PO (a) and TE (b) for each level (on x axis) and condition (separate lines). Capped vertical bars show 95% CI. The *, **, *** signs indicate differences with p-values respectively < 0.05, < 0.005, < 0.001.

Presence

Considering presence (Table 4.2), the analysis revealed that the difference in the means for the general item about the sense of being there (IPQ-G1P) was statistically significant, $p < 0.05$. The mean was higher with FGs than with HCs. The differences in the means for the spatial presence subscale (IPQ-SPP), the involvement subscale (IPQ-INVP), the experienced realism subscale (IPQ-REALP), and the global value (IPQ-TOTP) were not statistically significant.

Usability

Considering usability (Table 4.2), the analysis revealed no statistically significant difference for the overall SUS score.

Table 4.2: Mean, Standard Deviation, and paired t-Test results for IPQ scales and usability.

Measures	HC		FG		Paired t-Test		
	M	SD	M	SD	t	df	p bilateral
IPQ-G1P	4.50	1.14	4.92	1.02	-2.10	23	< .05
IPQ-SPP	4.48	1.06	4.85	0.51	-1.63	23	> .05
IPQ-INVP	3.96	1.19	4.14	1.09	-0.95	23	> .05
IPQ-REALP	2.66	0.97	3.05	0.85	-1.94	23	> .05
IPQ-TOTP	3.81	0.84	4.14	0.61	-1.85	23	> .05
SUS	83.54	11.28	83.13	11.23	0.14	23	> .05

Table 4.3: Mean, Standard Deviation, and Wilcoxon Signed-Ranks Test results for the items of the DAQ questionnaire.

Item		HC		FG		Wilcoxon Signed-Ranks Test	
		M	SD	M	SD	Z	p
1	Required force	1.08	0.28	1.58	0.65	-2.97	< 0.005
2	Smoothness of movements	4.21	0.59	3.63	0.65	-2.73	< 0.01
3	Mental effort	2.25	0.99	1.96	1.00	-1.71	> 0.05
4	Physical effort	1.25	0.44	1.29	0.55	-0.33	> 0.05
5	Accuracy of the movements	2.25	0.85	2.46	0.93	-0.82	> 0.05
6	Speed of the movements	3.00	0.51	3.04	0.46	-0.33	> 0.05
7	Finger fatigue	1.13	0.34	1.79	0.88	-3.18	< 0.005
8	Wrists fatigue	1.08	0.28	1.54	0.93	-0.24	< 0.05
9	Arms fatigue	1.17	0.38	1.38	0.65	-1.51	> 0.05
10	Shoulder fatigue	1.13	0.34	1.33	0.70	-1.67	> 0.05
11	Neck fatigue	1.29	0.75	1.29	0.75	0.00	> 0.05
12	General comfort	4.33	0.70	3.92	0.72	-1.85	> 0.05
13	Overall ease of use	4.79	0.42	4.54	0.59	-1.90	> 0.05

Fatigue

Regarding perceived fatigue (Table 4.3), the analysis of the DAQ questionnaire reported that the differences in the means for Item 1 ($p < 0.005$), Item 2 ($p < 0.01$), Item 7 ($p < 0.005$), and Item 8 ($p < 0.05$) were statistically significant. Instead, the differences in the other DAQ items were not statistically significant.

4.2.3 Discussion

Considering H1, we expected that users would remember more object locations with the FGs than HCs, thanks to the FG's capability of suggesting the shape of the objects and the possible influence of haptic feedback in updating the spatial image of the environment[195]. We could not confirm our hypothesis since there were no statistically significant differences in the PO.

Hypothesis H2 was instead confirmed by statistically significant differences in the means that showed how, on average, the PO was higher in the first two levels than in the third one for both interaction methods. This result aligns with the literature about the limited capacity of the WM. More precisely, users had reached higher values of PO in levels 1 (3 items) and 2 (5 items) than in level 3 (10 items) because, in the last level, the total number of the objects was over the upper bound of Miller’s magic number [222], and thus users’ WM might not have been capable of memorizing all the locations.

Considering H3, we expected that users would employ less TE to perform the task with the FGs than HCs because they could use the abilities already acquired during their lifetime to grab and move objects in the VE. Since results in terms of TE showed that HCs led participants to spend much less time performing the task, we rejected H3. The differences in the TE means were statistically significant in favor of the HCs for all levels. The HC interaction method took, in general, about 60% of the time to complete the OLSM task compared to the FGs.

Regarding H4, we expected a higher TE in the first level than the last two because of the need for familiarization with a new interaction method. The main effect of level supported our hypothesis. However, the significant interaction and the analysis of simple effects showed that the difference between the first and the second level was statistically significant for both HCs and FGs, and the difference between the first and the third was significant only when using the FGs.

A possible explanation for these results is that participants could not fully take advantage of the FG capability of providing users with haptic inputs about the shape of the objects over HCs because of the fatigue and discomfort of using those FGs, as shown by the DAQ results. In particular, it is possible that the fatigue from using the FGs could impede some users’ movements, leading them to spend more effort for opening and closing their hands rather than simply clicking the “grip” or “trigger” button of the HCs. Furthermore, the statistically significant difference in the means for Item 2 of the DAQ questionnaire showed how the perceived smoothness of movements is greater for the HCs.

Interestingly, in an object identification task, the combined results of Kreimeier and Götzelmann [172] and Kreimeier et al. [173] showed that FGs significantly increased the execution time with respect to HCs, contradicting the authors’ hypothesis as in our OLSM task. In [173], authors explained this discrepancy between hypothesis and results with the bulkiness of the HGs, which are the same that we used in our study.

The absence of differences between the interaction methods for the OLSM task regarding PO and the lower TE could also indicate that this type of OLSM task does not benefit from specialized FGs. A possible explanation is that participants remembered the items regardless of the haptic capabilities of the devices because they focused more on visual than haptic feedback to remember the shape of the objects to be placed at the different memorized locations. Even if the combination of visual and haptics are useful to build a representation of the near-body space [199], in our particular type of task with the employed FGs, the haptic input could not improve the spatial image of the environment in a significant way.

Considering hypothesis H5 regarding presence, we expected that users would feel a greater sense of presence with the FGs because of the more natural interaction with the VE, in line with the results seen in [226]. Indeed, we found a statistically significant

difference in favor of the FGs for the general item about the sense of “being there” (G1P). However, we could not confirm our hypothesis about the sense of presence for the global value (IPQ-TOTP) and the subscales of IPQ. The differences suggested that using the FGs could increase presence, but the data collected did not reach statistical significance.

Regarding hypothesis H6 about system usability, both interaction methods received high mean SUS scores with a negligible difference between the two. Therefore, we rejected the hypothesis about the greater usability of the FGs compared to HCs. This contrasts with the results in [226]. Interestingly, compared to the usability results in [255] for the same commercial FGs, users in our study reported higher usability values. A possible interpretation is that our task consisted of remembering the object locations rather than performing different actions that could lead users to focus more on grabbing and moving the objects in the VE.

Results confirmed hypothesis H7: the participants felt more fatigue in the fingers and wrists with FGs, as suggested by [255]. Statistically significant results concerning Item 1, 7, and 8 of the DAQ questionnaire showed that using FGs required participants to use more force to interact with the system and caused participants more fatigue in their wrists and fingers. In addition, despite statistical significance was not reached, the difference between the means for fatigue of the other body parts (Item 9 and 10) might suggest that FGs could increase the fatigue also of the arms and the shoulders. However, for both interaction methods, it must be noted that the average reported values for fatigue were very low.

4.2.4 Limitations

This user study is a preliminary evaluation of a system for OLSM training conducted with young people with no history of Alzheimer’s disease. This evaluation serves as a preliminary step to identify which characteristics of the system could affect OLSM training. Therefore, future work with patients is needed to assess the effectiveness of the system on the target users.

4.2.5 Conclusion

In this experiment, we investigated the influence of haptic feedback on an OLSM task in iVR by comparing a pair of FGs and a pair of HCs to perform the same OLSM task. Regarding presence, statistically significant differences were found in favor of interacting with FGs for the general item about the sense of being there. Regarding usability, participants judged the system almost equally usable when interacting with HGs and FGs. There was a statistically significant difference in the fatigue experienced after the task with the FGs, possibly because of the position of the weight of the gloves. The fatigue and discomfort of the FGs could have influenced the OLSM task results, in which there were no statistically significant differences about PO and participants employed less time to do the task with the HCs. This opens the possibility that other, more comfortable, and lighter FGs may lead to different results.

We plan to carry out a study on a OLSM task using a different, new model of FGs to check if it could lead to better results. If new studies would instead confirm the lack of differences between the FGs and HCs, we could conclude that a basic commercial VR

kit may be sufficient to perform this type of OLSM tasks without requiring specialized hardware. The absence of a requirement for specialized hardware could be helpful in the rehabilitation context, where patients, especially those with locomotion disabilities, would be able to do the rehabilitation at home without necessarily needing costly special hardware that could be available only in specialized clinics.

Another important aspect that needs to be evaluated in rehabilitation is the transfer of the improved spatial skills from the VE to the real world. In this direction, we plan to develop an OLSM task for rehabilitation in iVR concerning an everyday activity to investigate if the training in the VE improves spatial memory in the physical world.

A final aspect that could be evaluated for rehabilitation is the influence of performing the OLSM task in a multi-user setting, where users could collaborate or compete with each other. Indeed, the results of [185] showed different behaviors for single and pairs of users in an OLSM task not concerning rehabilitation, so it would be interesting to explore if rehabilitation could be more effective in a multi-user setting than doing the rehabilitation exercises individually.

The work presented in this section was published in the Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter [98].

5

XR Training for Executive Functions

Building on evidence that Executive Function performance can be enhanced through targeted cognitive training and physical exercise [88, 115], this chapter focuses on the use of XR Serious Games aimed at training executive functions. Through studies that include healthy validation and clinical applications, the chapter aims to understand how different design choices influence performance and user experience across populations, and how theoretical models enable rigorous analysis of user and patient behavior during gameplay.

The first study examines the effects of adaptive difficulty and multi-user interaction in a sorting task, serving as a preliminary validation with healthy participants before engaging patients with mild Neurocognitive Disorder (mNCD). This section evaluates performance and engagement by comparing adaptive versus linear difficulty progression and single-player versus multi-player modes.

The subsequent section introduces a theoretical model of an XR SG that proposes a Go/No-Go task based on the single-player adaptive difficulty-scaling version of the previously proposed SG. The theoretical model aims to predict user behavior across different cognitive health conditions.

In conclusion, the chapter presents the design, preliminary implementation, and preliminary results of a clinical protocol established to evaluate an enhanced version of the SG featuring expanded levels and interaction fidelity (immersive vs non-immersive VR), with older adults exhibiting mNCDs.

5.1 Effects of Difficulty Adaptation and Multi-user Setting in an XR SG for Sorting Tasks

Recent literature indicates that exergames can improve global cognition, highlighting their potential to enhance physical fitness while also providing cognitive improvements

[55, 88, 115]. Specifically, studies have shown improvements in executive functions, attentional processing, and visuospatial skills [294].

Building upon this premise, we propose an immersive Serious Game that involves a sorting task. The system aims to train Executive Functions in older adults diagnosed with mNCD within a fully immersive VE. Following standard clinical validation protocols, the system will be evaluated first with healthy participants and subsequently with the target patient population.

Driven by positive findings in the literature regarding DDA (Section 2.4) and multi-user settings (Section 2.5), this study investigates the optimal conditions for maximizing both user performance and experience.

To this end, we propose a system that implements: i) distinct difficulty scaling modes, a linear progression versus an adaptive mode based on real-time performance data; and ii) two gameplay modalities, single-player and multi-player. The adaptive system aims to maintain the player in the flow state [77], thereby minimizing boredom or frustration that can arise when the challenge level does not match the player's current skill. The multi-user setting allows us to investigate how social presence and collaboration may influence performance and perceived motivation during gameplay.

5.1.1 The Proposed Serious Game

Task in the Game

The core gameplay consists of a sorting task in which players categorize and throw balls that emerge from two lateral sliders. Throwable balls are divided into three categories: red, blue, and junk (identified by distinct stains, regardless of their base color).

To earn points, players must match the color of the ball to the corresponding container. Junk items must be discarded by throwing them onto the floor. Figure 5.1 illustrates the four types of objects and their respective target containers.

Players can grab a ball by moving their hand near it and pressing the "Trigger" button on the controller. To throw, players mimic a throwing motion and release the button.

Game Virtual Environment and Navigation

The SG is set in a minimalist virtual gym designed to minimize cognitive overload (Fig. 5.2). The player is positioned in front of a central table equipped with two lateral sliders that continuously generate objects. By alternating the generation of items (e.g., if a ball is created on the left side, the next one will be on the right), the game prevents any physical advantage for right- or left-handed users.

The central table features a concave design to gather the balls, serving as the primary interaction zone from which players grab objects before throwing them into the rear containers.

The SG includes a familiarization phase in which all ball types are continuously generated, allowing the player to practice the throwing mechanic. Figure 5.1 illustrates the VE during this phase.

To prioritize safety and mitigate the risk of cybersickness, the SG is designed for a seated experience. This limits the risky behaviors that can arise when the user has to

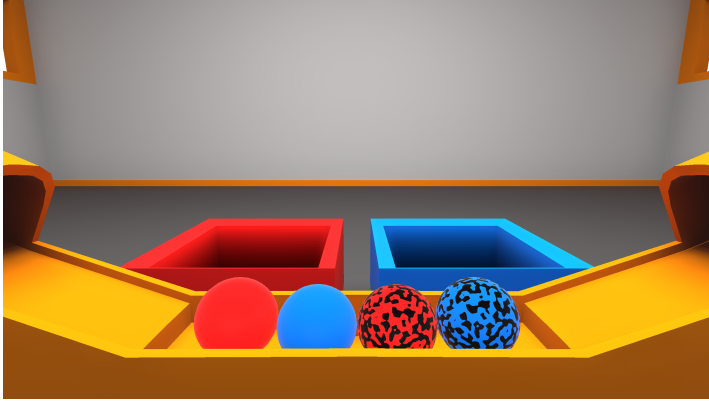


Figure 5.1: Types of balls to throw: red, blue, stained red (junk), and stained blue (junk). Behind them are the containers in which the balls have to be thrown (red on the left, blue on the right).

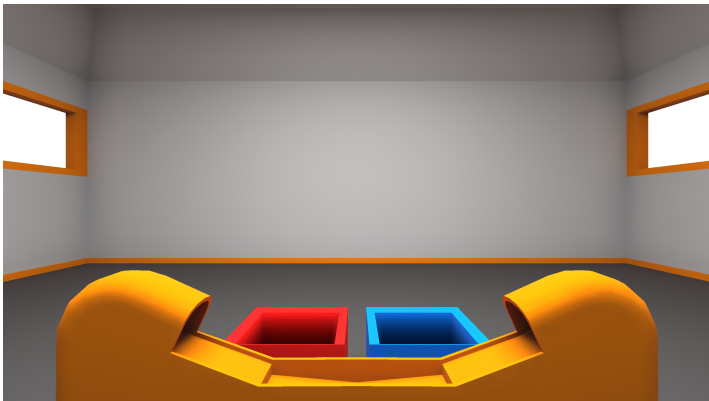


Figure 5.2: Virtual Environment of the sorting task. Objects move down the sliders toward the central table, where the player sorts them into the red or blue container or onto the floor.

walk in the physical world but cannot see it because they are wearing a VR headset.

Scoring and Progression Systems

A correct throw rewards the player with a base score of 5 points. While incorrect throws (e.g., placing a red ball in a blue container) do not deduct points, they do disrupt the combo system. Discarding junk on the floor is a required task, but does not yield points.

To increase player engagement and encourage skillful play, we implemented a combo system:

- **Multipliers:** Every sequence of five consecutive correct throws increases the score multiplier (1x to a maximum of 4x).
- **Resets:** The multiplier resets to 1x if a wrong throw occurs or if a valid ball is dropped on the floor.

Furthermore, a competitive ranking system features simulated scores gained by other players. To provide a realistic sense of competition, these scores are spaced approximately 500 points apart, with the top rank positioned about 1,000 points ahead. The gameplay interface, including the combo tracker and scoreboard, is shown in Figure 5.3.

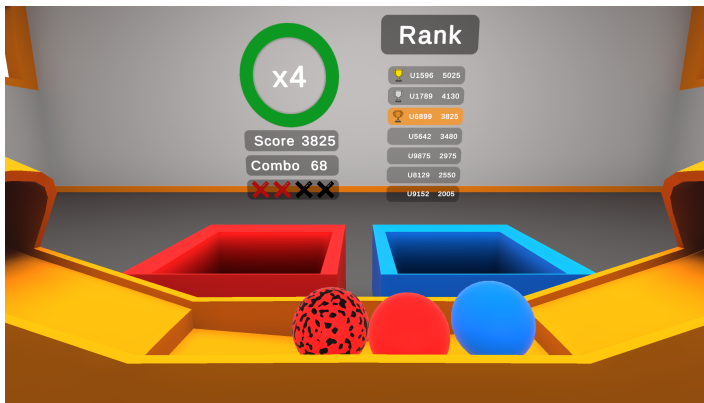


Figure 5.3: Gameplay user interface. On the left: current combo multiplier, total score, current combo, and error icons. On the right: real-time leaderboard highlighting the player's current rank.

Game Over conditions

The game-play session terminates under two conditions:

1. **Table Overflow:** If more than four balls accumulate on the table because the player struggles to throw the balls.
2. **Accuracy Threshold:** If the player commits four sorting errors (e.g., misplacing a ball in the wrong container). Dropping a ball on the floor is treated as a

"miss" rather than a "sorting error." This distinction prevents penalizing players for physical dexterity issues or accidental drops while still enforcing the rules of the categorization task.

Difficulty Scaling

The difficulty of the game sessions is determined by three key parameters:

- Spawn Interval: The time elapsed between the generation of consecutive balls.
- Visual Similarity: The density of stains on junk objects, which dictates how closely they resemble standard balls. Figure 5.4 shows the extremes of this similarity range.
- Junk Frequency: The probability that a generated object will be a "junk" item rather than a standard red or blue ball.

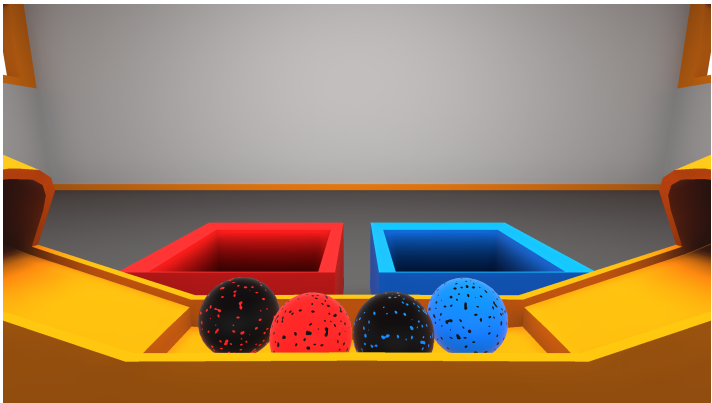


Figure 5.4: Comparison of junk object similarity. The figure shows the minimum and maximum stain densities for both red and blue variants.

The SG features 10 levels of difficulty, with parameters scaling linearly from Level 1 to Level 10. Following pilot testing, the boundary values were established as shown in Table 5.1.

Table 5.1: Difficulty parameter scaling from Level 1 to Level 10.

Parameter	Level 1 (Minimum)	Level 10 (Maximum)
Spawn Interval	3.00 seconds	0.75 seconds
Visual Similarity	20%	80%
Junk Frequency	0%	33.3%

At Level 1, there is a 50/50 probability of spawning either a red or blue ball. By Level 10, the distribution shifts so that there is an equal 1/3 probability for a red ball, a blue ball, or a junk object. The linear progression of these three variables across all

levels is visualized in the difficulty plots (Fig. 5.5) for spawn interval, visual similarity, and junk frequency.

Progression Modalities and Difficulty Adaptation

The SG features two distinct progression modes: a fixed linear progression and a DDA system. Difficulty adjustments occur every 30 seconds, a period defined as the *SessionDuration*.

Fixed Progression In the default mode, the difficulty increases from Level 1 to Level 10 at fixed intervals. The system triggers a level increment at the end of each *SessionDuration*. Under this regime, a player reaches the maximum difficulty level after 270 seconds of gameplay, provided that the game-over conditions (i.e., table clutter or excessive errors) are not met.

DDA To better accommodate varying player skill levels, we implemented an adaptive modality based on player performance using two normalized metrics within the range $[0, 1]$:

- **Throw Accuracy (*TA*):** The ratio of correct actions (e.g., throwing the blue ball in the blue container, or the ball with stains on the floor) to the total number of balls thrown.
- **Table Management Efficiency (*TME*):** The percentage of the *SessionDuration* during which two or fewer objects are present on the table. A high *TME* suggests the player is clearing objects faster than they are generated, indicating that the current difficulty level may be insufficient to maintain the flow state.

To determine the difficulty adjustment, these two metrics are averaged and scaled to a 100-point Performance Score (*PS*):

$$PS = \frac{TA + TME}{2} \times 100$$

At the conclusion of each *SessionDuration*, the difficulty level is adjusted according to the logic presented in Table 5.2. The values were determined after pilot testing.

Table 5.2: DDA Level Adjustment Logic

Performance Score (<i>PS</i>)	Level Adjustment
$PS < 80$	-2
$80 \leq PS < 85$	-1
$85 \leq PS < 90$	0
$90 \leq PS < 95$	+1
$PS \geq 95$	+2

This logic aims to maintain the player within the flow state by dynamically responding to their proficiency. If a player demonstrates near-perfect performance, the system

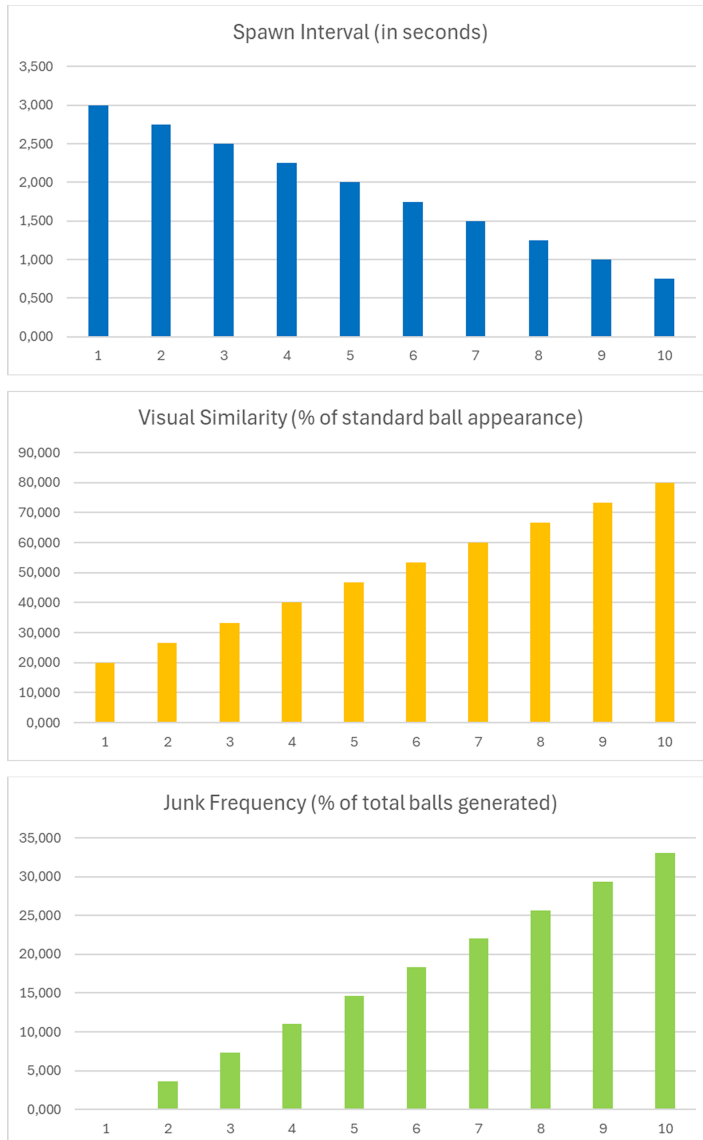


Figure 5.5: Bar plot of each variable that defines a level's difficulty.

accelerates the difficulty to reach a challenging state more quickly. Conversely, if the player struggles, the system reduces the pressure to prevent frustration and disengagement.

Multi-user Implementation

In the multi-user mode, the SG employs a cooperative multi-player format, where two players simultaneously perform the sorting task. In this mode, the individual points earned by each player are aggregated into a single cumulative score displayed on a shared scoreboard. The SG implements a collective loss condition, meaning that the failure of a single player results in a simultaneous loss for both players.

In the virtual environment, an additional set of baskets and sliders is added, visualized in front of each player, mirroring the workspace and real-time actions of their partner. This allows both participants to remain aware of each other's actions without leaving their designated stations. Figure 5.6 illustrates this shared virtual environment.

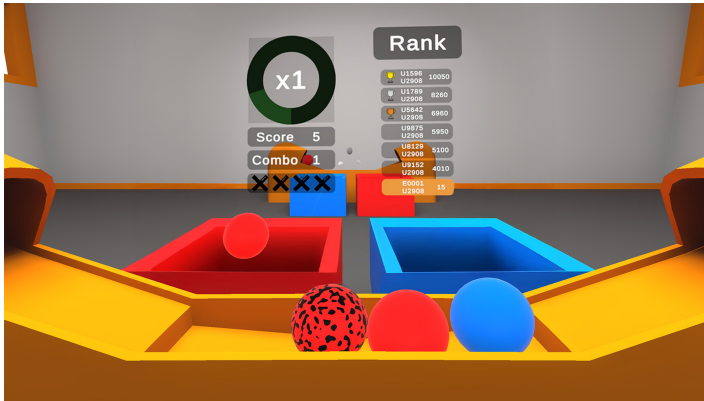


Figure 5.6: Multi-player setting in the VR. Players are placed in front of each other but far enough away to limit distractions.

The difficulty scaling logic varies based on the selected mode:

- **Linear Progression:** Difficulty is globally synchronized. Both players experience identical level increments simultaneously, regardless of individual performance.
- **DDA Progression:** Difficulty is calculated independently for each participant. While both players follow the DDA logic, their current difficulty levels may diverge based on their specific performance scores (*PS*). For instance, a highly proficient player may reach Level 8, while their partner remains at Level 5. This approach aims to maintain the game at an appropriately challenging level for both users, regardless of their individual skill levels.

5.1.2 User Study

The primary objective of this study is to determine which experimental condition yields better user experience and performance within the SG.

Design and Hypothesis

We employed a 2×2 between-subjects factorial design with two variables: Difficulty Scaling (Linear or Adaptive) and Game Mode (Single-player or Multi-player). The four groups are defined as depicted in Table 5.3.

Table 5.3: Experimental Conditions and Group Assignments

Group	Game Mode	Difficulty Scaling
G1	Single-player	Linear
G2	Single-player	Adaptive
G3	Multi-player	Linear
G4	Multi-player	Adaptive

To ensure consistency, the same designated experimenter acted as the second player for every participant. As an expert, this experimenter maintained a uniform level of play. This approach aimed to ensure that match outcomes depended solely on the participant’s performance, thereby avoiding the variability introduced by different partners.

We formulated the following hypotheses:

- **H1:** Participants in the *Adaptive Difficulty* condition will exhibit greater progress in the SG. By mitigating difficulty during moments of struggle, the system is expected to prevent premature failure, thereby extending gameplay duration.
- **H2:** We posit that dynamic adjustments will better maintain players within the flow state [77] by continuously balancing challenge against player skill preventing boredom or frustration. Consequently, we hypothesize that participants in the *Adaptive Difficulty* condition will report a higher quality of user experience compared to those in the *Linear Difficulty* condition, as suggested by Chen [58].

Finally, the study adopts an explorative approach regarding the Game Mode. On one hand, the cooperative multi-player mode could enhance user experience and performance due to the Köhler effect, which occurs when a less capable individual performs better in a team situation than they would if performing individually [234]. On the other hand, the single-player environment prioritizes user autonomy, which is a critical factor in player motivation [271].

Materials

Software and Hardware The SG was developed for desktop devices using Unity 2022.3 LTS¹. All 3D assets and environment models were created using Blender². To interact with the SG, a virtual reality headset is required.

Player performance was monitored by an internal logging system that recorded various metrics upon game termination, including the final score and game duration. In multi-player mode, the system tracked the progress of each participant individually, generating a distinct log file for each user.

¹<https://unity.com/>

²<https://www.blender.org/>

During the study, the SG was executed on a workstation (CPU: i9-10900X, GPU: NVIDIA Quadro RTX 4000, RAM: 32GB) and played via a Meta Quest Pro headset.

Questionnaires Data collection was conducted using the online software PsyToolkit [296, 297]. The study employed the following questionnaires:

- *Initial Questionnaire*. Gathered demographic data (gender, age) and assessed prior experience (in hours) with Head-Mounted Displays (HMDs) and weekly hours of sports and video games.
- *Self-Assessment Manikin (SAM)* [36]. A 3-item non-verbal pictorial assessment measuring the Pleasure (SAM_P), Arousal (SAM_A), and Dominance (SAM_D) associated with the participant's affective reaction to stimuli. For each scale, there are 5 images representing the intensity of an emotion. Participants are asked to indicate how they feel on that scale. As intermediate states are also valid, participants have to select one of the 9 possible values for each of the three scales.
- *Igroup Presence Questionnaire (IPQ)* [279]. A 13-item questionnaire measuring the sense of presence in a Virtual Environment (VE). Items are scored on a scale from 0 to 6 (where 6 indicates maximum presence). It comprises a general item on the sense of "being there" (IPQ-G1) and three subscales: (i) Spatial Presence (IPQ-SP), (ii) Involvement (IPQ-INV), and (iii) Realness (IPQ-REAL). A global presence score (IPQ-TOT) is calculated by averaging all items after reverse-scoring negative items.
- *Game Experience Questionnaire (GEQ)* [141]. This questionnaire requires the participant to indicate the degree of agreement with several items in a 5-point Likert Scale. The GEQ includes three distinct modules:
 - (i) The *Core Module* (33 items) assesses feelings during gameplay across seven dimensions: Competence (GEQ_CO), Sensory and Imaginative Immersion (GEQ_SII), Flow (GEQ_F), Tension/Annoyance (GEQ_TA), Challenge (GEQ_Ch), Negative Affect (GEQ_NA), and Positive Affect (GEQ_PA).
 - (ii) The *Social Presence Module* (17 items) assesses Empathy (GEQ_SP_E), Negative Feelings (GEQ_SP_NF), and Behavioral Involvement (GEQ_SP_BI).
 - (iii) The *Post-Game Module* (17 items) assesses participants' feelings after play: Positive Experience (GEQ_PE), Negative Experience (GEQ_NE), Tiredness (GEQ_Ti), and Return to Reality (GEQ_RtR).

Measures

Performance within the SG was analyzed using data extracted from the internal logger. Two primary metrics were used to quantify progress:

- **Game Duration:** The total time elapsed before game termination.
- **Final Score:** The maximum score achieved by the participant during the session.

User experience was evaluated across four distinct dimensions using the questionnaires detailed in Section 5.1.2:

- **Affective Reaction:** Measured via the SAM, assessing Pleasure, Arousal, and Dominance.
- **Presence:** Measured via the IPQ to determine the sense of presence within the VE.
- **Game Experience:** Measured using the *Core* and *Post-Game* modules of the GEQ, to assess how participants felt during gameplay and right after the end of it.
- **Social Presence:** Measured using the *Social Presence* module of the GEQ. This was administered exclusively to participants in the Multi-player conditions (G3 and G4).

Ethical Approval

The study was approved by the Institutional Review Board of the Department of Mathematics, Computer Science, and Physics of the University of Udine.

Participants

A total of 146 students were involved in the study. The students were enrolled in the Bachelor's degree programs in Digital Philosophy or Computer Science at the University of Udine.

Data from 34 participants were excluded from the final analysis. The exclusion criteria were as follows: participation in pilot evaluations ($n=23$), hardware malfunctions ($n=4$), and environmental distractions (noise) during the session ($n=2$). Additionally, we excluded a participant who did not provide written informed consent ($n=1$), an extreme outlier in VR experience (>300 hours, $n=1$), the only non-binary participant (because of the impossibility to balance this gender among the four groups having only one participant, $n=1$), a participant who failed to understand game rules ($n=1$), and a participant who was unable to play due to a physical injury (hand brace, $n=1$).

The final sample consisted of 112 participants evenly distributed across the four experimental conditions ($n = 28$ per group). The number of female participants was similar across groups (G1: 6, G2: 5, G3: 6, G4: 6). Table 5.4 presents the descriptive statistics derived from the initial questionnaire. To ensure that no statistically significant differences were present among the groups, a one-way ANOVA was conducted, and the test revealed no statistically significant results.

Procedure

The experimenter informed participants that they would be testing a SG. Participants were informed that VR usage might occasionally cause nausea or headaches, and they were free to withdraw from the experiment at any time. After providing written informed consent, participants were asked to silence their smartphones to minimize distractions and completed the initial questionnaire detailed in Section 5.1.2.

Table 5.4: Participant demographics including age, HMD experience, weekly sport activity, and weekly video game usage (Mean $\pm$ SD).

Group	Age	HMD Exp	Sport	Videogame
G1	23.80 $\pm$ 8.67	0.95 $\pm$ 2.11	3.39 $\pm$ 3.62	7.53 $\pm$ 9.03
G2	24.50 $\pm$ 10.40	1.44 $\pm$ 5.67	4.05 $\pm$ 3.12	5.31 $\pm$ 6.68
G3	23.00 $\pm$ 5.18	0.64 $\pm$ 1.12	4.31 $\pm$ 2.80	6.89 $\pm$ 10.30
G4	23.50 $\pm$ 6.71	1.14 $\pm$ 3.78	4.45 $\pm$ 3.20	6.61 $\pm$ 9.53

Participants were then seated and introduced to the VE. They received instructions on how to interact with virtual objects via the controller. The experimenter explained the game mechanics, which required participants to sort colored balls into corresponding baskets while discarding "dirty" balls onto the floor. Participants were informed that accumulating excessive sorting errors or allowing the counter to become overcrowded would result in a "Game Over." In such instances, participants were instructed to select "Play Again" to start a new gameplay session.

Before performance data collection began, the familiarization phase of the SG allowed participants to practice the task in the VE. They were encouraged to ask clarifying questions during this time.

For participants in the multi-player conditions (G3 and G4), the researcher explained that they would be playing cooperatively.

Once ready, participants started the gameplay session by pressing "Play" and engaged in consecutive game sessions for a total of 20 minutes.

Finally, the experimenter asked participants to complete the SAM, the IPQ, the GEQ Core module, the GEQ Post-game module, and the PXI questionnaires (see Section 5.1.2). Participants in groups G3 and G4 additionally completed the GEQ Social Presence module after the GEQ Core questionnaire.

5.1.3 Results

Statistical analyses were performed using SPSS Version 28.0.1.1³. A 2×2 ANOVA examined the effects of Game Mode (GM) and Difficulty Scaling (DS) on all dependent variables. In case of interactions, we analyzed simple effects of DS separately at the two levels of GM, and simple effects of GM separately at the two levels of DS using Bonferroni correction. Effect sizes are reported as partial eta squared η_p^2 . Scores from the GEQ Social Presence Module, which was administered exclusively to multi-player groups (G3 and G4), were compared using a Mann-Whitney U test as the data did not meet the assumption of normality.

Progress within the SG

Figure 5.7 shows the results of the measures for the progress within the SG, the duration of the longest game match and the maximum final score reached. No statistically significant results were found for either the main effects or the interaction (all $p > 0.05$).

