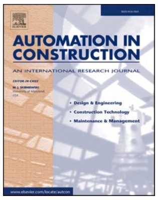
Automation in Construction 162 (2024) 105371

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon



Optical see-through augmented reality fire safety training for building occupants

Daniel Paes [a,*], Zhenan Feng [a], Maddy King [b], Hesam Khorrami Shad [a], Prasanth Sasikumar [c], Diego Pujoni [d], Ruggiero Lovreglio [a]

[a] School of Built Environment, Massey University, Auckland, New Zealand
[b] School of Music and Creative Media Production, Massey University, Wellington, New Zealand
[c] Bioengineering Institute, University of Auckland, Auckland, New Zealand
[d] Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, Brazil

ARTICLE INFO

Keywords:
Augmented reality
Fire safety
Emergency evacuation
Training
Human-computer interaction
Human factors

ABSTRACT

Fire safety training is crucial to increase building occupants' chances of surviving a fire emergency. Traditional training methods, such as lectures using video presentations, have limitations that can hinder learning performance. This article describes the development and testing of an alternative training solution using Augmented Reality (AR) technology. Through a controlled between-subject experiment, the AR-based fire safety training method was compared against a conventional video-based one based on participants' knowledge acquisition and retention, intrinsic motivation, and self-efficacy. Results suggest that the AR system was well-designed and as effective as the traditional method in terms of knowledge acquisition and retention and overall learning experience. However, it was found to be superior in terms of intrinsic motivation gain and self-efficacy retention. These findings demonstrate the potential of AR-based training methods to enhance building occupants' safety and provide directions for future developments and research in the field.

1. Introduction

Over the last decades, fire safety engineering has provided new design solutions to reduce the occurrence of fires and mitigate their impact on human lives and properties. However, the statistics highlight that fire emergencies remain a real threat to people and buildings [1]. Therefore, it is imperative to train building occupants on what to do during these emergencies [2]. Various methods have been adopted to train building occupants. These differ depending on building typology and the type of fire safety challenges faced in an emergency [3,4]. Among them, one of the commonly used methods adopted to train building occupants is evacuation drills [2].

Fire evacuation drills are mandatory in many countries. They are designed to test the evacuation performance of a building and the fire safety system in place and train building occupants and staff on what to do in the event of a fire [5,2]. However, there are several limitations to the use of traditional fire drills. For instance, holding drills repeatedly can have a diminishing impact on the response of building occupants as they might start associating notification systems with a drill rather than a real incident [6]. Further, traditional drills are often seen as a "tick-box exercise" rather than an opportunity to train people and gather safety-critical information [7]. However, one of their main limitations is pedagogical, as it is difficult to ensure they can provide effective training to all the trainees involved [2]. In fact, participants of drills often receive no feedback whatsoever to help them assess their evacuation choices retrospectively.

In the last decade, digital technologies have provided alternatives to overcome many of the existing challenges posed by traditional methods for fire safety training. Among them are Extended Reality (XR) solutions. XR is an umbrella term encompassing Virtual Reality (VR) and Augmented Reality (AR) technologies. Multiple VR-based solutions have been proposed in the literature and tested for different building typologies and fire safety guidelines [8,9,10]. Recent studies have shown that VR-based safety training can be more effective than traditional training in terms of knowledge acquisition and retention [11,4]. However, developing VR-based training solutions is time-consuming and labour-intensive. VR requires the development of an entire virtual world – it is not in any way grounded in the physical surroundings of the user. As

---

* Corresponding author.
E-mail addresses: d.paes@massey.ac.nz (D. Paes), z.feng1@massey.ac.nz (Z. Feng), m.king1@massey.ac.nz (M. King), hesam.khorramishad@massey.ac.nz (H. Khorrami Shad), psas598@aucklanduni.ac.nz (P. Sasikumar), diegopujoni@ufmg.br (D. Pujoni), r.lovreglio@massey.ac.nz (R. Lovreglio).

https://doi.org/10.1016/j.autcon.2024.105371
Received 3 May 2023; Received in revised form 3 March 2024; Accepted 6 March 2024

such, to be useful for location-based fire safety training, it requires the creation of a digital replica of the user's building, as well as the mitigation of motion sickness effects inherent to the technology [12].

Augmented Reality (AR) has been proposed as a method that could help overcome many of the limitations of VR training systems [13]. While VR has been utilised in the past, it falls short in providing a seamless integration with real-world environments. AR, on the other hand, offers a unique solution by overlaying digital information onto the physical surroundings, creating a contextually relevant training experience. AR is defined as an advanced visualisation technology that incorporates digital elements – 3D models, animated characters, sound, text, *etc.* – into views of the real world as mediated through a device's camera or other data-capturing instrument such as built-in infrared scanners. The digital elements augment and enhance the real world by adding an extra layer of information [14,15], resulting in an *augmented experience of reality*. This extra information can assist the user in performing tasks.

Research institutions and industry professionals have been investigating the advantages and challenges involved in utilising AR technology across a range of disciplines. Due to its ability to merge physical and virtual elements, AR technology has been primarily leveraged as a training or task performance aid tool. Butaslac et al. [15] note that the benefits of AR-based training include the fact that it also "offers the trainee the opportunity to do rehearsals and practice the process repeatedly". However, "conventional workforce education and upskilling do not take much advantage of applying emerging technologies to boost safety performance" [16]. Indeed, the use of AR technology in fire safety is quite rare, with only a handful of applications designed for fire evacuation preparedness (training) and response (assistance).

Testing the effectiveness of new AR solutions is of utmost importance before encouraging users and society to invest in them. The literature provides some insights regarding AR technology; certain system attributes and user characteristics appear to play a key role in determining its effectiveness. Although findings are still in the early stages [17], they consistently suggest that AR could be leveraged to create fire safety training tools that are cost-effective and relatively easy to develop and customise.

This study's motivation, therefore, stems from the need identified in the literature to further advance fire safety training methods and assess their effectiveness from the user standpoint. To date, there have been very few reports on the effectiveness of AR-based training tools, likely due to the newness of the topic. Researchers have yet to determine whether and how AR systems can be leveraged to develop more effective training methods. This topic requires further exploration through user-centred experimental research.

Thus, the objective of this research is to prototype an AR-based fire safety training system and test whether and to what extent it improves trainees' learning performance compared to traditional training regarding knowledge acquisition and retention, intrinsic motivation (performing an activity for the inherent satisfaction with the activity itself), and self-efficacy (one's perceived ability to learn or perform actions). More specifically, the experiment aims to find out whether and how changes in a single independent variable (training mode) of two different values, namely, traditional training (instructional video on a laptop) and AR training (guided mission on a head-mounted display), induce changes in the user-related dependent variables, namely, knowledge, intrinsic motivation, and self-efficacy. Measures of dependent variables in each experimental condition at different periods are used to perform a direct comparison between training modes. Additional technology-related measures will also be collected and analysed.

Ultimately, this research aims to contribute to the body of knowledge on the suitability and effectiveness of AR technology as a training tool in the fire safety domain and provide directions for developing and implementing increasingly effective systems to enhance building occupants' safety.

2. Research background

AR technology has been demonstrated as a promising training tool in many fields. This section provides the theoretical background and rationale for the present study through an overview of recent developments in the AR-assisted training and learning space.

Since late 2014, advances in computer processing power, display and sensing technologies have enabled the development and spread of affordable off-the-shelf XR solutions. These have different setups and specifications but generally encompass a display, built-in infrared sensors or cameras, body movement and gesture tracking sensors, interaction and navigation devices, and software applications to stitch all these components together. These components can vary largely, especially regarding the display type [18]. In AR, the displays can be optical-see-through (such as the Microsoft HoloLens) or video-see-through (smartphones and tablets), handheld or head-mounted. It should be noted that modern AR headsets (optical see-through AR) are equipped with built-in infrared sensors that can capture spatial data to construct and dynamically update a 3D map of an unknown environment – a technique named Simultaneous Localisation And Mapping (SLAM). The AR software application then uses this 3D map as a reference to place augmentations (the digital objects) in the correct location, orientation, and scale onto the live views of the real environment through the transparent display. In turn, the technique used in tablet- and smartphone-based AR systems (video see-through AR) to build those dynamic spatial maps is similar but uses the video cameras of such devices in a process technically known as photogrammetry. In video see-through AR, the augmentations are shown on the tablet/smartphone display.

The literature suggests that AR technology may provide increased learning performance compared to traditional teaching methods due to its ability to engage learners, leading to increased knowledge acquisition and retention. Kamal et al. [19], for example, found that an AR app "could hold trainees' attention, establish the relevance of materials, boost trainee confidence, and make their learning achievements more satisfactory". To examine the effectiveness of AR on learning, Bork et al. [20] used multiplayer AR to teach anatomy and found that it stimulated collaborative anatomy learning. These show that there are many indices on which to measure learning efficacy, but studies on AR in learning seem to consistently find that its ability to engage plays a key role. In a study of the ability of AR to engage, Yang et al. [21] found that AR increased user curiosity and attention towards AR advertisements. While Chen et al. [22] found that AR captions did not improve the learning performance of English-speaking learners, they did observe that the use of AR appeared to increase learner motivation. They note that "curiosity and desire to know more about the AR interactions may have increased their achievement goal motivation". Both studies noted that the increased engagement of learners was related to the novelty of the AR experience, which had a high probability of wearing off as users became more familiar with the technology.

In terms of sheer outcomes-based learner performance, Chang et al. [23] used AR to teach interior design and found that the learning performance of students who used AR improved significantly over a control group. Wahyu et al. [24] found that the use of AR in STEM education improved scientific literacy and student learning achievement in comparison to a control group. Chen [25] studied improvements in piano performance using AR. Test scores for students who used AR training improved 25% more than a control group who followed the traditional curriculum. These evaluations across different fields consistently found that AR-assisted learning outperformed traditional methods [15].

Many studies have identified the opportunity to use AR for training purposes, with the aim of achieving increased learning effectiveness while also providing cost savings. Butaslac et al. [15] note additional advantages: "it can provide a risk-free environment that can be considered to be very risky if it is performed in a real-life environment". This has distinct advantages for safety training in particular. In fact, the use of AR for safety training has been analysed in various domains, including in the built environment and construction [26]. Preliminary studies were limited to the development of frameworks and proofs-of-concept. For example, Hou et al. [16] developed and tested a framework of a system integrating BIM (Building Information Modelling), VR, and AR to support the learning of complex procedural skills and concluded that "it has great potential to improve the identification of safety hazards". More recently, Kim et al. [27] found that students preferred a prototype AR tool for training because it provided a "realistic visual context" for real-world hazards. Kamal et al. [19] used AR for occupational health and safety training and found an improvement in participants' test scores. Tatić and Tešić [28] created an AR system for industrial training that was machine-specific and intended to reduce accidents by enforcing the following of safe sequential instructions while breaking up the monotony of certain tasks. In each case, AR was proposed as a method for applying additional context, leading to better outcomes for participants.