³<https://www.ibm.com/products/spss>

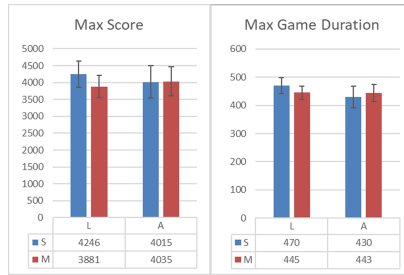


Figure 5.7: Means of maximum score gained and maximum duration (seconds) in the four groups. Capped vertical bars indicate $\pm$ SE.

Player Experience

Affective Reaction Figure 5.8 shows the results for affective reaction, measured using the SAM questionnaire. No statistically significant results were found for either the main effects or the interaction (all $p > 0.05$).

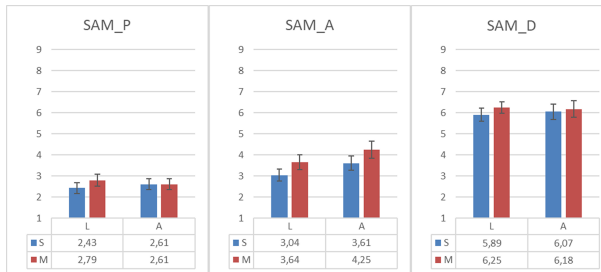


Figure 5.8: Means of affective reaction for Pleasure, Arousal, and Dominance in the four groups. Capped vertical bars indicate $\pm$ SE.

Presence in the VE Figure 5.9 shows the results for presence perceived in the VE, measured using the IPQ questionnaire. No statistically significant results were found for either the main effects or the interaction (all $p > 0.05$).

Game Experience Figure 5.10 shows the results for the player experience measured with the GEQ questionnaire Core Module. No statistically significant results were found for either the main effects or the interaction for almost all dependent variables. The only statistically significant result is for the variable GEQ_PA of the GEQ Core Module, for which there was a significant interaction between GM and DS, $F(1, 108) = 4.263$, $p < 0.05$, $\eta_p^2 = 0.038$. Despite the statistically significant interaction, none of the pairwise comparisons reached statistical significance after applying the Bonferroni correction (all $p > 0.05$).

Figure 5.11 shows the results for player experience, measured with the GEQ Post-Game Module. No statistically significant results were found for either the main effects or the interaction for all dependent variables (all $p > 0.05$).

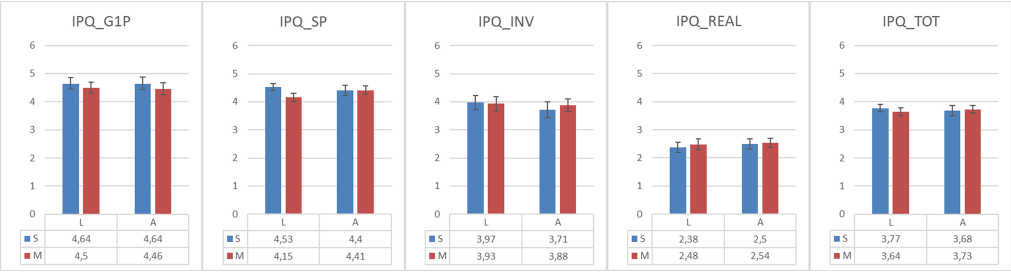


Figure 5.9: Means of presence total score, general item about the sense of “being there,” spatial subscale, involvement subscale, and experience realism subscale in the four groups. Capped vertical bars indicate $\pm$ SE.

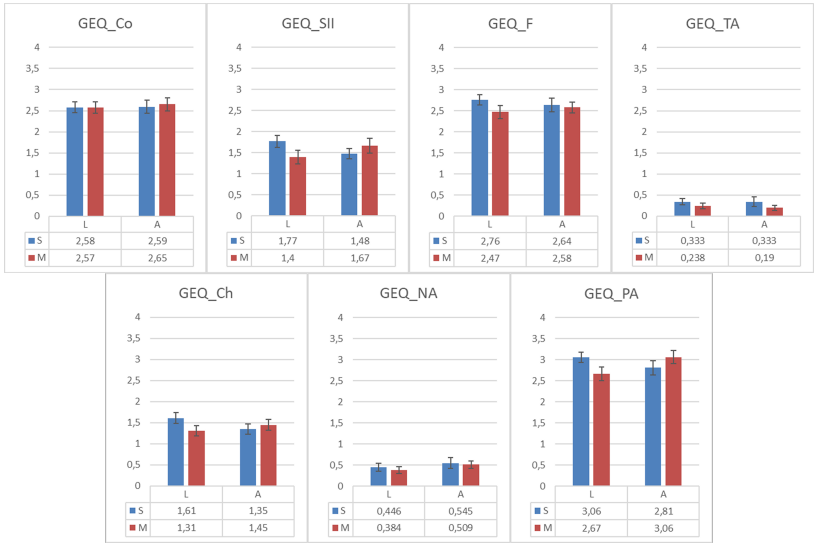


Figure 5.10: Means of Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative Affect, and Positive Affect in the four groups. Capped vertical bars indicate $\pm$ SE.

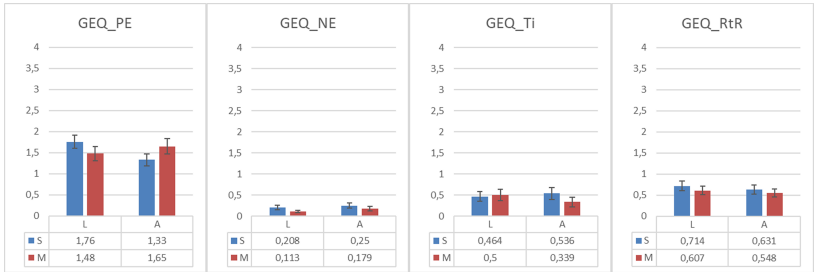


Figure 5.11: Means of Positive Experience, Negative Experience, Tiredness, and Return to Reality in the four groups. Capped vertical bars indicate $\pm$ SE.

Social Presence Figure 5.12 shows the results for the player experience measured with the GEQ questionnaire Social Presence Module. No statistically significant results were found for all depended variables (all $p > 0.05$).

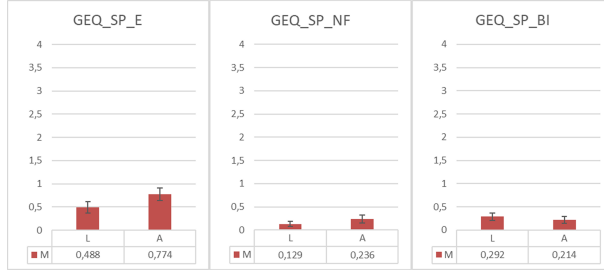


Figure 5.12: Means of Psychological Involvement in Empathy, Negative Feeling, and Behavioural Involvement in the four groups. Capped vertical bars indicate $\pm$ SE.

5.1.4 Discussion

Difficulty Scaling

Regarding H1, we hypothesized that adaptive difficulty scaling would facilitate better player progress. However, analyses of maximum scores and gameplay duration yielded no statistically significant differences between the two modalities. Consequently, H1 is not supported.

We had posited that DDA would enhance the user experience by maintaining players in a flow state; however, without the extended gameplay or mitigated failure predicted in H1, it may be that participants in both conditions had a similar user experience. Results of the comparative analyses regarding user experience metrics showed no statistically significant differences between the groups. Thus, H2 cannot be confirmed.

Although we involved more than 100 users, the effect sizes for both the maximum score ($F(1, 108) = 0.009, p > 0.05, \eta_p^2 < 0.001$) and max game duration ($F(1, 108) = 0.482, p > 0.05, \eta_p^2 = 0.004$) were negligible and the observed power was low (0.051 for max score and 0.106 for duration). Regarding user experience, nearly all comparisons yielded negligible effect sizes ($\eta_p^2 < 0.01$); the exceptions were SAM_A ($\eta_p^2 = 0.025$), GEQ_NA ($\eta_p^2 = 0.011$), and GEQ_NE ($\eta_p^2 = 0.01$), which had small effect sizes. Similar to the performance measures, the observed power for user experience metrics was low (all < 0.2 , except SAM_A at 0.381). Consequently, while we cannot definitively reject differences between the adaptive and linear difficulty groups, the effect sizes are so small that even if statistical significance were achieved with a larger sample, the difference would likely lack practical relevance.

A plausible explanation for these results is a ceiling effect, as participants appeared sufficiently skilled to reach and sustain the maximum difficulty level across the scaling modalities. To further investigate the ceiling effect, we analyzed an additional measure derived from the logs recorded during gameplay: the mean throw accuracy (TA) maintained over the 20 minutes of gameplay. While the maximum score and maximum time duration were virtually uncapped, since the condition for termination was based on

maximum errors and fulfillment of the table, the TA is strictly bounded within the $[0, 1]$ interval. This standardization permits a more precise evaluation of the player's skill, as a player who plays perfectly will achieve a TA of 1. Although an ANOVA revealed a statistically significant main effect of Difficulty Scaling favoring the Linear modality ($F(1, 108) = 5.31, p < .05, \eta_p^2 = 0.047$), the practical difference was small as mean accuracy was high across all groups: G1 ($M = 0.96, SD = 0.02$), G2 ($M = 0.95, SD = 0.03$), G3 ($M = 0.96, SD = 0.02$), and G4 ($M = 0.95, SD = 0.03$). With accuracy consistently hovering around 95%, it suggests that participants encountered minimal difficulty performing the task correctly in either condition. This lack of a proper challenge likely flattened potential performance and user experience variance, resulting in similar final scores and gameplay duration, as well as comparable subjective ratings, given the uniform proficiency in the core game mechanic.

Related to the ceiling effect is the calibration of the adaptive scaling parameters. The SG was developed with the intent of future deployment for older adults with mild NCD. Consequently, the initial parameter configuration prioritized accessibility, anticipating that the target demographic would face significant challenges. For instance, the original design featured a lower baseline pacing (one ball every 4 seconds) and a progression system of 7 difficulty levels. Although a pilot phase with younger participants suggested a recalibration for which we expanded the range to 10 levels and modified the adjustment logic (e.g., the original thresholds had a Performance Score < 60 that triggered the "-2" level reduction), the results indicate that these adjustments were insufficient for the young participants involved in the formal experiment. The consistent gameplay among participants suggests that, despite the "harder" configuration, the difficulty curve remained too simple for this demographic. These findings align with the systematic review by Souza et al. [292], who noted the trade-off that while heuristic-based parameter manipulation benefits from not requiring extensive background data, these methods remain heavily dependent on the game designer's ability to accurately predict player behavior.

Game Mode

The exploratory analysis regarding Game Mode yielded no statistically significant differences between the Single-player and Multi-player conditions for either performance or user experience metrics.

Similar to the findings for Difficulty scaling and despite the substantial sample size, effect sizes remained negligible for most variables ($\eta_p^2 < 0.01$). Exceptions showing small effect sizes included SAM_A ($\eta_p^2 = 0.028$), IPQ_SP ($\eta_p^2 = 0.013$), GEQ_F ($\eta_p^2 = 0.014$), GEQ_TA ($\eta_p^2 = 0.019$), and GEQ_NE ($\eta_p^2 = 0.024$). Observed power was consistently low (< 0.3), with the exception of SAM_A (0.42) and GEQ_NE (0.364). Consequently, while we cannot definitively reject the existence of differences between Single-player and Multi-player modes, the observed effect sizes are so small that, even if statistical significance were achieved with a larger sample, the differences are likely to lack practical relevance.

The lack of significant differences between the single-player and multi-player conditions can be attributed to several factors involving game design and participant characteristics. Since each player managed a personal pool of balls with no functional requirement to compete for shared resources or coordinate actions with the partner to

succeed, the sorting task was inherently individualistic. Therefore, the specific nature of the game mechanics favored parallel play [244] rather than truly cooperative interdependence [151]. Consequently, players likely prioritized their individual sorting performance over social interaction, thereby limiting the potential added value of playing the task with other players.

The limited social interaction was also likely facilitated by the VE's spatial layout. To minimize potential visual or cognitive distractions, the cooperative partner was positioned at a significant distance in front of the active player. Although participants could see the other player by checking the pair's ranking as the avatar of the other player was placed behind the user interfaces of the SG (See Section 5.1.1), it is reasonable to assume that players were focused on to the sorting task since a sustained concentration is required to maintain, especially at higher difficulty levels, the high precision found in the log files across all groups (about 95%). Consequently, despite the cumulative scoring mechanic, the combination of physical distance and task demand likely limited the sensation of social presence [32], preventing participants from perceiving the session as a genuinely cooperative gameplay.

This interpretation is also supported by the GEQ Social Presence Module scores, which show a floor effect. Behavioral Involvement ($M = 0.25, SD = 0.41$) and Empathy ($M = 0.63, SD = 0.67$) were notably low, indicating that participants neither coordinated their actions nor felt an emotional connection to the other player. For example, in the item 'What the other did affected what I did,' the majority of participants selected 'Not at all.' When combined with the near-absence of Negative Feelings ($M = 0.18, SD = 0.37$), the results point to a dynamic of parallel play rather than collaboration. It is possible that reducing the physical distance between players would alter the sense of social presence.

Overall User Experience

In the absence of statistically significant differences between groups, we discuss the aggregate scores for the user experience.

Considering affective reactions, the results from the SAM indicate that after playing with the SG, participants had a generally positive, stimulating emotional response. On the Pleasure dimension (SAM_P), the mean score ($M = 2.61, SD = 1.36$) suggests that users felt pleased and satisfied. This was accompanied by a moderately high level of activation, as indicated by the average SAM_A score ($M = 3.63, SD = 1.89$). Furthermore, the SAM_D score ($M = 6.10, SD = 1.77$) places participants in the upper range of perceived dominance. Collectively, these results suggest that participants felt stimulated and happy while retaining a sense of control over the experience [36].

Regarding presence in the VE, we applied the benchmarks proposed by Melo et al. [216]. The results indicate that the SG achieved marginally acceptable levels for Overall Presence ($M = 3.70, SD = 0.776$), Spatial Presence ($M = 4.37, SD = 0.82$), and Involvement ($M = 3.87, SD = 1.33$). However, the Realism subscale ($M = 2.48, SD = 0.94$) was rated as not acceptable. This specific lower rating may be attributed to the decision to minimize player distraction during the task by excluding detailed textures from the environment.

Considering user engagement more broadly, the in-game results of the GEQ further support the generally positive experience. Participants reported above-average Positive

Affect ($M = 2.90, SD = 0.81$) and Flow ($M = 2.61, SD = 0.76$) scores, indicating that the activity was both enjoyable and compelling. Similarly, the Competence dimension ($M = 2.60, SD = 0.77$) indicates that participants felt capable and effective during the task. In contrast, measures such as Tension/Annoyance ($M = 0.27, SD = 0.43$) and Negative Affect ($M = 0.47, SD = 0.53$) remained low, reflecting minimal frustration or discomfort. The Challenge subscale ($M = 1.43, SD = 0.66$) indicates that the task was perceived as moderately demanding but not overwhelming, consistent with the sense of control described in the SAM results. The lower rating on Sensory and Imaginative Immersion ($M = 1.58, SD = 0.81$) may align with the intentionally simplified visual design noted earlier, which prioritized task focus over environmental realism.

Additional post-experience ratings echo this pattern. While participants reported modest Positive Experience scores ($M = 1.56, SD = 0.86$), indicators of discomfort or fatigue, such as Negative Experience ($M = 0.19, SD = 0.27$) and Tiredness ($M = 0.46, SD = 0.67$), remained relatively low. The moderate Return to Reality score ($M = 0.63, SD = 0.55$) suggests that players transitioned smoothly back to their everyday context after the task, indicating an adequate but contained level of immersion.

Overall, these findings report that the SG provided a balanced, enjoyable, and non-fatiguing experience, characterized by a positive emotional tone.

5.1.5 Conclusions

The evaluation presented in this section investigated the impact of adaptive difficulty and multi-player interaction in an XR serious game for executive function training on both objective performance metrics and subjective user experience with a healthy demographic. Although quantitative analysis of game performance and questionnaire scores showed no statistically significant differences across conditions, participants reported the experience as engaging, emotionally positive, and physically sustainable.

The results highlighted that, while heuristic-based difficulty adaptation offers implementation simplicity through not requiring extensive background data, it lacks the generalization required to maintain a flow state across different demographics without recalibrating the parameters. Furthermore, in the context of the proposed task, the multi-user setting introduced technological and logistical overhead without demonstrating clear added value.

To further investigate the adaptive difficulty system, we will use probabilistic player modeling. By modeling the gameplay evolution across different player categories, it would be possible to predict how various parameters would affect gameplay for players ranging from healthy individuals to patients with mNCD, as described in the next section, focusing on a single-player version of the SG.

5.2 Prediction of User Behavior in an XR SG for Inhibitory Control Training using Model Checking

Neurodegenerative diseases like Alzheimer’s and Parkinson’s often impair CFs, necessitating early detection and ongoing monitoring by clinicians. Accurate diagnosis typically requires extensive neuropsychological and biomarker tests, which can be time-intensive

for practitioners and patients. To address this, there is growing interest in using SGs to identify quick, objective behavioral markers that complement traditional assessments and aid early detection of cognitive decline (see Section 2.8).

Building upon the conclusions of the previous work (see Section 5.1.5), the lack of distinct clinical benefit of the multi-user component of the developed SG, the organizational difficulties in scheduling synchronous sessions for patients, and the interest in the literature in using SGs as a assessments for cognitive decline, in this work we introduce a single-player SG based on the Go-NoGo Task. This cognitive task requires participants to press a button when a stimulus appears, except in the cases where they should refrain from pressing because a different stimulus is presented [129]. The new version of the SG is probabilistically modeled as a Discrete-Time Markov Chain using the tool PRISM [177]. While PRISM has already been used in the literature to model and verify SGs [80, 198], the novelty of this work is that we consider a game that evolves dynamically based on the patient’s performance.

The idea is that when patients play with the SG, some scenarios occur more frequently than others that are rarer. To quantify these behavioral variations, probabilities reflecting the patient’s Cognitive Impairment stage (healthy, mild NCD, major NCD) are assigned to key game actions. The probabilistic model checker PRISM is then used to validate the models with respect to key temporal properties and to determine the probabilities associated with crucial classes of paths patients take during gameplay. This system aims to serve as a support tool for diagnosing cognitive impairments by comparing expected with observed behavior in the SG.

5.2.1 Proposed Serious Game

Based on the SG presented in Section 5.1.1, a variation of the exergame was developed that employs a Go-NoGo task instead of a sorting task. This was meant to tune the SG specifically for inhibitory control, whereas the previous version was more generally for executive functions. Following the scoping review by Sharifzadeh et al. [282], which emphasizes the need for multidisciplinary teams to create effective and engaging SGs for health, during the development of this new version of the SG, we involved medical staff at the Cognition Behavior Technology Laboratory (CoBTek Lab) of the Université Côte d’Azur experienced with inhibitory control disorders.

The task implemented in the SG requires players to throw balls, replacing the button-press mechanism found in traditional Go-NoGo tasks. Specifically, the players’ task is to throw balls into a basket in front of them (Go-Basket) if the ball’s appearance matches the active rule (Go-Ball), or onto the floor (NoGo-Basket) if it does not match the rule (NoGo-Ball). Figure 5.13 shows an example of gameplay in which the active rule is to throw only the red balls in the Go-Basket.

The game ends when the player throws more NoGo-Balls in the Go-Basket than the maximum allowed (MaxNoGoInGoBasket) or when the number of balls available to throw exceeds the limit on the container (MaxBallsOnTable). A match lasts for a fixed number of TotalGameSessions, each consisting in *SessionDuration* seconds. At the end of each session, the Game Manager evaluates the PlayerScore and adjusts difficulty (LevelIncrement). The PlayerScore is calculated similarly to the Performance Score of the previous version of the SG (see Section 5.1.1). It is based on the ratio of correct throws (TotalCorrectThrows) relative to the total throws (TotalThrows) and the time



Figure 5.13: Example of SG gameplay. The rule is to throw red balls into the container and black-dotted ones away. The user interface on the left displays the score, combo streak, and NoGo-Balls mistakenly thrown in the Go-Basket (marked with red crosses) and the user interface on the right shows player ranking.

with few objects available to throw (FewObjectsOnTableLimit) relative to the session duration using the following formula:

$$PlayerScore = \left(\frac{TotalCorrectThrows}{TotalThrows} + \frac{TotalTimeWithFewObjects}{SessionDuration} \right) * \frac{1}{2}$$

As in the previous version of the SG, the LevelIncrement is calculated to modify the game difficulty by matching the player’s current skill level: i) increasing the difficulty if the PlayerScore is over the IncDiffBound, ii) decreasing the difficulty if it is lower than the DecDiffBound, iii) or staying at the same level if it is between the bounds. Table 5.5 summarizes how the LevelIncrement is calculated.

Table 5.5: Level Increment calculation logic.

PlayerScore	Level Increment
$PlayerScore < DecBound$	-1
$DecBound \leq PlayerScore \leq IncBound$	0
$PlayerScore > IncBound$	+1

The game’s difficulty involves three levels where the changing parameter is the time elapsed before creating a new ball (TimeToGenerateBall), ranging from the first level with one ball every 5 seconds to the last one with an ball every 3 seconds. Each new ball has a GoBallProb to be a Go-Ball and a NoGoBallProb to be a No-Go.

These actual parameters used in the proposed game have been set after discussing them with the medical team and conducting pilot evaluations and are presented in table 5.6

Table 5.6: Parameters used in the SG and in the model with their values.

Variable Name	Value
GoBallProb	0.7
NoGoBallProb	0.3
MaxNoGoInGoBasket	3
MaxBallsOnTable	4
SessionDuration	30
TotalGameSessions	40
FewObjectsOnTableLimit	2
DecDiffBound	0.7
IncDiffBound	0.8

5.2.2 Preliminaries on PRISM and Temporal Logics

PRISM [177] is a state-of-the-art tool for formal modeling and verification of systems with a random or probabilistic behavior. It supports various types of probabilistic models, including both discrete and continuous ones. In this study, we focus on discrete-time Markov chains (DTMCs), which are transition systems extended with probabilities. In DTMC, the likelihood of transitioning between states is determined by discrete probability distributions. DTMCs are characterized by their memoryless property, meaning that their current state encompasses all the information required to compute future states.

PRISM features a state-based modeling language inspired by the reactive modules formalism from [11]. A model consists of a set of *modules* that can interact with each other. The behavior of each module is defined through a set of commands of the form: $[] \text{ guard} \rightarrow \text{prob}_1 : \text{update}_1 + \dots + \text{prob}_n : \text{update}_n$; where the *guard* is a predicate over the model's variables, representing a condition that must be satisfied for the command to be executed. Each *update* specifies a potential state transition, defined by assigning new values to the variables. Each *update* is associated with a probability, and for each command, the sum of these probabilities must equal 1. The square brackets at the beginning of a command can either be empty or contain labels denoting *actions*. These actions enable synchronization, allowing two or more modules to transition simultaneously. PRISM models can also incorporate *rewards* [176], which associate real-valued rewards with specific states or transitions in the model.

The behavior of DTMCs can be expressed in PRISM using PCTL (Probabilistic Computation Tree Logic) [121], an extension of CTL (Computation Tree Logic) [69] that incorporates probabilities.

PCTL provides the following temporal operators: **X** (next time), **F** (eventually or sometimes in the future), **G** (always in the future), and **U** (until). PCTL replaces the CTL path quantifiers **A** (forall) and **E** (exists) with probabilities. This means that properties are expressed in terms of the fraction of paths for which they hold, rather than for all or some paths [121]. The key operator in PCTL is **P**, which is used to specify probabilities of events. For example, the PCTL property $P= 0.5 [X (y = 2)]$ holds in a state if the probability that $y = 2$ in the next state is 0.5. Additionally, the **P** operator can be used to compute the actual probability of a behavior occurring with

the form $P=?$. For instance, the property $P=? [G (y = 1)]$ represents the probability that y is always equal to 1.

PRISM also supports reward-based properties. The **R** operator is used to retrieve expected reward values. To validate DTMCs against PCTL properties and reward-based properties, PRISM provides model-checking algorithms [70]. These algorithms can also compute the probability of certain behaviors when required.

5.2.3 Prism Model of the proposed SG

Representation in the Prism Language

The model encodes the game match of a patient playing for 20 minutes and it is represented with three modules that manage the aspects of the SG: the **Manager** tracks how many sessions have been played and decides the difficulty to use for the sessions given the above rules; the **Gameplay** encodes the behavior of every session of the SG with the Go and NoGo balls available to throw, the timers for the total time with few balls and the countdown to create a new ball; the **Patient** describes how the player reacts to what happens in the SG by encoding the throws of the balls (Go and NoGo, correct and wrong), the recovery time and the possibility of stopping the game match prematurely.

The adaptation strategy is encoded in the Manager module that calculates how the difficulty level changes between the sessions.

Each transition of the global model (all the three modules) represents a time elapsed in the game of 1 second. In every state of the model, the guards of the set of commands (belonging to any of the modules) are satisfied in such a way that only one transition can happen, thus removing the possibility of having a non-deterministic choice between different transitions.

Description of the model

The three modules have local variables that encode the possible states of the SG and commands that represent the behavior of the model. Table 5.7 presents the different local variables of each module.

Table 5.7: Local variables of the three modules and their purpose.

Manager	
CurrenSession	Integer representing the total sessions played
HaveToPrepareSession	Boolean indicating if the Manager has to change the difficulty and update the session counter
RunningSession	Boolean indicating if the session is been played
GameMatchIsOver	Boolean indicating if the game match is over
ActiveTicks	Integer variable that influences the spawn rate of the balls, affecting the difficulty simulated
Patient	
HaveToBegin	Boolean indicating the patient have to start play

RecoveryT	Integer representing the countdown before the patient is ready to throw a ball again
PrematureEnd	Boolean indicating if the patient has left the game before the end or technical problems
CorrectThrowsC	Integer representing how many total correct throws were made during the current session
WrongThrowsC	Integer representing how many total wrong throw were made during the current session
ThrowMadeType	Integer representing the current throw made by the patient. The values it assumes indicate respectively: 0 - None, 1 - Correct Go Ball, 2 - Wrong Go Ball, 3 - Correct NoGo Ball, 4 - Wrong NoGo Ball
NoGoWrongThrowsC	Integer representing how many NoGo balls were mistakenly thrown in the Go container
Gameplay	
TimerToSpawn	Integer representing the countdown before creating a new ball
FewBallsOnTableT	Integer that represents how much time for the current session there were less or equal to FewObjectsOnTableLimit
InGameGoBallsC	Integer representing how many Go balls are available to be thrown
InGameNoGoBallsC	Integer representing how many NoGo balls are available to be thrown

There are 9 different labelled commands for synchronizing the behavior of the modules:

- **StartGameMatch:** Encodes the beginning of the game match, when the patient clicks on the start button. This transition can happen only once.
- **StartNewSession:** Encodes the starting of each game session by enabling the Gameplay and Patient modules and resetting their local variables.
- **EndSession:** Encodes the ending of each game session by disabling the Gameplay and Patient modules while the Manager transits to the next session and updates the level of difficulty or ends the simulation.
- **EndGameMatchManager:** Encodes the ending of the game match because the patient has performed all the scheduled game sessions.
- **EndGameMatchGameplay:** Encodes the ending of the game match because of excessive NoGo Throws in the GO Basket or because the number of balls available to throw exceeds the limit.
- **EndGameMatchPatient:** Encodes the ending of the game match because of a technical issue or because the patient decided to end the game prematurely.
- **ThrowGoBall** and **ThrowNoGoBall:** Encode ball's throw by the patient and: i) increase the relative counter (i.e., TotalCorrectThrows and TotalWrongThrows for the Go and NoGo balls); ii) decrease the total balls available to throw; iii) decides the RecoveryTime the patient needs.

- **Update Timers:** Encodes the updating of the timers used to create a new ball and the patient’s recovery time.

The Manager has seven commands to adjust game difficulty: remaining at any of the three levels, switching between the first and second, or switching between the second and third. The Gameplay includes a command for generating a new throwable ball. The Patient selects which ball to throw, prioritizing the most abundant type when multiple options are available.

Patient Parameters

The player’s behavior in the game match simulation is driven by probabilities encoded in the model, determining the likelihood of specific transitions when conditions are met. These probabilities influence the chances of making a correct or incorrect throw (Go or NoGo), the required recovery time before throwing again, and the likelihood of ending the game early. Recovery time can be: fast (immediate readiness to throw); normal (3-second rest); slow (6-second rest). The probability of premature game termination accounts for both technical issues and the player’s decision to stop due to fatigue or loss of interest.

To estimate these probabilities, we consulted medical staff at the Cognition Behavior Technology Laboratory (CoBTeK Lab) of Université Côte d’Azur, specifically seven doctors experienced with inhibitory control disorders who provided us probability estimates for each disease level (healthy, mild, major). The collected data is summarized in Table 5.8.

Table 5.8: Percentage probabilities calculated as the mean of the medical staff’s responses regarding the likelihood of patients performing specific actions in the SG.

	Healthy	Mild NCD	Major NCD
throw_correct_go	96.571	81.143	67.143
throw_wrong_go	3.429	18.857	32.857
throw_correct_no_go	83.571	64.286	47.143
throw_wrong_no_go	16.429	35.714	52.857
fast_recovery	58.286	17.143	0.000
normal_recovery	32.571	60.857	37.143
slow_recovery	7.571	19.571	58.286
premature_end	1.571	2.429	4.571

Description of the flow of the model

Figure 5.14 illustrates the flow of the model. The synchronized StartGameMatch command resets the session counter, sets the initial difficulty, and initiates the first session. As described earlier, commands are structured so that only one occurs at a time. After execution, all guards are reevaluated to determine the next action, ensuring mutual exclusivity. Guard evaluation follows the SG rules: i) check if the player has lost the game; ii) execute a throw if possible; iii) generate a new ball; iv) update the timers.

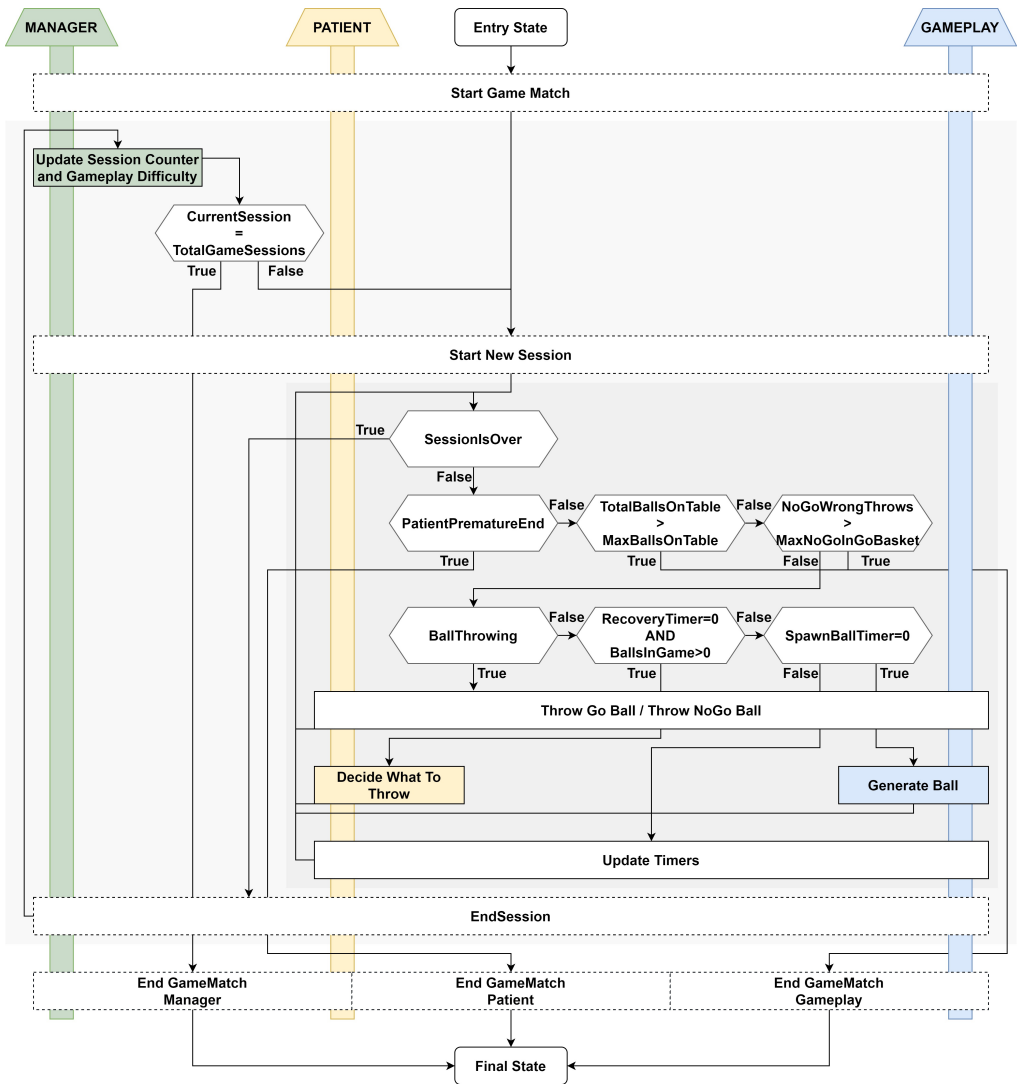


Figure 5.14: Representation of the flow of the model. The three trapezoid shapes at the top of the figure contain the names of the modules. The white boxes with dotted borders mean a transition that involve all the modules, and the ones with continued borders represent a synchronized transition between two modules. The colored boxes represent a transition that involve only that particular model. The hexagons encode a guard, showing how the flow changes between the possible outputs enabling a particular transition. The light-grey background groups the parts included in the game loop or the session loop.

5.2.4 Probabilistic queries and results

The model built depends on the patient’s disease level, as reflected in the probabilities defined in Section 5.2.3, since some transitions cannot occur when the probabilities are zero. All PRISM models are accessible on GitHub ⁴.

The model construction and query execution were performed using PRISM software on a desktop computer with an AMD Ryzen 3700X processor and 32GB RAM @ 3200MHz. The structure of a PRISM model is automatically compiled into a transition system, a graph comprising states and transitions. Table 5.9 presents the total states and transitions of the different models built.

Table 5.9: Total states and transitions of the different models built.

Model Type	Total States	Total Transitions
Healthy	3,257,476	5,244,133
Mild NCD	3,257,476	5,244,133
Major NCD	2,158,555	3,200,512

We defined probabilistic queries to test the model with two separate objectives: to validate the model and to study the evolution of the patient’s gameplay based on their disease level, as defined in Section 5.2.3.

Queries 1 and 2 ensure that the model behaves as expected and must be independent of the patient’s properties defined in Section 5.2.3, otherwise it indicates an error in the model’s encoding.

Queries 3 and 4 calculate how the patient’s behavior changes within the model based on the parameters defined in Table 5.8. To assess this, we use PRISM rewards to count the average number of states in which the conditions are met before the end of the game.

Query 1 What is the probability of reaching the final state of the model?

$P=? [F (gameOverCondition) \ \& \ (X \ gameMatchIsOver)]$

One property of the model is that all model scenarios must reach the final state, meaning that at least one transition among *EndGameMatchManager*, *EndGameMatchGameplay*, and *EndGameMatchPatient* must occur, updating the variable *gameMatchIsOver* to true in the next state. If the result is below 1, there exists a possibility of never reaching the final state.

Query 2 What is the probability of throwing a ball when there is one available?

$P=? [G(! (gameRunning \ \& \ patientCanThrow) | (X \ ballThrowing))]$

We want to confirm that in every possible state of the model, when the patients are playing and they can throw (i.e., there is at least a ball, the patients are not throwing a ball and have their *recoveryTimer* equal to zero), in the next state the variable *throwMadeType* will be assigned to a value indicating the throw made. If the result is below 1, there exists a possibility to not throwing a ball when possible.

⁴<https://github.com/AlessandroForgiarini/InhibitoryControlSGModel>

Query 3 How long will the patient be playing on average?

$R\{\text{"runningSession"}\}=? \quad [\quad F \quad \text{dmGameMatchIsOver} \quad]$

We can provide an estimation of the average time a patient will play with the game.

Query 4 How many times will the difficulty of the game be reduced?

$R\{\text{"difficultyDecrease"}\}=? \quad [\quad F \quad \text{dmGameMatchIsOver} \quad]$

We can provide the estimation of the average total times the Manager will reduce the game difficulty in a game match.

Results The results of the defined queries are shown in Table 5.10.

Table 5.10: Results of the considered queries and relative times to compute expressed in seconds.

	Q1		Q2		Q3		Q4	
	Result	Time	Result	Time	Result	Time	Result	Time
Healthy	1.00	3.55	1.00	285.79	396.15	30.88	0.0002	11.59
Mild NCD	1.00	3.28	1.00	267.74	266.38	28.20	0.09	29.20
Major NCD	1.00	2.67	1.00	121.78	159.50	12.83	0.42	15.00

5.2.5 Discussion and Conclusions

The work presented in this paper extends the literature on the use of SG in combination with theoretical modeling and verification of SG and the patient's behavior within it to support medical doctors in diagnosing the patient's level of cognitive impairment.

The results of Queries 1 and 2 (result=1 for all the considered classes of impairment) show that the model behaves as expected. Concerning the queries on the patient's behavior, the results of Query 3 allow one to predict that, on average, healthy people will play a single game match for about 6-7 minutes, while mild NCD patients will play for about 4-5 minutes, and, lastly, major NCD patients will play for only 2-3 minutes. The results of Query 4 anticipate that, on average, the game difficulty will be increased and then reduced significantly more often for individuals with mild and major NCDs than for healthy individuals. These changes indicate how the SG will adapt its difficulty to accommodate patients who could struggle more due to their impairments.

The results of the presented probabilistic queries reveal differences in the predicted patient behavior in the SG across the three disease levels considered in this work, suggesting that the tool can help medical doctors interpret patient performance and assess the stage of their disease. More precisely, we can deduce how relevant the patient's scenario is. For instance, if a healthy individual plays the SG as expected for a mild NCD patient, our methodology can spot this discrepancy. The same goes if a major NCD patient plays like a healthy individual. Furthermore, the results of probabilistic queries can help customize the game for different patient groups.

We created a simple 3-level SG model to verify its feasibility and viability as a support tool. We are developing a more complex version of the SG and probabilistic model to handle more difficulty levels and different ball color combinations.

This proof-of-concept model and the gathered results will allow us to discuss and refine the model with the medical team, adding more user queries that can be encoded and used to provide insight into the patient's condition. Furthermore, in the next Section, a user study with mild NCD patients is presented. Data collected from real patients could be used to adjust doctors' probabilities in the model, thereby improving its results.

The work presented in this section was published in the Proceedings of the 23rd International Conference AIME 2025 [100].

5.3 Acceptance and Efficacy of an XR SG for Inhibitory Control Training of Old Patients

The number of people aged 65 and over is projected to double to 1.5 billion in 2050⁵, and by 2030 one in six people worldwide will be aged 60 or over. At the biological level, aging results from the impact of the accumulation of a wide variety of molecular and cellular damage over time that leads to a gradual decrease in physical and mental capacity and a growing risk of disease⁶.

Building on the need for interventions aimed at strengthening or slowing age-related decline of Executive Functions (EF) and that XR SGs have shown positive outcomes across various fields (see Chapter 2), this work explores their application specifically for older adults with mNCD.

A high level of immersion could bring the participants to a deeper level of engagement than conventional therapy [137]. Furthermore, VR systems for cognitive impairments tend to use nVR interfaces and have yet to demonstrate the potential of iVR [137].

Preliminary evidence indicates that older adults could enjoy the experience of iVR technologies [139]. For instance, Mascaret et al. [210] examined with a survey among older adults the acceptance before use of a headset designed to prevent falls. Their results showed that intention to use the headset was positively predicted by perceived usefulness, perceived enjoyment, and perceived ease of use.

The acceptability of SGs with older people with mNCD symptoms is still being assessed. For instance, Hassandra et al. [122] developed and evaluated a SG regarding Alzheimer's disease, suggesting that the SG is acceptable and usable, but its efficacy need to be assessed.

Therefore, given the positive outcomes of the application of SGs in other fields, the present study aims to assess the use and acceptability by older people with mNCD of a VR SG for training and rehabilitation of IC by comparing immersive and non-immersive versions of the game.

⁵<https://www.ohchr.org/en/older-persons>

⁶<https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>

5.3.1 The proposed SG

Task in the Game

Based on the game presented in Section 5.2.1, we created a different version of the SG by adding another environment, difficulty levels, and additional rules.

The player's task is to throw objects in a basket in front of him/her and match the active rule. These objects can be of three colors (light red, yellow, and purple) and may have stains. The object's color is hidden until the player interacts with it. Fig. 5.15 shows an example of the gameplay for all the environments and conditions created. Focusing on Fig. 5.15C, it shows the active rule to throw in the basket only the objects red, yellow with stains, and purple. To earn points, the player must match the rule presented to them and throw it away if it does not match.

The ratio of Go objects to NoGo is 70/30. On average, for every 100 objects, 70 should be placed in the container and 30 discarded on the floor.

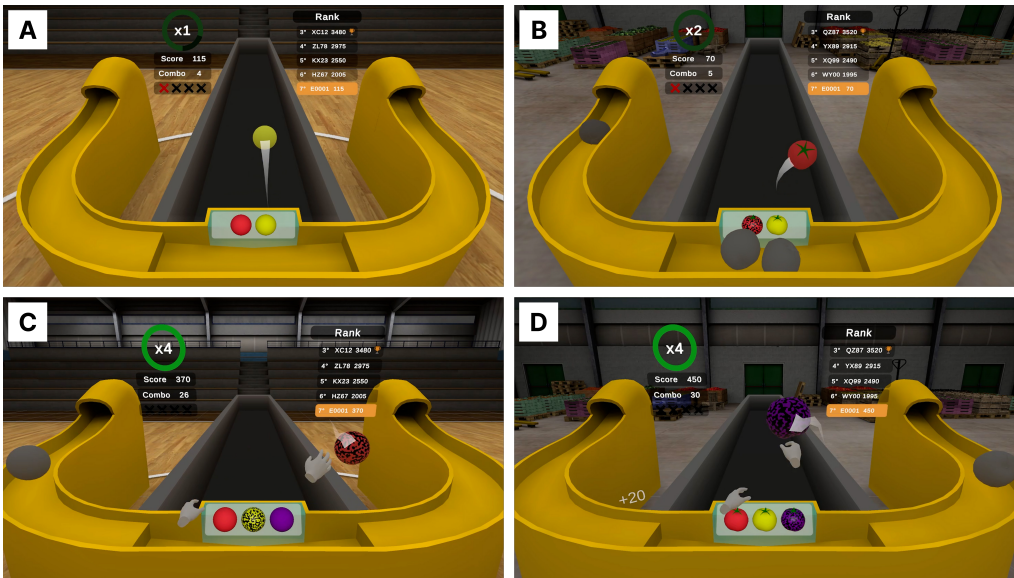


Figure 5.15: SG environments for the proposed IC training task. The top row (A, B) displays the nVR versions, and the bottom row (C, D) displays the iVR versions. Columns represent the Gym environment (A, C) and the Warehouse environment (B, D).

As in the original version of the SG presented in Section 5.1.1 and in the one presented in Section 5.2.1, the conditions for game over are table overflow (when more than four objects are present on the table) and accuracy threshold (when player commits four sorting errors by throwing a NoGo object in the Go container). Similarly, we have implemented the scoring system, combo system, and the ranking systems work as presented in Section 5.1.1. These systems aim to improve the player's willingness to perform skillfully in the SG.

Consistent with the previous version of the SG and following the results of Arlati

et al. [15], we ensure the safety and reduce the risk of cybersickness of the player by designing the SG so that it is possible to play with it without the need to move inside the VR by walking in the physical world. This limits the risk of behaviors that can arise when the user has to walk in the physical world but cannot see it because they are wearing a VR headset.

Game Virtual Environment

To mitigate potential player disengagement, we designed two versions of the same SG: one is set in a gym, as shown in Fig. 5.15 on the right, and the other in a food warehouse where the player has to throw vegetables instead of balls, as shown in Fig. 5.15 on the left.

Difficulty Scaling

To further investigate the efficacy of DDA techniques based on heuristics and given the positive results of adaptivity in the context of rehabilitation [137], we designed the SG so that the task’s challenge can be changed in real time. The SG uses the same approach as in previous versions; every SessionDuration (30s), the GameManager evaluates the player’s score and updates the difficulty of the task. After attempting the difficulty by changing linearly the time to spawn the objects of the level, we noticed that the difficulty increased faster than expected. Therefore, in this version of the SG we changed the approach to it by changing linearly the total objects to spawn in each level. Two parameters define the task difficulty:

- **Object Count:** The total number of objects generated within the session interval.
- **Color Variety:** The number of distinct object colors active in the scene.

The SG features 10 levels of difficulty, with parameters scaling from Level 1 to Level 10. Following pilot testing and discussions with medical staff at the Cognition Behavior Technology Laboratory (CoBTek Lab) at Université Côte d’Azur, which is experienced with inhibitory control disorders, the boundary values were established as shown in Table 5.11.

Table 5.11: Difficulty parameter scaling from Level 1 to Level 10.

Parameter	Level 1 (Minimum)	Level 10 (Maximum)
Object Count	1	3
Color Variety	5	20

At Level 1, the player must sort a total of five objects. These objects are limited to a single color variation (e.g., only red, or only stained red). By Level 10, the density increases to twenty objects, involving all three colors (red, yellow, and purple). In this highest difficulty setting, the active rule can be any permutation of the three colors in either their clean or stained states (e.g., *Stained Red – Yellow – Stained Purple*).

The progression of these variables across all levels is visualized in the difficulty plots (Fig. 5.5) for object count and color variety.

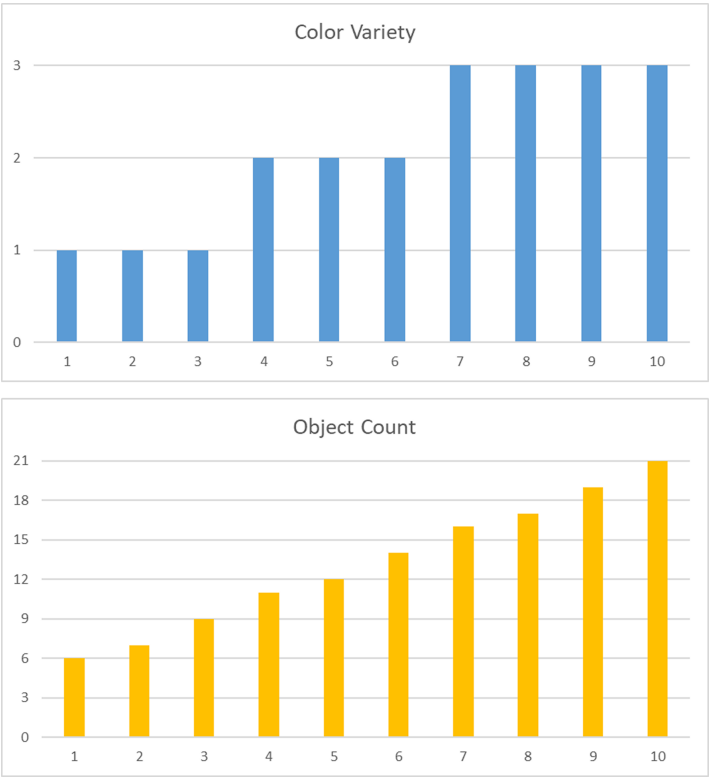


Figure 5.16: Bar plot of each variable that defines a level’s difficulty.

Interactions with the SG in iVR and nVR

In the iVR version (Figure 5.15, bottom row), players can see the VE via head tracking. To throw an object, the player must reach towards it, grab it by holding the controller’s trigger button, mimic a physical throwing gesture, and release the trigger to launch the object. User interface elements are rendered as floating canvases within the 3D space; interaction is performed by pointing a ray-cast from the controller and pressing the trigger.

In the nVR version (Figure 5.15, top row), the camera view is fixed. Interaction relies on touchscreen inputs: players can either touch and drag an object to a target position, releasing to drop it, or swipe towards the desired location to simulate a throw. Unlike the iVR version, UI elements are interacted with directly via touch.

5.3.2 Methodology

Our primary objective is to evaluate the acceptability of iVR and nVR serious games for the training and rehabilitation of inhibitory control (IC) in patients with Mild Neurocognitive Disorder (mNCD).

Our secondary objectives are to:

- Compare user experience and performance evolution between the iVR and nVR conditions.
- Assess the effects of iVR and nVR on patients’ IC compared to a control condition.
- Compare the longitudinal evolution of patients’ IC between the iVR and nVR conditions.

For these purposes, we conduct a randomized controlled trial with 4 arms: immersive VR (iVR), non-immersive VR (nVR), waiting immersive VR (W-iVR), and waiting non-immersive VR (W-nVR). Specifically, half of the participants will try immediately one of the two conditions (iVR or nVR) while the other half will wait to try one of the two conditions two months later. In this way we can compare the two VR conditions with a control condition and at the same time have all participants try one of the VR conditions.

Materials

Software and Hardware The SG was developed for desktop platforms using Unity 2022.3 LTS⁷. All 3D assets and environment models were created using Blender⁸. To assess player performance, an internal logging system recorded metrics (e.g., the final score and game duration) upon the termination of each game session.

The software ran on a local laptop (CPU: i7-6700HQ, GPU: NVIDIA GeForce GTX 1080, RAM: 16GB). The VR version of the SG will be played by participants with a Meta Quest 3⁹, while the nVR version will be engaged by participants with an external 10.1-inch touchscreen display (1920 × 1200 resolution, 16:10 aspect ratio) connected to the laptop.

⁷<https://unity.com/>

⁸<https://www.blender.org/>

⁹<https://www.meta.com/it/quest/quest-3/>

Neuropsychological assessment A neuropsychological assessment focused on inhibitory control will be performed, including:

- The Go/No-Go task from the TAP battery [341], which assesses response inhibition by measuring the ability to suppress a prepotent motor response (established through the frequent occurrence of Go trials).
- The Stroop test [200], which evaluates selective attention, processing speed, and the ability to inhibit the automatic tendency to read words while naming ink color (the primary inhibitory mechanism in the task).
- The Sustained Attention to Response Task (SART) [125], a high-frequency Go/No-Go task designed to assess sustained attention and response inhibition. The task presents participants with a rapid sequence of digits (1–9), requiring responses to frequent non-targets (Go trials) and response withholding to rare targets (No-Go trials, typically 3).
- The Frontal Assessment Battery (FAB) [91] is a brief tool assessing executive functions, yielding a total score (FAB_TOT) obtained by summing six subtests, each scored from 0–3 points (maximum 18): (i) Conceptualization (FAB_Co): evaluates the ability to conceptualize the links between two objects from the same category (e.g., an apple and a banana). (ii) Mental Flexibility (FAB_MF): assesses lexical fluency (phonemic fluency task) by asking subjects to recall as many words as they can beginning with a given letter. (iii) Motor Programming (FAB_MP): tests the ability to perform series of motor acts (Luria motor series); (iv) Sensitivity to Interference (FAB_SI): measures conflict resolution using conflicting instructions by asking subjects to provide an opposite response to the examiner's alternating signal, (e.g., tapping once when the examiner taps twice); (v) Inhibitory Control (FAB_IC): assesses the ability to suppress an automatic response (go–no go task) by subjects must inhibit a response that was previously given to the same stimulus (e.g., not tapping when the examiner taps twice); (vi) Environmental Autonomy (FAB_EA): evaluates the ability of being independent from environmental cues by assessing prehension behavior (the sight or sensory perception of the examiner's hands compels the patient to take them)

Questionnaires Several questionnaires are used in the user study:

- *Initial Questionnaire.* Gathered demographic data (gender, age) and assessed prior experience (in hours) with Head-Mounted Displays and weekly hours of sports and video games.
- *Questionnaire for the unified theory of acceptability and use of technology 2.* The acceptability of the proposed solution was assessed using the French version [123] of the Unified Theory of Acceptance and Use of Technology (UTAUT) [314] and its extension (UTAUT2) [315]. The questionnaire comprises 22 items rated on a 7-point Likert scale and evaluates eHealth acceptability across seven subscales: (i) Performance Expectancy (UTAUT_PE), the degree to which an individual believes that using the system will help them achieve gains in job performance or

daily activities; (ii) Effort Expectancy (UTAUT_EE), the degree of ease associated with using the system; (iii) Social Influence (UTAUT_SI), the extent to which an individual perceives that important others (e.g., family, friends, or peers) believe they should use the new system; (iv) Facilitating Conditions (UTAUT_FC), the belief that organizational and technical infrastructures exist to support system use; (v) Hedonic Motivation (UTAUT_HM), the fun or pleasure derived from using the technology; (vi) Price Value (UTAUT_PV), the price value is positive when the benefits of using a technology are perceived to be greater than the monetary cost and such price value has a positive impact on intention; and (vii) Habit (UTAUT_H), the extent to which people perform behaviors automatically due to learning or prior experience.

- *Self-Assessment Manikin (SAM)* [36]. A 3-item non-verbal pictorial assessment measuring the Pleasure (SAM_P), Arousal (SAM_A), and Dominance (SAM_D) associated with the participant's affective reaction to stimuli. For each scale, there are 5 images representing the intensity of an emotion. Participants are asked to indicate how they feel on that scale by choosing the corresponding image.
- *Igroup Presence Questionnaire (IPQ)* [279]. A 13-item questionnaire measuring the sense of presence in a Virtual Environment. Items are scored on a scale from 0 to 6 (where 6 indicates maximum presence). It comprises a general item on the sense of "being there" (IPQ-G1) and three subscales: (i) Spatial Presence (IPQ-SP), (ii) Involvement (IPQ-INV), and (iii) Realness (IPQ-REAL). A global presence score (IPQ-TOT) is calculated by averaging all items after reverse-scoring negative items.
- *Game Experience Questionnaire (GEQ)* [141]. This questionnaire asks participants to indicate their level of agreement with several items on a 5-point Likert scale. The GEQ comprises distinct modules:
 - (i) The *Core Module* (33 items), which assesses feelings during gameplay across seven dimensions: Competence (GEQ_CO), Sensory and Imaginative Immersion (GEQ_SII), Flow (GEQ_F), Tension/Annoyance (GEQ_TA), Challenge (GEQ_Ch), Negative Affect (GEQ_NA), and Positive Affect (GEQ_PA).
 - (ii) The *Post-game Module* (17 items), which assesses participants' feelings after play: Positive Experience (GEQ_PE), Negative Experience (GEQ_NE), Tiredness (GEQ_Ti), and Return to Reality (GEQ_RtR).
- - *Device Assessment Questionnaire (DAQ)* [90], which includes 13 items rated on a 5-point scale. Item 1 rates force required for actuation ("too low" to "too high"). Item 2 assesses smoothness during operation ("very rough" to "very smooth"). Item 3 measures mental effort for operation ("too low" to "too high"). Item 4 assess physical effort ("too low" to "too high"). Item 5 evaluates accurate pointing ("easy" to "difficult"). Item 6 rates operation speed ("too fast" to "too slow"). Items 7-11 quantify fatigue in specific areas ("none" to "very high"): fingers, wrist, arm, shoulder, and neck. Item 12 assesses general comfort ("very uncomfortable" to "very comfortable"), and Item 13 overall usability ("very difficult to use" to "very easy to use").

Semi-structured interviews To evaluate the subjective perceptions and opinions regarding the XR SG experience, participants were asked to answer questions in semi-structured interviews at two distinct stages. First, an interview (I1) was administered after each game session, as illustrated in Figure 5.17a. Furthermore, to gather cumulative information about the participants' overall acceptability and experience, a final interview (I2) will be conducted at the conclusion of the experiment, as depicted in Figure 5.17b.

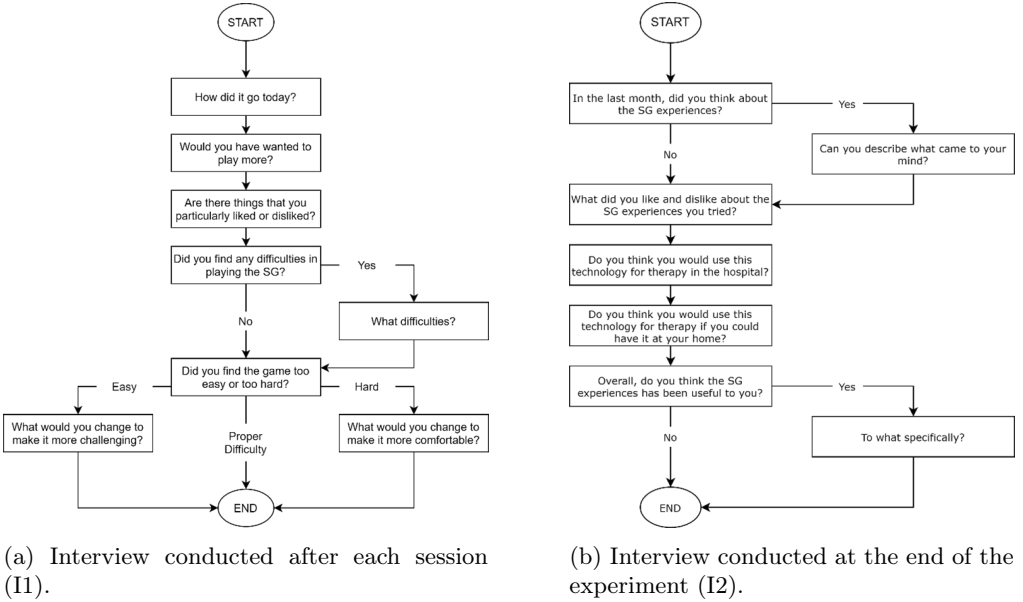


Figure 5.17: Flow diagrams of questions in the semi-structured interviews. The interviews were conducted in the language of the participants; sentences have been translated here for readers' convenience.

Measures

The acceptability of the use of iVR and nVR SGs by patients with mNCD will be assessed using scores gathered with questionnaires regarding the acceptability of the technology [123] and answers to the questions in qualitative interviews regarding the experience.

The user experience will be evaluated in terms of affective reaction with the SAM questionnaire [36], presence with the IPQ questionnaire [279], game experience with the GEQ questionnaire [141], device assessment with the DAQ [90], and answers to the questions in qualitative interviews regarding the experience.

The performance in the SG will be used with the data gathered by the logger inside the application.

The patients' inhibitory control level will be assessed with the metrics introduced in Section 5.3.2, namely: The Go/No-Go task from the TAP battery, the Stroop test, the SART test, and the FAB test.

Participants

Our main objective is to evaluate the acceptability of the technology in a quantitative and in a qualitative way, for which we aim at collecting the data of a total of 40 participants, 20 for the iVR condition and 20 for the nVR. Assuming that there will be a dropout of about 20%, we will enroll 48 participants.

The estimated sample size was based on previous studies assessing acceptability of VR interventions for elderly people, such as in the work of Campo-Prieto et al. [48] and Liepa et al. [188].

Participants were recruited from the Nice Hospital Memory Center. To be eligible for the study, individuals were required to be more than 60 years of age, fluent in French, and diagnosed with either Mild Neurocognitive Disorder or slight Major Neurocognitive Disorder according to DSM-5-TR diagnostic criteria [213]. Additionally, inclusion required the presence of deficits in inhibitory control, defined as a score of < 3 on the Conflicting Instructions subtest of the FAB. All participants are required to demonstrate the capacity to understand the informed consent form and provide voluntary consent.

Criteria for non-inclusion comprise significant self-reported motion sickness, or the presence of vision or motor impairments substantial enough to hinder task performance. Furthermore, individuals protected by law, under guardianship or curatorship, or otherwise ineligible to participate under Article L.1121-16 of the French Public Health Code are not enrolled. Participants retained the right to withdraw consent at any time; additionally, the protocol provides exclusion in the event of significant side effects (e.g., cybersickness) following the sessions.

Study Design

Participants in the iVR and w-iVR group will engage with the SG using the headset, while participants in nVR and w-nVR conditions will play using the touchscreen monitor connected to the laptop.

Group assignment of the participant Each participant will receive an anonymous user code so that all the stored information will be anonymous. A pseudo-randomization will be performed in order to balance the groups in terms of demographic and clinical features.

A combined value representing the participant's Disease Level (DL) will be calculated based on the results of the Neuropsychological assessment described in Section 5.3.2 as follows:

$$DL = w_1 \frac{MMSE}{30} + w_2 \frac{FAB}{18} + w_3 \frac{(GO/NO-GO)_z + 1}{2}$$

The weighting constants are set to $w_1 = 0.2$, $w_2 = 0.3$, and $w_3 = 0.5$ to prioritize the tests most specific to the proposed IC task.

The first 8 participants will be pseudo-randomly assigned to the four groups (iVR, nVR, W-iVR, W-nVR) until every group has 2 participants. Then, following Buttussi and Chittaro [42] the data gathered by the initial questionnaire and the DL will be used to assign the subsequent participants to the groups. Specifically, each parameter has a weight based on its possible influence on the proposed task (1 for the DL and previous

experiences with VR, 0.75 for sports activities and video games weekly, 0.5 for gender and age).

After calculating the weighted parameters, the assignment is made in such a way to minimize the weighted sum of the differences in the means and variances between the groups. Moreover, the assignment will maintain a maximum difference in number of participants between groups of 2 (for example, if at some point the iVR group has 8 participants and the W-iVR group has 10, a new participant will be assigned to W-iVR group only after assigning one the iVR group), and no group will have more than 12 participants at the end.

Visit 1 - Inclusion During the first visit informed consent will be collected, and inclusion and non-inclusion criteria verified. The scores at the Mini-Mental State Examination (MMSE) will be retrieved from the patient's medical records at the memory center. After filling out the Initial Questionnaire and the UTAUT2 questionnaire, participants are asked to perform the neuropsychological assessments focused on inhibitory control described in Section 5.3.2.

The data of the UTAUT2 questionnaire and the results of the neuropsychological assessments are used as a baseline for comparisons with the subsequent gathered data.

Game Sessions The trial of the SG consists of eight sessions (2 times per week for a total of 4 weeks) that last one hour each in which the participants will play the SG for 20 minutes, fill out some questionnaires, and answer some questions in an interview (I1) following the flow in Figure 5.17a.

In the odd sessions (S1, S3, S5, S7), participants will play the gym version of the SG, and in the even sessions (S2, S4, S6, S8), they will play the vegetable warehouse version. The participant in the waiting list will have no session planned during these 4 weeks.

The SG provides a familiarization stage to be performed before playing, during which participants can learn how to play the SG. The first time a participant plays in an environment (S1 and S2), this stage is mandatory. Instead, in the following sessions, the experimenter will ask participants if they want to do the familiarization before starting to play.

Following each gameplay session, participants will assess their experience by completing the SAM and responding to questions from Interview 1. Additionally, after the first and last sessions conducted within a Virtual Environment (gym or food warehouse), participants are required to complete the IPQ, GEQ, and DAQ

Visit 2 - session 8 After completing the Interview 1, participants are asked to fill up the UTAUT2 questionnaire, and to perform the neuropsychological assessments focused on inhibitory control described in Section 5.3.2.

Visit 3 Then, after waiting 28 days, the experimenter will administer the UTAUT2 questionnaire to participants, conducting the second interview (I2) that ask the final questions of the study and assessing perform the neuropsychological assessments focused on inhibitory control described in Section 5.3.2.

This will be the last visit for participants in iVR and nVR group. The study ends at this visit for these participants.

Participants in the WL groups will begin the game session as described before and have 2 additional visits:

- Visit 4 (the same as visit 2)
- Visit 5 (the same as visit 3)

The diagram in Fig. 5.18 summarizes the recording protocol, showing which questionnaires will be administered to the participants.

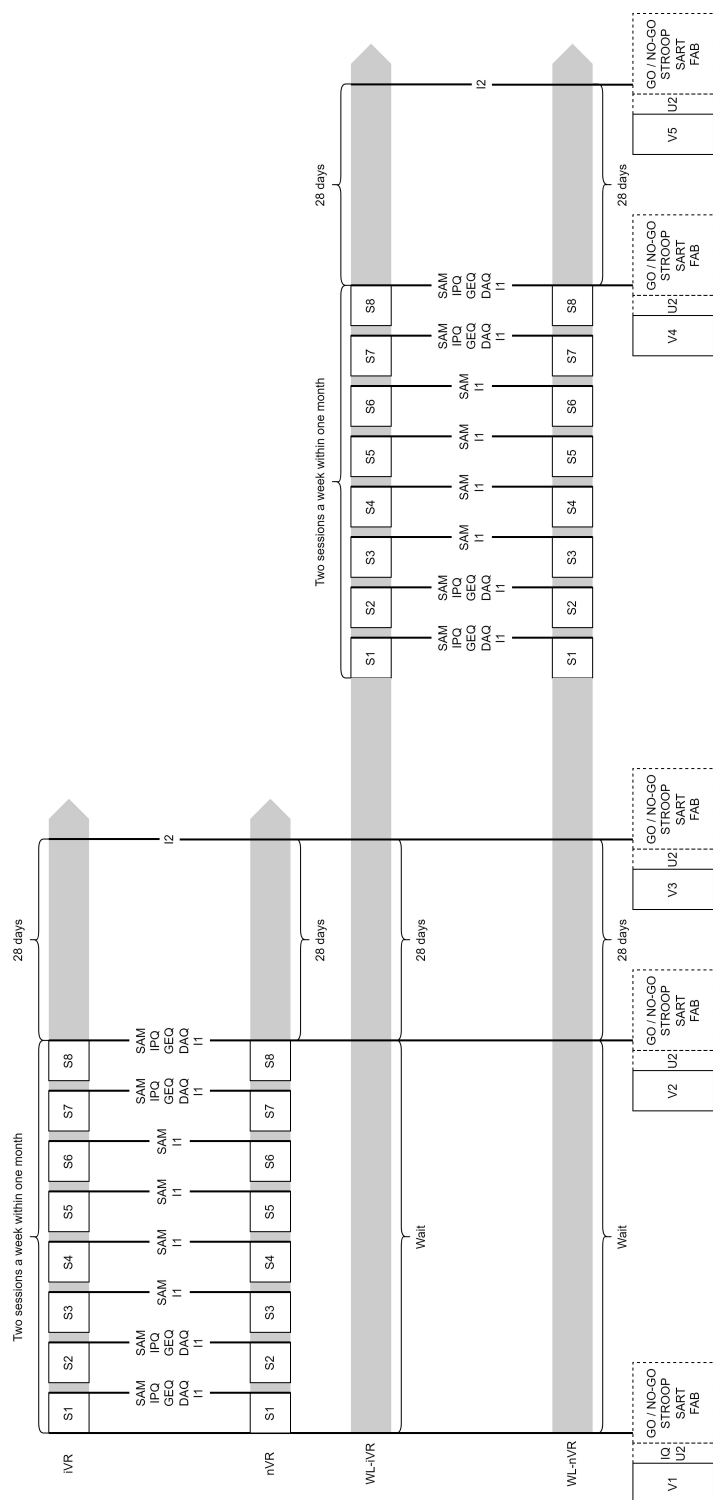


Figure 5.18: Protocol of the user study.

5.3.3 Preliminary results

Due to challenges in recruiting participants meeting the particular inclusion criteria, the user study, which started in October 2024, is still ongoing. This section reports preliminary findings. This section reports questionnaire data covering acceptance, immersion, game experience, and device assessment.

To date, 30 participants have been enrolled out of the targeted 48, with 15 having fully completed the experimental protocol. The results reported are restricted to these 15 completers, distributed across groups as follows: iVR ($n = 4$), nVR ($n = 4$), w-iVR ($n = 4$), and w-nVR ($n = 3$). Data are reported as mean values and standard deviations by group.

Participants demographic

Table 5.12 reports the average values and standard deviations for the demographic data and disease levels calculated.

Table 5.12: Participants characteristics (initial questionnaire) and Disease Level. Dichotomous variables are described as absolute frequencies; continuous variables are described as Mean $\pm$ SD.

	iVR	nVR	w-iVR	w-nVR
Gender				
Male	1	2	1	0
Female	3	2	3	3
Age Range				
60-69	1	1	0	0
70-79	1	3	0	2
≥ 80	2	0	4	1
DiseaseLevel	0.46 ± 0.10	0.52 ± 0.09	0.49 ± 0.18	0.49 ± 0.11
VRHours	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
SportHours	8.13 ± 9.06	2.50 ± 1.22	5.13 ± 2.84	3.67 ± 2.52
GameHours	3.50 ± 7.00	3.75 ± 6.85	3.50 ± 7.00	0.00 ± 0.00

Acceptability

The acceptability of the proposed solution was assessed using the french version [123] of the Unified Theory of Acceptance and Use of Technology and its extension (UTAUT2). These models evaluate user acceptance with a 7-point Likert scale across seven distinct constructs: (i) Performance Expectancy (UTAUT_PE), (ii) Effort Expectancy (UTAUT_EE), (iii) Social Influence (UTAUT_SI), (iv) Facilitating Conditions (UTAUT_FC), (v) Hedonic Motivation (UTAUT_HM), (vi) Price Value (UTAUT_PV), (vii) Habit (UTAUT_H). Table 5.13 details the descriptive statistics collected during the visits.