AR has also been tested and found useful in emergency evacuation performance and training. There have been a few attempts when it comes to AR-assisted emergency evacuation. For example, Ahn and Han [29] used AR to provide walking routes for indoor evacuation. Their RescueMe system enabled participants to escape in 75% less time than the control group and was found to be highly effective for evacuation in a crowd. Tsai and Yau [30] used AR to provide instant escape routes to people in the event of a radioactive accident. The tool shaved 30 s off participants' escape time, compared to a previous system that pointed participants to the nearest shelter [31]. These early experiments showed the great potential of AR to assist in efficient evacuation. A more recent prototype application by Kanangkaew et al. [32] combines BIM and marker-based video see-through AR to offer interactive real-time guidance information to building occupants so they can find the best evacuation route during a fire. The system displays a person's location within the building, the exit location, the shortest exit route, and animations for evacuation guidance. Yoo and Choi [33] developed a prototype of a smart video see-through AR navigation and emergency evacuation system to aid building occupants in identifying the optimal escape path during a fire. This innovative system combines Internet of Things (IoT) and Machine Learning (ML) technologies, allowing it to capture sensing data, detect emergencies, and predict disaster area, propagation, and occupants' location to derive the best evacuation path for each individual. Chen et al. [34] also developed and tested a system combining BIM, IoT, VR and optical see-through AR to improve building fire safety. They found that AR can improve firefighting and fire evacuation efficiency by providing users with timely and intuitive wayfinding instructions at the fire scene. However, they suggest that further research is still needed to establish the effectiveness of AR tools in improving user situational awareness and task performance in general.

A few other studies have proposed AR evacuation training systems. Iguchi et al. [35] developed an AR system that lets teachers practice issuing instructions to virtual children in a simulated disaster situation. Overall, teachers found the tool useful and acceptable. Kawai et al. [36] created a game-based AR evacuation system in which participants had to make decisions during a disaster. Results showed that participants felt the tool was superior to conventional evacuation training. More recently, Catal et al. [37] built a mobile application using video see-through AR to provide users with building evacuation training, teaching them how to locate and reach the nearest exit in the event of a fire. Their application includes various scenarios and features to guide and train users in different situations and was developed utilising 3D modelling and video game development software. They found their system to be "very effective" for evacuation drills and proposed that "this kind of system can be preferred instead of theoretical training sessions".

However, the effectiveness of AR training systems over traditional training methods requires further research. In a meta-analysis of 25 studies that used XR tools for training enhancement, Kaplan et al. [38] found no significant improvement in the outcomes of AR training over traditional training techniques (*i.e.*, in classrooms). Determining the effectiveness of AR training systems is complex, as studies use different means and measures of evaluation. Face validity, *i.e.*, the degree to which an AR application "resembles the real working situation", is the aspect most often evaluated by AR training studies [39]. "Face evaluation is performed to assess the degree of resemblance between training with the system and the educational construct with a questionnaire or a small interview" [15]. However, as Suresh et al. [40] point out, "face validity is recognised as being a poor measure due to the subjective nature of 'perceived realism' and lack of quantifiability, thus diminishing the overall impact of results". Conversely, predictive evaluation "reflects the training effects gained from the system to the actual practice itself, which proves skills acquisition" [15]. Nonetheless, predictive evaluations are still limited in this field due to the high cost of resources and the potential danger of real-life scenarios. This points to the need for further comparative control group-based research into the effectiveness of AR training systems.

The literature provides evidence that game-based and interactive elements that create more realistic scenarios, such as NPCs (non-player characters), could improve AR training effectiveness. For example, Mitsuhara et al. [41] proposed an AR head-mounted evacuation training system that could simulate the urgency of a real disaster by showing computer-generated characters running to escape. In line with Schmidt's [42] schema theory that proposes that the body unconsciously updates its learned movements in response to observing the movements of others, incorporating the observation of evacuees into AR training could be an effective means of teaching the body how to move in an evacuation.

3. Research method

The research method is based on previous studies – described in detail in Sections 2 and 3.3 – on developing and testing XR-based training tools, particularly those experiment-based ones that have developed their own prototypes, conducted user-centred prototype testing, and included a comparison between conditions (with *versus* without XR technology support).

The method involves developing and testing an AR-based fire safety training system designed to instruct building occupants on how to respond in the event of a fire. This includes performing specific tasks such as identifying and isolating the fire, locating and attempting to activate the fire alarm, and finding and putting on a safety vest.

Following its development, the testing involved a controlled between-subject experiment utilising pre- and post-training survey questionnaires with 50 participants split into two groups. Each group was exposed to different training conditions: either AR-based or video-based. The experiment aimed to investigate the impact of the AR-based training method compared to the conventional video-based one on participants' knowledge, intrinsic motivation, and self-efficacy levels. The method framework is provided in Fig. 1 below.

3.1. Experimental design

This study adopts a between-subject experimental design, exposing participants to one of the two experimental conditions and comparing their responses. A between-subject design was deemed the best design option, as the experiment involves assessing participants' cognitive gains (knowledge, intrinsic motivation, and self-efficacy levels) through task performance. In other words, participants are expected to learn the tasks in each condition. By splitting participants into separate conditions, the limitations of a within-subject experimental design – where all participants would complete the first and second conditions – are avoided. These include participants experiencing improved performance in the second condition because they have learned from the first one. Despite the smaller sample size and greater statistical power of within-subject designs, this limitation would have a distinctly negative

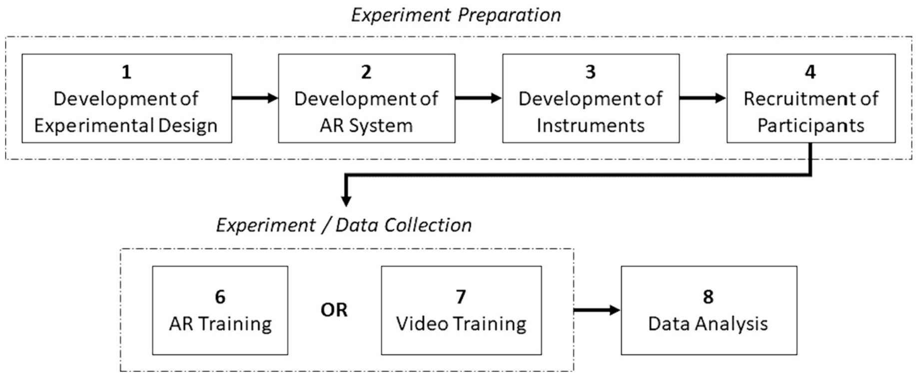


Fig. 1. Research method framework.

impact on this study's data validity [43,44].

The number and values of the independent variable (training mode) create two different experimental treatments or conditions, namely, the traditional training (instructional video on a laptop) and the AR training (guided mission on a head-mounted display) [44]. The traditional training involves the stationary participant sitting in a closed room, watching a video on a conventional laptop. The video consists of a first-person view (FPV) recording of the AR training guided mission through the AR head-mounted display. In other words, the video shows exactly what a person would see and experience from their viewpoint as they go through the entire guided mission in the AR training. One of the researchers recorded the video, which combined screen and audio capture from the AR application running in the head-mounted display. The video shows the correct execution of tasks, all performed by the researcher. This video training approach (*i.e.*, an FPV video recording of the application) has been adopted in previous VR training studies [45]. In contrast, the AR training (described in detail in Section 3.2) involves the participant wearing an AR head-mounted display and walking through the building while following on-screen instructions. Through the headset, the participants can see both the real world in front of them and the digital objects and instructions on the screen.

The dependent variables relating to the participant's learning performance include a) knowledge, b) intrinsic motivation, and c) self-efficacy, collected in each condition. These measures are collected at three stages: immediately before, immediately after, and four weeks after the training experience. Additional training-related measures are collected immediately after the training session, including a) recommendations efficacy, b) recommendations simplicity, c) task load, d) perceived realism (only AR training), e) usability (only AR training), and f) AR preference (only AR training). Data collection makes use of multiple survey questionnaires (see Section 3.3). Knowledge acquisition or gain is given by the difference in knowledge levels before and immediately after training. In turn, knowledge retention is given by the difference between knowledge levels immediately after the training session and four weeks later. The same applies to the other two dependent variables (intrinsic motivation and self-efficacy). While it is possible to assess retention at various intervals after the training session, a four-week gap is the most prevalent period in similar studies, including Sacks et al. [46], Lovreglio et al. [11], Rahouti et al. [4], and Feng et al. [47]. The experimental design structure is provided below (Fig. 2).

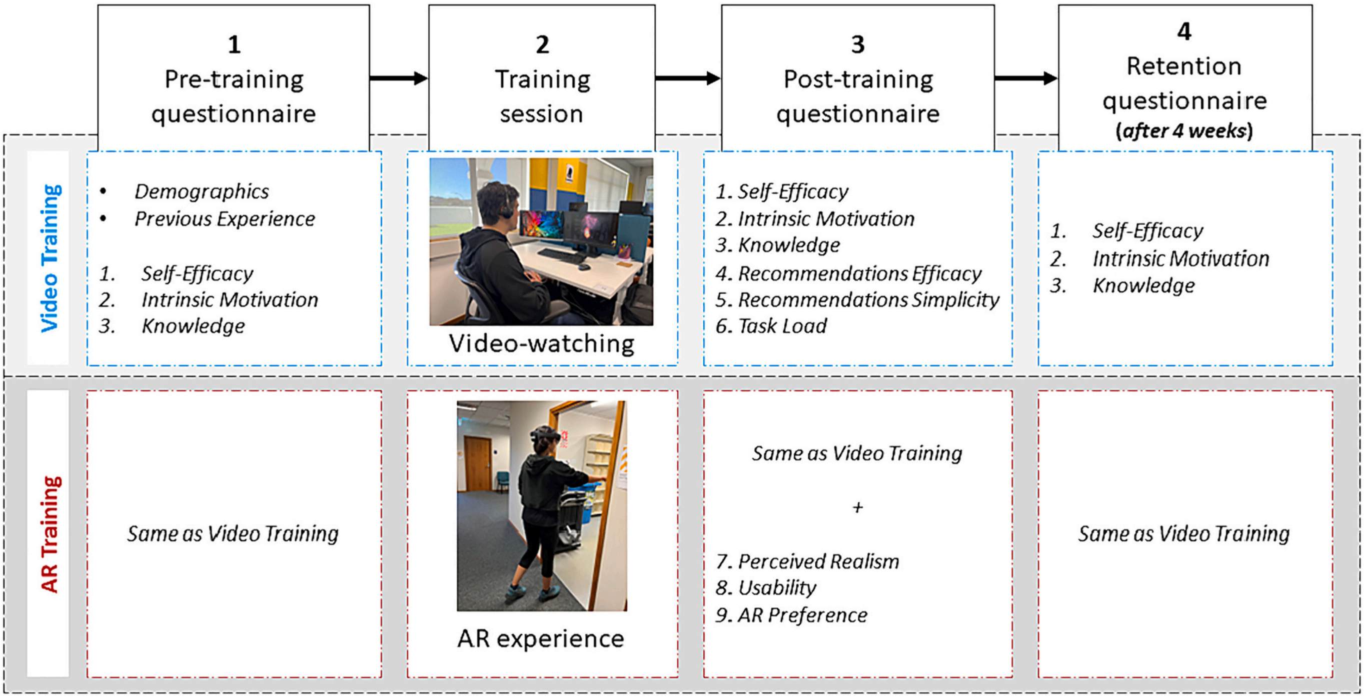


Fig. 2. Experimental design structure.