Table 5.13: Descriptive Statistics for UTAUT Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
VISIT 1				
UTAUT_PE	4.58 $\pm$ 1.07	3.92 $\pm$ 0.50	4.58 $\pm$ 0.74	4.11 $\pm$ 0.84
UTAUT_EE	2.50 $\pm$ 1.24	3.25 $\pm$ 1.10	3.63 $\pm$ 0.66	3.75 $\pm$ 2.46
UTAUT_SI	4.25 $\pm$ 0.50	5.00 $\pm$ 0.72	3.92 $\pm$ 1.13	4.89 $\pm$ 2.01
UTAUT_FC	3.75 $\pm$ 0.74	4.67 $\pm$ 0.82	4.17 $\pm$ 0.96	3.67 $\pm$ 2.40
UTAUT_HM	3.25 $\pm$ 1.37	4.33 $\pm$ 1.19	5.33 $\pm$ 0.72	5.22 $\pm$ 1.35
UTAUT_PV	3.75 $\pm$ 1.20	4.58 $\pm$ 1.10	3.58 $\pm$ 1.83	4.33 $\pm$ 1.53
UTAUT_H	2.00 $\pm$ 0.82	2.67 $\pm$ 0.27	2.42 $\pm$ 1.13	2.44 $\pm$ 1.39
VISIT 2				
UTAUT_PE	4.00 $\pm$ 1.36	5.58 $\pm$ 1.17	5.67 $\pm$ 0.00	4.67 $\pm$ 1.53
UTAUT_EE	4.44 $\pm$ 1.55	4.06 $\pm$ 2.11	4.08 $\pm$ 1.13	3.58 $\pm$ 1.42
UTAUT_SI	4.58 $\pm$ 1.40	5.92 $\pm$ 0.83	3.22 $\pm$ 1.95	5.56 $\pm$ 1.50
UTAUT_FC	5.58 $\pm$ 0.32	5.50 $\pm$ 0.69	5.67 $\pm$ 1.45	4.00 $\pm$ 1.33
UTAUT_HM	6.08 $\pm$ 0.50	4.17 $\pm$ 2.22	5.00 $\pm$ 1.15	4.56 $\pm$ 1.58
UTAUT_PV	4.58 $\pm$ 0.69	4.00 $\pm$ 0.72	3.78 $\pm$ 1.26	4.33 $\pm$ 1.53
UTAUT_H	3.42 $\pm$ 1.71	3.50 $\pm$ 1.00	3.78 $\pm$ 2.46	2.67 $\pm$ 1.67
VISIT 3				
UTAUT_PE	5.00 $\pm$ 1.63	4.17 $\pm$ 1.17	5.42 $\pm$ 1.40	5.44 $\pm$ 0.51
UTAUT_EE	3.63 $\pm$ 1.39	4.75 $\pm$ 1.19	3.25 $\pm$ 1.71	3.17 $\pm$ 1.53
UTAUT_SI	3.50 $\pm$ 2.65	5.75 $\pm$ 0.96	3.50 $\pm$ 2.08	4.33 $\pm$ 1.53
UTAUT_FC	5.83 $\pm$ 0.96	5.00 $\pm$ 1.59	4.50 $\pm$ 1.14	3.00 $\pm$ 1.53
UTAUT_HM	2.75 $\pm$ 1.91	4.08 $\pm$ 1.73	4.92 $\pm$ 0.92	3.44 $\pm$ 1.35
UTAUT_PV	3.42 $\pm$ 0.83	4.00 $\pm$ 1.63	3.67 $\pm$ 1.41	5.22 $\pm$ 0.69
UTAUT_H	4.00 $\pm$ 1.87	3.17 $\pm$ 0.58	4.25 $\pm$ 1.52	3.67 $\pm$ 1.53
VISIT 4				
UTAUT_PE	/	/	5.17 $\pm$ 1.45	4.67 $\pm$ 2.60
UTAUT_EE	/	/	4.38 $\pm$ 0.75	2.58 $\pm$ 1.51
UTAUT_SI	/	/	3.50 $\pm$ 2.74	2.67 $\pm$ 2.89
UTAUT_FC	/	/	3.83 $\pm$ 1.23	2.11 $\pm$ 1.39
UTAUT_HM	/	/	5.67 $\pm$ 1.89	6.67 $\pm$ 0.58
UTAUT_PV	/	/	4.00 $\pm$ 0.82	5.00 $\pm$ 1.00
UTAUT_H	/	/	4.83 $\pm$ 0.96	2.33 $\pm$ 1.20
VISIT 5				
UTAUT_PE	/	/	5.83 $\pm$ 1.04	4.56 $\pm$ 2.50
UTAUT_EE	/	/	4.75 $\pm$ 2.30	3.00 $\pm$ 0.87
UTAUT_SI	/	/	4.00 $\pm$ 2.80	3.44 $\pm$ 1.35
UTAUT_FC	/	/	5.42 $\pm$ 0.96	3.78 $\pm$ 1.02
UTAUT_HM	/	/	6.00 $\pm$ 0.82	3.67 $\pm$ 2.03
UTAUT_PV	/	/	4.92 $\pm$ 0.83	4.44 $\pm$ 0.39
UTAUT_H	/	/	4.42 $\pm$ 1.23	2.44 $\pm$ 1.26

Inhibitory Control

Go/No-Go task from the TAP battery

Table 5.14: Descriptive Statistics for the Inhibitory Control test of the TAP battery at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
VISIT 1				
Nb_correct_responses	20.00 $\pm$ 0.00	20.00 $\pm$ 0.00	19.50 $\pm$ 1.00	18.30 $\pm$ 1.53
Nb_errors	0.75 $\pm$ 0.96	1.00 $\pm$ 0.82	2.25 $\pm$ 1.26	0.33 $\pm$ 0.58
VISIT 2				
Nb_correct_responses	19.00 $\pm$ 1.41	18.80 $\pm$ 1.89	19.00 $\pm$ 1.00	15.70 $\pm$ 7.51
Nb_errors	2.25 $\pm$ 3.86	0.50 $\pm$ 0.58	2.67 $\pm$ 2.52	0.33 $\pm$ 0.58
VISIT 3				
Nb_correct_responses	19.00 $\pm$ 1.41	16.80 $\pm$ 5.85	19.80 $\pm$ 0.50	19.00 $\pm$ 1.00
Nb_errors	1.75 $\pm$ 2.87	2.25 $\pm$ 2.87	1.00 $\pm$ 1.41	0.33 $\pm$ 0.58
VISIT 4				
Nb_correct_responses	/	/	20.00 $\pm$ 0.00	19.70 $\pm$ 0.58
Nb_errors	/	/	0.75 $\pm$ 0.96	0.00 $\pm$ 0.00
VISIT 5				
Nb_correct_responses	/	/	19.30 $\pm$ 0.96	19.30 $\pm$ 1.15
Nb_errors	/	/	1.00 $\pm$ 1.15	0.67 $\pm$ 0.58

Stroop test

Table 5.15: Descriptive Statistics for Stroop test at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
VISIT 1				
Stroop_indice_if	1.05 $\pm$ 0.18	1.16 $\pm$ 0.21	1.04 $\pm$ 0.20	1.33 $\pm$ 0.28
Stroop_indice_IF	1.84 $\pm$ 0.45	1.24 $\pm$ 0.83	1.93 $\pm$ 0.81	2.23 $\pm$ 0.17
Stroop_errors	2.25 $\pm$ 2.06	1.50 $\pm$ 1.91	5.50 $\pm$ 4.36	2.33 $\pm$ 2.08
VISIT 2				
Stroop_indice_if	1.44 $\pm$ 0.22	1.21 $\pm$ 0.44	1.13 $\pm$ 0.26	1.36 $\pm$ 0.18
Stroop_indice_IF	2.74 $\pm$ 0.17	1.19 $\pm$ 0.95	2.94 $\pm$ 1.09	2.76 $\pm$ 0.75
Stroop_errors	3.00 $\pm$ 1.83	3.75 $\pm$ 1.26	3.67 $\pm$ 2.31	2.67 $\pm$ 3.06
VISIT 3				
Stroop_indice_if	1.19 $\pm$ 0.13	1.23 $\pm$ 0.20	1.10 $\pm$ 0.34	1.23 $\pm$ 0.16
Stroop_indice_IF	2.84 $\pm$ 1.26	1.39 $\pm$ 0.94	2.23 $\pm$ 0.90	2.66 $\pm$ 1.20
Stroop_errors	1.00 $\pm$ 0.82	2.75 $\pm$ 2.06	2.25 $\pm$ 1.71	2.33 $\pm$ 2.52
VISIT 4				
Stroop_indice_if	/	/	0.83 $\pm$ 0.18	1.46 $\pm$ 0.22

Table 5.15: Descriptive Statistics for Stroop test at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
Stroop_indice_IF	/	/	1.92 $\pm$ 1.02	2.44 $\pm$ 0.45
Stroop_errors	/	/	3.50 $\pm$ 3.11	2.00 $\pm$ 1.00
VISIT 5				
Stroop_indice_if	/	/	1.09 $\pm$ 0.21	1.29 $\pm$ 0.04
Stroop_indice_IF	/	/	2.10 $\pm$ 0.67	1.97 $\pm$ 0.18
Stroop_errors	/	/	1.25 $\pm$ 0.96	1.67 $\pm$ 2.08

SART test

Table 5.16: Descriptive Statistics for SART test at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
VISIT 1				
errors_Go	22.50 $\pm$ 14.70	25.30 $\pm$ 6.13	16.80 $\pm$ 22.50	32.30 $\pm$ 7.37
errors_NoGo	4.75 $\pm$ 4.57	5.50 $\pm$ 3.42	4.75 $\pm$ 4.03	4.00 $\pm$ 2.65
temps_Go	490 $\pm$ 52.70	605 $\pm$ 109	396 $\pm$ 266	629 $\pm$ 42.10
nb_anticipatory	4.00 $\pm$ 3.74	5.25 $\pm$ 4.79	1.75 $\pm$ 2.06	6.33 $\pm$ 3.79
temps_anticipatory	29.90 $\pm$ 20.30	34.20 $\pm$ 23.10	17.40 $\pm$ 26.40	42.70 $\pm$ 8.96
VISIT 2				
errors_Go	42.80 $\pm$ 63.20	24.50 $\pm$ 7.33	21.70 $\pm$ 24.50	32.00 $\pm$ 8.00
errors_NoGo	4.50 $\pm$ 2.65	6.75 $\pm$ 3.95	6.00 $\pm$ 3.61	6.00 $\pm$ 4.58
temps_Go	509 $\pm$ 48.80	584 $\pm$ 107	502 $\pm$ 53.30	622 $\pm$ 46.20
nb_anticipatory	2.50 $\pm$ 2.08	4.75 $\pm$ 2.36	1.67 $\pm$ 2.08	5.33 $\pm$ 4.73
VISIT 3				
errors_Go	34.30 $\pm$ 46.30	18.50 $\pm$ 14.50	13.00 $\pm$ 12.10	33.30 $\pm$ 16.90
errors_NoGo	5.75 $\pm$ 3.59	3.25 $\pm$ 2.22	7.25 $\pm$ 4.50	5.33 $\pm$ 3.21
temps_Go	480 $\pm$ 49.70	611 $\pm$ 95.60	388 $\pm$ 242	641 $\pm$ 30.90
nb_anticipatory	3.50 $\pm$ 4.43	8.88 $\pm$ 14.50	2.50 $\pm$ 2.65	3.67 $\pm$ 2.08
VISIT 4				
errors_Go	/	/	10.30 $\pm$ 13.40	19.70 $\pm$ 14.40
errors_NoGo	/	/	4.75 $\pm$ 2.06	5.67 $\pm$ 4.62
temps_Go	/	/	464 $\pm$ 34.10	526 $\pm$ 7.24
nb_anticipatory	/	/	2.00 $\pm$ 2.71	5.67 $\pm$ 4.04
VISIT 5				
errors_Go	/	/	15.50 $\pm$ 15.20	24.7 $\pm$ 7.64
errors_NoGo	/	/	6.50 $\pm$ 2.89	9.00 $\pm$ 2.00
temps_Go	/	/	469 $\pm$ 22.70	485 $\pm$ 46.50
nb_anticipatory	/	/	2.25 $\pm$ 4.50	6.67 $\pm$ 5.51

Frontal Assessment Battery The last metric used to assess inhibitory control was the Frontal Assessment Battery (FAB) with its six subscales and total score (FAB_TOT): (i) Conceptualization (FAB_Co), (ii) Mental Flexibility (FAB_MF), (iii) Motor Programming (FAB_MP), (iv) Sensitivity to Interference (FAB_SI), (v) Inhibitory Control (FAB_IC), and (vi) Environmental Autonomy (FAB_EA).

Table 5.17 reports the descriptive statistics collected across visits.

Table 5.17: Descriptive Statistics for FAB Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
VISIT 1				
FAB_TOT	11.80 $\pm$ 0.96	13.80 $\pm$ 3.30	11.50 $\pm$ 4.43	12.30 $\pm$ 1.53
FAB_Co	1.75 $\pm$ 1.50	2.50 $\pm$ 0.58	2.75 $\pm$ 0.50	3.00 $\pm$ 0.00
FAB_MF	2.75 $\pm$ 0.50	2.25 $\pm$ 1.50	1.50 $\pm$ 1.00	2.33 $\pm$ 0.58
FAB_MP	1.25 $\pm$ 1.50	2.25 $\pm$ 0.96	1.75 $\pm$ 1.50	1.00 $\pm$ 0.00
FAB_SI	2.50 $\pm$ 0.58	2.50 $\pm$ 0.58	2.25 $\pm$ 1.50	2.33 $\pm$ 0.58
FAB_IC	0.50 $\pm$ 0.58	0.75 $\pm$ 0.50	0.75 $\pm$ 0.96	0.67 $\pm$ 0.58
FAB_EA	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00	2.50 $\pm$ 1.00	3.00 $\pm$ 0.00
VISIT 2				
FAB_TOT	14.00 $\pm$ 3.37	15.30 $\pm$ 0.96	12.30 $\pm$ 3.06	11.30 $\pm$ 1.53
FAB_Co	2.25 $\pm$ 0.96	3.00 $\pm$ 0.00	2.67 $\pm$ 0.58	3.00 $\pm$ 0.00
FAB_MF	2.75 $\pm$ 0.50	2.25 $\pm$ 1.50	1.67 $\pm$ 1.15	2.00 $\pm$ 1.00
FAB_MP	2.50 $\pm$ 1.00	2.50 $\pm$ 0.58	1.67 $\pm$ 1.15	1.00 $\pm$ 1.00
FAB_SI	2.50 $\pm$ 0.58	3.00 $\pm$ 0.00	2.67 $\pm$ 0.58	1.67 $\pm$ 1.53
FAB_IC	1.00 $\pm$ 1.41	1.25 $\pm$ 1.26	0.67 $\pm$ 1.15	0.67 $\pm$ 0.58
FAB_EA	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00
VISIT 3				
FAB_TOT	15.00 $\pm$ 0.82	14.80 $\pm$ 0.50	11.80 $\pm$ 4.03	12.30 $\pm$ 0.58
FAB_Co	2.75 $\pm$ 0.50	2.75 $\pm$ 0.50	2.75 $\pm$ 0.50	3.00 $\pm$ 0.00
FAB_MF	2.50 $\pm$ 0.58	2.50 $\pm$ 1.00	1.75 $\pm$ 1.50	1.33 $\pm$ 0.58
FAB_MP	2.50 $\pm$ 0.58	2.50 $\pm$ 0.58	1.75 $\pm$ 0.96	2.00 $\pm$ 1.73
FAB_SI	2.75 $\pm$ 0.50	2.75 $\pm$ 0.50	2.25 $\pm$ 1.50	2.33 $\pm$ 0.58
FAB_IC	1.50 $\pm$ 1.00	1.25 $\pm$ 1.50	1.00 $\pm$ 1.41	0.67 $\pm$ 0.58
FAB_EA	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00	2.25 $\pm$ 1.50	3.00 $\pm$ 0.00
VISIT 4				
FAB_TOT	/	/	13.80 $\pm$ 0.96	11.30 $\pm$ 1.53
FAB_Co	/	/	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00
FAB_MF	/	/	1.75 $\pm$ 0.96	1.67 $\pm$ 0.58
FAB_MP	/	/	2.75 $\pm$ 0.50	1.33 $\pm$ 1.53
FAB_SI	/	/	1.50 $\pm$ 1.29	2.00 $\pm$ 1.73
FAB_IC	/	/	1.75 $\pm$ 0.96	0.33 $\pm$ 0.58
FAB_EA	/	/	3.00 $\pm$ 0.00	3.00 $\pm$ 0.00
VISIT 5				
FAB_TOT	/	/	13.30 $\pm$ 1.71	14.30 $\pm$ 2.52

Table 5.17: Descriptive Statistics for FAB Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
FAB_Co	/	/	2.75 ± 0.50	3.00 ± 0.00
FAB_MF	/	/	1.75 ± 0.96	1.33 ± 1.15
FAB_MP	/	/	2.50 ± 1.00	2.33 ± 0.58
FAB_SI	/	/	2.50 ± 0.58	3.00 ± 0.00
FAB_IC	/	/	0.75 ± 1.50	1.67 ± 1.15
FAB_EA	/	/	3.00 ± 0.00	3.00 ± 0.00

User Experience regarding SG

Affective Reaction The affective reaction was measured with the SAM[36] pictorial assessment measuring the Pleasure (SAM_P), Arousal (SAM_A), and Dominance (SAM_D). Table 5.18 details the descriptive statistics collected after each game session.

Table 5.18: Descriptive Statistics of affective reaction for Pleasure, Arousal, and Dominance at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 1				
SAM_P	2.50 ± 1.29	2.50 ± 1.29	2.75 ± 1.71	2.00 ± 1.00
SAM_A	2.75 ± 0.50	3.25 ± 1.26	3.75 ± 0.96	4.33 ± 1.15
SAM_D	3.00 ± 1.41	3.00 ± 0.00	3.25 ± 1.26	4.00 ± 1.00
SESSION 2				
SAM_P	3.25 ± 1.26	2.25 ± 0.50	3.25 ± 1.50	1.33 ± 0.58
SAM_A	3.75 ± 0.96	3.50 ± 1.29	4.00 ± 0.82	3.67 ± 0.58
SAM_D	3.25 ± 0.96	3.50 ± 0.58	4.25 ± 0.96	3.33 ± 1.15
SESSION 3				
SAM_P	3.00 ± 0.82	3.00 ± 1.15	2.25 ± 0.96	1.33 ± 0.58
SAM_A	3.50 ± 1.29	3.75 ± 1.26	2.00 ± 1.41	4.00 ± 1.00
SAM_D	3.25 ± 1.71	3.00 ± 1.63	3.00 ± 1.41	4.00 ± 1.00
SESSION 4				
SAM_P	2.75 ± 0.50	1.67 ± 0.58	2.50 ± 0.58	2.33 ± 0.58
SAM_A	4.50 ± 0.58	3.67 ± 1.53	2.25 ± 0.96	4.33 ± 1.15
SAM_D	3.75 ± 1.89	4.00 ± 1.00	4.75 ± 0.50	3.67 ± 1.15
SESSION 5				
SAM_P	3.00 ± 1.41	3.00 ± 0.82	2.00 ± 0.82	2.33 ± 0.58
SAM_A	4.25 ± 0.96	3.25 ± 1.26	2.75 ± 1.71	4.67 ± 0.58
SAM_D	2.75 ± 2.06	2.75 ± 1.71	4.25 ± 0.50	3.33 ± 2.08
SESSION 6				
SAM_P	2.75 ± 0.50	2.50 ± 1.00	2.25 ± 1.89	1.33 ± 0.58
SAM_A	4.00 ± 0.82	4.00 ± 1.15	2.50 ± 1.73	3.67 ± 1.53
SAM_D	4.25 ± 0.96	3.75 ± 0.96	4.25 ± 0.50	4.67 ± 0.58

Table 5.18: Descriptive Statistics of affective reaction for Pleasure, Arousal, and Dominance at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 7				
SAM_P	2.50 $\pm$ 0.58	2.00 $\pm$ 1.41	1.75 $\pm$ 0.50	1.00 $\pm$ 0.00
SAM_A	3.50 $\pm$ 1.29	3.50 $\pm$ 1.91	3.75 $\pm$ 0.50	4.00 $\pm$ 1.41
SAM_D	3.75 $\pm$ 1.89	3.25 $\pm$ 1.50	4.25 $\pm$ 0.50	4.50 $\pm$ 0.71
SESSION 8				
SAM_P	3.00 $\pm$ 0.00	2.50 $\pm$ 0.58	2.00 $\pm$ 0.82	2.00 $\pm$ 1.00
SAM_A	4.00 $\pm$ 0.82	3.25 $\pm$ 1.26	2.75 $\pm$ 1.50	3.67 $\pm$ 0.58
SAM_D	3.75 $\pm$ 1.26	3.75 $\pm$ 0.96	4.75 $\pm$ 0.50	4.33 $\pm$ 0.58

Presence The sense of presence in the VE was measured using the metrics provided by the IPQ questionnaire [279], which are a total score (IPQ-TOTP), a general item assessing the sense of "being there" (G1P), and three subscales: spatial presence (IPQ-SPP), involvement (IPQ-INVP), and experienced realism (IPQ-REALP). Table 5.19 reports the descriptive statistics collected after game sessions. Sessions 1 and 7 were conducted in the gym VE, while Sessions 2 and 8 used the warehouse VE.

Table 5.19: Descriptive Statistics for the sense of presence at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 1				
G1P	5.25 $\pm$ 0.96	3.75 $\pm$ 0.96	4.75 $\pm$ 2.50	5.00 $\pm$ 1.73
IPQ-SPP	4.60 $\pm$ 0.52	3.15 $\pm$ 0.77	5.15 $\pm$ 1.14	3.47 $\pm$ 0.76
IPQ-INVP	4.56 $\pm$ 1.13	3.31 $\pm$ 1.39	4.63 $\pm$ 1.16	3.75 $\pm$ 1.52
IPQ-REALP	2.63 $\pm$ 0.97	1.56 $\pm$ 1.07	2.75 $\pm$ 1.27	1.58 $\pm$ 2.13
IPQ-TOTP	4.07 $\pm$ 0.54	2.79 $\pm$ 0.62	4.29 $\pm$ 1.19	3.12 $\pm$ 1.34
SESSION 2				
G1P	4.25 $\pm$ 2.87	4.00 $\pm$ 2.45	5.50 $\pm$ 1.00	5.33 $\pm$ 0.58
IPQ-SPP	4.25 $\pm$ 1.11	3.80 $\pm$ 2.08	5.50 $\pm$ 0.66	4.13 $\pm$ 0.83
IPQ-INVP	4.38 $\pm$ 1.71	3.25 $\pm$ 1.90	4.56 $\pm$ 1.13	4.25 $\pm$ 2.41
IPQ-REALP	3.31 $\pm$ 0.77	1.63 $\pm$ 1.13	3.19 $\pm$ 0.85	0.83 $\pm$ 0.63
IPQ-TOTP	4.02 $\pm$ 0.85	3.04 $\pm$ 1.27	4.57 $\pm$ 0.47	3.31 $\pm$ 1.16
SESSION 7				
G1P	5.00 $\pm$ 1.41	4.25 $\pm$ 2.87	5.75 $\pm$ 0.50	6.00 $\pm$ 0.00
IPQ-SPP	4.50 $\pm$ 1.25	3.85 $\pm$ 1.82	5.10 $\pm$ 0.48	4.80 $\pm$ 0.00
IPQ-INVP	4.00 $\pm$ 0.74	4.63 $\pm$ 1.45	5.06 $\pm$ 0.83	5.25 $\pm$ 1.06
IPQ-REALP	3.38 $\pm$ 1.36	2.38 $\pm$ 1.23	2.25 $\pm$ 1.37	0.13 $\pm$ 0.18
IPQ-TOTP	4.07 $\pm$ 0.67	3.68 $\pm$ 1.54	4.32 $\pm$ 0.43	3.68 $\pm$ 0.25
SESSION 8				
G1P	5.50 $\pm$ 1.00	5.00 $\pm$ 0.82	5.75 $\pm$ 0.50	4.33 $\pm$ 1.53
IPQ-SPP	4.95 $\pm$ 0.76	4.20 $\pm$ 0.67	4.85 $\pm$ 0.77	3.47 $\pm$ 0.23

Table 5.19: Descriptive Statisticstistics for the sense of presence at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
IPQ-INVP	5.31 $\pm$ 0.55	3.56 $\pm$ 0.47	4.19 $\pm$ 1.96	2.75 $\pm$ 2.05
IPQ-REALP	2.75 $\pm$ 1.57	1.88 $\pm$ 1.53	2.69 $\pm$ 1.55	1.00 $\pm$ 1.73
IPQ-TOTP	4.46 $\pm$ 0.39	3.41 $\pm$ 0.74	4.11 $\pm$ 0.24	2.62 $\pm$ 0.95

Game Experience during gameplay The game experience during gameplay was measured using the GEQ Core Module [141] with its seven subscales: (i) Competence (GEQ_CO), (ii) Sensory and Imaginative Immersion (GEQ_SII), (iii) Flow (GEQ_F), (iv) Tension/Annoyance (GEQ_TA), (v) Challenge (GEQ_Ch), (vi) Negative Affect (GEQ_NA), (vii) and Positive Affect (GEQ_PA). Table 5.20 reports the descriptive statistics collected after game sessions. Sessions 1 and 7 were conducted in the gym VE, while Sessions 2 and 8 used the warehouse VE.

Table 5.20: Descriptive Statistics for GEQ constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 1				
GEQ_Co	3.25 $\pm$ 1.05	3.35 $\pm$ 0.66	3.80 $\pm$ 1.21	3.40 $\pm$ 0.60
GEQ_SII	2.29 $\pm$ 1.03	2.67 $\pm$ 0.72	2.58 $\pm$ 0.62	2.89 $\pm$ 1.13
GEQ_F	4.20 $\pm$ 0.63	3.80 $\pm$ 1.21	4.70 $\pm$ 0.38	4.20 $\pm$ 0.92
GEQ_TA	1.50 $\pm$ 0.64	1.25 $\pm$ 0.32	1.08 $\pm$ 0.17	1.11 $\pm$ 0.19
GEQ_Ch	2.35 $\pm$ 0.53	2.60 $\pm$ 0.33	3.85 $\pm$ 1.59	1.87 $\pm$ 0.46
GEQ_NA	1.44 $\pm$ 0.59	1.50 $\pm$ 0.54	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
GEQ_PA	3.20 $\pm$ 1.48	3.30 $\pm$ 1.31	4.15 $\pm$ 0.70	4.20 $\pm$ 0.80
SESSION 2				
GEQ_Co	2.85 $\pm$ 0.38	3.30 $\pm$ 0.68	3.95 $\pm$ 0.60	3.27 $\pm$ 0.42
GEQ_SII	2.17 $\pm$ 1.21	2.83 $\pm$ 0.76	2.79 $\pm$ 0.69	2.28 $\pm$ 0.77
GEQ_F	4.15 $\pm$ 1.23	3.65 $\pm$ 1.00	4.35 $\pm$ 1.05	4.20 $\pm$ 0.40
GEQ_TA	1.25 $\pm$ 0.17	1.08 $\pm$ 0.17	1.33 $\pm$ 0.47	1.00 $\pm$ 0.00
GEQ_Ch	2.65 $\pm$ 0.53	2.00 $\pm$ 0.52	2.15 $\pm$ 0.55	2.07 $\pm$ 0.92
GEQ_NA	1.56 $\pm$ 1.13	1.06 $\pm$ 0.13	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00
GEQ_PA	2.60 $\pm$ 0.63	4.05 $\pm$ 1.00	4.15 $\pm$ 0.60	4.20 $\pm$ 0.53
SESSION 7				
GEQ_Co	3.15 $\pm$ 0.34	3.75 $\pm$ 0.50	4.05 $\pm$ 0.25	4.50 $\pm$ 0.71
GEQ_SII	1.96 $\pm$ 0.42	2.33 $\pm$ 0.69	2.25 $\pm$ 0.32	2.08 $\pm$ 0.12
GEQ_F	4.40 $\pm$ 0.59	3.40 $\pm$ 1.10	4.55 $\pm$ 0.90	5.00 $\pm$ 0.00
GEQ_TA	1.58 $\pm$ 0.42	1.33 $\pm$ 0.67	1.08 $\pm$ 0.17	1.00 $\pm$ 0.00
GEQ_Ch	2.45 $\pm$ 1.32	2.20 $\pm$ 0.28	2.05 $\pm$ 1.00	2.10 $\pm$ 1.56
GEQ_NA	1.44 $\pm$ 0.72	1.94 $\pm$ 1.23	1.13 $\pm$ 0.25	1.00 $\pm$ 0.00
GEQ_PA	2.75 $\pm$ 0.97	3.60 $\pm$ 1.08	4.20 $\pm$ 0.65	4.50 $\pm$ 0.14
SESSION 8				

Table 5.20: Descriptive Statistics for GEQ constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
GEQ_Co	3.40 $\pm$ 0.49	3.90 $\pm$ 0.53	4.25 $\pm$ 0.76	4.07 $\pm$ 0.81
GEQ_SII	2.58 $\pm$ 0.55	2.75 $\pm$ 0.65	2.46 $\pm$ 0.46	2.39 $\pm$ 0.54
GEQ_F	5.05 $\pm$ 0.74	3.40 $\pm$ 0.91	4.20 $\pm$ 0.78	3.73 $\pm$ 0.83
GEQ_TA	1.50 $\pm$ 0.64	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
GEQ_Ch	2.45 $\pm$ 1.00	1.50 $\pm$ 0.62	2.60 $\pm$ 0.94	1.67 $\pm$ 0.99
GEQ_NA	1.25 $\pm$ 0.29	1.69 $\pm$ 0.77	1.06 $\pm$ 0.13	1.17 $\pm$ 0.29
GEQ_PA	3.30 $\pm$ 0.68	4.05 $\pm$ 0.76	4.25 $\pm$ 0.62	4.67 $\pm$ 0.12

Game Experience after gameplay The Post-game Module of the GEQ [141] was used to assess game experience after gameplay with its four components: (i) Positive Experience (GEQ_PE), (ii) Negative Experience (GEQ_NE), (iii) Tiredness (GEQ_Ti), (iv) and Return to Reality (GEQ_RtR) Table 5.21 reports the descriptive statistics collected after game sessions. Sessions 1 and 7 were conducted in the gym VE, while Sessions 2 and 8 used the warehouse VE.

Table 5.21: Statistics for GEQ Post Game Module constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 1				
GEQ_PE	1.95 $\pm$ 0.57	2.30 $\pm$ 1.16	2.90 $\pm$ 0.58	2.87 $\pm$ 0.61
GEQ_NE	1.25 $\pm$ 0.32	1.25 $\pm$ 0.29	1.04 $\pm$ 0.08	1.06 $\pm$ 0.10
GEQ_Ti	1.75 $\pm$ 0.50	2.38 $\pm$ 1.11	2.13 $\pm$ 1.31	1.67 $\pm$ 0.29
GEQ_RtR	1.25 $\pm$ 0.50	1.25 $\pm$ 0.32	1.25 $\pm$ 0.32	1.67 $\pm$ 1.15
SESSION 2				
GEQ_PE	1.85 $\pm$ 0.66	2.70 $\pm$ 0.38	3.35 $\pm$ 0.44	2.53 $\pm$ 0.31
GEQ_NE	1.38 $\pm$ 0.55	1.04 $\pm$ 0.08	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
GEQ_Ti	2.00 $\pm$ 1.35	1.25 $\pm$ 0.50	1.63 $\pm$ 0.48	1.50 $\pm$ 0.87
GEQ_RtR	1.00 $\pm$ 0.00	1.08 $\pm$ 0.17	1.42 $\pm$ 0.50	1.00 $\pm$ 0.00
SESSION 7				
GEQ_PE	2.35 $\pm$ 1.12	2.35 $\pm$ 1.08	3.45 $\pm$ 0.64	3.10 $\pm$ 0.99
GEQ_NE	1.13 $\pm$ 0.16	1.42 $\pm$ 0.83	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
GEQ_Ti	2.75 $\pm$ 1.85	2.13 $\pm$ 1.65	1.75 $\pm$ 0.65	2.00 $\pm$ 1.41
GEQ_RtR	1.17 $\pm$ 0.33	1.17 $\pm$ 0.33	1.50 $\pm$ 0.43	1.00 $\pm$ 0.00
SESSION 8				
GEQ_PE	2.10 $\pm$ 0.66	2.80 $\pm$ 0.91	3.65 $\pm$ 0.92	2.87 $\pm$ 0.64
GEQ_NE	1.13 $\pm$ 0.25	1.17 $\pm$ 0.19	1.17 $\pm$ 0.33	1.00 $\pm$ 0.00
GEQ_Ti	3.25 $\pm$ 1.50	1.00 $\pm$ 0.00	1.63 $\pm$ 0.48	2.33 $\pm$ 2.31
GEQ_RtR	1.00 $\pm$ 0.00	1.17 $\pm$ 0.33	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00

Device Assessment The user experience in using the technology was assessed using the Device Assessment Questionnaire (DAQ) [90]. Table 5.22 reports the descriptive

statistics collected after game sessions. Sessions 1 and 7 were conducted in the gym VE, while Sessions 2 and 8 used the warehouse VE.

Table 5.22: Descriptive statistics for the items of the DAQ questionnaire at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
SESSION 1				
Required force	2.50 $\pm$ 0.58	2.75 $\pm$ 0.50	2.75 $\pm$ 0.50	2.67 $\pm$ 0.58
Smoothness of movements	4.25 $\pm$ 0.96	4.00 $\pm$ 0.82	3.75 $\pm$ 0.50	4.33 $\pm$ 0.58
Mental effort	3.25 $\pm$ 0.50	3.50 $\pm$ 0.58	3.50 $\pm$ 1.29	3.33 $\pm$ 0.58
Physical effort	2.75 $\pm$ 0.50	2.50 $\pm$ 1.00	2.50 $\pm$ 0.58	2.67 $\pm$ 0.58
Accuracy of the movements	1.25 $\pm$ 0.50	2.25 $\pm$ 0.96	3.50 $\pm$ 1.00	2.00 $\pm$ 1.73
Speed of the movements	2.50 $\pm$ 0.58	2.75 $\pm$ 0.50	2.25 $\pm$ 0.96	2.33 $\pm$ 1.15
Finger fatigue	1.00 $\pm$ 0.00	1.75 $\pm$ 1.50	1.75 $\pm$ 1.50	1.00 $\pm$ 0.00
Wrists fatigue	1.00 $\pm$ 0.00	1.50 $\pm$ 1.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Arms fatigue	1.25 $\pm$ 0.50	1.50 $\pm$ 1.00	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00
Shoulder fatigue	2.00 $\pm$ 1.15	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Neck fatigue	1.50 $\pm$ 0.58	2.00 $\pm$ 1.41	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
General comfort	3.00 $\pm$ 1.41	3.75 $\pm$ 0.50	3.50 $\pm$ 0.58	4.33 $\pm$ 0.58
Overall ease of use	4.25 $\pm$ 0.96	3.50 $\pm$ 0.58	4.00 $\pm$ 0.00	4.00 $\pm$ 0.00
SESSION 2				
Required force	3.25 $\pm$ 0.50	2.75 $\pm$ 0.50	3.25 $\pm$ 0.50	2.67 $\pm$ 0.58
Smoothness of movements	4.00 $\pm$ 1.41	3.50 $\pm$ 0.58	4.00 $\pm$ 0.82	4.00 $\pm$ 1.00
Mental effort	3.50 $\pm$ 1.29	3.25 $\pm$ 0.50	3.50 $\pm$ 1.00	3.33 $\pm$ 0.58
Physical effort	3.75 $\pm$ 0.50	3.50 $\pm$ 0.58	2.50 $\pm$ 1.29	2.00 $\pm$ 1.00
Accuracy of the movements	2.75 $\pm$ 1.26	2.25 $\pm$ 1.50	3.25 $\pm$ 1.71	2.00 $\pm$ 1.73
Speed of the movements	2.00 $\pm$ 0.82	2.25 $\pm$ 0.96	2.25 $\pm$ 0.96	1.67 $\pm$ 1.15
Finger fatigue	1.75 $\pm$ 1.50	1.75 $\pm$ 1.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Wrists fatigue	1.00 $\pm$ 0.00	1.75 $\pm$ 1.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Arms fatigue	1.00 $\pm$ 0.00	1.50 $\pm$ 1.00	2.00 $\pm$ 1.41	1.00 $\pm$ 0.00
Shoulder fatigue	3.25 $\pm$ 1.71	1.75 $\pm$ 1.50	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00
Neck fatigue	1.00 $\pm$ 0.00	2.00 $\pm$ 2.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
General comfort	3.00 $\pm$ 0.82	3.75 $\pm$ 0.96	4.25 $\pm$ 0.50	4.33 $\pm$ 0.58
Overall ease of use	3.50 $\pm$ 1.29	3.75 $\pm$ 1.26	4.25 $\pm$ 0.96	4.33 $\pm$ 0.58
SESSION 7				
Required force	2.75 $\pm$ 0.50	3.00 $\pm$ 0.00	2.25 $\pm$ 0.50	2.00 $\pm$ 1.41
Smoothness of movements	3.50 $\pm$ 1.29	4.00 $\pm$ 1.15	4.00 $\pm$ 0.00	4.00 $\pm$ 1.41
Mental effort	3.75 $\pm$ 1.26	3.00 $\pm$ 0.00	2.25 $\pm$ 0.50	1.50 $\pm$ 0.71
Physical effort	4.25 $\pm$ 0.50	3.00 $\pm$ 0.00	2.00 $\pm$ 0.00	2.50 $\pm$ 0.71
Accuracy of the movements	1.25 $\pm$ 0.50	3.25 $\pm$ 1.71	2.25 $\pm$ 0.96	1.00 $\pm$ 0.00
Speed of the movements	2.25 $\pm$ 0.96	3.00 $\pm$ 0.00	2.50 $\pm$ 0.58	2.50 $\pm$ 0.71
Finger fatigue	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	2.00 $\pm$ 1.41
Wrists fatigue	1.00 $\pm$ 0.00	2.00 $\pm$ 2.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Arms fatigue	2.50 $\pm$ 1.91	1.75 $\pm$ 1.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00

Table 5.22: Descriptive statistics for the items of the DAQ questionnaire at different time instants. Values are reported as Mean $\pm$ Standard Deviation.

	iVR	nVR	w-iVR	w-nVR
Shoulder fatigue	3.50 $\pm$ 1.73	1.00 $\pm$ 0.00	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00
Neck fatigue	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
General comfort	2.25 $\pm$ 0.96	3.75 $\pm$ 0.96	4.00 $\pm$ 0.00	5.00 $\pm$ 0.00
Overall ease of use	4.25 $\pm$ 0.50	4.00 $\pm$ 1.15	3.75 $\pm$ 0.50	5.00 $\pm$ 0.00
SESSION 8				
Required force	2.50 $\pm$ 0.58	3.25 $\pm$ 0.50	2.75 $\pm$ 0.96	2.33 $\pm$ 1.53
Smoothness of movements	4.00 $\pm$ 1.15	3.75 $\pm$ 0.50	4.00 $\pm$ 0.82	4.33 $\pm$ 0.58
Mental effort	3.50 $\pm$ 0.58	3.00 $\pm$ 0.00	2.75 $\pm$ 0.96	3.33 $\pm$ 1.53
Physical effort	3.00 $\pm$ 0.82	3.50 $\pm$ 1.00	2.25 $\pm$ 1.26	3.33 $\pm$ 1.53
Accuracy of the movements	2.00 $\pm$ 1.41	2.75 $\pm$ 1.26	1.50 $\pm$ 0.58	1.67 $\pm$ 1.15
Speed of the movements	2.00 $\pm$ 0.82	3.00 $\pm$ 0.00	2.00 $\pm$ 0.82	2.00 $\pm$ 1.00
Finger fatigue	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Wrists fatigue	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
Arms fatigue	1.50 $\pm$ 1.00	1.00 $\pm$ 0.00	1.50 $\pm$ 1.00	1.00 $\pm$ 0.00
Shoulder fatigue	3.00 $\pm$ 1.41	1.00 $\pm$ 0.00	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00
Neck fatigue	1.25 $\pm$ 0.50	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00	1.00 $\pm$ 0.00
General comfort	3.50 $\pm$ 0.58	3.50 $\pm$ 0.58	4.25 $\pm$ 0.50	4.67 $\pm$ 0.58
Overall ease of use	4.25 $\pm$ 0.96	4.00 $\pm$ 1.41	4.00 $\pm$ 0.00	4.00 $\pm$ 1.00

6

Conclusions and Future Work

This thesis investigated the potential of XR and SGs in healthcare. Specifically, it addressed the need for effective procedural training for nurses and doctors, as well as cognitive training for healthy and clinical populations. Through a series of user studies and protocol developments, this work has examined how specific design elements (level of immersion, use of haptic devices, feedback modalities, use of adaptive techniques to calibrate in real-time the difficulty of the task, and multi-user settings) impact learning outcomes, user experience, and performance in procedural and cognitive tasks.

6.1 XR Procedural Training for Nurses and Doctors

The thesis investigated two distinct design elements within the context of training healthcare professionals.

To assess how different feedback modalities within the VE affect user experience and BLSD performance, we evaluated an immersive XR system that combines VR with a physical mannequin. The XR system provided two types of feedback: real-time feedback on compressions and post-performance debriefing on the correctness of procedural steps and compression series. Our study highlighted that this dual-feedback approach, when combined with refresher training sessions, is effective for improving and maintaining BLSD skills. However, the results highlighted a critical need for continuous practice; even after retraining, the median percentage of correct compressions was 45.00%, with one-quarter of participants performing 11.70% or fewer correct compressions. This result underscores that while XR is a viable tool for professional retraining, systems that facilitate frequent, accessible repetition are fundamental to maintaining CPR proficiency.

To investigate the influence of multi-user settings on pre-hospital STEMI management training, this thesis presented a multi-user XR system involving an instructor, a medical doctor simulating a Case Manager, and a nurse executing the rescue within the VE. The system was used in a blended learning Continuing Medical Education course for nurses and doctors. Qualitative analysis revealed that learners highly valued the system, particularly its collaborative nature and its provision of a safe and engaging en-

vironment for high-stakes rehearsal. However, quantitative analysis showed that, despite a positive trend, nurses trained in the blended course achieved procedural performance comparable to that of a control group, with no statistically significant differences in performance scores during a STEMI simulation rescue.

Overall, the positive subjective feedback for both BLSD and STEMI systems suggests that the value of iVR extends beyond immediate knowledge retention scores. XR systems can serve as effective motivational tools, possibly encouraging more frequent voluntary training.

Based on the findings of this thesis, future research could:

- **Enhanced Feedback Mechanisms:** Investigate nesting post-performance feedback within a simulation replay. This approach would allow trainees to review their performance from an external perspective (third-person view), reducing the likelihood of skipping feedback. Such replays should emphasize positive reinforcement, highlighting correct behaviors while offering constructive advice to rectify ineffective actions.
- **Collaborative Multi-Learner Scenarios:** Support multiple active learners simultaneously, i.e., two or more nurses wearing the VR headset and performing interventions on the same virtual patient. In real-world dynamics, nurses typically rescue patients as part of a team, coordinating with team members to perform several tasks in parallel (e.g., one nurse speaks with the Case Manager doctor while another applies the IV to the patient). A system that supports synchronous, multi-role intervention would enable the simulation of these real-world dynamics and facilitate the training of the soft skills required for team coordination.
- **Automation and Adaptive Difficulty:** Integrate algorithms capable of identifying protocol deviations in real-time (e.g., if the learner did not follow a standardized procedure). This data could be used to provide immediate feedback on the correctness of actions and to drive an adaptive difficulty system that dynamically calibrates the simulation by tailoring the experience to the learner's specific expertise level, without requiring an instructor.

6.2 XR Training for Object Location Spatial Memory

In the context of training OLSM of healthy people, this thesis focused on two design elements: the level of immersion and the use of force-feedback haptic gloves to enhance the realism of VR experiences.

The level of immersion was investigated in a SG aimed at training OLSM, developed for both VR headsets and mobile devices. The user study with healthy participants showed that both versions were highly usable, and the game experience was positively perceived. While the iVR group reported a significantly higher sense of presence, the nVR group demonstrated faster in-game progress. Regardless of the SG version participants played with, results showed that playing the SG led to a statistically significant improvement in a subsequent real-world OLSM task. These results suggest that while immersion enhances the sense of presence, the cognitive benefits of the SG may also be

delivered effectively through non-immersive platforms, thereby avoiding the additional hardware and development costs required by VR headsets.

The second design element investigated was how different haptic feedback modalities influence user experience and performance in an iVR OLSM task. Specifically, a user study comparing force-feedback gloves with standard handheld controllers found that while force-feedback gloves significantly enhanced the “sense of being there,” they also led to higher fatigue and discomfort, likely due to the device’s weight and ergonomics. Notably, these ergonomic factors appeared to influence efficiency; participants using handheld controllers completed tasks in less time, with no statistically significant difference in OLSM performance between the two conditions. These findings suggest that for standard OLSM tasks, common commercial VR hardware may be sufficient, avoiding the fatigue and superior costs associated with force haptic devices.

Overall, these experiments highlight that while immersion and haptic realism significantly enhance the subjective user experience, they deliver similar acquisition of OLSM skills compared to more common devices.

Based on these findings, future research could:

- **Adaptive Difficulty:** Implement DDA techniques to calibrate the task difficulty in real-time. By dynamically adjusting parameters (e.g., the number of objects or the time available to remember their positions), based on the user’s performance, the system could maintain the user in a flow state by constantly providing a task difficulty that is neither too easy nor too hard.
- **Haptic Hardware:** Investigate the benefits of haptic feedback by replicating the OLSM task with different force-feedback gloves. Using different haptic gloves may reveal whether they can enhance memory performance in ways previously masked by physical discomfort.
- **Multi-User Setting:** Implement multi-user settings enhanced with haptic gloves, to explore whether performing OLSM tasks in these settings may influence engagement and training efficacy compared to individual exercises.

6.3 XR Training for Executive Functions

Focusing on training Executive Functions in both healthy participants and older adults with mNCD, this thesis investigated three design elements: DDA, multi-user settings, and immersion levels.

In the first system, an XR SG involving a sorting task is proposed to investigate the effects of adaptive difficulty and multi-user settings. Specifically, a user study compared an adaptive difficulty progression based on heuristics against linear difficulty increments, and single-player against multi-player modes. The results highlighted that while heuristic-based difficulty adaptation offers implementation simplicity by not requiring extensive background data, it lacks the necessary scalability to maintain a flow state across diverse demographics without frequent parameter recalibration. Furthermore, in the specific context of this sorting task, the multi-user setting introduced technological and logistical overhead without demonstrating clear added value compared to the single-user experience.

To further investigate how adaptive difficulty progression affects gameplay for users with varying degrees of cognitive impairment (healthy, mild NCD, and major NCD), a variant of the SG was developed based on the Go/No-Go task. This variant of the SG was modeled as a Discrete-Time Markov Chain, allowing for the evaluation of probabilistic queries that reveal distinct predicted behaviors across the three disease levels. The results suggest that the system can serve as a valuable instrument for clinicians to interpret patient performance and assess disease staging. Specifically, with probabilistic query data, it may be possible to identify discrepancies between expected and actual performance when playing with the SG (e.g., a healthy individual exhibiting behavior characteristic of an mNCD patient, or vice versa).

Building on these insights, we developed an enhanced version of the Go/No-Go SG aimed at training Inhibitory Control. This system features ten levels of difficulty, two distinct environments, and supports both iVR and nVR modalities. The primary objective of the evaluation protocol is to assess the acceptability of XR SG technologies as training tools for patients with mild NCD. Secondary objectives include a comparative analysis of user experience, in-game performance, and cognitive metrics between iVR and nVR conditions, utilizing a control group for cognitive benchmarking. This system is currently undergoing validation with mild NCD patients at the Cognition Behaviour Technology research institute of the Université Côte d’Azur.

Overall, these investigations suggest that XR offers promising tools for executive function training. Results from DDA techniques suggest that simple heuristics may be sufficient for general gameplay but lack the precision required for clinical patients, where, if needed, continuous updating of difficulty parameters may not be feasible. Instead, probabilistic modeling offers a framework for predicting how changes in parameters can affect user performance without involving participants. Additionally, the inclusion of multi-user elements requires careful consideration of the trade-off between social engagement and system complexity, as its benefits in repetitive cognitive training scenarios are less immediate.

Based on such findings, future research could:

- **Haptic-Enhanced Training:** Assess how integrating haptic gloves into XR training for executive functions affects user experience and training efficacy. Haptic gloves can simulate physical resistance and provide more tactile cues than standard controllers, potentially increasing immersion and engagement. Future studies should test whether differentiated haptic feedback for incorrect actions (e.g., triggering a distinct vibration when grabbing a “No-Go” object) provides a stronger reinforcement signal than visual or auditory cues alone.
- **Physiological-Based Adaptation:** Integrate physiological sensors (e.g., EEG, heart rate variability, eye tracking) into the DDA loop to capture physiological data in real time. By correlating these markers with in-game performance, the system could detect frustration or disengagement before performance declines and enable more precise difficulty calibration based on physiological data.
- **Machine Learning-Driven DDA:** Use an alternative approach to relying on the DTMC-based model by investigating the application of machine learning algorithms to improve the DDA logic of the SG. Unlike heuristic or static probabilistic models, reinforcement learning techniques can learn how specific difficulty changes

affect users in real-time. By continuously analyzing metrics such as reaction times and in-game performance, a reinforcement learning agent can dynamically adjust difficulty parameters. This approach may maintain a flow state more effectively throughout the rehabilitation period than static rules.

6.4 Overall Conclusion

By evaluating different XR design elements (level of immersion, use of haptics, feedback modalities, dynamic difficulty adjustment, and multi-user setting), this thesis offers insights into the application of XR technologies in healthcare training. The collective findings do not support the assumption that higher technological fidelity automatically yields superior learning outcomes. For example, in the context of OLSM training, while high-immersion VR and force-feedback devices consistently enhanced user engagement and presence, they did not outperform less immersive devices in skill acquisition. In fact, lower-fidelity options reduced fatigue and the need for specialized hardware, increasing the availability of such systems.

Moreover, multi-user settings proved valuable for the validity of procedural knowledge training for healthcare professionals on a rescue team in the context of STEMI procedure. However, in the context of an individual cognitive task, the multi-user setting did not report clear added value compared to the single-user experience. Furthermore, the investigation into adaptive difficulty shows that simple heuristics are limited in effectively tailoring difficulty to player skill, suggesting to experiment with more robust, data-driven modeling.

List of Tables

3.1	Participants characteristics (initial questionnaire). Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹ Chi-square test; ² Wilcoxon rank sum test; ³ Fisher's exact test.	31
3.2	Comparison between G1 and G2 at Post training (Session 1; all participants) and among Post training, Retention, and Post refresh (completers only). Variables are described as mean $\pm$ standard deviation, median (first-third quartiles). ¹ Wilcoxon-Mann-Whitney test; ² Wilcoxon signed rank sum test; ³ Friedman's test; ⁴ Cochran's Q test; ⁵ McNemar test; ⁶ Chi square test.	34
3.2	Comparison between G1 and G2 at Post training (Session 1; all participants) and among Post training, Retention, and Post refresh (completers only). Variables are described as mean $\pm$ standard deviation, median (first-third quartiles). ¹ Wilcoxon-Mann-Whitney test; ² Wilcoxon signed rank sum test; ³ Friedman's test; ⁴ Cochran's Q test; ⁵ McNemar test; ⁶ Chi square test.	35
3.2	Comparison between G1 and G2 at Post training (Session 1; all participants) and among Post training, Retention, and Post refresh (completers only). Variables are described as mean $\pm$ standard deviation, median (first-third quartiles). ¹ Wilcoxon-Mann-Whitney test; ² Wilcoxon signed rank sum test; ³ Friedman's test; ⁴ Cochran's Q test; ⁵ McNemar test; ⁶ Chi square test.	36
3.3	Comparison between completers and non completers, by group, in the initial questionnaire, post-training measures, and USE questionnaire scores. Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹ Chi-square test; ² Wilcoxon rank sum test; ³ Fisher's exact test.	37
3.3	Comparison between completers and non completers, by group, in the initial questionnaire, post-training measures, and USE questionnaire scores. Dichotomous variables are described as absolute frequencies (percentage); continuous variables are described as mean $\pm$ SD, median (first quartile-third quartile). ¹ Chi-square test; ² Wilcoxon rank sum test; ³ Fisher's exact test.	38
3.4	Participants' gender, age range, professional role, years of experience in the role and in the emergency area, and previous use of VR.	54

3.5	Identified themes and sub-themes with their description and participants who talked about them.	56
3.5	Identified themes and sub-themes with their description and participants who talked about them.	57
3.5	Identified themes and sub-themes with their description and participants who talked about them.	58
3.6	Checklist of the steps the nurse has to perform in the simulation with score assigned that reflects the clinical importance. Steps with “p” suffix refers to the parameter component of the related step.	67
3.6	Checklist of the steps the nurse has to perform in the simulation with score assigned that reflects the clinical importance. Steps with “p” suffix refers to the parameter component of the related step.	68
3.7	Overview of Assessment Areas and total score obtainable by each nurse during the simulation.	68
3.8	Summary of Performance Metrics at the step and nurse level.	69
3.9	Demographic characteristics, professional background, prior training, and self-rated knowledge of STEMI procedures of participants. Data are reported as absolute values with percentages relative to each group, and self-rated knowledge is presented as Mean / SD.	69
3.9	Demographic characteristics, professional background, prior training, and self-rated knowledge of STEMI procedures of participants. Data are reported as absolute values with percentages relative to each group, and self-rated knowledge is presented as Mean / SD.	70
3.10	Number of nurses who completed the step of the procedure. ¹ Fisher’s exact test; ² Chi-square test.	71
3.10	Number of nurses who completed the step of the procedure. ¹ Fisher’s exact test; ² Chi-square test.	72
3.11	Number of nurses who met the required parameter in the parameter specific step. ¹ Fisher’s exact test.	72
3.12	Mean, Standard deviation, and Wilcoxon-Mann-Whitney test results for the metrics at nurse level.	74
4.1	Mean, standard deviation, t-test results, and effect size for presence (IPQ scales), usability (SUS), player experience (PXFUN and PXPSY), and progress in the SG (MLW).	89
4.2	Mean, Standard Deviation, and paired t-Test results for IPQ scales and usability.	100
4.3	Mean, Standard Deviation, and Wilcoxon Signed-Ranks Test results for the items of the DAQ questionnaire.	100
5.1	Difficulty parameter scaling from Level 1 to Level 10.	109
5.2	DDA Level Adjustment Logic	110
5.3	Experimental Conditions and Group Assignments	113
5.4	Participant demographics including age, HMD experience, weekly sport activity, and weekly video game usage (Mean $\pm$ SD).	116
5.5	Level Increment calculation logic.	124
5.6	Parameters used in the SG and in the model with their values.	125

5.7	Local variables of the three modules and their purpose.	126
5.8	Percentage probabilities calculated as the mean of the medical staff's responses regarding the likelihood of patients performing specific actions in the SG.	128
5.9	Total states and transitions of the different models built.	130
5.10	Results of the considered queries and relative times to compute expressed in seconds.	131
5.11	Difficulty parameter scaling from Level 1 to Level 10.	134
5.12	Participants characteristics (initial questionnaire) and Disease Level. Dichotomous variables are described as absolute frequencies; continuous variables are described as Mean $\pm$ SD.	144
5.13	Descriptive Statistics for UTAUT Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	145
5.14	Descriptive Statistics for the Inhibitory Control test of the TAP battery at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.	146
5.15	Descriptive Statistics for Stroop test at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.	146
5.15	Descriptive Statistics for Stroop test at different time instants. Categorical variables are described as absolute frequencies. Continuous variables are reported as Mean $\pm$ Standard Deviation.	147
5.16	Descriptive Statistics for SART test at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	147
5.17	Descriptive Statistics for FAB Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	148
5.17	Descriptive Statistics for FAB Constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	149
5.18	Descriptive Statistics of affective reaction for Pleasure, Arousal, and Dominance at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	149
5.18	Descriptive Statistics of affective reaction for Pleasure, Arousal, and Dominance at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	150
5.19	Descriptive Statistics for the sense of presence at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	150
5.19	Descriptive Statistics for the sense of presence at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	151
5.20	Descriptive Statistics for GEQ constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	151
5.20	Descriptive Statistics for GEQ constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	152
5.21	Statistics for GEQ Post Game Module constructs at different time instants. Values are reported as Mean $\pm$ Standard Deviation.	152

5.22	Descriptive statistics for the items of the DAQ questionnaire at different time instants. Values are reported as Mean $\pm$ Standard Deviation. . . .	153
5.22	Descriptive statistics for the items of the DAQ questionnaire at different time instants. Values are reported as Mean $\pm$ Standard Deviation. . . .	154

List of Figures

3.1	The software simulating the virtual scenario and a user wearing the devices to use it: A) the software as seen by the user, B) the VR headset, C) the wrist trackers, D) the manikin with trackers.	23
3.2	Interfaces displayed by the XR system: A) semi-transparent hands suggesting how to perform the step, B) directional arrows suggesting where to perform the step, C) panel with the compression depth and frequency, D) pop-up window with written text and illustrations describing each step, E) debriefing panel. All interfaces are localized in the participants' native language.	26
3.3	Comparisons of the measures between G1 and G2 at the end of S1. Capped vertical bars indicate the standard error for non-dichotomous measures. The horizontal grey bar in the "Frequency (cpr/min)" and "Depth (cm)" graphs indicate the correct ranges. The * and *** signs indicate differences with p-values < 0.05 and < 0.005 respectively. . . .	39
3.4	Comparisons of the measures for all completers between post_training, retention and post_refresh. Capped vertical bars indicate the standard error for non-dichotomous measures. The horizontal grey bar in the "Frequency (cpr/min)" and "Depth (cm)" graphs indicate the correct ranges. The *, ***, and **** signs indicate differences with p-values < 0.05, < 0.005, and <= 0.001 respectively.	41
3.5	The user interface of the Instructor Application, localized in the language of course instructors.	47
3.6	The user interface of the Instructor Application, localized in the language of course instructors.	48
3.7	The user interface of the Instructor Application, localized in the language of course instructors.	49
3.8	The user interface of the Instructor Application, localized in the language of course instructors.	50
3.9	The user interface of the Instructor Application, localized in the language of course instructors.	51
3.10	Flow diagram of questions in the interview. The interview was conducted in the language of participants. Sentences have been translated here for readers' convenience.	52
3.11	Number of nurses who completed the specific step in groups G1 (blue) and G2 (orange). The y-axis indicates the total number of nurses per group.	73

3.12	Number of nurses who completed the specific step meeting the required parameter in groups G1 (blue) and G2 (orange). The y-axis indicates the total number of nurses per group.	73
3.13	Mean scores for nurse level metrics in groups G1 (blue) and G2 (orange). Error bars represent $\pm$ SE. The y-axis indicates the full obtainable score range for the scale (minimum to maximum).	74
4.1	Setting of the VE (left) and of the real environment (right) where the user study was conducted.	80
4.2	User playing with the iVR version of the SG during the memorization phase (on the left) and selecting an object from the shelf inventory during the restoration phase (on the right).	81
4.3	User interfaces of the Serious Game showing the quick inventory and the shelf in the iVR (left) and nVR (right) versions.	81
4.4	Visual feedback on the locations given to the player when the object placed is correct (on the left) or wrong (on the right).	82
4.5	Player's target time in the SG to obtain the best trophy for the different levels (bottom blue line) and maximum time to complete the level (top orange line). The X axis indicates the levels, the Y axis indicates the time in seconds.	83
4.6	A user playing the SG developed in the iVR version (on the left) and nVR version (on the right).	84
4.7	Object configurations for the first (A, B) and second (C, D) real-world OLSM tasks. Configurations A and C show the initial setups to be memorized by the participant, while B and D show the room arrangements after the experimenter's modifications.	87
4.8	Number of errors (left part) and time elapsed (right part) in the pre and post (on x axis) OLSM tasks in the real world and version of the SG (separate columns). Capped vertical bars show standard error. The * and **** signs indicate differences with p-values respectively < 0.05 and < 0.001	90
4.9	Virtual Environment used for the task composed by an empty room, the two tables, and the instruction in the users' native language.	96
4.10	The two interaction methods used to handle virtual objects. On the left side, the user has the HC (HTC Vive Pro Controller) in his hand, and on the right side, he is using the FG (Senseglove DK1).	96
4.11	PO (a) and TE (b) for each level (on x axis) and condition (separate lines). Capped vertical bars show 95% CI. The *, **, *** signs indicate differences with p-values respectively < 0.05 , < 0.005 , < 0.001	99
5.1	Types of balls to throw: red, blue, stained red (junk), and stained blue (junk). Behind them are the containers in which the balls have to be thrown (red on the left, blue on the right).	107
5.2	Virtual Environment of the sorting task. Objects move down the sliders toward the central table, where the player sorts them into the red or blue container or onto the floor.	107

5.3	Gameplay user interface. On the left: current combo multiplier, total score, current combo, and error icons. On the right: real-time leaderboard highlighting the player's current rank.	108
5.4	Comparison of junk object similarity. The figure shows the minimum and maximum stain densities for both red and blue variants.	109
5.5	Bar plot of each variable that defines a level's difficulty.	111
5.6	Multi-player setting in the VR. Players are placed in front of each other but far enough away to limit distractions.	112
5.7	Means of maximum score gained and maximum duration (seconds) in the four groups. Capped vertical bars indicate $\pm$ SE.	117
5.8	Means of affective reaction for Pleasure, Arousal, and Dominance in the four groups. Capped vertical bars indicate $\pm$ SE.	117
5.9	Means of presence total score, general item about the sense of "being there," spatial subscale, involvement subscale, and experience realism subscale in the four groups. Capped vertical bars indicate $\pm$ SE.	118
5.10	Means of Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative Affect, and Positive Affect in the four groups. Capped vertical bars indicate $\pm$ SE.	118
5.11	Means of Positive Experience, Negative Experience, Tiredness, and Return to Reality in the four groups. Capped vertical bars indicate $\pm$ SE.	118
5.12	Means of Psychological Involvement in Empathy, Negative Feeling, and Behavioural Involvement in the four groups. Capped vertical bars indicate $\pm$ SE.	119
5.13	Example of SG gameplay. The rule is to throw red balls into the container and black-dotted ones away. The user interface on the left displays the score, combo streak, and NoGo-Balls mistakenly thrown in the Go-Basket (marked with red crosses) and the user interface on the right shows player ranking.	124
5.14	Representation of the flow of the model. The three trapezoid shapes at the top of the figure contain the names of the modules. The white boxes with dotted borders mean a transition that involve all the modules, and the ones with continued borders represent a synchronized transition between two modules. The colored boxes represent a transition that involve only that particular model. The hexagons encode a guard, showing how the flow changes between the possible outputs enabling a particular transition. The light-grey background groups the parts included in the game loop or the session loop.	129
5.15	SG environments for the proposed IC training task. The top row (A, B) displays the nVR versions, and the bottom row (C, D) displays the iVR versions. Columns represent the Gym environment (A, C) and the Warehouse environment (B, D).	133
5.16	Bar plot of each variable that defines a level's difficulty.	135
5.17	Flow diagrams of questions in the semi-structured interviews. The interviews were conducted in the language of the participants; sentences have been translated here for readers' convenience.	139
5.18	Protocol of the user study.	143

Bibliography

- [1] Jonathan R. Abbas, Michael M. H. Chu, Ceyon Jeyarajah, et al. Virtual reality in simulation-based emergency skills training: A systematic review with a narrative synthesis. *Resuscitation Plus*, 16:100484, December 2023. ISSN 2666-5204. doi: 10.1016/j.resplu.2023.100484. URL <https://www.sciencedirect.com/science/article/pii/S2666520423001273>.
- [2] Jonathan R Abbas, Alexander O'Connor, Eshwar Ganapathy, et al. What is Virtual Reality? A healthcare-focused systematic review of definitions. *Health Policy and Technology*, 12(2):100741, June 2023. ISSN 2211-8837. doi: 10.1016/j.hlpt.2023.100741. URL <https://www.sciencedirect.com/science/article/pii/S2211883723000199>.
- [3] Alaa Abd-alrazaq, Mohannad Alajlani, Dari Alhuwail, et al. The Effectiveness and Safety of Serious Games for Improving Cognitive Abilities Among Elderly People With Cognitive Impairment: Systematic Review and Meta-Analysis. *JMIR Serious Games*, 10(1):e34592, March 2022. doi: 10.2196/34592. URL <https://games.jmir.org/2022/1/e34592>.
- [4] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, et al. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies*, 135:102370, March 2020. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2019.102370. URL <https://www.sciencedirect.com/science/article/pii/S1071581919301302>.
- [5] Accreditation Council for Continuing Medical Education. CME Content: Definition and Examples | ACCME, June 2023. URL <https://www.accme.org/accreditation-rules/policies/cme-content-definition-and-examples>.
- [6] Adilzhan Adilkanov, Matteo Rubagotti, and Zhanat Kappassov. Haptic Devices: Wearability-Based Taxonomy and Literature Review. *IEEE Access*, 10:91923–91947, 2022. ISSN 2169-3536. doi: 10.1109/ACCESS.2022.3202986.
- [7] Kamran Ahmed, Tim T. Wang, Hutan Ashrafian, et al. The effectiveness of continuing medical education for specialist recertification. *Can Urol Assoc J*, 7 (7-8):266–272, 2013. ISSN 1911-6470. doi: 10.5489/cuaj.378.
- [8] Sun Joo (Grace) Ahn, Kristine L. Nowak, and Jeremy N. Bailenson. Unintended consequences of spatial presence on learning in virtual reality. *Computers & Education*, 186:104532, September 2022. ISSN 0360-1315. doi: 10.1016/

- j.compedu.2022.104532. URL <https://www.sciencedirect.com/science/article/pii/S0360131522001038>.
- [9] Gary L. Allen, editor. *Human Spatial Memory: Remembering Where*. Erlbaum, Mahwah, NJ, 2004. ISBN 978-0-8058-4218-0.
 - [10] Haya Almaree, Roland Fischer, René Weller, et al. Enhancing anatomy learning through collaborative VR? An advanced investigation. *Computers & Graphics*, 123:104019, October 2024. ISSN 0097-8493. doi: 10.1016/j.cag.2024.104019. URL <https://www.sciencedirect.com/science/article/pii/S0097849324001547>.
 - [11] Rajeev Alur and Thomas A. Henzinger. Reactive Modules. *Formal Methods in System Design*, 15(1):7–48, July 1999. ISSN 1572-8102. doi: 10.1023/A:1008739929481. URL <https://doi.org/10.1023/A:1008739929481>.
 - [12] Robert Anderson, Alexandre Sebaldt, Yiqun Lin, et al. Optimal training frequency for acquisition and retention of high-quality CPR skills: A randomized trial. *Resuscitation*, 135:153–161, February 2019. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2018.10.033. URL <https://www.sciencedirect.com/science/article/pii/S0300957218310682>.
 - [13] Lesley Andrew, Donna Barwood, Julie Boston, et al. Serious games for health promotion in adolescents – a systematic scoping review. *Educ Inf Technol*, 28(5): 5519–5550, May 2023. ISSN 1573-7608. doi: 10.1007/s10639-022-11414-9. URL <https://doi.org/10.1007/s10639-022-11414-9>.
 - [14] J. A. Anguera, J. Boccanfuso, J. L. Rintoul, et al. Video game training enhances cognitive control in older adults. *Nature*, 501(7465):97–101, September 2013. ISSN 1476-4687. doi: 10.1038/nature12486. URL <https://www.nature.com/articles/nature12486>.
 - [15] Sara Arlati, Simona Gabriella Di Santo, Flaminia Franchini, et al. Acceptance and Usability of Immersive Virtual Reality in Older Adults with Objective and Subjective Cognitive Decline. *JAD*, 80(3):1025–1038, April 2021. ISSN 13872877, 18758908. doi: 10.3233/JAD-201431. URL <https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-201431>.
 - [16] Oscar Arrogante, José Ríos-Díaz, Laura Carrión-García, et al. Deliberate practice in resuscitation training using a feedback device, and the effects of the physical characteristics of the rescuer on the acquisition and retention of cardiopulmonary resuscitation skills: Randomized clinical trial. *International Emergency Nursing*, 58:101037, September 2021. ISSN 1755-599X. doi: 10.1016/j.ienj.2021.101037. URL <https://www.sciencedirect.com/science/article/pii/S1755599X21000756>.
 - [17] Winfred Arthur Jr., Winston Bennett Jr., Pamela L. Stanush, et al. Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Hum. Perform.*, 11(1):57–101, March 1998. ISSN 0895-9285. doi: 10.1207/s15327043hup1101_3. URL https://doi.org/10.1207/s15327043hup1101_3.

- [18] Nasim Aslani, Ali Behmanesh, Ali Garavand, et al. The Virtual Reality Technology Effects and Features in Cardiology Interventions Training: A Scoping Review. *Med J Islam Repub Iran*, 36:77, July 2022. ISSN 1016-1430. doi: 10.47176/mjiri.36.77. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9448494/>.
- [19] Maryam Asoodar, Fatemeh Janesarvatan, Hao Yu, et al. Theoretical foundations and implications of augmented reality, virtual reality, and mixed reality for immersive learning in health professions education. *Adv Simul*, 9(1): 36, September 2024. ISSN 2059-0628. doi: 10.1186/s41077-024-00311-5. URL <https://doi.org/10.1186/s41077-024-00311-5>.
- [20] Karam Atli, Warren Selman, and Abhishek Ray. A Comprehensive Multicomponent Neurosurgical Course with use of Virtual Reality: Modernizing the Medical Classroom. *Journal of Surgical Education*, 78(4):1350–1356, July 2021. ISSN 1931-7204. doi: 10.1016/j.jsurg.2020.11.003. URL <https://www.sciencedirect.com/science/article/pii/S1931720420304281>.
- [21] João B Augusto, Miguel B Santos, Daniel Faria, et al. Real-Time Visual Feedback Device Improves Quality Of Chest Compressions: A Manikin Study. *Bull Emerg Trauma*, 8(3):135–141, July 2020. ISSN 2322-2522. doi: 10.30476/BEAT.2020.83080. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7468227/>.
- [22] Sayed Azher, Aralia Mills, Jinzhi He, et al. Findings Favor Haptics Feedback in Virtual Simulation Surgical Education: An Updated Systematic and Scoping Review. *Surg Innov*, 31(3):331–341, June 2024. ISSN 1553-3506. doi: 10.1177/15533506241238263. URL <https://doi.org/10.1177/15533506241238263>.
- [23] Alan Baddeley. Working Memory. *Science*, 255(5044):556–559, 1992. ISSN 0036-8075. URL <https://www.jstor.org/stable/2876819>.
- [24] Alan Baddeley. Working memory. *Current Biology*, 20(4):R136–R140, February 2010. ISSN 0960-9822. doi: 10.1016/j.cub.2009.12.014. URL <https://www.sciencedirect.com/science/article/pii/S0960982209021332>.
- [25] Nilufar Baghaei, Hai-Ning Liang, John Naslund, et al. Games for Mental Health. *Games Health J.*, 11(6):337–340, December 2022. ISSN 2161-783X. doi: 10.1089/g4h.2022.0199. URL <https://www.liebertpub.com/doi/full/10.1089/g4h.2022.0199>.
- [26] Albert Bandura. Social Cognitive Theory: An Agentic Perspective. *Annu. Rev. Psychol.*, 52(1):1–26, 2001. doi: 10.1146/annurev.psych.52.1.1. URL <https://doi.org/10.1146/annurev.psych.52.1.1>.
- [27] Aaron Bangor, Kortum Philip T., and James T. and Miller. An Empirical Evaluation of the System Usability Scale. *Int. J. Human–Computer Interact.*, 24(6):574–594, July 2008. ISSN 1044-7318. doi: 10.1080/10447310802205776. URL <https://doi.org/10.1080/10447310802205776>.

- [28] Sandra Barteit, Lucia Lanfermann, Till Bärnighausen, et al. Augmented, Mixed, and Virtual Reality-Based Head-Mounted Devices for Medical Education: Systematic Review. *JMIR Serious Games*, 9(3):e29080, July 2021. doi: 10.2196/29080. URL <https://games.jmir.org/2021/3/e29080>.
- [29] Julia Belger, Marie Blume, Mert Akbal, et al. The immersive virtual memory task: Assessing object-location memory in neurological patients using immersive virtual reality. *Neuropsychol. Rehabil.*, 34(6):870–898, July 2024. ISSN 0960-2011. doi: 10.1080/09602011.2023.2256957. URL <https://doi.org/10.1080/09602011.2023.2256957>.
- [30] Melanie L. Bell, Michael G. Kenward, Diane L. Fairclough, et al. Differential dropout and bias in randomised controlled trials: When it matters and when it may not. *BMJ*, 346:e8668, January 2013. ISSN 1756-1833. doi: 10.1136/bmj.e8668. URL <https://www.bmj.com/content/346/bmj.e8668>.
- [31] Farhan Bhanji, Aaron J. Donoghue, Margaret S. Wolff, et al. Part 14: Education: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, 132(18 Suppl 2):S561–573, November 2015. ISSN 1524-4539. doi: 10.1161/CIR.0000000000000268.
- [32] Frank Biocca, Chad Harms, and Judee K. Burgoon. Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators and Virtual Environments*, 12(5):456–480, October 2003. doi: 10.1162/105474603322761270. URL <https://doi.org/10.1162/105474603322761270>.
- [33] Antonia Blanié, Sophie Gorse, Philippe Roulleau, et al. Impact of learners’ role (active participant-observer or observer only) on learning outcomes during high-fidelity simulation sessions in anaesthesia: A single center, prospective and randomised study. *Anaesthesia Critical Care & Pain Medicine*, 37(5):417–422, October 2018. ISSN 2352-5568. doi: 10.1016/j.accpm.2017.11.016. URL <https://www.sciencedirect.com/science/article/pii/S2352556817301753>.
- [34] Gonul Bodur, Zehra Turhan, Yunus Emre Altun, et al. Evaluating the effectiveness of virtual reality simulation in cpr training for nursing students: A randomized controlled trial. *Nurse Education in Practice*, 87:104486, August 2025. ISSN 1471-5953. doi: 10.1016/j.nepr.2025.104486. URL <https://www.sciencedirect.com/science/article/pii/S1471595325002422>.
- [35] Marie-Stéphanie Bracq, Estelle Michinov, and Pierre Jannin. Virtual Reality Simulation in Nontechnical Skills Training for Healthcare Professionals: A Systematic Review. *Simul. Healthc.*, 14(3):188, June 2019. ISSN 1559-2332. doi: 10.1097/SIH.0000000000000347. URL https://journals.lww.com/simulationinhealthcare/fulltext/2019/06000/Creating_Effective_Simulation_Scenarios.9.aspx.
- [36] Margaret M. Bradley and Peter J. Lang. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, 25(1):49–59, March 1994. ISSN 0005-7916. doi: 10.1016/0005-7916(94)90063-9. URL <https://www.sciencedirect.com/science/article/pii/S0005791694900639>.