3.2. AR system design

The prototype system was developed using an AR head-mounted display (Microsoft HoloLens II) and the Unity 3D game engine (2019.3.2f1) on a Windows computer. The AR head-mounted display comes equipped with several features, including a transparent holographic screen with 2 k 3:2 resolution, display optimization for 3D eye position, and unrestricted field of view. The device has a built-in computer powered by a Qualcomm Snapdragon 850 processor, complemented by a holographic processor, 4GB of memory, and 64GB of internal storage. Connectivity options include Wi-Fi and Bluetooth. Its functionality extends to advanced six-degree-of-freedom tracking capabilities, head tracking facilitated by four visible light cameras, hand tracking, eye tracking utilising two infrared cameras, and a 1MP time-of-flight depth sensor, enabling spatial mapping. For visual capture, the device features an 8MP camera for stills and 1080p30 videos. Further enhancing user interaction, the AR display supports voice command and control (using a built-in microphone) and iris recognition. It also includes speakers, an accelerometer, a gyroscope, and a magnetometer. It is designed to accommodate glass wearers and weighs 566 g (for additional information, please visit the manufacturer's website: https://www.microsoft.com/en-us/hololens/hardware#document-experiences).

The Windows computer was equipped with an Intel Core i7–8700 3.2GHz CPU with 6 Cores, 32GB DDR4 RAM, NVIDIA GeForce GTX 1080 GPU, and Windows 10 operational system. The user interface (UI) was designed according to the MRTK (Microsoft Mixed Reality Toolkit 2.7.3) design guidelines. The decision to use Unity 2019.3 LTS was made for its reputation as the most reliable long-term support version available at the time of prototype development. Sample templates from the MRTK design guidelines were used to streamline the prototyping process and ensure compliance with industry best practices for AR content.

The AR system was implemented in a section of the circulation areas on the third floor of the Quadrangle Building A at Massey University in Auckland, NZ. The selected corridor is about 20 m long and 1.5 m wide, with several office doors. This area was chosen because it is a typical contemporary office building in terms of layout and construction technology. Fig. 3 shows the study environment layout (circulation area), the route taken by participants during the experience, and the location of tasks performed by the participants.

The AR application consisted of a guided mission with three sequential tasks prompted by a virtual assistant, represented by a tiny floating firefighter animation. The firefighter animation was chosen given the application's theme (fire safety) to enhance user experience [48]. The firefighter model used in the prototype was found on the Unity asset store. The three tasks were 1) identify and isolate the fire by closing the door of the room where the fire is located, 2) locate and attempt to activate the fire alarm right beside the floor exit door, and 3) locate and put on a safety vest found on the wall opposite the floor exit door (see Fig. 3 for the location of tasks in the study environment). The tasks were generated from and reflect the learning outcomes of standard fire safety training, which include acquired knowledge on how to: a) contain the fire (by closing the room door), b) activate the fire alarm, c) call emergency services, d) wear the safety vest e) evacuate people (if fire warden), f) evacuate, and g) report to the safety manager.

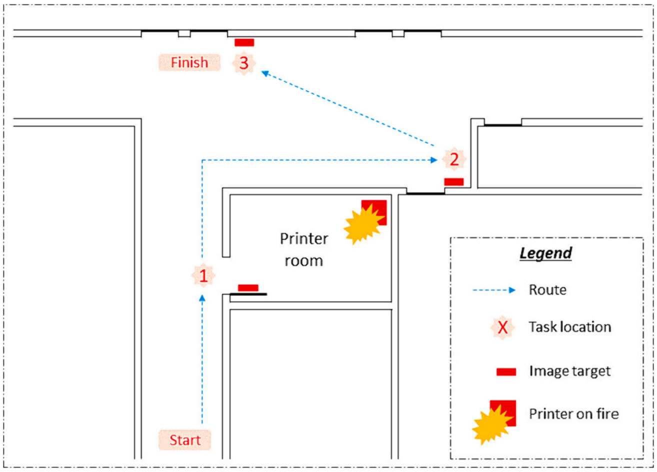


Fig. 3. Study environment layout (circulation area in the office building).

At the beginning of the experience, the participant was greeted with a text UI screen (see Fig. 4 (a)) and then introduced to the animated firefighter character, who provided further instructions on the instructional method and guided them through the three tasks. The tasks were delivered as floating message boxes and speech instructions (as per the MRTK design guidelines). Before each task, the firefighter provided initial instructions on what to do and how to complete a task. The instructions delivered by the firefighter were developed with consultation from Massey University's fire warden. The tasks required the user to perform a visual search while exploring the environment to identify and interact with targets of interest, such as flames, fire alarms, and safety vests.

Vuforia image tracking was used to create anchor points (image targets) across the study environment. Image targets were university posters and flyers printed on A4-sized paper sheets affixed to walls to ensure that the AR experience seamlessly blended with the physical environment, as these types of pictures and poster sizes are commonly found in corporate settings. When the user reached an anchor point, the current task was completed, and the system proceeded to the next task. For instance, when the user stopped in front of the fire alarm beside the floor exit door (task 2), the AR application was able to recognise the image target placed beside it, which served as a trigger to initiate the task (see Fig. 4 (d)). At that moment, the firefighter appeared and delivered instructions on how to use the fire alarm to complete the task. Similarly, when the user stopped in front of the safety vest on the wall opposite the floor exit door (task 3), the AR application recognised the image target beside it and triggered the firefighter to deliver new instructions. Fig. 4 shows a set of screenshots of the AR environment.

Before the experiment, calibration was carried out to ensure the visual augmentations (virtual fire and smoke) and image targets were correctly positioned. In this study, the virtual fire was placed inside a printer room, and posters were placed in three different places: on the printer room door, beside the fire alarm next to the floor exit door, and on the wall opposite the floor exit door covered by a safety vest (see Fig. 3). It should be noted that the prototype allows for customisation through the calibration procedure. This configuration stage also allows for customisation of the user experience for each participant. The location of the virtual fire and image targets can be easily modified to fit any floor layout and experience requirements (duration, difficulty, *etc.*). Fig. 5 provides the AR system prototype operational framework.

To make the application customisable, a configurable game manager was implemented. This manager allowed for the customisation of elements, such as the list of tasks and the actions to be taken after each scene. The game manager handled the presentation of text boxes and the playback of audio clips in each scene. The initial scene served as the configuration stage, where the fire could be placed, parameters such as the flame size adjusted, and Vuforia's status verified. Unity's particle system was used to create and configure the fire and smoke effects. The game manager allowed for adjusting parameters dynamically, enabling the control of the size and intensity of flames and smoke. For audio, clips were recorded that corresponded to different scenes within the experience (flame and smoke sound, fire alarm sound). The source code of the developed AR system prototype has been uploaded to the Gitea DevOps platform and can be made available upon request.

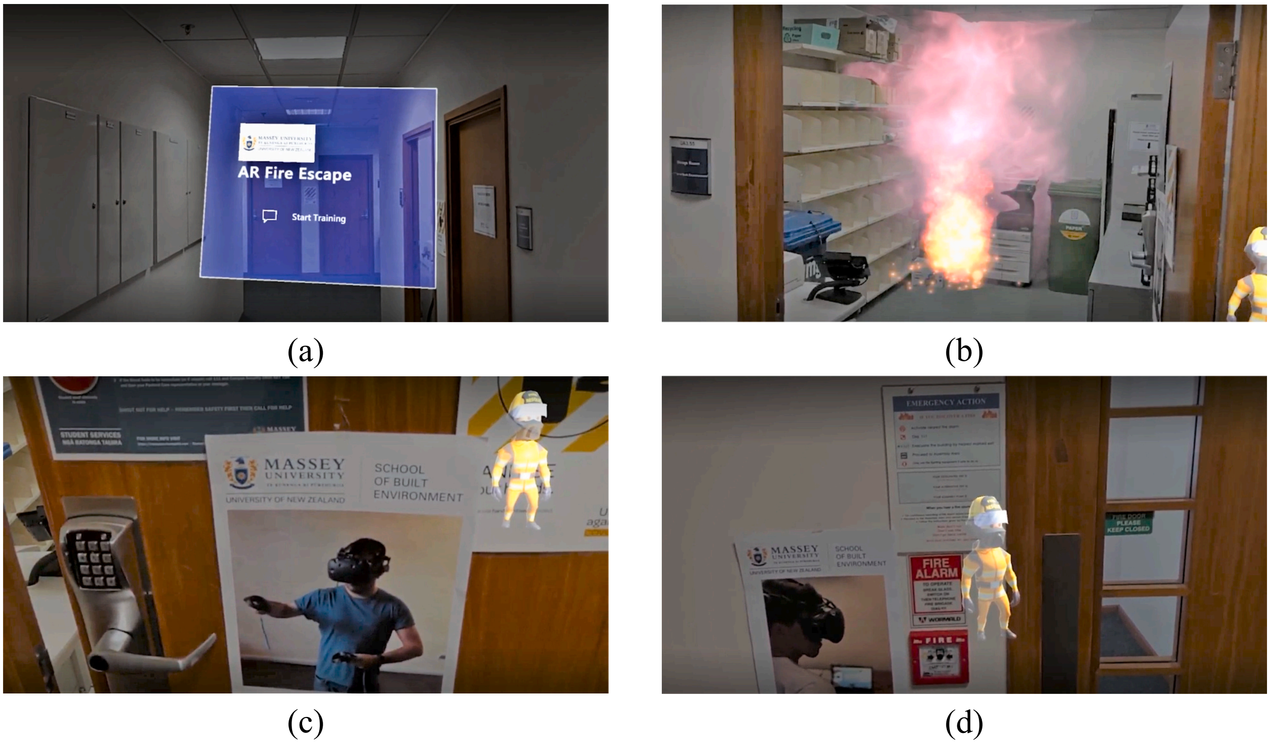


Fig. 4. Screenshots of the AR environment: (a) a floating greeting text box; (b) virtual fire in the printer room; (c) a university poster attached to the printer room door and the animated firefighter; (d) a university poster beside the fire alarm and the animated firefighter.

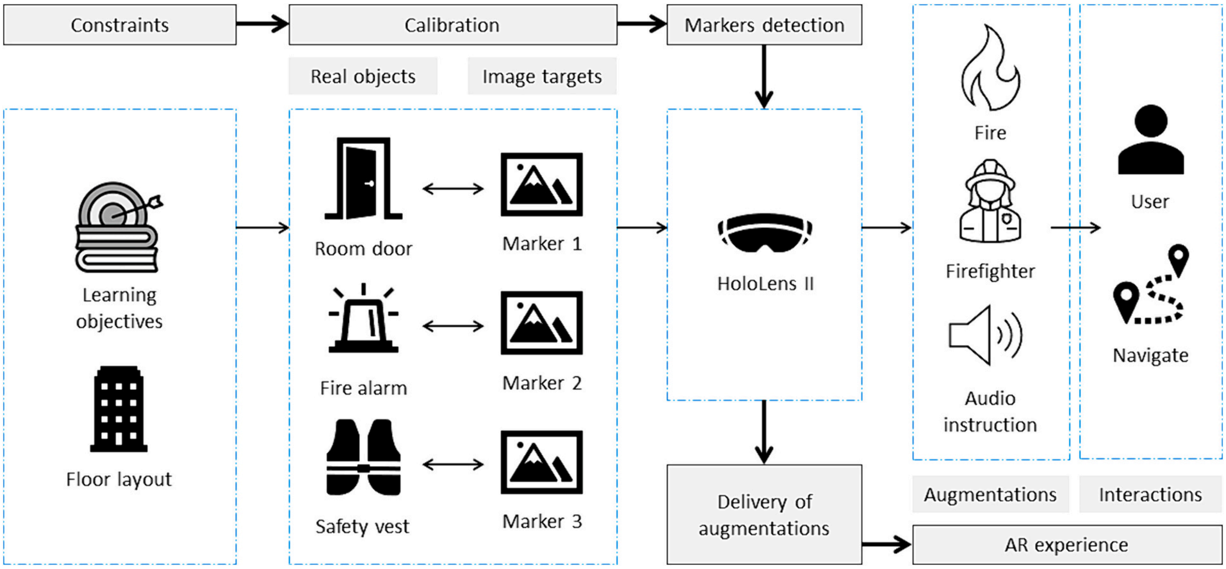


Fig. 5. AR system prototype operational framework.

3.3. *Data collection instruments*

The data collection instruments consist of three questionnaires applied at different stages of the experiment, namely, pre-training, post-training, and retention questionnaires. Each questionnaire includes a different combination of clusters of items aimed at gathering participants' information and responses. These questionnaires were based on previous studies that have developed and used similar instruments (*e.g.*, [49,50,51,4,52,53,54]). Most constructs are based on the Protection Motivation Theory [55] and Makransky and Petersen's [51] learning model. The Protection Motivation Theory is a useful theory that describes how individuals are motivated to protect themselves from threats. This theory has been adopted in several previous studies to assess VR training solutions and compare them against traditional training methods for counterterrorism, earthquake, and fire safety [49,56,11]. In turn, Makransky and Petersen's [51] learning model explains the factors affecting the learning process when using VR solutions, and it has been adopted to assess previous VR safety training solutions [57,4]. This study represents the first attempt to combine these two theories to assess an AR-based fire safety training solution and compare it with a traditional method, as explained in Section 3.1.

The clusters of items included in this work collect information regarding: 1) Demographics, 2) Previous Experience, 3) Self-Efficacy, 4) Intrinsic Motivation, 5) Knowledge, 6) Recommendations Efficacy, 7) Recommendations Simplicity, 8) Task Load, 9) Perceived Realism, 10) Usability, 11) AR Preference. The pre-training questionnaire (PreQ) encompasses clusters 1 to 5. The post-training questionnaire (PostQ) encompasses clusters 3 to 8 in the video training condition and 3 to 11 in the AR training condition. The retention questionnaire (RetQ) encompasses clusters 3 to 5. The clusters of items are described in more detail as follows.

1) the *Demographics* section comprises 3 items used to collect participants' characteristics, including gender and educational level. This information is collected to check that the sample of the AR training group is comparable to the sample of the video training group.
2) the *Previous Experience* section comprises 2 items used to collect participants' previous fire safety training experience and experience with video games. This data is collected to ensure that the two groups have similar previous experiences to avoid any experience bias.
3) the *Self-Efficacy* section comprises 5 items used to collect participants' self-efficacy, defined as the "belief in one's ability to perform a specific action" [49,58]. This measure is a predictor of real-world performance [59]. It was measured using the items used by Rahouti et al. [4] for fire safety training. The five factors of self-efficacy questions include: 1) confidence in understanding basic concepts of fire safety, 2) confidence in understanding complex concepts of fire safety, 3) confidence in the ability to perform well on assignments and tasks in fire safety exercises, 4) expectations to perform well in fire safety training, and 5) confidence in the skills acquired from previous fire safety training sessions. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.
4) the *Intrinsic Motivation* section comprises 5 items used to collect participants' intrinsic motivation, defined as the satisfaction derived from the learning activity itself, rather than from any of its subsequent consequences [60,61]. This measure has been found to be a significant predictor of learning outcomes through self-efficacy [51]. In addition, AR has been found to be a significant contributor to learner motivation [62]. Intrinsic motivation was measured using the items proposed by Rahouti et al. [4] for fire safety training. The five items in this section assess the following aspects: 1) enjoyment of fire safety training activities, 2) whether fire safety training activities are considered fun to perform, 3) the potential boredom associated with fire safety training activities, 4) the level of attention captured by fire safety training activities, and 5) how interesting participants find fire safety training activities. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.
5) the *Knowledge* section encompasses 2 open-ended items used to assess participants' knowledge regarding the actions to take in case of a fire emergency in a university building. The open-ended questions were preferred as they do not suggest to participants any reasonable answer as a multiple-choice questionnaire might do [49]. Participants were asked to answer the following two questions: 1) "What are you supposed to do if you spot a fire in a room of your office building?" and 2) "What are you supposed to do if you are the fire warden for the area where you identify a fire?". Based on the number of distinct actions/tasks – tasks a to g, listed in Section 3.2 – mentioned in their answers to both questions, participants received a score between 0 and 7. Repeated mentions of the same action across questions are considered as a single occurrence. For example, if in question 1, participants mention actions a and b, and in question 2, they mention actions b and c, this results in a total of three distinct actions mentioned (both answers combined). Consequently, they would receive a score of 3.
6) the *Recommendations Efficacy* section comprises 2 items aimed at collecting participants' perceptions on the usefulness and efficacy of recommendations taught during the training. This construct was measured because recommendations provided by a training solution must be effective by clearly communicating the best actions to mitigate any risks related to a threat [49]. It was measured using the items proposed by Lovreglio et al. [11] and Lovreglio et al. [57]. These items cover the following aspects: 1) the perceived usefulness of the recommendations for personal safety, and 2) the belief that the recommendations will enable effective action in real-life fire emergencies. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.
7) the *Recommendations Simplicity* section includes 3 items used to collect participants' perceptions on how easily they could learn, remember, and execute the recommendations taught during the training. This construct was included as recommendations provided during training must be perceived as simple to understand, remember, and implement. In fact, the effectiveness of fear appeals might drop when low scores are found for this construct [49]. It was measured using the items proposed by Lovreglio et al. [11] and Lovreglio et al. [57]. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.
8) the *Task Load* section comprises the 6 items from the NASA Task Load Index (TLX), a well-established and validated instrument used to collect participants' perceived task load during the training experience. This measure has been used in a previous study to assess an AR safety training application [63] and an AR flight training application [64], and it represents a valuable construct to compare the two training solutions in this work. The results of the TLX survey conducted by Arjoni et al. [64] showed a clear increase in the perceived workload with the use of AR glasses, which is not completely unexpected and may be due to the limited field of view and ergonomic aspects of the headset. The Task Load items address the following factors: 1) the mental demand of the training, 2) the physical demand of the training, 3) the perceived pace of the training, whether rushed or slow, 4) the participants' self-assessment of success in completing the training, 5) the level of effort required to accomplish the training, and 6) the extent to which participants experienced feelings of insecurity, discouragement, irritation, stress, and annoyance during the training. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree").
9) the *Perceived Realism* section comprises 3 items used to collect participants' perceptions on the realism of the flames, smoke, and sound effects in the AR training condition. This measure was collected to assess how well the augmentations blended with the real world. Butaslac et al. [15] argue that "the simulator's quality of realism affects the enhancement of the learning transfer", and therefore, the efficacy of AR training is often deduced using face and content evaluation methods. The items were adapted from Le et al. [65]. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.
10) the *Usability* section comprises 8 items that capture participants' perceived capabilities to use and interact with the AR training application and assess if they experienced motion sickness during or after the training. The items were adapted from the ones used in multiple VR and AR studies (*e.g.*, [37,66,65,67,68]). Ratings

were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.

11) the *AR Preference* section consists of 3 items used to collect participants' perceptions regarding the relative advantages of AR training over traditional training methods, including how much more engaging, easier to remember, and preferable the AR training experience would be. These items have been previously used in VR research by Lovreglio et al. [11] and Lovreglio et al. [57] and were adapted to fit this AR study. Ratings were collected using a 7-point Likert scale (−3 = "strongly disagree", +3 = "strongly agree"), and answers were averaged to form the final score for each participant.

3.4. *Participants and experiment session*

The nonprobabilistic sample comprises 50 participants and complies with the following qualification/inclusion criteria: over 18 years of age and a minimum educational level equal to or over completed high school. This study's sample size aligns with the ones used in previous between-subject studies in the field comprising two experimental conditions. For instance, Sacks et al. [46] conducted two experiments in which the samples were 20 and 25 participants; Chen et al. [34] and Lam et al. [50] worked with samples comprising 30 participants; Lin et al. [69] recruited a sample of 40 participants; Chittaro and Sioni [49] had a sample size of 44; Yip et al. [54] recruited 46 individuals. The samples in the studies by Tarng et al. [52] and Zhang and Robb [70] were slightly larger, with 56 and 60 participants, respectively.

Ethics approval (protocol: Southern A, Application 18/21) and participants' consent (consent forms) were obtained prior to the commencement of the study. Participants were randomly assigned to one of the two conditions. Data collection took place in the experiment sessions. Each session took approximately 30 min, resulting in a total of 25 h of data collection over seven weeks. Fig. 6 shows participants undergoing video-based and AR-based training during experiment sessions.

4. Results

The inferential data analysis aims at checking: a) whether differences in knowledge, intrinsic motivation, and self-efficacy responses – immediately before, immediately after, and four weeks after the training session – between experimental conditions are significant, and b) whether differences in training-related measures – recommendations efficacy, recommendations simplicity, task load – between experimental conditions are significant (Sections 4.2 to 4.5).

Generalised Estimating Equation (GEE) models were employed to check for statistically significant differences between the training groups (*i.e.*, AR and Video) at various stages of the experiment (*i.e.*, Pre, Post, and Ret). These regression models are suitable for analysing longitudinal data from repeated measurements of the same participants over time [71]. For each response variable under examination, a distinct GEE model was adjusted with the training group, experiment period, and their interaction as predictors (independent variables). Exchangeable working correlation structures were adopted. All response variables exhibited well-behaved probability distributions and were adequately modelled by a Gaussian distribution. To check the significance of the difference between training groups within a given period – before the training (Pre), after the training (Post), or after four weeks (Ret) – and the difference between periods for the same training group, it was necessary to apply multiple comparison tests using Tukey's correction. Additionally, to quantify the magnitude of these differences, Cohen's d was utilised as the measure of effect size. Cohen's d measures the difference between two means in terms of standard deviations, representing the magnitude of the difference between the groups. It is important to note that there are no upper or lower limits to the value it can take. An effect size $d > 1$ (or lower than −1) indicates a very large effect. A negative value does not diminish the significance or the magnitude of the effect; instead, it provides directionality to the difference [72]. For all tests, a significance level of 5% was assumed.

In turn, the descriptive analysis provides the participants' perceptions of the realism and usability of the AR training solution as well as their preference for AR training over traditional training methods (Section 4.6). The number of participants remained the same through all three data collection stages, with no dropouts occurring throughout the study. This was achieved by consistently following up with participants throughout the study. That is, both the AR and Video training groups had 25 participants before the training (Pre), after the training (Post), and after four weeks (Ret). All participants completed the retention questionnaires exactly four weeks after the training session. No participants reported experiencing motion sickness or any discomfort while using the AR system prototype.