- [37] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qual. Res. Psychol.*, 3(2):77–101, January 2006. ISSN 1478-0887. doi: 10.1191/1478088706qp063oa. URL <https://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa>.
- [38] Victor Bréhault, Emmanuel Dubois, Arnaud Prouzeau, et al. A Systematic Literature Review to Characterize Asymmetric Interaction in Collaborative Systems. In *Proc. 2025 CHI Conf. Hum. Factors Comput. Syst.*, CHI '25, pages 1–19, New York, NY, USA, April 2025. Association for Computing Machinery. ISBN 979-8-4007-1394-1. doi: 10.1145/3706598.3713129. URL <https://dl.acm.org/doi/10.1145/3706598.3713129>.
- [39] John Brooke. SUS: A 'Quick and Dirty' Usability Scale. In *Usability Evaluation In Industry*. CRC Press, 1996. ISBN 978-0-429-15701-1.
- [40] Fabio Buttussi and Luca Chittaro. Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Trans. Vis. Comput. Graph.*, 24(2):1063–1076, February 2018. ISSN 1941-0506. doi: 10.1109/TVCG.2017.2653117.
- [41] Fabio Buttussi and Luca Chittaro. A Comparison of Procedural Safety Training in Three Conditions: Virtual Reality Headset, Smartphone, and Printed Materials. *IEEE Trans. Learn. Technol.*, 14(1):1–15, February 2021. ISSN 1939-1382. doi: 10.1109/TLT.2020.3033766.
- [42] Fabio Buttussi and Luca Chittaro. Acquisition and retention of spatial knowledge through virtual reality experiences: Effects of VR setup and locomotion technique. *International Journal of Human-Computer Studies*, 177:103067, September 2023. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2023.103067. URL <https://www.sciencedirect.com/science/article/pii/S1071581923000769>.
- [43] Fabio Buttussi, Luca Chittaro, and Francesca Valent. A virtual reality methodology for cardiopulmonary resuscitation training with and without a physical mannequin. *Journal of Biomedical Informatics*, 111:103590, November 2020. ISSN 1532-0464. doi: 10.1016/j.jbi.2020.103590. URL <https://www.sciencedirect.com/science/article/pii/S1532046420302197>.
- [44] Robert A Byrne, Xavier Rossello, J J Coughlan, et al. 2023 ESC Guidelines for the management of acute coronary syndromes: Developed by the task force on the management of acute coronary syndromes of the European Society of Cardiology (ESC). *Eur Heart J*, 44(38):3720–3826, October 2023. ISSN 0195-668X. doi: 10.1093/eurheartj/ehad191. URL <https://doi.org/10.1093/eurheartj/ehad191>.
- [45] L  na  c B. Cadet and Hanna Chainay. Memory of virtual experiences: Role of immersion, emotion and sense of presence. *International Journal of Human-Computer Studies*, 144:102506, December 2020. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2020.102506. URL <https://www.sciencedirect.com/science/article/pii/S1071581920301087>.

- [46] Robert M. Califf, Eric D. Peterson, Raymond J. Gibbons, et al. Integrating quality into the cycle of therapeutic development. *J Am Coll Cardiol*, 40(11):1895–1901, December 2002. ISSN 0735-1097. doi: 10.1016/s0735-1097(02)02537-8.
- [47] Nicko R. Caluya, Alexander Plopski, Jayzon F. Ty, et al. Transferability of Spatial Maps: Augmented Versus Virtual Reality Training. In *2018 IEEE Conf. Virtual Real. 3D User Interfaces VR*, pages 387–393, March 2018. doi: 10.1109/VR.2018.8447561. URL <https://ieeexplore.ieee.org/abstract/document/8447561>.
- [48] Pablo Campo-Prieto, José M^a Cancela-Carral, and Gustavo Rodríguez-Fuentes. Wearable Immersive Virtual Reality Device for Promoting Physical Activity in Parkinson’s Disease Patients. *Sensors*, 22(9):3302, January 2022. ISSN 1424-8220. doi: 10.3390/s22093302. URL <https://www.mdpi.com/1424-8220/22/9/3302>.
- [49] Giulio Canapa, Fabio Paternò, and Carmen Santoro. Interactive Serious Games for Cognitive Training of Older Adults: A Systematic Review. *IEEE Trans. Comput. Soc. Syst.*, pages 1–24, 2025. ISSN 2329-924X. doi: 10.1109/TCSS.2025.3532307. URL <https://ieeexplore.ieee.org/abstract/document/10859161>.
- [50] Lorainne Tudor Car, Bhone Myint Kyaw, Andrew Teo, et al. Outcomes, Measurement Instruments, and Their Validity Evidence in Randomized Controlled Trials on Virtual, Augmented, and Mixed Reality in Undergraduate Medical Education: Systematic Mapping Review. *JMIR Serious Games*, 10(2):e29594, April 2022. doi: 10.2196/29594. URL <https://games.jmir.org/2022/2/e29594>.
- [51] Jordi Castillo, Encarna Rodríguez-Higueras, Ricardo Belmonte, et al. Efficacy of Virtual Reality Simulation in Teaching Basic Life Support and Its Retention at 6 Months. *Int J Environ Res Public Health*, 20(5):4095, February 2023. ISSN 1661-7827. doi: 10.3390/ijerph20054095. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10001443/>.
- [52] Shih-Lin Chang, Ming-Jen Kuo, Yenn-Jiang Lin, et al. Virtual reality informative aids increase residents’ atrial fibrillation ablation procedures-related knowledge and patients’ satisfaction. *J. Chin. Med. Assoc.*, 84(1):25, January 2021. ISSN 1726-4901. doi: 10.1097/JCMA.0000000000000464. URL https://journals.lww.com/jcma/Fulltext/2021/01000/Virtual_reality_informative_aids_increase.6.aspx.
- [53] Wen Chang, Chun-Chih Lin, Julia Crilly, et al. Virtual Reality Simulation for Undergraduate Nursing Students for Care of Patients With Infectious Diseases: Mixed Methods Study. *JMIR Med. Educ.*, 11(1):e64780, February 2025. doi: 10.2196/64780. URL <https://mededu.jmir.org/2025/1/e64780>.
- [54] Ying-Cheng Chao, Sophia H. Hu, Hsiao-Yean Chiu, et al. The effects of an immersive 3d interactive video program on improving student nurses’ nursing skill competence: A randomized controlled trial study. *Nurse Education Today*, 103: 104979, August 2021. ISSN 0260-6917. doi: 10.1016/j.nedt.2021.104979. URL <https://www.sciencedirect.com/science/article/pii/S0260691721002367>.

- [55] Ying-Yu Chao, Yvonne K. Scherer, and Carolyn A. Montgomery. Effects of Using Nintendo Wii™ Exergames in Older Adults: A Review of the Literature. *J Aging Health*, 27(3):379–402, April 2015. ISSN 0898-2643. doi: 10.1177/0898264314551171. URL <https://doi.org/10.1177/0898264314551171>.
- [56] David Checa and Andres Bustillo. A review of immersive virtual reality serious games to enhance learning and training. *Multimed Tools Appl*, 79(9):5501–5527, March 2020. ISSN 1573-7721. doi: 10.1007/s11042-019-08348-9. URL <https://doi.org/10.1007/s11042-019-08348-9>.
- [57] Feng-Qin Chen, Yu-Fei Leng, Jian-Feng Ge, et al. Effectiveness of Virtual Reality in Nursing Education: Meta-Analysis. *J. Med. Internet Res.*, 22(9):e18290, September 2020. doi: 10.2196/18290. URL <https://www.jmir.org/2020/9/e18290>.
- [58] Jenova Chen. Flow in games (and everything else). *Commun. ACM*, 50(4):31–34, April 2007. ISSN 0001-0782. doi: 10.1145/1232743.1232769. URL <https://dl.acm.org/doi/10.1145/1232743.1232769>.
- [59] Adam Cheng, Vinay M. Nadkarni, Mary Beth Mancini, et al. Resuscitation Education Science: Educational Strategies to Improve Outcomes From Cardiac Arrest: A Scientific Statement From the American Heart Association. *Circulation*, 138(6):e82–e122, August 2018. doi: 10.1161/CIR.0000000000000583. URL <https://www.ahajournals.org/doi/10.1161/CIR.0000000000000583>.
- [60] Adam Cheng, David J. Magid, Marc Auerbach, et al. Part 6: Resuscitation Education Science: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. *Circulation*, 142(16_suppl_2):S551–S579, October 2020. doi: 10.1161/CIR.0000000000000903. URL <https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000903>.
- [61] Adam Cheng, Nino Fijacko, Andrew Lockey, et al. Use of augmented and virtual reality in resuscitation training: A systematic review. *Resuscitation Plus*, 18:100643, June 2024. ISSN 2666-5204. doi: 10.1016/j.resplu.2024.100643. URL <https://www.sciencedirect.com/science/article/pii/S2666520424000948>.
- [62] Pengfei Cheng, Yangxi Huang, Pengyu Yang, et al. The Effects of Serious Games on Cardiopulmonary Resuscitation Training and Education: Systematic Review With Meta-Analysis of Randomized Controlled Trials. *JMIR Serious Games*, 12(1):e52990, February 2024. doi: 10.2196/52990. URL <https://games.jmir.org/2024/1/e52990>.
- [63] Dung-Hung Chiang, Chia-Chang Huang, Shu-Chuan Cheng, et al. Immersive virtual reality (VR) training increases the self-efficacy of in-hospital healthcare providers and patient families regarding tracheostomy-related knowledge and care skills: A prospective pre-post study. *Medicine (Baltimore)*, 101(2):e28570, January 2022. doi: 10.1097/MD.00000000000028570. URL https://journals.lww.com/md-journal/fulltext/2022/01140/immersive_virtual_reality__vr__training_increases.59.aspx.

- [64] Feng-Kuang Chiang, Xiaojing Shang, and Lu Qiao. Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, 129:107125, April 2022. ISSN 0747-5632. doi: 10.1016/j.chb.2021.107125. URL <https://www.sciencedirect.com/science/article/pii/S0747563221004489>.
- [65] Luca Chittaro and Fabio Buttussi. Learning Safety Through Public Serious Games: A Study of “Prepare for Impact” on a Very Large, International Sample of Players. *IEEE Trans. Vis. Comput. Graph.*, 28(3):1573–1584, March 2022. ISSN 1941-0506. doi: 10.1109/TVCG.2020.3022340.
- [66] Luca Chittaro and Roberto Ranon. Web3D technologies in learning, education and training: Motivations, issues, opportunities. *Computers & Education*, 49(1): 3–18, August 2007. ISSN 0360-1315. doi: 10.1016/j.compedu.2005.06.002. URL <https://www.sciencedirect.com/science/article/pii/S0360131505000813>.
- [67] Jeeyae Choi, C. Elise Thompson, Jeungok Choi, et al. Effectiveness of Immersive Virtual Reality in Nursing Education: Systematic Review. *Nurse Educ.*, 47(3):E57, May/June 2022. ISSN 0363-3624. doi: 10.1097/NNE.0000000000001117. URL https://journals.lww.com/nurseeducatoronline/fulltext/2022/05000/Effectiveness_of_Immersive_Virtual_Reality_in.23.aspx.
- [68] Douglas B. Clark, Emily E. Tanner-Smith, and Stephen S. Killingsworth. Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Rev. Educ. Res.*, 86(1):79–122, March 2016. ISSN 0034-6543. doi: 10.3102/0034654315582065. URL <https://doi.org/10.3102/0034654315582065>.
- [69] E. M. Clarke, E. A. Emerson, and A. P. Sistla. Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Trans. Program. Lang. Syst.*, 8(2):244–263, April 1986. ISSN 0164-0925. doi: 10.1145/5397.5399. URL <https://dl.acm.org/doi/10.1145/5397.5399>.
- [70] Edmund M. Clarke, Orna Grumberg, and Doron A. Peled. *Model Checking*. MIT Press, Cambridge, Mass, January 1999. ISBN 978-0-262-03270-4.
- [71] Jean-Philippe Collet, Holger Thiele, Emanuele Barbato, et al. 2020 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *European Heart Journal*, 42(14):1289–1367, April 2021. ISSN 0195-668X. doi: 10.1093/eurheartj/ehaa575. URL <https://doi.org/10.1093/eurheartj/ehaa575>.
- [72] J. P. Collins and R. M. Harden. AMEE Medical Education Guide No. 13: Real patients, simulated patients and simulators in clinical examinations. *Medical Teacher*, 20(6):508–521, January 1998. ISSN 0142-159X. doi: 10.1080/01421599880210. URL <https://doi.org/10.1080/01421599880210>.
- [73] Philip M. Corsi and Philip Michael. Human memory and the medial temporal region of the brain. 1972.

- [74] Italian Resuscitation Council. Linee Guida ERC per la Rianimazione Cardiopolmonare durante pandemia Covid-19 – IRC, September 2023. URL <https://www.ircouncil.it/linee-guida-erc-la-rianimazione-cardiopolmonare-pandemia-covid-19/>.
- [75] Irene Cristofori, Shira Cohen-Zimmerman, and Jordan Grafman. Executive functions. In *Handbook of Clinical Neurology*, volume 163, pages 197–219. Elsevier, January 2019. doi: 10.1016/B978-0-12-804281-6.00011-2. URL <https://www.sciencedirect.com/science/chapter/handbook/abs/pii/B9780128042816000112>.
- [76] Mihaly Csikszentmihalyi. FLOW: The Psychology of Optimal Experience. page 6, 2000.
- [77] Mihaly Csikszentmihalyi. *Flow: The Psychology of Happiness*. Random House, November 2013. ISBN 978-1-4481-7770-7.
- [78] Fazlida Dahalan, Norlidah Alias, and Mohd Shahril Nizam Shaharom. Gamification and Game Based Learning for Vocational Education and Training: A Systematic Literature Review. *Educ Inf Technol*, 29(2):1279–1317, February 2024. ISSN 1573-7608. doi: 10.1007/s10639-022-11548-w. URL <https://doi.org/10.1007/s10639-022-11548-w>.
- [79] Eric Anthony Day, Bennett Jr Winston, and Antoinette M. Portrey. *Individual and Team Skill Decay: The Science and Implications for Practice*. Taylor & Francis Group, 2013.
- [80] Elisabetta De Maria, Thibaud L’Yvonnet, Sabine Moisan, et al. Probabilistic Activity Recognition for Serious Games with Applications in Medicine. In Osman Hasan and Frédéric Mallet, editors, *Form. Tech. Saf.-Crit. Syst.*, pages 106–124, Cham, 2020. Springer International Publishing. ISBN 978-3-030-46902-3. doi: 10.1007/978-3-030-46902-3_7.
- [81] Jessica T. DeCuir-Gunby, Patricia L. Marshall, and Allison W. McCulloch. Developing and Using a Codebook for the Analysis of Interview Data: An Example from a Professional Development Research Project. *Field Methods*, 23(2): 136–155, May 2011. ISSN 1525-822X. doi: 10.1177/1525822X10388468. URL <https://doi.org/10.1177/1525822X10388468>.
- [82] Mathias Delahaye, Boris Zbinden, Bruno Herbelin, et al. On the Importance of Providing a Tangible Haptic Response for Training Cardiopulmonary Resuscitation in Virtual Reality. *ICAT-EGVE 2020 - Int. Conf. Artif. Real. Telexistence Eurographics Symp. Virtual Environ.*, page 9 pages, 2020. ISSN 1727-530X. doi: 10.2312/EGVE.20201255. URL <https://diglib.eg.org/handle/10.2312/egve20201255>.
- [83] Samantha Dequanter, Marie-Pierre Gagnon, Mame-Awa Ndiaye, et al. The Effectiveness of e-Health Solutions for Aging With Cognitive Impairment: A Systematic Review. *The Gerontologist*, 61(7):e373–e394, October 2021. ISSN 0016-9013. doi: 10.1093/geront/gnaa065. URL <https://doi.org/10.1093/geront/gnaa065>.

- [84] Sebastian Deterding, Dan Dixon, Rilla Khaled, et al. From game design elements to gamefulness: Defining "gamification". In *Proc. 15th Int. Acad. MindTrek Conf. Envisioning Future Media Environ.*, MindTrek '11, pages 9–15, New York, NY, USA, September 2011. Association for Computing Machinery. ISBN 978-1-4503-0816-8. doi: 10.1145/2181037.2181040. URL <https://dl.acm.org/doi/10.1145/2181037.2181040>.
- [85] Alice Dewhirst, Richard Laugharne, and Rohit Shankar. Therapeutic use of serious games in mental health: Scoping review. *BJPsych Open*, 8(2):e37, March 2022. ISSN 2056-4724. doi: 10.1192/bjo.2022.4. URL <https://www.cambridge.org/core/journals/bjpsych-open/article/therapeutic-use-of-serious-games-in-mental-health-scoping-review/46F4444870E994D8124AC398286F48C4>.
- [86] Eshita Dhar, Umashankar Upadhyay, Yaoru Huang, et al. A scoping review to assess the effects of virtual reality in medical education and clinical care. *Digit. Health*, 9:20552076231158022, January 2023. ISSN 2055-2076. doi: 10.1177/20552076231158022. URL <https://doi.org/10.1177/20552076231158022>.
- [87] Sakshi Dhir, Wei-Peng Teo, Samuel R. Chamberlain, et al. The Effects of Combined Physical and Cognitive Training on Inhibitory Control: A Systematic Review and Meta-Analysis. *Neuroscience & Biobehavioral Reviews*, 128:735–748, September 2021. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2021.07.008. URL <https://www.sciencedirect.com/science/article/pii/S0149763421003079>.
- [88] Adele Diamond. Executive Functions. *Annu. Rev. Psychol.*, 64(1): 135–168, January 2013. ISSN 0066-4308, 1545-2085. doi: 10.1146/annurev-psych-113011-143750. URL <https://www.annualreviews.org/doi/10.1146/annurev-psych-113011-143750>.
- [89] Cindy Dong, Chanam Shin, Joan McDonagh, et al. Immersive virtual reality simulation versus screen-based virtual simulation: An examination of learning outcomes in nursing education. *Clinical Simulation in Nursing*, 102:101710, May 2025. ISSN 1876-1399. doi: 10.1016/j.ecns.2025.101710. URL <https://www.sciencedirect.com/science/article/pii/S1876139925000271>.
- [90] Sarah A. Douglas, Arthur E. Kirkpatrick, and I. Scott MacKenzie. Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, CHI '99, pages 215–222, New York, NY, USA, May 1999. Association for Computing Machinery. ISBN 978-0-201-48559-2. doi: 10.1145/302979.303042. URL <https://doi.org/10.1145/302979.303042>.
- [91] B. Dubois, A. Slachevsky, I. Litvan, et al. The FAB. *Neurology*, 55(11):1621–1626, December 2000. doi: 10.1212/WNL.55.11.1621. URL <https://www.neurology.org/doi/abs/10.1212/WNL.55.11.1621>.
- [92] Randolph D. Easton and M. Jeanne Sholl. Object-array structure, frames of reference, and retrieval of spatial knowledge. *J. Exp. Psychol. Learn. Mem. Cogn.*, 21(2):483–500, 1995. ISSN 1939-1285. doi: 10.1037/0278-7393.21.2.483.

- [93] Barrett Ens, Joel Lanir, Anthony Tang, et al. Revisiting collaboration through mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98, November 2019. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2019.05.011. URL <https://www.sciencedirect.com/science/article/pii/S1071581919300606>.
- [94] ERC. Bringing resuscitation to the world, July 2023. URL <https://www.erc.edu/courses/advanced-life-support>.
- [95] Timothy Everson, Matthew Joordens, Helen Forbes, et al. Virtual Reality and Haptic Cardiopulmonary Resuscitation Training Approaches: A Review. *IEEE Syst. J.*, 16(1):1391–1399, March 2022. ISSN 1937-9234. doi: 10.1109/JSYST.2020.3048140. URL <https://ieeexplore.ieee.org/abstract/document/9330756>.
- [96] Liv Fegran, Wilma ten Ham-Baloyi, Mariann Fossum, et al. Simulation debriefing as part of simulation for clinical teaching and learning in nursing education: A scoping review. *Nurs. Open*, 10(3):1217–1233, 2023. ISSN 2054-1058. doi: 10.1002/nop2.1426. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/nop2.1426>.
- [97] Zhenan Feng, Vicente A. González, Robert Amor, et al. Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education*, 127:252–266, December 2018. ISSN 0360-1315. doi: 10.1016/j.compedu.2018.09.002. URL <https://www.sciencedirect.com/science/article/pii/S0360131518302380>.
- [98] Alessandro Forgiarini, Fabio Buttussi, and Luca Chittaro. Virtual Reality for Object Location Spatial Memory: A Comparison of Handheld Controllers and Force Feedback Gloves. In *Proc. 15th Biannu. Conf. Ital. SIGCHI Chapter, CHIItaly '23*, pages 1–9, New York, NY, USA, September 2023. Association for Computing Machinery. ISBN 979-8-4007-0806-0. doi: 10.1145/3605390.3605403. URL <https://doi.org/10.1145/3605390.3605403>.
- [99] Alessandro Forgiarini, Laura Deroma, Fabio Buttussi, et al. Introducing Virtual Reality in a STEMI Coronary Syndrome Course: Qualitative Evaluation with Nurses and Doctors. *Cyberpsychology, Behavior, and Social Networking*, 27(6):387–398, June 2024. ISSN 2152-2715, 2152-2723. doi: 10.1089/cyber.2023.0414. URL <https://www.liebertpub.com/doi/10.1089/cyber.2023.0414>.
- [100] Alessandro Forgiarini, Elisabetta De Maria, and Fabio Buttussi. Probabilistic Modeling and Verification of an Adaptive VR Serious Game for Patients with Cognitive Impairment. In Riccardo Bellazzi, José Manuel Juárez Herrero, Lucia Sacchi, et al., editors, *Artif. Intell. Med.*, pages 160–164, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-95841-0. doi: 10.1007/978-3-031-95841-0_30.
- [101] The R Foundation. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, July 2023. URL <https://www.r-project.org/>.
- [102] Tord Hettervik Froland, Werner Ravyse, Mika Luimula, et al. Single- and Multiuser Virtual Reality Training for Forklift Driving. In *2023 IEEE 2nd Int. Conf.*

- Cogn. Asp. Virtual Real. CVR*, pages 000073–000080, October 2023. doi: 10.1109/CVR58941.2023.10395371. URL <https://ieeexplore.ieee.org/abstract/document/10395371>.
- [103] Roberto Galvez, Robert C. Wallon, Laura Shackelford, et al. Use of Virtual Reality to Educate Undergraduate Medical Students on Cardiac Peripheral and Collateral Circulation. *Med.Sci.Educ.*, 31(1):19–22, February 2021. ISSN 2156-8650. doi: 10.1007/s40670-020-01104-x. URL <https://doi.org/10.1007/s40670-020-01104-x>.
- [104] Abrar Gani, Oliver Pickering, Caroline Ellis, et al. Impact of haptic feedback on surgical training outcomes: A Randomised Controlled Trial of haptic versus non-haptic immersive virtual reality training. *Annals of Medicine and Surgery*, 83:104734, November 2022. ISSN 2049-0801. doi: 10.1016/j.amsu.2022.104734. URL <https://www.sciencedirect.com/science/article/pii/S2049080122014947>.
- [105] Meiyuzi Gao, Philip Kortum, and Frederick Oswald. Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, 62(1):1414–1418, September 2018. ISSN 2169-5067. doi: 10.1177/1541931218621322. URL <https://doi.org/10.1177/1541931218621322>.
- [106] A. Garson. President’s page: The Great Circle: A target for better patient care. *J Am Coll Cardiol*, 34(1):294–295, July 1999. ISSN 0735-1097. doi: 10.1016/s0735-1097(99)00240-5.
- [107] Ivana Gazova, Kamil Vlček, Jan Laczó, et al. Spatial navigation-A unique window into physiological and pathological aging. *Frontiers in aging neuroscience*, 4:16, June 2012. doi: 10.3389/fnagi.2012.00016.
- [108] Patrick Gebhardt, Xingyao Yu, Andreas Köhn, et al. MolecuSense: Using Force-Feedback Gloves for Creating and Interacting with Ball-and-Stick Molecules in VR. In *Proc. 15th Int. Symp. Vis. Inf. Commun. Interact.*, VINCI ’22, pages 1–5, New York, NY, USA, October 2022. Association for Computing Machinery. ISBN 978-1-4503-9806-0. doi: 10.1145/3554944.3554956. URL <https://dl.acm.org/doi/10.1145/3554944.3554956>.
- [109] Katie A. Gilligan, Eirini Flouri, and Emily K. Farran. The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology*, 163:107–125, November 2017. ISSN 0022-0965. doi: 10.1016/j.jecp.2017.04.016. URL <https://www.sciencedirect.com/science/article/pii/S0022096516302028>.
- [110] Maurits Graafland, Mary Dankbaar, Agali Mert, et al. How to Systematically Assess Serious Games Applied to Health Care. *JMIR Serious Games*, 2(2):e3825, November 2014. doi: 10.2196/games.3825. URL <https://games.jmir.org/2014/2/e11>.
- [111] David A. Grant and Esta A. Berg. Wisconsin Card Sorting Test, July 2014. URL <https://doi.apa.org/doi/10.1037/t31298-000>.

- [112] Jan-Thorsten Gräsner, Jan Wnent, Johan Herlitz, et al. Survival after out-of-hospital cardiac arrest in Europe - Results of the EuReCa TWO study. *Resuscitation*, 148:218–226, March 2020. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2019.12.042. URL <https://www.sciencedirect.com/science/article/pii/S0300957220300460>.
- [113] Anthony G. Greenwald. Within-subjects designs: To use or not to use? *Psychol. Bull.*, 83(2):314–320, 1976. ISSN 1939-1455. doi: 10.1037/0033-2909.83.2.314.
- [114] Peidi Gu, Yihang Li, Hailin Ji, et al. The Role of Haptic Interaction in Embodied Extended-Reality Learning: A Three-Level Meta-Analysis. *Educ Psychol Rev*, 37(4):97, September 2025. ISSN 1573-336X. doi: 10.1007/s10648-025-10072-w. URL <https://doi.org/10.1007/s10648-025-10072-w>.
- [115] Angela Guarino, Giuseppe Forte, Jasmine Giovannoli, et al. Executive functions in the elderly with mild cognitive impairment: A systematic review on motor and cognitive inhibition, conflict control and cognitive flexibility. *Aging Ment. Health*, 24(7):1028–1045, July 2020. ISSN 1360-7863. doi: 10.1080/13607863.2019.1584785. URL <https://doi.org/10.1080/13607863.2019.1584785>.
- [116] Debora Gugelmin-Almeida, Lucia Tobase, Thatiane Facholi Polastri, et al. Do automated real-time feedback devices improve CPR quality? A systematic review of literature. *Resuscitation Plus*, 6:100108, June 2021. ISSN 2666-5204. doi: 10.1016/j.resplu.2021.100108. URL <https://www.sciencedirect.com/science/article/pii/S2666520421000333>.
- [117] Youn Suk Ham, Young Mi Park, and Young Sook Roh. Comparison of Learning Outcomes Between Active Participants and Observer Roles in Simulation-Based Learning. *Clinical Simulation in Nursing*, 87:101497, February 2024. ISSN 1876-1399. doi: 10.1016/j.ecns.2023.101497. URL <https://www.sciencedirect.com/science/article/pii/S1876139923001111>.
- [118] Benjamin M. Hampstead, Krish Sathian, Pamela A. Phillips, et al. Mnemonic strategy training improves memory for object location associations in both healthy elderly and patients with amnesic mild cognitive impairment: A randomized, single-blind study. *Neuropsychology*, 26:385–399, 2012. ISSN 1931-1559. doi: 10.1037/a0027545.
- [119] Laura Erika Maria Hannen, Betül Toprak, Jessica Weimann, et al. Clinical characteristics, causes and predictors of outcomes in patients with in-hospital cardiac arrest: Results from the SURVIVE-ARREST study. *Clin Res Cardiol*, 112(2): 258–269, February 2023. ISSN 1861-0692. doi: 10.1007/s00392-022-02084-1. URL <https://doi.org/10.1007/s00392-022-02084-1>.
- [120] Camilla Hansen, Camilla Bang, Mathilde Stærk, et al. Certified Basic Life Support Instructors Identify Improper Cardiopulmonary Resuscitation Skills Poorly: Instructor Assessments Versus Resuscitation Manikin Data. *Simul. Healthc.*, 14(5):281, October 2019. ISSN 1559-2332. doi: 10.1097/SIH.0000000000000386. URL https://journals.lww.com/simulationinhealthcare/fulltext/2019/10000/certified_basic_life_support_instructors_identify.1.aspx.

- [121] Hans Hansson and Bengt Jonsson. A logic for reasoning about time and reliability. *Formal Aspects of Computing*, 6(5):512–535, September 1994. ISSN 1433-299X. doi: 10.1007/BF01211866. URL <https://doi.org/10.1007/BF01211866>.
- [122] Mary Hassandra, Evangelos Galanis, Antonis Hatzigeorgiadis, et al. A Virtual Reality App for Physical and Cognitive Training of Older People With Mild Cognitive Impairment: Mixed Methods Feasibility Study. *JMIR Serious Games*, 9(1):e24170, March 2021. doi: 10.2196/24170. URL <https://games.jmir.org/2021/1/e24170>.
- [123] Meggy Hayotte, Pierre Thérouanne, Laura Gray, et al. The French eHealth Acceptability Scale Using the Unified Theory of Acceptance and Use of Technology 2 Model: Instrument Validation Study. *J. Med. Internet Res.*, 22(4):e16520, April 2020. doi: 10.2196/16520. URL <https://www.jmir.org/2020/4/e16520>.
- [124] Lukas Hejtmanek, Michael Starrett, Emilio Ferrer, et al. How Much of What We Learn in Virtual Reality Transfers to Real-World Navigation? *Multisensory Res.*, 33(4-5):479–503, March 2020. ISSN 2213-4808, 2213-4794. doi: 10.1163/22134808-20201445. URL https://brill.com/view/journals/msr/33/4-5/article-p479_6.xml.
- [125] William S. Helton. Impulsive responding and the sustained attention to response task. *J. Clin. Exp. Neuropsychol.*, 31(1):39–47, January 2009. ISSN 1380-3395. doi: 10.1080/13803390801978856. URL <https://doi.org/10.1080/13803390801978856>.
- [126] Lena Günterberg Heyn, Espen Andreas Brembo, Kirsten Røland Byermoen, et al. Exploring facilitation in virtual simulation in nursing education: A scoping review. *PEC Innovation*, 3:100233, December 2023. ISSN 2772-6282. doi: 10.1016/j.pecinn.2023.100233. URL <https://www.sciencedirect.com/science/article/pii/S2772628223001139>.
- [127] C. R. Hind. Continuing medical education. *Postgrad Med J*, 71(833):129–131, March 1995. ISSN 0032-5473. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2398167/>.
- [128] Thomas Hinterecker, Caroline Leroy, Mintao Zhao, et al. No advantage for remembering horizontal over vertical spatial locations learned from a single viewpoint. *Mem Cogn*, 46(1):158–171, January 2018. ISSN 1532-5946. doi: 10.3758/s13421-017-0753-9. URL <https://doi.org/10.3758/s13421-017-0753-9>.
- [129] Wilhelm Hofmann, Brandon J. Schmeichel, and Alan D. Baddeley. Executive functions and self-regulation. *Trends in Cognitive Sciences*, 16(3):174–180, March 2012. ISSN 1364-6613. doi: 10.1016/j.tics.2012.01.006. URL <https://www.sciencedirect.com/science/article/pii/S1364661312000289>.
- [130] Bruce D. Homer, Teresa M. Ober, Maya C. Rose, et al. Speed Versus Accuracy: Implications of Adolescents’ Neurocognitive Developments in a Digital Game to Train Executive Functions. *Mind Brain Educ.*, 13(1):41–52, 2019. ISSN 1751-228X. doi: 10.1111/mbe.12189. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mbe.12189>.

- [131] Chengang Hong and Liping Wang. Virtual Reality Technology in Nursing Professional Skills Training: Bibliometric Analysis. *JMIR Serious Games*, 11(1):e44766, August 2023. doi: 10.2196/44766. URL <https://games.jmir.org/2023/1/e44766>.
- [132] Rebecca J. Hood, Steven Maltby, Angela Keynes, et al. Development and Pilot Implementation of TACTICS VR: A Virtual Reality-Based Stroke Management Workflow Training Application and Training Framework. *Front. Neurol.*, 12, 2021. ISSN 1664-2295. doi: 10.3389/fneur.2021.665808. URL <https://www.frontiersin.org/articles/10.3389/fneur.2021.665808>.
- [133] Steven J. Howard, Janice Johnson, and Juan Pascual-Leone. Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, 31:1–21, July 2014. ISSN 0885-2014. doi: 10.1016/j.cogdev.2014.03.001. URL <https://www.sciencedirect.com/science/article/pii/S088520141400032X>.
- [134] Panpan Huai, Yao Li, Xiaomeng Wang, et al. The effectiveness of virtual reality technology in student nurse education: A systematic review and meta-analysis. *Nurse Education Today*, 138:106189, July 2024. ISSN 0260-6917. doi: 10.1016/j.nedt.2024.106189. URL <https://www.sciencedirect.com/science/article/pii/S0260691724000996>.
- [135] Jiawei Huang and Alexander Klippel. The Effects of Visual Realism on Spatial Memory and Exploration Patterns in Virtual Reality. In *Proc. 26th ACM Symp. Virtual Real. Softw. Technol.*, VRST '20, pages 1–11, New York, NY, USA, November 2020. Association for Computing Machinery. ISBN 978-1-4503-7619-8. doi: 10.1145/3385956.3418945. URL <https://doi.org/10.1145/3385956.3418945>.
- [136] Wen Huang, Rod D. Roscoe, Mina C. Johnson-Glenberg, et al. Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion. *J. Comput. Assist. Learn.*, 37(3):745–758, 2021. ISSN 1365-2729. doi: 10.1111/jcal.12520. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12520>.
- [137] Emilie Hummel, Mélanie Cogné, Marie Lange, et al. VR for Vocational and Ecological Rehabilitation of Patients With Cognitive Impairment: A Survey. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 31:4167–4178, 2023. ISSN 1558-0210. doi: 10.1109/TNSRE.2023.3324131. URL <https://ieeexplore.ieee.org/document/10288537>.
- [138] Christophe Hurter, Bernice Rogowitz, Guillaume Truong, et al. Memory Recall for Data Visualizations in Mixed Reality, Virtual Reality, 3D and 2D. *IEEE Trans. Vis. Comput. Graph.*, 30(10):6691–6706, October 2024. ISSN 1941-0506. doi: 10.1109/TVCG.2023.3336588. URL <https://ieeexplore.ieee.org/document/10473176/keywords>.
- [139] Hanne Huygelier, Brenda Schraepen, Raymond van Ee, et al. Acceptance of immersive head-mounted virtual reality in older adults. *Sci Rep*, 9(1):4519, March 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-41200-6. URL <https://www.nature.com/articles/s41598-019-41200-6>.