4.1. *Sample demographics*

Table 1 provides a breakdown of the sample characteristics. It can be seen that the distribution of all parameters is relatively balanced across the two groups. In both groups, most participants have postgraduate degrees. Most have undergone fire safety training either over a year ago or have never done so. Most participants in both groups also reported that they either play video games once a year or have never played them.

Fig. 6. Participants undergoing video-based (left) and AR-based training (right).

Table 1
Sample demographics.

Parameter	Sample (n = 50)			
	AR training group (n = 25)		Video training group (n = 25)	
	#	%	#	%
Gender				
Woman	13	52%	12	48%
Man	12	48%	13	52%
Others/Prefer not to disclose	0	0%	0	0%
Educational level				
No university degree	0	0%	2	8%
Undergraduate degree	3	12%	3	12%
Diploma degree	0	0%	0	0%
Postgraduate degree	22	88%	20	80%
Previous fire safety training experience				
Over a week ago	0	0%	0	0%
Over a month ago	1	4%	3	12%
Over 3 months ago	2	8%	4	16%
Over 6 months ago	4	16%	2	8%
Over a year ago	8	32%	6	24%
Never	10	40%	10	40%
Experience with video games				
More than once a day	1	4%	1	4%
Once a day	0	0%	2	8%
Once a week	2	8%	3	12%
Once a month	8	32%	4	16%
Once a year	7	28%	6	24%
Never	7	28%	9	36%

4.2. *Analyses of knowledge*

The knowledge scores obtained by the participants of the Video training and AR training groups are analysed in this section. The analysis is carried out considering when the knowledge assessment took place in the experiment: before the training (Pre), after the training (Post), and after four weeks (Ret). The results of the knowledge scores are presented in Fig. 7 using boxplots.

The data show a significant increase in knowledge when comparing the scores before and after the training (Pre *vs* Post) for both groups. There is a significant decrease in knowledge when comparing the scores after the training and four weeks later (Post *vs* Ret) for both groups as well (see Table 2). The participants' knowledge before the training is not statistically different between the AR and Video groups, showing that both groups started the experiment with comparable fire safety knowledge. The results also show no evidence of a significant difference between the knowledge scores of the two groups after the training and after four weeks (see Table 3).

Table 2
Comparison of the knowledge scores before the training (Pre), after the training (Post), and after four weeks (Ret) for the AR and Video training conditions.

	Pre *vs* Post		Post *vs* Ret	
	Video	AR	Video	AR
Z	−7.44	−7.28	3.66	2.54
P-value	**<0.001**	**<0.001**	**0.001**	**0.030**
d	−1.77	−1.42	0.83	0.56

Table 3
Comparison between knowledge scores of AR and Video training conditions for the knowledge assessment done before the training (Pre), after the training (Post), and after four weeks (Ret).

AR *vs* Video	Pre	Post	Ret
Z	−0.88	−1.60	−0.79
P-value	0.377	0.110	0.431
d	−0.17	−0.52	−0.24

4.3. *Analyses of intrinsic motivation*

The intrinsic motivation scores obtained by the participants of the Video training and AR training groups are analysed in this section. The analysis is carried out considering that the intrinsic motivation was measured before the training (Pre), after the training (Post), and after four weeks (Ret). The results of the average intrinsic motivation scores are shown in Fig. 8.

The data show a significant increase in intrinsic motivation when comparing the scores before and after the training (Pre *vs* Post) only for the AR group. There is no evidence of a significant increase in intrinsic motivation for the Video group. There is no significant decrease in intrinsic motivation when comparing the scores after the training and four weeks later (Post *vs* Ret) for both groups (see Table 4). The participants' intrinsic motivation before the training is not statistically different between the AR and Video groups, showing that both groups started the experiment with a comparable level of intrinsic motivation. The results also show no evidence of a significant difference between the intrinsic motivation scores of the two groups after the training and after four weeks (see Table 5).

4.4. *Analyses of self-efficacy*

The self-efficacy scores obtained by the participants of the Video training and AR training groups are analysed in this section. The analysis is carried out taking into account that the self-efficacy was measured

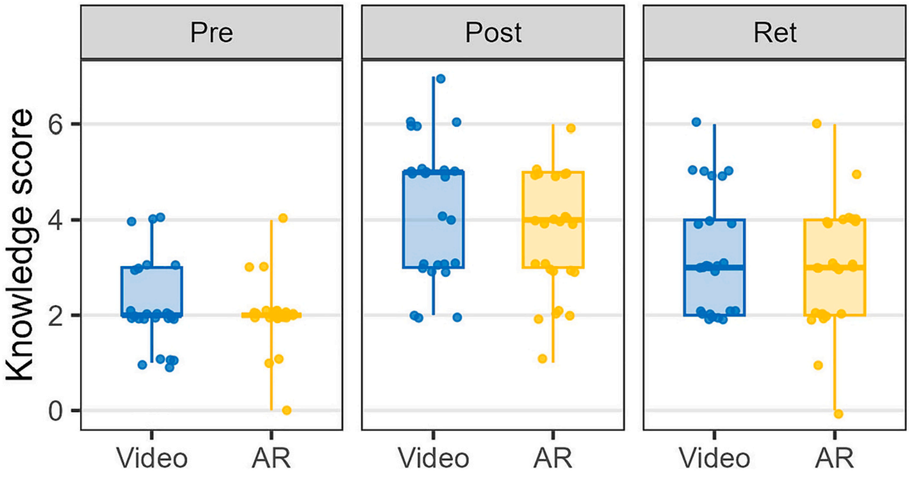


Fig. 7. Participants' knowledge scores before the training (Pre), after the training (Post), and after four weeks (Ret) in the AR and Video training conditions.

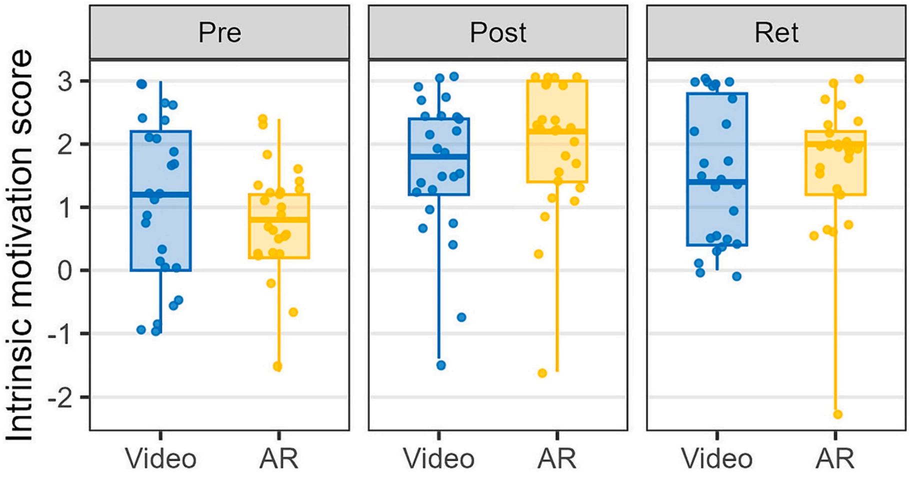


Fig. 8. Participants' intrinsic motivation scores before the training (Pre), after the training (Post), and after four weeks (Ret) in the AR and Video training conditions.

Table 4
Comparison of the intrinsic motivation scores before the training (Pre), after the training (Post), and after four weeks (Ret) for the AR and Video training conditions.

	Pre *vs* Post		Post *vs* Ret	
	Video	AR	Video	AR
Z	−1.62	−5.92	0.39	1.64
P-value	0.239	**<0.001**	0.918	0.227
d	−0.51	−1.06	0.10	0.24

Table 5
Comparison between the intrinsic motivation scores of AR and Video training conditions for the intrinsic motivation assessment done before the training (Pre), after the training (Post), and after four weeks (Ret).

AR *vs* Video	Pre	Post	Ret
Z	−0.94	0.97	0.48
P-value	0.346	0.332	0.633
d	−0.27	0.28	0.13

before the training (Pre), after the training (Post), and after four weeks (Ret). The results are the average self-efficacy scores shown in Fig. 9.

The data show a significant increase in self-efficacy when comparing the scores before and after the training (Pre *vs* Post) for both groups. There is a significant decrease in self-efficacy when comparing the scores after the training and four weeks later (Post *vs* Ret) only for the Video group. There is no evidence of a significant decrease in self-efficacy for the AR group (see Table 6). The participants' self-efficacy before the training is not statistically different between the AR and Video groups, showing that both groups started the experiment with a comparable level of self-efficacy. The results also show no evidence of a significant difference between the self-efficacy scores of the two groups after the training and after four weeks (see Table 7).

4.5. *Analyses of recommendations efficacy, simplicity, and task load*

The analyses of the efficacy and simplicity of recommendations and task load are carried out in this section. The analyses are conducted considering that these constructs and related clusters of items were

Table 6
Comparison of the self-efficacy scores before the training (Pre), after the training (Post), and after four weeks (Ret) for the AR and Video training conditions.

	Pre *vs* Post		Post *vs* Ret	
	Video	AR	Video	AR
Z	−4.18	−3.38	2.36	1.05
P-value	**<0.001**	**0.002**	**0.048**	0.547
d	−1.11	−0.99	0.46	0.30

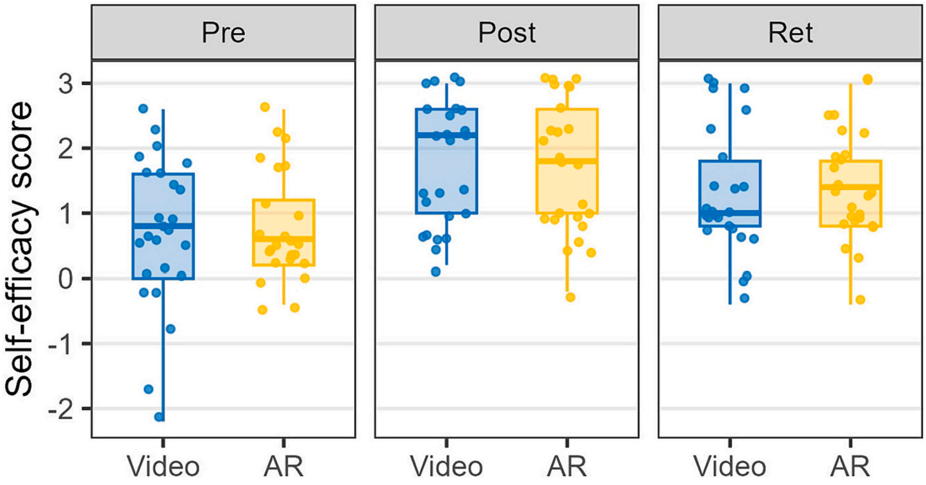


Fig. 9. Participants' self-efficacy scores before the training (Pre), after the training (Post), and after four weeks (Ret) in the AR and Video training conditions.

Table 7
Comparison between the self-efficacy scores of AR and Video training conditions for the self-efficacy assessment done before the training (Pre), after the training (Post), and after four weeks (Ret).

AR *vs* Video	Pre	Post	Ret
Z	0.29	−0.15	0.44
P-value	0.775	0.882	0.659
d	0.08	−0.04	0.12

measured only after the training (Post). The scores for the efficacy and simplicity of recommendations for both groups are shown in Fig. 10, while the scores for the different items of the task load test are provided in Fig. 11. The results of the statistical tests show no significant differences in recommendations efficacy, simplicity, and task load between training groups.

4.6. *Analyses of perceived realism, usability, and AR preference*

This section provides the analyses of participants' views about the realism and usability of the AR training system as well as their preference for AR training over traditional methods. This data was collected from the participants of the AR training group. Fig. 12 shows the scores for the three constructs. On the one hand, the results indicate that the realism of visual augmentations (flame and smoke "holograms") and sound effects of the AR system prototype can still be improved, as over 25% of the sample provided a score below zero. On the other hand, the majority of the sample (over 75%) were either satisfied or very satisfied with the usability of the prototype. Finally, the data show that most participants would prefer using an AR-based fire safety training tool over traditional training methods in the future.

5. Discussion

This paper demonstrates the effectiveness of AR in delivering fire safety knowledge. Results align with prior research that identified a moderate positive effect of AR on learning gains [73]. The acquired fire safety knowledge following training was found to be equivalent between the AR group and the video group. This finding also aligns with existing literature in that XR technologies may not necessarily surpass traditional methods in terms of knowledge acquisition [38]. For example, findings by Thees et al. [74] indicated that the AR-assisted approach did not surpass the traditional method in effectively conveying conceptual labwork knowledge. Similarly, Feng et al. [47] compared immersive VR against safety manuals for excavation safety training and found that both methods had comparable performance regarding knowledge acquisition.

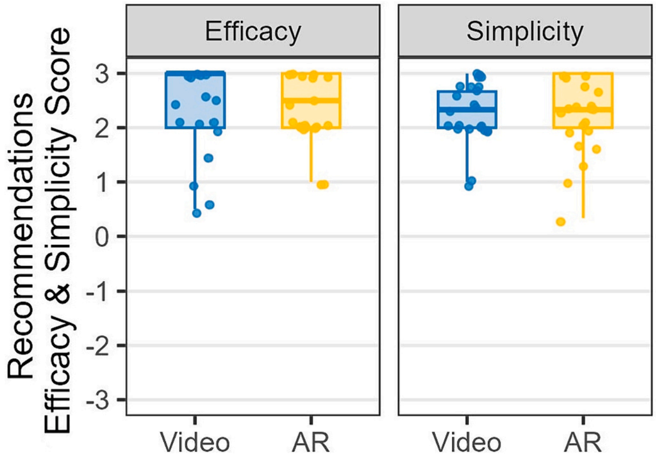


Fig. 10. Recommendations efficacy and simplicity scores in the AR and Video training conditions.

The results of the current study diverge from earlier research that identified AR as more effective than traditional methods for training more complex tasks. Wolf et al. [75] and Zhang and Robb [70] found significantly fewer errors in AR conditions for complex tasks, while for simpler tasks, AR performed comparably to flat screen- or video-based methods. It seems that the positive effect of AR is amplified the more challenging the training tasks are. One possible explanation is that AR might help reduce cognitive load compared to other instructional approaches [76] through enhanced information clarity, knowledge acquisition speed, and stimulation [75]. Butaslac et al. [15] note that cognitive load can be reduced by presenting the user with adequate cues and instructions. In this regard, the present study detected no significant differences in recommendations efficacy, simplicity, and task load between training conditions, which suggests an effective AR application design. Scores of recommendations efficacy and simplicity were notably high in both conditions, demonstrating that participants found the "guided mission" approach in both video and AR formats to be highly efficient and useful, regardless of whether it was experienced passively (*i.e.*, watching a video) or actively (*i.e.*, performed through AR). The results from the task load survey indicate that neither training method was physically or mentally demanding, and both training sessions were similarly easy to complete. Participants experienced no difficulties and did not become stressed or annoyed during training in either condition. These results suggest that the training tasks were easy to complete and, therefore, not complex. This could explain why participants performed similarly in knowledge acquisition across the two training methods. According to findings from Wolf et al. [75] and Zhang and Robb [70], AR may not make a significant difference in training less complex or cognitively demanding tasks.

In line with the above, key lessons learned from the development of the AR system prototype include the importance of ensuring system adaptability and robust bandwidth, given the diversity of fire safety training scenarios within buildings. Since each scenario is unique, the possibility of adapting the system to varying protocols seamlessly is crucial. Eventual necessary adaptations – such as changing from text-to-speech to a voice actor – should be easily implementable. Secondly, this study demonstrated a need for a user-centred system design ensuring that all visual and auditory stimuli are clear, concise, and understandable. This approach should account for individuals with poor eyesight or impaired hearing. Lastly, there is a need to provide comprehensive documentation and operational instructions accompanying the device, making it accessible and user-friendly. This is important because even seemingly simple tasks may pose challenges to some users without proper guidance.

Interestingly, unlike the video group, the AR group exhibited a significant increase in intrinsic motivation when comparing the scores before and after the training and did not exhibit a significant decline in self-efficacy four weeks after training. Furthermore, it should be noted that the statistically significant decline in knowledge for the AR group four weeks after training was marginal, and the effect size of such a decline was much smaller than that of the video group. This means people may be less prone to forgetting after undergoing AR training compared to traditional video training and implies that AR training may hold an advantage in terms of knowledge retention. This finding is in line with evidence from previous research. Gargrish et al. [77] investigated the effectiveness of AR in geometry learning and found that trainees who learned using AR retained more knowledge after 2–4 weeks than those who learned with a desktop application. The participants of their study exhibited higher excitement and motivation while engaging with AR than with the desktop application. The researchers argued that the more engaging and immersive learning experience offered by AR facilitated memory retention. Similarly, Lam et al. [50] found that AR surpassed paper-based manuals in terms of knowledge retention (after 48 h) about product parts and disassembly processes. The study participants preferred AR's usefulness and user-friendliness over paper-

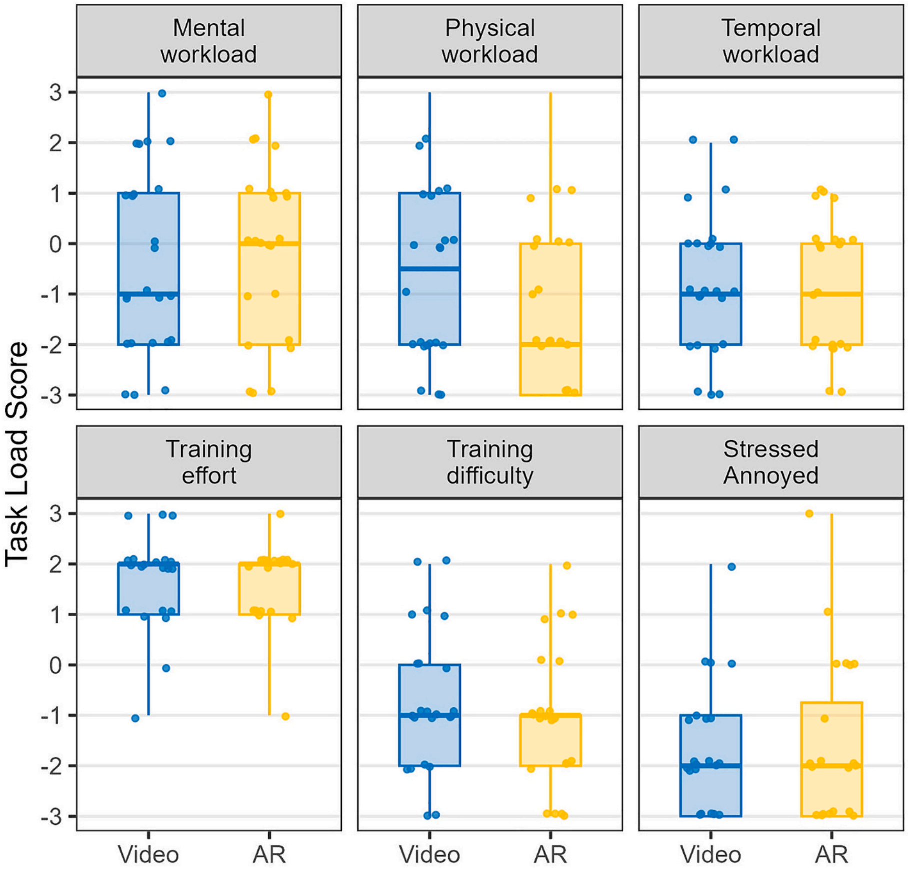


Fig. 11. Task load scores in the AR and Video training conditions.

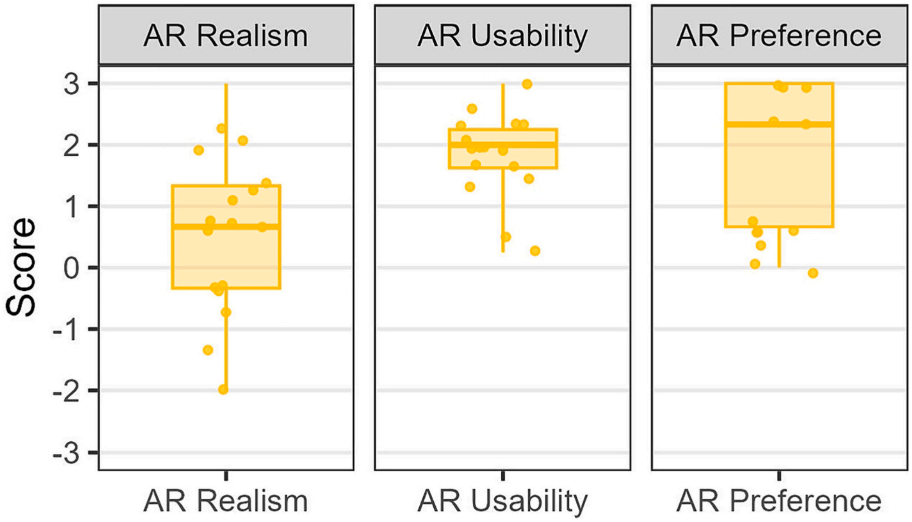


Fig. 12. Perceived realism, usability, and AR preference scores.

based manuals.

In the present study, the perceived realism, usability and AR preference scores suggest that participants are highly motivated and interested in exploring this technology for training. Therefore, in summary, factors such as motivation, engagement, and immersion may contribute to the superior learning experience observed with AR training. The literature corroborates this potential link to cognition and memory [78,79,80]. In addition, it seems that memory recall can be improved through AR's feature of intuitive visualisation and interaction with real-world objects from a first-person perspective [81].

6. Conclusions

The assumption that AR-based training is more effective than traditional methods has been fueled despite the absence of empirical evidence to support it. Many studies demonstrating AR training effectiveness in built environment disciplines often deliver an anecdotal report of users' experiences with AR systems, disregarding the more obvious learning gains eventually enabled by such technology. These studies provided the foundations of research in the field and revealed the need for further comparative experiment-based research into the effectiveness of AR training systems. After all, experimental research appears to be the only approach that enables researchers to judge beliefs and assumptions with systematically measured confidence and reliability [44].