- [140] Gwo-Jen Hwang, Chun-Chun Chang, and Shu-Yun Chien. A motivational model-based virtual reality approach to prompting learners' sense of presence, learning achievements, and higher-order thinking in professional safety training. *Br. J. Educ. Technol.*, 53(5):1343–1360, September 2022. ISSN 0007-1013. doi: 10.1111/bjet.13196. URL <https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13196>.
- [141] W.A. IJsselsteijn, Y.A.W. de Kort, and K. Poels. *The Game Experience Questionnaire*. Technische Universiteit Eindhoven, Eindhoven, 2013.
- [142] SAS Institute Inc. SAS: Analytics, Artificial Intelligence and Data Management, September 2023. URL https://www.sas.com/en_us/home.html.
- [143] Arham I. Iqbal, Ali Aamir, Abdullah Hammad, et al. Immersive Technologies in Healthcare: An In-Depth Exploration of Virtual Reality and Augmented Reality in Enhancing Patient Care, Medical Education, and Training Paradigms. *J Prim Care Community Health*, 15:21501319241293311, January 2024. ISSN 2150-1319. doi: 10.1177/21501319241293311. URL <https://doi.org/10.1177/21501319241293311>.
- [144] Noah Ivers, Gro Jamtvedt, Signe Flottorp, et al. Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*, (6):CD000259, June 2012. ISSN 1469-493X. doi: 10.1002/14651858.CD000259.pub3.
- [145] Noah M. Ivers, Jeremy M. Grimshaw, Gro Jamtvedt, et al. Growing Literature, Stagnant Science? Systematic Review, Meta-Regression and Cumulative Analysis of Audit and Feedback Interventions in Health Care. *J GEN INTERN MED*, 29(11):1534–1541, November 2014. ISSN 1525-1497. doi: 10.1007/s11606-014-2913-y. URL <https://doi.org/10.1007/s11606-014-2913-y>.
- [146] Samar Thabet Jallad and Burçin Işık. The effectiveness of virtual reality simulation as learning strategy in the acquisition of medical skills in nursing education: A systematic review. *Ir J Med Sci*, 191(3):1407–1426, June 2022. ISSN 1863-4362. doi: 10.1007/s11845-021-02695-z. URL <https://doi.org/10.1007/s11845-021-02695-z>.
- [147] Gro Jamtvedt, Signe Flottorp, and Noah Ivers. Audit and Feedback as a Quality Strategy. In *Improving Healthcare Quality in Europe: Characteristics, Effectiveness and Implementation of Different Strategies*, page 265. European Observatory on Health Systems and Policies, October 2019. URL <https://www.ncbi.nlm.nih.gov/books/NBK549284/>.
- [148] M. Kathryn Jedrzejewski, Douglas C. Ewbank, Haidong Wang, et al. The Impact of Exercise, Cognitive Activities, and Socialization on Cognitive Function: Results From the National Long-Term Care Survey. *Am J Alzheimers Dis Other Dement*, 29(4):372–378, June 2014. ISSN 1533-3175. doi: 10.1177/1533317513518646. URL <https://doi.org/10.1177/1533317513518646>.

- [149] Lasse Jensen and Flemming Konradsen. A review of the use of virtual reality head-mounted displays in education and training. *Educ Inf Technol*, 23(4):1515–1529, July 2018. ISSN 1573-7608. doi: 10.1007/s10639-017-9676-0. URL <https://doi.org/10.1007/s10639-017-9676-0>.
- [150] Jason Jerald. *The VR Book: Human-Centered Design for Virtual Reality*. ACM Press, September 2015. ISBN 978-1-970001-13-6.
- [151] David W. Johnson and Roger T. Johnson. An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educational Researcher*, 38(5):365–379, June 2009. ISSN 0013-189X, 1935-102X. doi: 10.3102/0013189X09339057. URL <https://journals.sagepub.com/doi/10.3102/0013189X09339057>.
- [152] Matthew Johnson, Amanda Peat, Leanne Boyd, et al. The impact of quantitative feedback on the performance of chest compression by basic life support trained clinical staff. *Nurse Education Today*, 45:163–166, October 2016. ISSN 0260-6917. doi: 10.1016/j.nedt.2016.08.006. URL <https://www.sciencedirect.com/science/article/pii/S0260691716301538>.
- [153] Janine Jongbloed, Rawad Chaker, and Elise Lavoué. Immersive procedural training in virtual reality: A systematic literature review. *Computers & Education*, 221: 105124, November 2024. ISSN 0360-1315. doi: 10.1016/j.compedu.2024.105124. URL <https://www.sciencedirect.com/science/article/pii/S0360131524001386>.
- [154] Amiruddin Junli, Siti Nor Ismalina Isa, and Farrah Shafeera Ibrahim. Factors of cardiopulmonary resuscitation skills retention among healthcare providers: A scoping review. *Nurse Education in Practice*, 69:103617, May 2023. ISSN 1471-5953. doi: 10.1016/j.nepr.2023.103617. URL <https://www.sciencedirect.com/science/article/pii/S1471595323000793>.
- [155] Alexandra D. Kaplan, Jessica Cruit, Mica Endsley, et al. The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. *Hum Factors*, 63(4):706–726, June 2021. ISSN 0018-7208. doi: 10.1177/0018720820904229. URL <https://doi.org/10.1177/0018720820904229>.
- [156] Kim Kargut, Carl Gutwin, and Andy Cockburn. Effects of Device Environment and Information Layout on Spatial Memory and Performance in VR Selection Tasks. In *Proc. 2024 CHI Conf. Hum. Factors Comput. Syst.*, CHI '24, pages 1–17, New York, NY, USA, May 2024. Association for Computing Machinery. ISBN 979-8-4007-0330-0. doi: 10.1145/3613904.3642486. URL <https://doi.org/10.1145/3613904.3642486>.
- [157] William Kaye, Sharon F. Rallis, Mary E. Mancini, et al. The problem of poor retention of cardiopulmonary resuscitation skills may lie with the instructor, not the learner or the curriculum. *Resuscitation*, 21(1):67–87, February 1991. ISSN 0300-9572. doi: 10.1016/0300-9572(91)90080-I. URL <https://www.sciencedirect.com/science/article/pii/030095729190080I>.

- [158] Roy P. C Kessels, L Jaap Kappelle, Edward H. F de Haan, et al. Lateralization of spatial-memory processes: Evidence on spatial span, maze learning, and memory for object locations. *Neuropsychologia*, 40(8):1465–1473, January 2002. ISSN 0028-3932. doi: 10.1016/S0028-3932(01)00199-3. URL <https://www.sciencedirect.com/science/article/pii/S0028393201001993>.
- [159] Roy P. C. Kessels, Stefan Rijken, Liesbeth W. a. Joosten-Weyn Banningh, et al. Categorical spatial memory in patients with mild cognitive impairment and Alzheimer dementia: Positional versus object-location recall. *J. Int. Neuropsychol. Soc.*, 16(1):200–204, January 2010. ISSN 1469-7661, 1355-6177. doi: 10.1017/S1355617709990944. URL <https://www.cambridge.org/core/journals/journal-of-the-international-neuropsychological-society/article/abs/categorical-spatial-memory-in-patients-with-mild-cognitive-impairment-and-alzheimer-dementia/279008CED16020155D3C4D89ABE4614A>.
- [160] Roy P.C. Kessels, Edward H.F. De Haan, L.Jaap Kappelle, et al. Varieties of human spatial memory: A meta-analysis on the effects of hippocampal lesions. *Brain Res. Rev.*, 35(3):295–303, July 2001. ISSN 01650173. doi: 10.1016/S0165-0173(01)00058-3. URL <https://linkinghub.elsevier.com/retrieve/pii/S0165017301000583>.
- [161] Yoojin Kim and Ha-Young Park. Effects of virtual reality training on clinical skill performance in nursing students: A systematic review, meta-analysis and meta-regression. *Int. J. Nurs. Pract.*, 30(6):e13284, 2024. ISSN 1440-172X. doi: 10.1111/ijn.13284. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/ijn.13284>.
- [162] J. G. Kingma. Myocardial Infarction: An Overview of STEMI and NSTEMI Physiopathology and Treatment. *WJCD*, 08(11):498–517, 2018. ISSN 2164-5329, 2164-5337. doi: 10.4236/wjcd.2018.811049. URL <http://www.scirp.org/journal/doi.aspx?DOI=10.4236/wjcd.2018.811049>.
- [163] Femke Kirschner, Fred Paas, and Paul A. Kirschner. A Cognitive Load Approach to Collaborative Learning: United Brains for Complex Tasks. *Educ Psychol Rev*, 21(1):31–42, March 2009. ISSN 1573-336X. doi: 10.1007/s10648-008-9095-2. URL <https://doi.org/10.1007/s10648-008-9095-2>.
- [164] Nicholas E. Kman, Alan Price, Vita Berezina-Blackburn, et al. First Responder Virtual Reality Simulator to train and assess emergency personnel for mass casualty response. *JACEP Open*, 4(1):e12903, February 2023. ISSN 2688-1152. doi: 10.1002/emp2.12903. URL <https://www.sciencedirect.com/science/article/pii/S2688115224001541>.
- [165] Ying-Chih Ko, Ming-Ju Hsieh, Adam Cheng, et al. Faculty Development Approaches for Life Support Courses: A Scoping Review. *J. Am. Heart Assoc.*, 11(11):e025661, June 2022. doi: 10.1161/JAHA.122.025661. URL <https://www.ahajournals.org/doi/full/10.1161/JAHA.122.025661>.
- [166] Masaki Kobayashi, Miyuki Iwamoto, Saki Une, et al. Simulated communication skills training program effects using augmented reality with real-time feedback: A

- randomized control study. *Alzheimers Dement.*, 18(S8):e062055, 2022. ISSN 1552-5279. doi: 10.1002/alz.062055. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.062055>.
- [167] Philip Kortum and S. Camille Peres. The Relationship Between System Effectiveness and Subjective Usability Scores Using the System Usability Scale. *Int. J. Human-Computer Interact.*, 30(7):575–584, July 2014. ISSN 1044-7318. doi: 10.1080/10447318.2014.904177. URL <https://doi.org/10.1080/10447318.2014.904177>.
- [168] Aaron Kotranza, D. Scott Lind, and Benjamin Lok. Real-Time Evaluation and Visualization of Learner Performance in a Mixed-Reality Environment for Clinical Breast Examination. *IEEE Trans. Vis. Comput. Graph.*, 18(7):1101–1114, July 2012. ISSN 1941-0506. doi: 10.1109/TVCG.2011.132. URL <https://ieeexplore.ieee.org/abstract/document/5963665>.
- [169] Marileen M. T. E. Kouijzer, Hanneke Kip, Yvonne H. A. Bouman, et al. Implementation of virtual reality in healthcare: A scoping review on the implementation process of virtual reality in various healthcare settings. *Implement Sci Commun*, 4(1):67, June 2023. ISSN 2662-2211. doi: 10.1186/s43058-023-00442-2. URL <https://doi.org/10.1186/s43058-023-00442-2>.
- [170] Jules J. M. Kraaijkamp, Eléonore F. van Dam van Isselt, Anke Persoon, et al. eHealth in Geriatric Rehabilitation: Systematic Review of Effectiveness, Feasibility, and Usability. *J. Med. Internet Res.*, 23(8):e24015, August 2021. doi: 10.2196/24015. URL <https://www.jmir.org/2021/8/e24015>.
- [171] Jeanine Krath, Linda Schürmann, and Harald F. O. von Korflesch. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior*, 125:106963, December 2021. ISSN 0747-5632. doi: 10.1016/j.chb.2021.106963. URL <https://www.sciencedirect.com/science/article/pii/S0747563221002867>.
- [172] Julian Kreimeier and Timo Götzelmann. FeelVR: Haptic Exploration of Virtual Objects. In *Proc. 11th Pervasive Technol. Relat. Assist. Environ. Conf.*, PETRA '18, pages 122–125, New York, NY, USA, June 2018. Association for Computing Machinery. ISBN 978-1-4503-6390-7. doi: 10.1145/3197768.3201526. URL <https://dl.acm.org/doi/10.1145/3197768.3201526>.
- [173] Julian Kreimeier, Sebastian Hammer, Daniel Friedmann, et al. Evaluation of different types of haptic feedback influencing the task-based presence and performance in virtual reality. In *Proc. 12th ACM Int. Conf. Pervasive Technol. Relat. Assist. Environ.*, PETRA '19, pages 289–298, New York, NY, USA, June 2019. Association for Computing Machinery. ISBN 978-1-4503-6232-0. doi: 10.1145/3316782.3321536. URL <https://doi.org/10.1145/3316782.3321536>.
- [174] Eric Krokos, Catherine Plaisant, and Amitabh Varshney. Virtual memory palaces: Immersion aids recall. *Virtual Reality*, 23(1):1–15, March 2019. ISSN

- 1434-9957. doi: 10.1007/s10055-018-0346-3. URL <https://doi.org/10.1007/s10055-018-0346-3>.
- [175] Timothy Kubal, Doug G Letson, Alberto A Chiappori, et al. Longitudinal cohort study to determine effectiveness of a novel simulated case and feedback system to improve clinical pathway adherence in breast, lung and GI cancers. *BMJ Open*, 6(9):e012312, September 2016. ISSN 2044-6055. doi: 10.1136/bmjopen-2016-012312. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5030551/>.
- [176] Marta Kwiatkowska, Gethin Norman, and David Parker. Stochastic Model Checking. In Marco Bernardo and Jane Hillston, editors, *Formal Methods for Performance Evaluation: 7th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, SFM 2007, Bertinoro, Italy, May 28-June 2, 2007, Advanced Lectures*, pages 220–270. Springer, Berlin, Heidelberg, 2007. ISBN 978-3-540-72522-0. doi: 10.1007/978-3-540-72522-0_6. URL https://doi.org/10.1007/978-3-540-72522-0_6.
- [177] Marta Kwiatkowska, Gethin Norman, and David Parker. PRISM 4.0: Verification of Probabilistic Real-Time Systems. In Ganesh Gopalakrishnan and Shaz Qadeer, editors, *Comput. Aided Verification*, pages 585–591, Berlin, Heidelberg, 2011. Springer. ISBN 978-3-642-22110-1. doi: 10.1007/978-3-642-22110-1_47.
- [178] Bhone Myint Kyaw, Nakul Saxena, Pawel Posadzki, et al. Virtual Reality for Health Professions Education: Systematic Review and Meta-Analysis by the Digital Health Education Collaboration. *J. Med. Internet Res.*, 21(1):e12959, January 2019. doi: 10.2196/12959. URL <https://www.jmir.org/2019/1/e12959>.
- [179] Min Lang, Samir Ghandour, Blaire Rikard, et al. Medical Extended Reality for Radiology Education and Training. *Journal of the American College of Radiology*, 21(10):1583–1594, October 2024. ISSN 1546-1440. doi: 10.1016/j.jacr.2024.05.006. URL <https://www.sciencedirect.com/science/article/pii/S1546144024005167>.
- [180] Siew Tiang Lau, Chiew Jiat Rosalind Siah, Wen Liang Loh, et al. Enhancing professional competency in clinical procedures using head-mounted display virtual reality - a mixed method study. *Med. Educ. Online*, 28(1):2232134, December 2023. ISSN null. doi: 10.1080/10872981.2023.2232134. URL <https://doi.org/10.1080/10872981.2023.2232134>.
- [181] Patrick Lavoie, Louise-Andrée Brien, Isabelle Ledoux, et al. Immersive and non-immersive virtual reality: A quasi-experimental study in undergraduate nursing education. *Clinical Simulation in Nursing*, 99:101682, February 2025. ISSN 1876-1399. doi: 10.1016/j.ecns.2024.101682. URL <https://www.sciencedirect.com/science/article/pii/S1876139924001749>.
- [182] Jung Jae Lee, Vivien Wai Yin Tsang, Maggie Mee Kie Chan, et al. Virtual reality simulation-enhanced blood transfusion education for undergraduate nursing students: A randomised controlled trial. *Nurse Education Today*, 129:105903, October 2023. ISSN 0260-6917. doi: 10.1016/j.nedt.2023.105903. URL <https://www.sciencedirect.com/science/article/pii/S0260691723001971>.

- [183] Wenche Lervik, Mads Solberg, Astrid Camilla Wiig, et al. Turning nursing students' mistakes into resources for learning in simulation-based training: Facilitators' assumptions about providing feedback in debriefing. *BMC Medical Education*, 25(1):76, January 2025. ISSN 1472-6920. doi: 10.1186/s12909-024-06628-z. URL <https://doi.org/10.1186/s12909-024-06628-z>.
- [184] Lu Li, Sisi Fan, Yuxin Xing, et al. Application of mixed reality in prehospital emergency nursing education: A scoping review. *Nurse Education in Practice*, 86:104415, July 2025. ISSN 1471-5953. doi: 10.1016/j.nepr.2025.104415. URL <https://www.sciencedirect.com/science/article/pii/S1471595325001714>.
- [185] Hai-Ning Liang, Feiyu Lu, Yuwei Shi, et al. Evaluating the effects of collaboration and competition in navigation tasks and spatial knowledge acquisition within virtual reality environments. *Future Generation Computer Systems*, 95: 855–866, June 2019. ISSN 0167-739X. doi: 10.1016/j.future.2018.02.029. URL <https://www.sciencedirect.com/science/article/pii/S0167739X17308324>.
- [186] Sabrina Licata, Annarita Tullio, and Francesca Valent. Audit and Feedback in emergency: A systematic review and an Italian project to investigate and improve quality of care. *Emerg. Care J.*, 16(3), December 2020. ISSN 2282-2054. doi: 10.4081/ecj.2020.9201. URL <https://www.pagepressjournals.org/index.php/ecj/article/view/9201>.
- [187] Silje Stangeland Lie, Nikolina Helle, Nina Vahl Sletteland, et al. Implementation of Virtual Reality in Health Professions Education: Scoping Review. *JMIR Med. Educ.*, 9(1):e41589, January 2023. doi: 10.2196/41589. URL <https://mededu.jmir.org/2023/1/e41589>.
- [188] Agris Liepa, Jason Tang, Ingemara Jaundaldere, et al. Feasibility randomized controlled trial of a virtual reality exergame to improve physical and cognitive functioning in older people. *Acta Gymnica*, June 2022. ISSN 23364912, 23364920. doi: 10.5507/ag.2022.007. URL <http://gymnica.upol.cz/doi/10.5507/ag.2022.007.html>.
- [189] Mei-Yu Lin, Mei-Zen Huang, and Pao-Chin Lai. Effect of virtual reality training on clinical skills of nursing students: A systematic review and meta-analysis of randomized controlled trials. *Nurse Education in Practice*, 81:104182, November 2024. ISSN 1471-5953. doi: 10.1016/j.nepr.2024.104182. URL <https://www.sciencedirect.com/science/article/pii/S1471595324003111>.
- [190] Justina Yat Wa Liu, Yue-Heng Yin, Patrick Pui Kin Kor, et al. The Effects of Immersive Virtual Reality Applications on Enhancing the Learning Outcomes of Undergraduate Health Care Students: Systematic Review With Meta-synthesis. *J. Med. Internet Res.*, 25(1):e39989, March 2023. doi: 10.2196/39989. URL <https://www.jmir.org/2023/1/e39989>.
- [191] Kai Liu, Weiwei Zhang, Wei Li, et al. Effectiveness of virtual reality in nursing education: A systematic review and meta-analysis. *BMC Med Educ*, 23(1):710, September 2023. ISSN 1472-6920. doi: 10.1186/s12909-023-04662-x. URL <https://doi.org/10.1186/s12909-023-04662-x>.

- [192] Zhejun Liu, Jin Yunshui, Ma Minhua, et al. A Comparison of Immersive and Non-Immersive VR for the Education of Filmmaking. *Int. J. Human-Computer Interact.*, 39(12):2478–2491, July 2023. ISSN 1044-7318. doi: 10.1080/10447318.2022.2078462. URL <https://doi.org/10.1080/10447318.2022.2078462>.
- [193] Tania Llana, Marta Mendez, M. Carmen Juan, et al. Navigational object-location memory assessment in real and virtual environments: A systematic review. *Behavioural Brain Research*, 480:115388, March 2025. ISSN 0166-4328. doi: 10.1016/j.bbr.2024.115388. URL <https://www.sciencedirect.com/science/article/pii/S0166432824005448>.
- [194] Tuva Fjærtøft Lønne, Håvard R. Karlsen, Eva Langvik, et al. The effect of immersion on sense of presence and affect when experiencing an educational scenario in virtual reality: A randomized controlled study. *Heliyon*, 9(6), June 2023. ISSN 2405-8440. doi: 10.1016/j.heliyon.2023.e17196. URL [https://www.cell.com/heliyon/abstract/S2405-8440\(23\)04404-3](https://www.cell.com/heliyon/abstract/S2405-8440(23)04404-3).
- [195] Jack M. Loomis, Roberta L. Klatzky, and Nicholas A. Giudice. Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language. In Simon Lacey and Rebecca Lawson, editors, *Multisensory Imagery*, pages 131–155. Springer, New York, NY, 2013. ISBN 978-1-4614-5879-1. doi: 10.1007/978-1-4614-5879-1_8. URL https://doi.org/10.1007/978-1-4614-5879-1_8.
- [196] Arnold Lund. Measuring Usability with the USE Questionnaire. *Usability and User Experience Newsletter of the STC Usability SIG*, 8, January 2001.
- [197] Rasmus Meyer Lyngby, Mina Nicole Händel, Anne Mielke Christensen, et al. Effect of real-time and post-event feedback in out-of-hospital cardiac arrest attended by EMS — A systematic review and meta-analysis. *Resuscitation Plus*, 6:100101, June 2021. ISSN 2666-5204. doi: 10.1016/j.resplu.2021.100101. URL <https://www.sciencedirect.com/science/article/pii/S2666520421000266>.
- [198] Thibaud L’Yvonnet, Elisabetta De Maria, Sabine Moisan, et al. Probabilistic model checking for human activity recognition in medical serious games. *Science of Computer Programming*, 206:102629, June 2021. ISSN 01676423. doi: 10.1016/j.scico.2021.102629. URL <https://linkinghub.elsevier.com/retrieve/pii/S0167642321000228>.
- [199] E. Macaluso and A. Maravita. The representation of space near the body through touch and vision. *Neuropsychologia*, 48(3):782–795, February 2010. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2009.10.010. URL <https://www.sciencedirect.com/science/article/pii/S0028393209004138>.
- [200] Colin M. MacLeod. The Stroop Task in Cognitive Research. In *Cognitive Methods and Their Application to Clinical Research*, pages 17–40. American Psychological Association, Washington, DC, US, 2005. ISBN 978-1-59147-185-1. doi: 10.1037/10870-002.
- [201] Maria Grazia Maggio, Francesca Baglio, Raffaella Maione, et al. The overlooked role of exergames in cognitive-motor neurorehabilitation: A systematic

- review. *npj Digit. Med.*, 8(1):419, July 2025. ISSN 2398-6352. doi: 10.1038/s41746-025-01843-4. URL <https://www.nature.com/articles/s41746-025-01843-4>.
- [202] Marc-André Maheu-Cadotte, Sylvie Cossette, Véronique Dubé, et al. Efficacy of Serious Games in Healthcare Professions Education: A Systematic Review and Meta-analysis. *Simul. Healthc.*, 16(3):199, June 2021. ISSN 1559-2332. doi: 10.1097/SIH.0000000000000512. URL https://journals.lww.com/simulationinhealthcare/abstract/2021/06000/efficacy_of_serious_games_in_healthcare.7.aspx.
- [203] Shachar Maidenbaum, Ansh Patel, Elisabeth Stein, et al. Spatial Memory Rehabilitation in Virtual Reality – Extending findings from Epilepsy Patients to the General Population. In *2019 Int. Conf. Virtual Rehabil. ICVR*, pages 1–7, July 2019. doi: 10.1109/ICVR46560.2019.8994573.
- [204] Henna Mäkinen, Elina Haavisto, Sara Havola, et al. User experiences of virtual reality technologies for healthcare in learning: An integrative review. *Behav. Inf. Technol.*, 41(1):1–17, January 2022. ISSN 0144-929X. doi: 10.1080/0144929X.2020.1788162. URL <https://doi.org/10.1080/0144929X.2020.1788162>.
- [205] Guido Makransky and Gustav B. Petersen. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality. *Educ Psychol Rev*, 33(3):937–958, September 2021. ISSN 1573-336X. doi: 10.1007/s10648-020-09586-2. URL <https://doi.org/10.1007/s10648-020-09586-2>.
- [206] Patrick Manser, Hanna Poikonen, and Eling D. de Bruin. Feasibility, usability, and acceptance of “Brain-IT”—A newly developed exergame-based training concept for the secondary prevention of mild neurocognitive disorder: A pilot randomized controlled trial. *Front Aging Neurosci*, 15:1163388, September 2023. ISSN 1663-4365. doi: 10.3389/fnagi.2023.1163388. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10557950/>.
- [207] Fabrizia Mantovani, Gianluca Castelnuovo, Andrea Gaggioli, et al. Virtual Reality Training for Health-Care Professionals. *Cyberpsychol. Behav.*, 6(4):389–395, August 2003. ISSN 1094-9313. doi: 10.1089/109493103322278772. URL <https://www.liebertpub.com/doi/abs/10.1089/109493103322278772>.
- [208] Spyridon S Marinopoulos, Todd Dorman, Neda Ratanawongsa, et al. Effectiveness of continuing medical education. *Evid Rep Technol Assess (Full Rep)*, (149):1–69, January 2007. ISSN 1530-4396. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4781050/>.
- [209] Steven Marks, Lynn Shaffer, David Zehnder, et al. Under pressure: What individual characteristics lead to performance of high-quality chest compressions during CPR practice sessions? *Resuscitation Plus*, 14:100380, June 2023. ISSN 2666-5204. doi: 10.1016/j.resplu.2023.100380. URL <https://www.sciencedirect.com/science/article/pii/S2666520423000231>.

- [210] Nicolas Mascret, Lisa Delbes, Amélie Voron, et al. Acceptance of a Virtual Reality Headset Designed for Fall Prevention in Older Adults: Questionnaire Study. *J. Med. Internet Res.*, 22(12):e20691, December 2020. doi: 10.2196/20691. URL <https://www.jmir.org/2020/12/e20691>.
- [211] Fabien Mathy and Jacob Feldman. What’s magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, 122(3):346–362, March 2012. ISSN 0010-0277. doi: 10.1016/j.cognition.2011.11.003. URL <https://www.sciencedirect.com/science/article/pii/S0010027711002733>.
- [212] Richard E. Mayer. *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, July 2014. ISBN 978-1-139-99248-0.
- [213] Michael B. First M.D. *DSM-5-TR® Handbook of Differential Diagnosis*. American Psychiatric Pub, January 2024. ISBN 978-1-61537-359-8.
- [214] Esmaeil Mehraeen, Mohsen Dashti, Pegah Mirzapour, et al. Serious games in nursing education: A systematic review of current evidence. *International Journal of Africa Nursing Sciences*, 22:100838, January 2025. ISSN 2214-1391. doi: 10.1016/j.ijans.2025.100838. URL <https://www.sciencedirect.com/science/article/pii/S2214139125000253>.
- [215] Anne-Sophie Melenhorst, Wendy A. Rogers, and Don G. Bouwhuis. Older adults’ motivated choice for technological innovation: Evidence for benefit-driven selectivity. *Psychol. Aging*, 21(1):190–195, 2006. ISSN 1939-1498. doi: 10.1037/0882-7974.21.1.190.
- [216] Miguel Melo, Guilherme Gonçalves, José Vasconcelos-Raposo, et al. How Much Presence is Enough? Qualitative Scales for Interpreting the Igroup Presence Questionnaire Score. *IEEE Access*, 11:24675–24685, 2023. ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3254892. URL <https://ieeexplore.ieee.org/abstract/document/10064261>.
- [217] Aline Menin, Rafael Torchelsen, and Luciana Nedel. An Analysis of VR Technology Used in Immersive Simulations with a Serious Game Perspective. *IEEE Comput. Graph. Appl.*, 38(2):57–73, March 2018. ISSN 1558-1756. doi: 10.1109/MCG.2018.021951633.
- [218] Marvin Mergen, Marcel Meyerheim, and Norbert Graf. Reviewing the current state of virtual reality integration in medical education – a scoping review protocol. *Systematic Reviews*, 12(1):97, June 2023. ISSN 2046-4053. doi: 10.1186/s13643-023-02266-6. URL <https://doi.org/10.1186/s13643-023-02266-6>.
- [219] Marvin Mergen, Norbert Graf, and Marcel Meyerheim. Reviewing the current state of virtual reality integration in medical education - a scoping review. *BMC Med Educ*, 24(1):788, July 2024. ISSN 1472-6920. doi: 10.1186/s12909-024-05777-5. URL <https://doi.org/10.1186/s12909-024-05777-5>.
- [220] Niamh A. Merriman, Eugenie Roudaia, Jan Ondřej, et al. “CityQuest,” A Custom-Designed Serious Game, Enhances Spatial Memory Performance in Older Adults.

- Front. Aging Neurosci.*, 14, March 2022. ISSN 1663-4365. doi: 10.3389/fnagi.2022.806418. URL <https://www.frontiersin.org/articles/10.3389/fnagi.2022.806418>.
- [221] Paul Milgram and Fumio Kishino. A Taxonomy of Mixed Reality Visual Displays. *IEICE Trans. Information Systems*, vol. E77-D, no. 12:1321–1329, December 1994.
- [222] George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.*, 63(2):81–97, 1956. ISSN 1939-1471. doi: 10.1037/h0043158.
- [223] Ari Min, Haeyoung Min, and Sujeong Kim. Effectiveness of serious games in nurse education: A systematic review. *Nurse Education Today*, 108:105178, January 2022. ISSN 0260-6917. doi: 10.1016/j.nedt.2021.105178. URL <https://www.sciencedirect.com/science/article/pii/S0260691721004354>.
- [224] Alessio Mocco, Lucia Valmaggia, Lara Bernardi, et al. Enhancing Physical Activity with Immersive Virtual Reality: A Systematic Review. *Cyberpsychology Behav. Soc. Netw.*, 27(5):303–317, May 2024. ISSN 2152-2715. doi: 10.1089/cyber.2023.0394. URL <https://www.liebertpub.com/doi/abs/10.1089/cyber.2023.0394>.
- [225] Daniel R. Montello, David Waller, Mary Hegarty, et al. Spatial Memory of Real Environments, Virtual Environments, and Maps. In *Human Spatial Memory*. Psychology Press, 2003. ISBN 978-1-4106-0998-4.
- [226] Hye Sung Moon, Grady Orr, and Myounghoon Jeon. Hand Tracking with Vibrotactile Feedback Enhanced Presence, Engagement, Usability, and Performance in a Virtual Reality Rhythm Game. *International Journal of Human-Computer Interaction*, pages 1–12, June 2022. ISSN 1044-7318, 1532-7590. doi: 10.1080/10447318.2022.2087000. URL <https://www.tandfonline.com/doi/full/10.1080/10447318.2022.2087000>.
- [227] Sarah Morélot, Alain Garrigou, Julie Dedieu, et al. Virtual reality for fire safety training: Influence of immersion and sense of presence on conceptual and procedural acquisition. *Computers & Education*, 166:104145, June 2021. ISSN 0360-1315. doi: 10.1016/j.compedu.2021.104145. URL <https://www.sciencedirect.com/science/article/pii/S0360131521000221>.
- [228] Miguel Antônio Moretti, Adriana de Oliveira Camboim, Caroline Awoki Ferrandez, et al. Retention of Cardiopulmonary Resuscitation Skills in Medical Students. *Arq Bras Cardiol*, 117(5):1030–1035, November 2021. ISSN 1678-4170. doi: 10.36660/abc.20200546.
- [229] Christian Moro, James Birt, Zane Stromberga, et al. Virtual and Augmented Reality Enhancements to Medical and Science Student Physiology and Anatomy Test Performance: A Systematic Review and Meta-Analysis. *Anat. Sci. Educ.*, 14(3):368–376, 2021. ISSN 1935-9780. doi: 10.1002/ase.2049. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/ase.2049>.
- [230] Fatemeh Mortazavi, Hadi Moradi, and Abdol-Hossein Vahabie. Dynamic difficulty adjustment approaches in video games: A systematic literature review. *Multimed*

- Tools Appl*, 83(35):83227–83274, October 2024. ISSN 1573-7721. doi: 10.1007/s11042-024-18768-x. URL <https://doi.org/10.1007/s11042-024-18768-x>.
- [231] Jeanne Nakamura and Mihaly Csikszentmihalyi. The Concept of Flow. In Mihaly Csikszentmihalyi, editor, *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*, pages 239–263. Springer Netherlands, Dordrecht, 2014. ISBN 978-94-017-9088-8. doi: 10.1007/978-94-017-9088-8_16. URL https://doi.org/10.1007/978-94-017-9088-8_16.
- [232] Takuro Nakao, Kai Kunze, Megumi Isogai, et al. FingerFlex: Shape Memory Alloy-based Actuation on Fingers for Kinesthetic Haptic Feedback. In *Proc. 19th Int. Conf. Mob. Ubiquitous Multimed.*, MUM '20, pages 240–244, New York, NY, USA, November 2020. Association for Computing Machinery. ISBN 978-1-4503-8870-2. doi: 10.1145/3428361.3428404. URL <https://doi.org/10.1145/3428361.3428404>.
- [233] Andrea N. Neher, Rafael Wespi, Benjamin D. Rapphold, et al. Interprofessional Team Training With Virtual Reality: Acceptance, Learning Outcome, and Feasibility Evaluation Study. *JMIR Serious Games*, 12(1):e57117, November 2024. doi: 10.2196/57117. URL <https://games.jmir.org/2024/1/e57117>.
- [234] David L. Neumann, Robyn L. Moffitt, Patrick R. Thomas, et al. A systematic review of the application of interactive virtual reality to sport. *Virtual Reality*, 22(3):183–198, September 2018. ISSN 1434-9957. doi: 10.1007/s10055-017-0320-5. URL <https://doi.org/10.1007/s10055-017-0320-5>.
- [235] J. P. Nolan, K. G. Monsieurs, L. Bossaert, et al. European Resuscitation Council COVID-19 guidelines executive summary. *Resuscitation*, 153:45–55, August 2020. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2020.06.001. URL <https://www.sciencedirect.com/science/article/pii/S030095722030232X>.
- [236] Theresa M. Olasveengen, Federico Semeraro, Giuseppe Ristagno, et al. European Resuscitation Council Guidelines 2021: Basic Life Support. *Resuscitation*, 161:98–114, April 2021. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2021.02.009. URL <https://www.sciencedirect.com/science/article/pii/S0300957221000629>.
- [237] Marcia K. O'Malley and Abhishek Gupta. Haptic interfaces (pp. 25–74). *Morgan-Kaufman Publ. Park JH Shea CH Wright DL 2000 Reduc.-Freq. Concurr. Termin. Feedback Test Guid. Hypothesis J. Mot. Behav.*, 32(3):287–296, 2008.
- [238] Dorothea Sze Min Ong, Melvyn Zhang Weibin, and Ranganath Vallabhajosyula. Serious games as rehabilitation tools in neurological conditions: A comprehensive review. *Technol. Health Care*, 29(1):15–31, January 2021. ISSN 0928-7329. doi: 10.3233/THC-202333. URL <https://journals.sagepub.com/doi/10.3233/THC-202333>.
- [239] Daniel Paes, Javier Irizarry, and Diego Pujoni. An evidence of cognitive benefits from immersive design review: Comparing three-dimensional perception and presence between immersive and non-immersive virtual environments. *Automation in Construction*, 130:103849, October 2021. ISSN 09265805. doi:

- 10.1016/j.autcon.2021.103849. URL <https://linkinghub.elsevier.com/retrieve/pii/S0926580521003009>.
- [240] Federica Pallavicini, Alessandro Pepe, and Maria Eleonora Minissi. Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? *Simul. Gaming*, 50(2):136–159, April 2019. ISSN 1046-8781. doi: 10.1177/1046878119831420. URL <https://doi.org/10.1177/1046878119831420>.
 - [241] Federica Pallavicini, Alessandro Pepe, Massimo Clerici, et al. Virtual Reality Applications in Medicine During the COVID-19 Pandemic: Systematic Review. *JMIR Serious Games*, 10(4):e35000, October 2022. doi: 10.2196/35000. URL <https://games.jmir.org/2022/4/e35000>.
 - [242] Panagiotis D. Paraschos and Dimitrios E. Koulouriotis. Game Difficulty Adaptation and Experience Personalization: A Literature Review. *Int. J. Human-Computer Interact.*, 39(1):1–22, January 2023. ISSN 1044-7318. doi: 10.1080/10447318.2021.2020008. URL <https://doi.org/10.1080/10447318.2021.2020008>.
 - [243] Jin-Hyuck Park. Effects of virtual reality-based spatial cognitive training on hippocampal function of older adults with mild cognitive impairment. *Int. Psychogeriatr.*, 34(2):157–163, February 2022. ISSN 1041-6102, 1741-203X. doi: 10.1017/S1041610220001131. URL <https://www.cambridge.org/core/journals/international-psychogeriatrics/article/abs/effects-of-virtual-realitybased-spatial-cognitive-training-on-hippocampal-function-of-older-adults/4FFAA190ABA3FC5148BD2E748F69535B>.
 - [244] Mildred B. Parten. Social participation among pre-school children. *J. Abnorm. Soc. Psychol.*, 27(3):243, 1932. doi: 10.1037/h0074524. URL <https://psycnet.apa.org/journals/abn/27/3/243/>.
 - [245] Shiva Pedram, Grace Kennedy, and Sal Sanzone. Toward the validation of VR-HMDs for medical education: A systematic literature review. *Virtual Reality*, May 2023. ISSN 1434-9957. doi: 10.1007/s10055-023-00802-2. URL <https://doi.org/10.1007/s10055-023-00802-2>.
 - [246] Gustav Bøg Petersen, Valdemar Stenberdt, Richard E. Mayer, et al. Collaborative generative learning activities in immersive virtual reality increase learning. *Computers & Education*, 207:104931, December 2023. ISSN 0360-1315. doi: 10.1016/j.compedu.2023.104931. URL <https://www.sciencedirect.com/science/article/pii/S0360131523002087>.
 - [247] Théo Pezel, Augustin Coisne, Guillaume Bonnet, et al. Simulation-based training in cardiology: State-of-the-art review from the French Commission of Simulation Teaching (Commission d’enseignement par simulation–COMSI) of the French Society of Cardiology. *Archives of Cardiovascular Diseases*, 114(1):73–84, January 2021. ISSN 1875-2136. doi: 10.1016/j.acvd.2020.10.004. URL <https://www.sciencedirect.com/science/article/pii/S187521362030262X>.