Traditional fire safety training approaches are designed to save lives, and any improvements made on their limitations (as discussed in Section 1) are beneficial in the ultimate protection of human life during emergencies. This research aimed to address some of these limitations by comparing the effectiveness of traditional approaches to novel XR-assisted methods using an optical see-through AR-based fire safety training system. The aim was to determine whether there are grounds for adopting or further exploring AR as a tool for fire safety training. If there is no empirical evidence of the superior effectiveness of a novel method over conventional ones, then there is no purpose in adopting or further exploring it. While there are many traditional training methods, this study focused on video-based training as a modern, widely-used one that could provide a comparable experience to AR. For the video-based training condition, the AR experience was recorded and shown to participants, thus replicating the training and only changing the delivery mechanism. This made video training the right control treatment for this study, as experimenting with all possible traditional training methods would be impractical. It should be noted that other XR solutions (such as VR) may also enhance safety training (as discussed in Section 1). Each technology presents drawbacks and opportunities. Ultimately, VR and AR may not be seen as competitors but as allies in providing alternative methods to enhance training experiences and learning.

Thus, a fundamental step for AR to be considered a more effective fire safety training technique is to understand whether and to what extent it improves trainees' learning experience and outcomes compared to conventional training approaches. In the current study – the first academic attempt to demonstrate AR-based safety training effectiveness in built environment discipline – this was achieved through a comparative experiment-based assessment of technology-related and user-related factors that underlie fire safety training effectiveness, including trainees' knowledge acquisition and retention, intrinsic motivation, and self-efficacy. This study hypothesised that the eventual superior effectiveness of AR reported in the literature would be associated with enhanced learning performance (given by the dependent variables mentioned above). Results suggest that the AR system was well-designed and as effective as the traditional method in terms of knowledge acquisition and retention and overall learning experience. However, it outperformed the video-based training in terms of intrinsic motivation gain and self-efficacy retention. By providing evidence that AR may facilitate better learning experiences and enable greater learning gains over conventional training approaches, this research corroborates the need to further advance AR as a tool for fire safety training.

It should be noted that non-significant differences detected in this study still provide valuable insights (as in any scientific investigation). For instance, Paes et al. [82] have recently published a similar study where they found no associations between dependent variables. As more similar studies are published, it will eventually be possible to combine their data into a meta-analysis and identify further trends. In line with current literature, the significant differences observed in both intrinsic motivation gain and self-efficacy retention are noteworthy effects of XR-based safety training methods (see, for instance, [11]). Thus, this study's preliminary findings are promising in the sense that any method capable of enhancing intrinsic motivation and self-efficacy levels represents a valuable alternative for improving safety training.

Nonetheless, unlike VR solutions, AR technology has received little attention in the fire safety field, with only a handful of applications designed for fire evacuation preparedness (training) and response (assistance). As discussed in Section 2, there are only a few AR-assisted emergency evacuation applications, such as the ones developed by Ahn and Han [29], Kanangkaew et al. [32], Yoo and Choi [33], and Chen et al. [34]. It should be noted that those are not for training purposes but for assistance during a fire. The only existing AR evacuation training application, which could eventually be adapted for fire evacuation training, is the one by Catal et al. [37], which uses a video see-through AR system (tablet-based). Their study, however, did not conduct a comparative effectiveness evaluation of the system. Moreover, to date, most studies assessing the effectiveness of XR-based safety training solutions have focused on evaluating knowledge acquisition only immediately after the training. However, it is also fundamental to evaluate what individuals will remember weeks or even years after the training in order to determine the effectiveness of a method.

Given the above, this study provides three major contributions: first, it represents the first formal attempt in the published academic literature to develop a prototype of an optical see-through AR-based fire safety training solution (using a transparent-visor head-mounted display instead of a tablet/smartphone) and test its effectiveness through a user-centred comparative assessment. It should be noted that the AR system proposed has been designed to offer customisation flexibility. As shown in Fig. 5, interaction with real-world objects was achieved through image targets attached next to them. This method offers the flexibility for the system to be adapted to different space layouts. During the calibration stage, the location of markers 1, 2, and 3 can be adjusted to the floor layout of the building where the training will take place. This enables the delivery of effective training using the developed AR system in any indoor environment. Second, this study represents the first attempt to combine the Protection Motivation Theory [55] and Makransk and Petersen's [51] learning model – which have been used before to assess VR safety training solutions – to assess an AR-based fire safety training solution and compare it with a traditional method. Lastly, this is the first time that knowledge, intrinsic motivation, and self-efficacy are assessed together in three stages: before, immediately after, and four weeks after training. Thus, this study contributes to the body of knowledge by expanding our understanding of the suitability and effectiveness of AR-based fire safety training methods from the user standpoint and providing directions for developing and implementing more effective systems to enhance building occupants' safety. Ultimately, user-centred research on AR-based fire safety training systems could lead to the development of increasingly efficient tools tailored to specific users and topics, improving trainees' knowledge and understanding of fire safety measures and their chances of surviving a fire emergency.

This study's findings are generalisable across the study population using the developed AR-based fire safety training system. It provides empirical evidence that if a similar AR system were to be developed to train a similar population on similar topics/tasks (for example, building evacuation in disaster events such as earthquakes, tsunamis and terrorist attacks), similar results could be expected. Naturally, as in most HCI studies, this study's findings are limited to the hardware and software combination used and the AR system design so that results could change, for example, when altering the display type, animations, interactions and user interface properties. Similarly, results may vary when using the AR system prototype to train a population with different characteristics, such as individuals who may not be as highly educated as the participants in this study.

Furthermore, the research results may be restricted to the nature of the training topic and tasks. Results could differ if different topics with different learning objectives are delivered in different locations. Nonetheless, it should be noted that the developed prototype offers customisation flexibility, allowing for easy adjustment of the location of image targets (see Fig. 5) to suit different floor layouts and experience preferences, such as the number and sequence of tasks, duration, and difficulty levels. In the future, the prototype AR application can be expanded to include additional tasks, augmentations and interactions, covering a wider range of learning outcomes of fire safety training in various contexts.

These limitations are not particular to this study but to any investigation in the HCI field – as its name suggests, it deals with how a specific group of people interact with certain technologies and what human and technological factors affect the quality of such interaction. Although this study could have elicited different results if conducted with different equipment, user-centred research should prioritise understanding the user experience rather than focusing on the systems, devices, or equipment used. Another study limitation related to equipment lies in the absence of an evaluation regarding the potential time and cost constraints associated with the development and implementation of the AR method, an aspect that is worth considering in understanding the practical implications and feasibility of adopting such technology for fire safety training. However, technology is advancing rapidly, and systems may quickly become outdated or be replaced by more cost-efficient solutions. As such, it is recommended that studies on the validation of digital technology focus on user experiences.

In this study, the process of selecting questions from existing questionnaires involved a qualitative assessment by the research team of the suitability of each of those questions in light of the study objectives and experimental conditions while ensuring that the final set of questions covered the main aspects of each construct measured. Adapting and combining questions from multiple existing instruments creates non-validated custom-designed surveys that fit this study's purposes. As discussed by Paes et al. [82], this practice is not uncommon in the XR research community. Similar intrinsic motivation and self-efficacy questions were used by Rahouti et al. [4]. Recommendations efficacy, recommendations simplicity, and AR preference items were proposed and used by Lovreglio et al. [11] and Lovreglio et al. [57]. Perceived realism questions were adapted from the ones used by Le et al. [65], while usability items were adapted from those used in multiple XR studies (*e.g.*, [37,66,65,67,68]). Nonetheless, although the questionnaires used in this study were derived from published instruments, they would still benefit from systematic validation. This validation was not conducted in this study and should be addressed in the future through an in-depth examination of three types of validity: construct, criterion, and content validity [83].

This study aimed to present preliminary retention findings within a commonly employed retention timeframe (as discussed in Section 3.1). Naturally, longer intervals would likely yield different results from the ones found in this study. While assessing retention at multiple intervals over longer periods – as observed in longitudinal studies – could provide more insights into the effectiveness of training methods, this approach poses numerous challenges, including maintaining participant engagement over extended data collection periods [84]. Consequently, conducting longer retention studies requires a larger initial sample size to ensure sufficient statistical power, as participant numbers are expected to decrease over time. Moreover, given that the time between training and when the knowledge is needed, such as during a fire event, can be quite lengthy (more than four weeks), future studies may explore the impact of training methods on the retention of knowledge, intrinsic motivation, and self-efficacy over longer periods.

Several other potential avenues are worth further investigation in future studies. First, the effectiveness of AR training may vary based on the nature of the learning outcomes, knowledge delivered (*i.e.*, procedural or declarative), and types of training tasks. Second, the complexity of knowledge and training tasks may also impact AR effectiveness compared to other training methods. Third, yet unexplored latent factors may influence knowledge acquisition and retention, intrinsic motivation, and self-efficacy in AR training, such as sense of immersion, degree of realism (and its multiple dimensions such as pictorial, spatial, and auditory realism), engagement, motivation, and personality traits. It is worth noting that spatial perception capabilities were intrinsic to the training objectives in Gargrish et al. [77] and Lam et al. [50] studies, who found AR training to provide superior knowledge retention. Studies on the relationship between spatial perception and performance have been extensively conducted in the VR community (see, for instance, [85,86]). Thus, future research may also delve into human-computer interaction aspects (*e.g.*, visual perception, interactions, user interface, human factors) to enhance trainees' learning experience and outcomes and maximise the effectiveness of AR-based fire safety training tools. Lastly, future comparative studies may evaluate the performance of AR against other traditional training modes (such as lecture-based training) and XR solutions (such as VR training systems) to identify their advantages and disadvantages in comparison to each other.

CRediT authorship contribution statement

Daniel Paes: Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Zhenan Feng:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Maddy King:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **Hesam Khorrami Shad:** Writing – original draft, Software, Resources, Investigation, Data curation. **Prasanth Sasikumar:** Writing – original draft, Visualization, Software, Resources, Investigation. **Diego Pujoni:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Data curation. **Ruggiero Lovreglio:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was partially supported by Massey University's College of Sciences Research Enhancement and Development Initiative (REaDI) Fund.

References

[1] NFPA. https://www.nfpa.org/~/media/Files/News%20and%20Research/Fire%20statistics%20and%20reports/US%20Fire%20Problem/osFireLoss.ashx, 2022.
[2] S. Gwynne, E.D. Kuligowski, K.E. Boyce, D. Nilsson, A. Robbins, R. Lovreglio, J. Thomas, A. Roy-Poirier, Enhancing egress drills: preparation and assessment of evacuee performance, Fire Mater. 43 (6) (2019) 613–631, https://doi.org/10.1002/fam.2448.
[3] A. All, B. Plovie, E.P. Nunez Castellar, J. Van Looy, Pre-test influences on the effectiveness of digital-game based learning : a case study of a fire safety game, Comput. Educ. 114 (2017) 24–37, https://doi.org/10.1016/j.compedu.2017.05.018.
[4] A. Rahouti, R. Lovreglio, S. Datoussaïd, T. Descamps, Prototyping and validating a non-immersive virtual reality serious game for healthcare fire safety training, Fire. Technol. 57 (2021) 3041–3078, https://doi.org/10.1007/s10694-021-01098-x.
[5] S. Gwynne, M. Amos, M. Kinateder, N. Benichou, K. Boyce, C.N. van Der Wal, E. Ronchi, The future of evacuation drills: assessing and enhancing evacuee performance, Saf. Sci. 129 (2020) 104767, https://doi.org/10.1016/j.ssci.2020.104767.
[6] A. Rigos, E. Mohlin, E. Ronchi, The cry wolf effect in evacuation: a game-theoretic approach, Physica A: Stat. Mech. Appl. 526 (2019) 120890, https://doi.org/10.1016/j.physa.2019.04.126.
[7] M. Ramirez, K. Kubicek, C. Peek-Asa, M. Wong, Accountability and assessment of emergency drill performance at schools, Family Commun. Health 32 (2) (2009) 105–114. http://www.jstor.org/stable/44954593.
[8] E. Gagliardi, G. Bernardini, E. Quagliarini, M. Schumacher, D. Calvaresi, Characterization and future perspectives of virtual reality evacuation drills for safe

built environments: a systematic literature review, Saf. Sci. 163 (2023) 106141, https://doi.org/10.1016/j.ssci.2023.106141.