- [248] W. A. Phillips. On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, 16(2):283–290, March 1974. ISSN 1532-5962. doi: 10.3758/BF03203943. URL <https://doi.org/10.3758/BF03203943>.
- [249] Christopher Picard, Bin Ge Yang, Colleen Norris, et al. Cardiopulmonary Resuscitation Feedback: A Comparison of Device-Measured and Self-Assessed Chest Compression Quality. *Journal of Emergency Nursing*, 47(2):333–341.e1, March 2021. ISSN 0099-1767. doi: 10.1016/j.jen.2020.10.003. URL <https://www.sciencedirect.com/science/article/pii/S0099176720303512>.
- [250] Jan L. Plass, Bruce D. Homer, and Charles K. Kinzer. Foundations of Game-Based Learning. *Educ. Psychol.*, 50(4):258–283, October 2015. ISSN 0046-1520. doi: 10.1080/00461520.2015.1122533. URL <https://www.tandfonline.com/doi/abs/10.1080/00461520.2015.1122533>.
- [251] Christian Plotzky, Ulrike Lindwedel, Michaela Sorber, et al. Virtual reality simulations in nurse education: A systematic mapping review. *Nurse Educ Today*, 101:104868, June 2021. ISSN 1532-2793. doi: 10.1016/j.nedt.2021.104868. URL <https://www.sciencedirect.com/science/article/pii/S0260691721001258>.
- [252] Albert Postma, Roy P. C. Kessels, and Marieke van Asselen. How the brain remembers and forgets where things are: The neurocognition of object–location memory. *Neuroscience & Biobehavioral Reviews*, 32(8):1339–1345, October 2008. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2008.05.001. URL <https://www.sciencedirect.com/science/article/pii/S0149763408000626>.
- [253] Jack Pottle. Virtual reality and the transformation of medical education. *Future Healthc J*, 6(3):181–185, October 2019. ISSN 2514-6645. doi: 10.7861/fhj.2019-0036. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6798020/>.
- [254] Mamtha Prajapati and Sudesh Kumar. Virtual reality revolution in healthcare: A systematic review. *Health Technol.*, 15(2):231–242, March 2025. ISSN 2190-7196. doi: 10.1007/s12553-025-00941-3. URL <https://doi.org/10.1007/s12553-025-00941-3>.
- [255] F. Gabriele Praticcò, Davide Calandra, Matteo Piviotti, et al. Assessing the User Experience of Consumer Haptic Devices for Simulation-based Virtual Reality. In *2021 IEEE 11th Int. Conf. Consum. Electron. ICCE-Berl.*, pages 1–6, November 2021. doi: 10.1109/ICCE-Berlin53567.2021.9719998.
- [256] Angelika Qvick, Manar Radif, Caroline Brever, et al. Survival of in-hospital cardiac arrest in men and women in a large Swedish cohort. *Scand J Trauma Resusc Emerg Med*, 26(1):108, December 2018. ISSN 1757-7241. doi: 10.1186/s13049-018-0576-0. URL <https://doi.org/10.1186/s13049-018-0576-0>.
- [257] Giulio Radeschi, Andrea Mina, Giacomo Berta, et al. Incidence and outcome of in-hospital cardiac arrest in Italy: A multicentre observational study in the Piedmont Region. *Resuscitation*, 119:48–55, October 2017. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2017.06.020. URL <https://www.sciencedirect.com/science/article/pii/S0300957217302630>.

- [258] Unnikrishnan Radhakrishnan, Konstantinos Koumaditis, and Francesco Chinello. A systematic review of immersive virtual reality for industrial skills training. *Behav. Inf. Technol.*, 40(12):1310–1339, September 2021. ISSN 0144-929X. doi: 10.1080/0144929X.2021.1954693. URL <https://doi.org/10.1080/0144929X.2021.1954693>.
- [259] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, et al. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, April 2020. ISSN 0360-1315. doi: 10.1016/j.compedu.2019.103778. URL <https://www.sciencedirect.com/science/article/pii/S0360131519303276>.
- [260] Rémi Rampin, Vicky Rampin, and Sarah DeMott. Taguette, the free and open-source qualitative data analysis tool, May 2023. URL <https://www.taguette.org/>.
- [261] Karan Rangarajan, Heather Davis, and Philip H. Pucher. Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool? *Journal of Surgical Education*, 77(2):337–347, March 2020. ISSN 1931-7204. doi: 10.1016/j.jsurg.2019.09.006. URL <https://www.sciencedirect.com/science/article/pii/S1931720419306609>.
- [262] Neda Ratanawongsa, Patricia A. Thomas, Spyridon S. Marinopoulos, et al. The reported validity and reliability of methods for evaluating continuing medical education: A systematic review. *Acad Med*, 83(3):274–283, March 2008. ISSN 1040-2446. doi: 10.1097/ACM.0b013e3181637925.
- [263] Ali Rezaei, Hossein Karimi, Reza Jafari, et al. Comparing virtual reality and traditional training in radiation safety practices over three years among cardiologists and scrub nurses. *JVS-Vascular Insights*, 3:100146, January 2025. ISSN 2949-9127. doi: 10.1016/j.jvsvi.2024.100146. URL <https://www.sciencedirect.com/science/article/pii/S2949912724000941>.
- [264] Philippe Robert, Alexandra König, Hélène Amieva, et al. Recommendations for the use of Serious Games in people with Alzheimer’s Disease, related disorders and frailty. *Front. Aging Neurosci.*, 6, March 2014. ISSN 1663-4365. doi: 10.3389/fnagi.2014.00054. URL <https://www.frontiersin.org><https://www.frontiersin.org/journals/aging-neuroscience/articles/10.3389/fnagi.2014.00054/full>.
- [265] G.G. Robertson, S.K. Card, and J.D. Mackinlay. Three views of virtual reality: Nonimmersive virtual reality. *Computer*, 26(2):81–, February 1993. ISSN 1558-0814. doi: 10.1109/2.192002.
- [266] Beth Rogers, Kathy A. Baker, and Ashley E. Franklin. Learning Outcomes of the Observer Role in Nursing Simulation: A Scoping Review. *Clinical Simulation in Nursing*, 49:81–89, December 2020. ISSN 1876-1399. doi: 10.1016/j.ecns.2020.06.003. URL <https://www.sciencedirect.com/science/article/pii/S1876139920300517>.

- [267] R. Le Grand Rogers, Yizza Narvaez, Arjun K. Venkatesh, et al. Improving emergency physician performance using audit and feedback: A systematic review. *The American Journal of Emergency Medicine*, 33(10):1505–1514, October 2015. ISSN 0735-6757. doi: 10.1016/j.ajem.2015.07.039. URL <https://www.sciencedirect.com/science/article/pii/S0735675715005951>.
- [268] P. Ropponen, M. Tomietto, S. Pramila-Savukoski, et al. Impacts of VR simulation on nursing students' competence, confidence, and satisfaction: A systematic review and meta-analysis of randomised controlled trials. *Nurse Education Today*, 152:106756, September 2025. ISSN 0260-6917. doi: 10.1016/j.nedt.2025.106756. URL <https://www.sciencedirect.com/science/article/pii/S0260691725001923>.
- [269] Grace Ryan, Shauna Callaghan, Anthony Rafferty, et al. Virtual reality in midwifery education: A mixed methods study to assess learning and understanding. *Nurse Education Today*, 119:105573, December 2022. ISSN 0260-6917. doi: 10.1016/j.nedt.2022.105573. URL <https://www.sciencedirect.com/science/article/pii/S0260691722003094>.
- [270] Grace V. Ryan, Shauna Callaghan, Anthony Rafferty, et al. Learning Outcomes of Immersive Technologies in Health Care Student Education: Systematic Review of the Literature. *J. Med. Internet Res.*, 24(2):e30082, February 2022. doi: 10.2196/30082. URL <https://www.jmir.org/2022/2/e30082>.
- [271] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motiv Emot*, 30(4): 344–360, December 2006. ISSN 1573-6644. doi: 10.1007/s11031-006-9051-8. URL <https://doi.org/10.1007/s11031-006-9051-8>.
- [272] Mohamad M. Saab, Josephine Hegarty, David Murphy, et al. Incorporating virtual reality in nurse education: A qualitative study of nursing students' perspectives. *Nurse Education Today*, 105:105045, October 2021. ISSN 0260-6917. doi: 10.1016/j.nedt.2021.105045. URL <https://www.sciencedirect.com/science/article/pii/S0260691721003026>.
- [273] Farveh Sabermahani, Mostafa Almasi-Dooghaee, and Abbas Sheikhtaheri. Effectiveness of Serious Games in Evaluating Cognitive Status of the Elderly: A Systematic Review And Meta-Analysis. *Games Health J.*, September 2024. ISSN 2161-783X. doi: 10.1089/g4h.2023.0106. URL <https://www.liebertpub.com/doi/abs/10.1089/g4h.2023.0106>.
- [274] Omar Sadek, Fiona Baldwin, Rebecca Gray, et al. Impact of Virtual and Augmented Reality on Quality of Medical Education During the COVID-19 Pandemic: A Systematic Review. *Journal of Graduate Medical Education*, 15(3): 328–338, June 2023. ISSN 1949-8349. doi: 10.4300/JGME-D-22-00594.1. URL <https://doi.org/10.4300/JGME-D-22-00594.1>.
- [275] Adriana Salatino, Claudio Zavattaro, Roberto Gammeri, et al. Virtual reality rehabilitation for unilateral spatial neglect: A systematic review of immersive, semi-immersive and non-immersive techniques. *Neuroscience & Biobehavioral Reviews*, 152:105248, September 2023. ISSN 0149-7634. doi: 10.1016/j.

- neubiorev.2023.105248. URL <https://www.sciencedirect.com/science/article/pii/S0149763423002178>.
- [276] Maria V. Sanchez-Vives and Mel Slater. From presence to consciousness through virtual reality. *Nat Rev Neurosci*, 6(4):332–339, April 2005. ISSN 1471-0048. doi: 10.1038/nrn1651. URL <https://www.nature.com/articles/nrn1651>.
- [277] Jonas Schild, Dieter Lerner, Sebastian Misztal, et al. EPICSAVE — Enhancing vocational training for paramedics with multi-user virtual reality. In *2018 IEEE 6th Int. Conf. Serious Games Appl. Health SeGAH*, pages 1–8, May 2018. doi: 10.1109/SeGAH.2018.8401353. URL <https://ieeexplore.ieee.org/abstract/document/8401353>.
- [278] Marc Schluep, Sanne Elisabeth Hoeks, Michiel Blans, et al. Long-term survival and health-related quality of life after in-hospital cardiac arrest. *Resuscitation*, 167: 297–306, October 2021. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2021.07.006. URL <https://www.sciencedirect.com/science/article/pii/S0300957221002550>.
- [279] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments*, 10(3):266–281, June 2001. doi: 10.1162/105474601300343603. URL <https://doi.org/10.1162/105474601300343603>.
- [280] Michal Sedlák, Čeněk Šašinka, Zdeněk Stachoň, et al. Collaborative and individual learning of geography in immersive virtual reality: An effectiveness study. *PLOS ONE*, 17(10):e0276267, October 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0276267. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0276267>.
- [281] Federico Semeraro, Antonio Frisoli, Claudio Loconsole, et al. Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality training: A randomised crossover mannequin pilot study. *Resuscitation*, 84(4):501–507, April 2013. ISSN 0300-9572. doi: 10.1016/j.resuscitation.2012.12.006. URL <https://www.sciencedirect.com/science/article/pii/S0300957212009410>.
- [282] Nahid Sharifzadeh, Hadi Kharrazi, Elham Nazari, et al. Health Education Serious Games Targeting Health Care Providers, Patients, and Public Health Users: Scoping Review. *JMIR Serious Games*, 8(1):e13459, March 2020. doi: 10.2196/13459. URL <https://games.jmir.org/2020/1/e13459>.
- [283] M. Jeanne Sholl and Tina L. Nolin. Orientation specificity in representations of place. *J. Exp. Psychol. Learn. Mem. Cogn.*, 23(6):1494–1507, 1997. ISSN 1939-1285. doi: 10.1037/0278-7393.23.6.1494.
- [284] Shefaly Shorey and Esperanza Debby Ng. The use of virtual reality simulation among nursing students and registered nurses: A systematic review. *Nurse Education Today*, 98:104662, March 2021. ISSN 0260-6917. doi: 10.1016/j.nedt.2020.104662. URL <https://www.sciencedirect.com/science/article/pii/S0260691720315124>.

- [285] Roland Sigrist, Georg Rauter, Robert Riener, et al. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychon Bull Rev*, 20(1):21–53, February 2013. ISSN 1531-5320. doi: 10.3758/s13423-012-0333-8. URL <https://doi.org/10.3758/s13423-012-0333-8>.
- [286] Samfriati Sinurat, Mestiana Br. Karo, Romida Simbolon, et al. The efficacy of virtual-reality teaching and learning method in enhancing interprofessional knowledge, clinical skill performance, and self-confidence in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 149: 106661, June 2025. ISSN 0260-6917. doi: 10.1016/j.nedt.2025.106661. URL <https://www.sciencedirect.com/science/article/pii/S0260691725000966>.
- [287] Lucas Siqueira Rodrigues, Timo Torsten Schmidt, John Nyakatura, et al. Assessing the Effects of Sensory Modality Conditions on Object Retention across Virtual Reality and Projected Surface Display Environments. *Proc. ACM Hum.-Comput. Interact.*, 8(ISS):537:255–537:282, October 2024. doi: 10.1145/3698137. URL <https://dl.acm.org/doi/10.1145/3698137>.
- [288] Ka-Chun Siu, Bradley J. Best, Jong Wook Kim, et al. Adaptive Virtual Reality Training to Optimize Military Medical Skills Acquisition and Retention. *Mil Med*, 181(suppl_5):214–220, May 2016. ISSN 0026-4075. doi: 10.7205/MILMED-D-15-00164. URL <https://doi.org/10.7205/MILMED-D-15-00164>.
- [289] Ioannis Skolidis, Olivier Muller, and Stephane Fournier. CardioVerse: The cardiovascular medicine in the era of Metaverse. *Trends in Cardiovascular Medicine*, May 2022. ISSN 1050-1738. doi: 10.1016/j.tcm.2022.05.004. URL <https://www.sciencedirect.com/science/article/pii/S1050173822000718>.
- [290] Mel Slater and Maria V. Sanchez-Vives. Enhancing Our Lives with Immersive Virtual Reality. *Front. Robot. AI*, 3, December 2016. ISSN 2296-9144. doi: 10.3389/frobt.2016.00074. URL <https://www.frontiersin.org/articles/10.3389/frobt.2016.00074>.
- [291] Sheryl Sorby, Norma Veurink, and Scott Streiner. Does spatial skills instruction improve STEM outcomes? The answer is ‘yes’. *Learning and Individual Differences*, 67:209–222, October 2018. ISSN 1041-6080. doi: 10.1016/j.lindif.2018.09.001. URL <https://www.sciencedirect.com/science/article/pii/S1041608018301456>.
- [292] Carlos Henrique R. Souza, Daniela F. Nascimento, Luciana O. Berretta, et al. Solutions for Dynamic Difficulty Adjustment in digital games: A Systematic Literature Review. *Entertainment Computing*, 55:101041, September 2025. ISSN 1875-9521. doi: 10.1016/j.entcom.2025.101041. URL <https://www.sciencedirect.com/science/article/pii/S1875952125001211>.
- [293] Mathilde Stærk, Lauge Vammen, Camilla Fuchs Andersen, et al. Basic life support skills can be improved among certified basic life support instructors. *Resuscitation Plus*, 6:100120, June 2021. ISSN 2666-5204. doi: 10.1016/j.resplu.2021.100120. URL <https://www.sciencedirect.com/science/article/pii/S266652042100045X>.

- [294] Emma Stanmore, Brendon Stubbs, Davy Vancampfort, et al. The effect of active video games on cognitive functioning in clinical and non-clinical populations: A meta-analysis of randomized controlled trials. *Neuroscience & Biobehavioral Reviews*, 78:34–43, July 2017. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2017.04.011. URL <https://www.sciencedirect.com/science/article/pii/S014976341730129X>.
- [295] Hans Stefan, Michael Mortimer, and Ben Horan. Evaluating the effectiveness of virtual reality for safety-relevant training: A systematic review. *Virtual Reality*, 27(4):2839–2869, December 2023. ISSN 1359-4338, 1434-9957. doi: 10.1007/s10055-023-00843-7. URL <https://link.springer.com/10.1007/s10055-023-00843-7>.
- [296] Gijsbert Stoet. PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4):1096–1104, November 2010. ISSN 1554-3528. doi: 10.3758/BRM.42.4.1096. URL <https://doi.org/10.3758/BRM.42.4.1096>.
- [297] Gijsbert Stoet. PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teach. Psychol.*, 44(1):24–31, January 2017. ISSN 0098-6283. doi: 10.1177/0098628316677643. URL <https://doi.org/10.1177/0098628316677643>.
- [298] Ewa Szczepocka, Łukasz Mokros, Jakub Kaźmierski, et al. Virtual reality-based training may improve visual memory and some aspects of sustained attention among healthy older adults – preliminary results of a randomized controlled study. *BMC Psychiatry*, 24(1):347, May 2024. ISSN 1471-244X. doi: 10.1186/s12888-024-05811-2. URL <https://doi.org/10.1186/s12888-024-05811-2>.
- [299] Tahere Talebi Azadboni, Somayeh Nasiri, Sahar Khenarinezhad, et al. Effectiveness of serious games in social skills training to autistic individuals: A systematic review. *Neuroscience & Biobehavioral Reviews*, 161:105634, June 2024. ISSN 01497634. doi: 10.1016/j.neubiorev.2024.105634. URL <https://linkinghub.elsevier.com/retrieve/pii/S0149763424001039>.
- [300] Corey E. Tatel and Phillip L. Ackerman. Procedural skill retention and decay: A meta-analytic review. *Psychol. Bull.*, 151(6):696–736, 2025. ISSN 1939-1455. doi: 10.1037/bul0000481.
- [301] R Core Team. R: A Language and Environment for Statistical Computing, 2023. URL <https://www.R-project.org/>.
- [302] Talia Tene, Diego Fabián Vique López, Paulina Elizabeth Valverde Aguirre, et al. Virtual reality and augmented reality in medical education: An umbrella review. *Front. Digit. Health*, 6, March 2024. ISSN 2673-253X. doi: 10.3389/fdgth.2024.1365345. URL <https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2024.1365345/full>.
- [303] Peter J Toghill. Continuing Medical Education for Physicians. *J R Coll Physicians Lond*, 28(2):155–156, 1994. ISSN 0035-8819. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5400910/>.

- [304] Tiffany Tong, Mark Chignell, Mary C. Tierney, et al. A Serious Game for Clinical Assessment of Cognitive Status: Validation Study. *JMIR Serious Games*, 4(1): e5006, May 2016. doi: 10.2196/games.5006. URL <https://games.jmir.org/2016/1/e7>.
- [305] Esmaeel Toni, Elham Toni, Mahsa Fereidooni, et al. Acceptance and use of extended reality in surgical training: An umbrella review. *Syst Rev*, 13(1): 299, December 2024. ISSN 2046-4053. doi: 10.1186/s13643-024-02723-w. URL <https://doi.org/10.1186/s13643-024-02723-w>.
- [306] Roberto Trevi, Stefania Chiappinotto, Alvisa Palese, et al. Virtual Reality for Cardiopulmonary Resuscitation Healthcare Professionals Training: A Systematic Review. *J Med Syst*, 48(1):50, May 2024. ISSN 1573-689X. doi: 10.1007/s10916-024-02063-1. URL <https://doi.org/10.1007/s10916-024-02063-1>.
- [307] Zeynep Turan and Sinem Cilligol Karabey. The use of immersive technologies in distance education: A systematic review. *Educ Inf Technol*, May 2023. ISSN 1573-7608. doi: 10.1007/s10639-023-11849-8. URL <https://doi.org/10.1007/s10639-023-11849-8>.
- [308] Mohib Ullah, Sareer Ul Amin, Muhammad Munsif, et al. Serious games in science education: A systematic literature. *Virtual Reality & Intelligent Hardware*, 4 (3):189–209, June 2022. ISSN 2096-5796. doi: 10.1016/j.vrih.2022.02.001. URL <https://www.sciencedirect.com/science/article/pii/S2096579622000201>.
- [309] David H. Uttal, David I. Miller, and Nora S. Newcombe. Exploring and Enhancing Spatial Thinking: Links to Achievement in Science, Technology, Engineering, and Mathematics? *Curr Dir Psychol Sci*, 22(5):367–373, October 2013. ISSN 0963-7214. doi: 10.1177/0963721413484756. URL <https://doi.org/10.1177/0963721413484756>.
- [310] Vanessa Vallejo, Patric Wyss, Luca Rampa, et al. Evaluation of a novel Serious Game based assessment tool for patients with Alzheimer’s disease. *PLOS ONE*, 12(5):e0175999, May 2017. ISSN 1932-6203. doi: 10.1371/journal.pone.0175999. URL <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0175999>.
- [311] Nesse van der Meer, Vivian van der Werf, Willem-Paul Brinkman, et al. Virtual reality and collaborative learning: A systematic literature review. *Front. Virtual Real.*, 4, May 2023. ISSN 2673-4192. doi: 10.3389/frvir.2023.1159905. URL <https://www.frontiersin.org/journals/virtual-reality/articles/10.3389/frvir.2023.1159905/full>.
- [312] Sebastian Vandermolen, Fabrizio Ricci, C Anwar A Chahal, et al. ‘The Digital Cardiologist’: How Technology Is Changing the Paradigm of Cardiology Training. *Current Problems in Cardiology*, 47(12):101394, December 2022. ISSN 0146-2806. doi: 10.1016/j.cpcardiol.2022.101394. URL <https://www.sciencedirect.com/science/article/pii/S0146280622002912>.
- [313] Sofia Barlocco De La Vega, Evelyne Guerif-Dubreucq, Jebrane Bouaoud, et al. Mixed reality versus manikins in basic life support simulation-based training for

- medical students in France: The mixed reality non-inferiority randomized controlled trial. *JEEHP*, 22, May 2025. ISSN 1975-5937. doi: 10.3352/jeehp.2025.22.15. URL http://www.jeehp.org/journal/view.php?number=561&utm_source=chatgpt.com.
- [314] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, et al. User Acceptance of Information Technology: Toward A Unified View1. *MIS Q.*, 27(3):425–478, September 2003. ISSN 0276-7783, 2162-9730. doi: 10.2307/30036540. URL <https://misq.umn.edu/misq/article/27/3/425/1340/User-Acceptance-of-Information-Technology-Toward-A>.
- [315] Viswanath Venkatesh, James Y. L. Thong, and Xin Xu. Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology1. *MIS Q.*, 36(1):157–178, March 2012. ISSN 0276-7783, 2162-9730. doi: 10.2307/41410412. URL <https://misq.umn.edu/misq/article/36/1/157/1470/Consumer-Acceptance-and-Use-of-Information>.
- [316] Sara Ventura, Eleonora Brivio, Giuseppe Riva, et al. Immersive Versus Non-immersive Experience: Exploring the Feasibility of Memory Assessment Through 360° Technology. *Front. Psychol.*, 10, 2019. ISSN 1664-1078. doi: 10.3389/fpsyg.2019.02509. URL <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02509>.
- [317] Brian N. Verdine, Casey M. Irwin, Roberta Michnick Golinkoff, et al. Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology*, 126:37–51, October 2014. ISSN 0022-0965. doi: 10.1016/j.jecp.2014.02.012. URL <https://www.sciencedirect.com/science/article/pii/S0022096514000484>.
- [318] Brian N. Verdine, Roberta Michnick Golinkoff, Kathy Hirsh-Pasek, et al. I. Spatial Skills, Their Development, and Their Links to Mathematics. *Monogr. Soc. Res. Child Dev.*, 82(1):7–30, 2017. ISSN 1540-5834. doi: 10.1111/mono.12280. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/mono.12280>.
- [319] Paul Vogt, Rik Boer, Marieke de Boer, et al. Designing and Evaluating a Virtual Reality Training for Paramedics to Practice Triage in Complex Situations. In Jessie Y. C. Chen and Gino Fragomeni, editors, *Virtual Augment. Mix. Real.*, pages 503–522, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-35634-6. doi: 10.1007/978-3-031-35634-6_36.
- [320] Qisong Wang, Bo Kang, and Per Ola Kristensson. Supporting Playful Rehabilitation in the Home using Virtual Reality Headsets and Force Feedback Gloves. In *2022 IEEE Conf. Virtual Real. 3D User Interfaces VR*, pages 504–513, March 2022. doi: 10.1109/VR51125.2022.00070.
- [321] Sara Wattenbarger, Annemarie Silver, Tifany Hoyne, et al. Real-Time Cardiopulmonary Resuscitation Feedback and Targeted Training Improve Chest Compression Performance in a Cohort of International Healthcare Providers. *The Journal of Emergency Medicine*, 58(1):93–99, January 2020. ISSN 0736-4679. doi: 10.1016/j.jemermed.2019.09.027. URL <https://www.sciencedirect.com/science/article/pii/S0736467919308108>.

- [322] David P. Way, Ashish R. Panchal, Alan Price, et al. Learner evaluation of an immersive virtual reality mass casualty incident simulator for triage training. *BMC Digit Health*, 2(1):56, September 2024. ISSN 2731-684X. doi: 10.1186/s44247-024-00117-5. URL <https://doi.org/10.1186/s44247-024-00117-5>.
- [323] Séamas Weech, Sophie Kenny, and Michael Barnett-Cowan. Presence and Cybersickness in Virtual Reality Are Negatively Related: A Review. *Front. Psychol.*, 10, 2019. ISSN 1664-1078. doi: 10.3389/fpsyg.2019.00158. URL <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00158>.
- [324] Ria Wolkorte, Dorothea J. Heersema, and Inge Zijdwind. Reduced Dual-Task Performance in MS Patients Is Further Decreased by Muscle Fatigue. *Neurorehabil Neural Repair*, 29(5):424–435, June 2015. ISSN 1545-9683. doi: 10.1177/1545968314552529. URL <https://doi.org/10.1177/1545968314552529>.
- [325] Adele Pei Ning Woon, Wen Qi Mok, Ying Jia Shermin Chieng, et al. Effectiveness of virtual reality training in improving knowledge among nursing students: A systematic review, meta-analysis and meta-regression. *Nurse Education Today*, 98:104655, March 2021. ISSN 0260-6917. doi: 10.1016/j.nedt.2020.104655. URL <https://www.sciencedirect.com/science/article/pii/S0260691720315057>.
- [326] World Health Organization. *Handbook for National Quality Policy and Strategy: A Practical Approach for Developing Policy and Strategy to Improve Quality of Care*. World Health Organization, Geneva, 2018. ISBN 978-92-4-156556-1. URL <https://apps.who.int/iris/handle/10665/272357>.
- [327] Bian Wu, Xiaoxue Yu, and Xiaoqing Gu. Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *Br. J. Educ. Technol.*, 51(6):1991–2005, 2020. ISSN 1467-8535. doi: 10.1111/bjet.13023. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13023>.
- [328] Myra H. Wyckoff, Robert Greif, Peter T. Morley, et al. 2022 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations: Summary From the Basic Life Support; Advanced Life Support; Pediatric Life Support; Neonatal Life Support; Education, Implementation, and Teams; and First Aid Task Forces. *Circulation*, 146(25):e483–e557, December 2022. doi: 10.1161/CIR.0000000000001095. URL <https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000001095>.
- [329] G. Wynne, T. Marteau, and T. R. Evans. Instructors—a weak link in resuscitation training. *J R Coll Physicians Lond*, 26(4):372–373, October 1992. ISSN 0035-8819.
- [330] Mengxin Xu, Maria Murcia-Lopez, and Anthony Steed. Object location memory error in virtual and real environments. In *2017 IEEE Virtual Real. VR*, pages 315–316, Los Angeles, CA, USA, 2017. IEEE. ISBN 978-1-5090-6647-6. doi: 10.1109/VR.2017.7892303. URL <http://ieeexplore.ieee.org/document/7892303/>.
- [331] Chih-Wei Yang, Zui-Shen Yen, Jane E. McGowan, et al. A systematic review of retention of adult advanced life support knowledge and skills in healthcare providers. *Resuscitation*, 83(9):1055–1060, September 2012. ISSN 0300-9572, 1873-1570. doi:

- 10.1016/j.resuscitation.2012.02.027. URL [https://www.resuscitationjournal.com/article/S0300-9572\(12\)00125-6/fulltext?utm_source=chatgpt.com](https://www.resuscitationjournal.com/article/S0300-9572(12)00125-6/fulltext?utm_source=chatgpt.com).
- [332] Sun-Yi Yang. Multi-user interactive neonatal resuscitation program using hand tracking in immersive virtual reality: A quasi-experimental design. *Nurse Education Today*, 153:106793, October 2025. ISSN 0260-6917. doi: 10.1016/j.nedt.2025.106793. URL <https://www.sciencedirect.com/science/article/pii/S0260691725002291>.
- [333] Mohsen Zahiri, Carl A. Nelson, Dmitry Oleynikov, et al. Evaluation of Augmented Reality Feedback in Surgical Training Environment. *Surg Innov*, 25(1):81–87, February 2018. ISSN 1553-3506. doi: 10.1177/1553350617739425. URL <https://doi.org/10.1177/1553350617739425>.
- [334] Olivia Zechner, Helmut Schrom-Feiertag, Rafael Wespi, et al. Enhancing Mixed Reality Simulation Training Technology With Real-Time Performance Visualization: Mixed Methods Study With Medical First Responders. *JMIR XR Spat. Comput. JMXR*, 1(1):e57655, December 2024. doi: 10.2196/57655. URL <https://xr.jmir.org/2024/1/e57655>.
- [335] Dandan Zhang, Hongwu Liao, Yitong Jia, et al. Effect of virtual reality simulation training on the response capability of public health emergency reserve nurses in China: A quasiexperimental study. *BMJ Open*, 11(9):e048611, September 2021. ISSN 2044-6055, 2044-6055. doi: 10.1136/bmjopen-2021-048611. URL <https://bmjopen.bmj.com/content/11/9/e048611>.
- [336] Dandan Zhang, MuLi Fu, Jianzhong Zhang, et al. Evaluating Whether Nonimmersion Virtual Reality Simulation Training Improves Nursing Competency in Isolation Wards: Randomized Controlled Trial. *J. Med. Internet Res.*, 27(1):e63131, January 2025. doi: 10.2196/63131. URL <https://www.jmir.org/2025/1/e63131>.
- [337] Xiao Zhang, Tuire Koponen, Pekka Räsänen, et al. Linguistic and Spatial Skills Predict Early Arithmetic Development via Counting Sequence Knowledge. *Child Dev.*, 85(3):1091–1107, 2014. ISSN 1467-8624. doi: 10.1111/cdev.12173. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/cdev.12173>.
- [338] Guanjie Zhao, Minjie Fan, Yibiao Yuan, et al. The comparison of teaching efficiency between virtual reality and traditional education in medical education: A systematic review and meta-analysis. *Ann Transl Med*, 9(3):252, February 2021. ISSN 2305-5839. doi: 10.21037/atm-20-2785. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7940910/>.
- [339] J.M. Zheng, K.W. Chan, and I. Gibson. Virtual reality. *IEEE Potentials*, 17(2): 20–23, April 1998. ISSN 1558-1772. doi: 10.1109/45.666641.
- [340] Kathrin Zimmermann, Claudia C. von Bastian, Christina Röcke, et al. Transfer after process-based object-location memory training in healthy older adults. *Psychol. Aging*, 31:798–814, 2016. ISSN 1939-1498. doi: 10.1037/pag0000123.

- [341] Peter Zimmermann and Bruno Fimm. A test battery for attentional performance. In *Applied Neuropsychology of Attention*. Psychology Press, 2002. ISBN 978-0-203-30701-4.
- [342] Xinyi Zuo, Yong Tang, Yifang Chen, et al. Effects of Electronic Serious Games on Older Adults With Alzheimer’s Disease and Mild Cognitive Impairment: Systematic Review With Meta-Analysis of Randomized Controlled Trials. *JMIR Serious Games*, 12(1):e55785, July 2024. doi: 10.2196/55785. URL <https://games.jmir.org/2024/1/e55785>.
- [343] M. Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9): 25–32, September 2005. ISSN 1558-0814. doi: 10.1109/MC.2005.297. URL <https://ieeexplore.ieee.org/abstract/document/1510565>.