[9] Z. Feng, V. Gonzalez, R. Amor, R. Lovreglio, G. Cabrera-Guerrero, Immersive virtual reality serious games for evacuation training and research: a systematic literature review, Comput. Educ. 127 (2018) 252–266, https://doi.org/10.1016/j.compedu.2018.09.002.

[10] U. Rüppel, K. Schatz, Designing a BIM-based serious game for fire safety evacuation simulations, Adv. Eng. Inform. 25 (4) (2011) 600–611, https://doi.org/10.1016/j.aei.2011.08.001.

[11] R. Lovreglio, X. Duan, A. Rahouti, R. Phipps, D. Nilsson, Comparing the effectiveness of fire extinguisher virtual reality and video training, Virtual Reality 25 (2021) 133–145, https://doi.org/10.1007/s10055-020-00447-5.

[12] R. Lovreglio, V. Gonzalez, Z. Feng, R. Amor, M. Spearpoint, J. Thomas, M. Trotter, R. Sacks, Prototyping virtual reality serious games for building earthquake preparedness: the Auckland City Hospital case study, Adv. Eng. Inform. 38 (2018) 670–682, https://doi.org/10.1016/j.aei.2018.08.018.

[13] R. Lovreglio, M. Kinateder, Augmented reality for pedestrian evacuation research: promises and limitations, Saf. Sci. (2020), https://doi.org/10.1016/j.ssci.2020.104750.

[14] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, B. MacIntyre, Recent advances in augmented reality, IEEE Comput. Graph. Appl. 21 (6) (2001) 34–47, https://doi.org/10.1109/38.963459.

[15] I.M. Butaslac, Y. Fujimoto, T. Sawabe, M. Kanbara, H. Kato, Systematic review of augmented reality training systems, in: IEEE Transactions on Visualization and Computer Graphics, 2022, https://doi.org/10.1109/TVCG.2022.3201120.

[16] L. Hou, H. Chi, W. Tarng, J. Chai, K. Panuwatwanich, X. Wang, A framework of innovative learning for skill development in complex operational tasks, Autom. Constr. 83 (2017) 29–40, https://doi.org/10.1016/j.autcon.2017.07.001.

[17] L.W. Menzemer, E. Ronchi, M.M.V. Karsten, S. Gwynne, J. Frederiksen, A scoping review and bibliometric analysis of methods for fire evacuation training in buildings, Fire Saf. J. 136 (2023) 103742, https://doi.org/10.1016/j.firesaf.2023.103742.

[18] A. Hussien, A. Waraich, D. Paes, A review of mixed-reality applications in construction 4.0, in: A. Sawhney, M. Riley, J. Irizarry (Eds.), Construction 4.0 - an Innovation Platform for the Built Environment, Taylor & Francis/Routledge, Abingdon, 2020.

[19] A.A. Kamal, S.N. Junaini, A.H. Hashim, F.S. Sukor, M.F. Said, The enhancement of OSH training with an augmented reality-based app, Int. J. Online Biomed. Eng. (iJOE) 17 (13) (2021) 120–134, https://doi.org/10.3991/ijoe.v17i13.24517.

[20] F. Bork, A. Lehner, U. Eck, N. Navab, J. Waschke, D. Kugelmann, The effectiveness of collaborative augmented reality in gross anatomy teaching: a quantitative and qualitative pilot study, Anat. Sci. Educ. 14 (2020) 590–604, https://doi.org/10.1002/ase.2016.

[21] S. Yang, J.R. Carlson, S. Chen, How augmented reality affects advertising effectiveness: the mediating effects of curiosity and attention toward the ad, J. Retail. Consum. Serv. 54 (2020) 102020, https://doi.org/10.1016/j.jretconser.2019.102020.

[22] M.-P. Chen, L.-C. Wang, D. Zou, S.-Y. Lin, H. Xie, C.-C. Tsai, Effects of captions and English proficiency on learning effectiveness, motivation and attitude in augmented-reality-enhanced theme-based contextualized EFL learning, Comput. Assist. Lang. Learn. 35 (3) (2022) 381–411, https://doi.org/10.1080/09588221.2019.1704787.

[23] Y.-S. Chang, K.-J. Hu, C.-W. Chiang, A. Lugmayr, Applying mobile augmented reality (AR) to teach interior design students in layout plans: evaluation of learning effectiveness based on the ARCS model of learning motivation theory, Sensors 20 (1) (2020) 105, https://doi.org/10.3390/s20010105.

[24] Y. Wahyu, I.W. Suastra, I.W. Sadia, N.K. Suarni, The effectiveness of Mobile augmented reality assisted STEM-based learning on scientific literacy and Students' achievement, Int. J. Instr. 13 (3) (2020) 343–356, https://doi.org/10.29333/iji.2020.13324a.

[25] Y. Chen, Interactive piano training using augmented reality and the internet of things, Educ. Inf. Technol. (2022), https://doi.org/10.1007/s10639-022-11443-4.

[26] H. Khorrami Shad, K. Tak Wing Yiu, R. Lovreglio, Z. Feng, State-of-the-art analysis of the integration of augmented reality with construction technologies to improve construction safety, Smart Sustain. Built Environ. (2022), https://doi.org/10.1108/SASBE-07-2022-0151.

[27] K. Kim, M. Alshair, B. Holtkamp, C. Yun, S. Khalafi, L. Song, M.J. Suh, Using immersive augmented reality to assess the effectiveness of construction safety training, J. Construct. Eng. Proj. Manag. 9 (4) (2019) 16–33, https://doi.org/10.6106/JCEPM.2019.9.4.016.

[28] D. Tatić, B. Tešić, The application of augmented reality technologies for the improvement of occupational safety in an industrial environment, Comput. Ind. 85 (2017) 1–10, https://doi.org/10.1016/j.compind.2016.11.004.

[29] J. Ahn, R. Han, An indoor augmented-reality evacuation system for the smartphone using personalized Pedometry, HCIS 2 (18) (2012), https://doi.org/10.1186/2192-1962-2-18.

[30] M.-K. Tsai, N.-J. Yau, Enhancing usability of augmented-reality-based mobile escape guidelines for radioactive accidents, J. Environ. Radioact. 118 (2013) 15–20, https://doi.org/10.1016/j.jenvrad.2012.11.001.

[31] M.-K. Tsai, Y.-C. Lee, C.-H. Lu, M.-H. Chen, T.-Y. Chou, N.-J. Yau, Integrating geographical information and augmented reality techniques for mobile escape guidelines on nuclear accident sites, J. Environ. Radioact. 109 (2012) 36–44, https://doi.org/10.1016/j.jenvrad.2011.12.025.

[32] S. Kanangkaew, N. Jokkaw, T. Tongthong, A real-time fire evacuation system based on the integration of building information modeling and augmented reality, J. Build. Eng. 67 (2023) 105883, https://doi.org/10.1016/j.jobe.2023.105883.

[33] S.-J. Yoo, S.-H. Choi, Indoor AR navigation and emergency evacuation system based on machine learning and IoT technologies, IEEE Internet Things J. 9 (21) (2022) 20853–20868, https://doi.org/10.1109/JIOT.2022.3175677.

[34] H. Chen, L. Hou, G. Zhang, S. Moon, Development of BIM, IoT and AR/VR technologies for fire safety and upskilling, Autom. Constr. 125 (2021) 103631, https://doi.org/10.1016/j.autcon.2021.103631.

[35] K. Iguchi, H. Mitsuhara, M. Shishibori, Evacuation instruction training system using augmented reality and a smartphone-based head mounted display, in: 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Vienna, Austria, 2016, 2016, pp. 1–6, https://doi.org/10.1109/ICT-DM.2016.7857220.

[36] J. Kawai, H. Mitsuhara, M. Shishibori, Game-based evacuation drill using augmented reality and head-mounted display, Interact. Technol. Smart Educ. 13 (3) (2016) 186–201, https://doi.org/10.1108/ITSE-01-2016-0001.

[37] C. Catal, A. Akbulut, B. Tunali, E. Ulug, E. Ozturk, Evaluation of augmented reality technology for the design of an evacuation training game, Virtual Reality 24 (2020) 359–368, https://doi.org/10.1007/s10055-019-00410-z.

[38] A.D. Kaplan, J. Cruit, M. Endsley, S.M. Beers, B.D. Sawyer, P.A. Hancock, The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a Meta-analysis, Hum. Factors 63 (4) (2021) 706–726, https://doi.org/10.1177/0018720820904229.

[39] E.Z. Barsom, M. Graafland, M.P. Schijven, Systematic review on the effectiveness of augmented reality applications in medical training, Surg. Endosc. 30 (10) (2016) 4174–4183, https://doi.org/10.1007/s00464-016-4800-6.

[40] D. Suresh, A. Aydin, S. James, K. Ahmed, P. Dasgupta, The role of augmented reality in surgical training: a systematic review, Surg. Innov. (2022), https://doi.org/10.1177/15533506221140506.

[41] H. Mitsuhara, C. Tanimura, J. Nemoto, M. Shishibori, Why Don't you evacuate speedily? Augmented reality-based evacuee visualisation in ICT-based evacuation Drill, in: 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE), 2018, pp. 893–899, https://doi.org/10.1109/TALE.2018.8615405.

[42] R.A. Schmidt, A schema theory of discrete motor skill learning, Psychol. Rev. 82 (4) (1975) 225–260, https://doi.org/10.1037/h0076770.

[43] G. Charness, U. Gneezyb, M. Kuhn, Experimental methods: between-subject and within-subject design, J. Econ. Behav. Organ. 81 (2012) 1–8, https://doi.org/10.1016/j.jebo.2011.08.009.

[44] J. Lazar, J.H. Feng, H. Hochheiser, Research Methods in Human-Computer Interaction, 2nd edition, Morgan Kaufmann, Amsterdam, 2017. ISBN: 978-0-12-805390-4.

[45] O.A. Meyer, M.K. Omdahl, G. Makransky, Investigating the effect of pre-training when learning through immersive virtual reality and video: a media and methods experiment, Comput. Educ. 140 (2019), https://doi.org/10.1016/j.compedu.2019.103603. Article 103603.

[46] R. Sacks, A. Perlman, R. Barak, Construction safety training using immersive virtual reality, Constr. Manag. Econ. 31 (9) (2013) 1005–1017, https://doi.org/10.1080/01446193.2013.828844.

[47] Z. Feng, R. Lovreglio, T.W. Yiu, D.M. Acosta, B. Sun, N. Li, Immersive virtual reality training for excavation safety and hazard identification, Smart Sustain. Built Environ. (2023), https://doi.org/10.1108/SASBE-10-2022-0235.

[48] G. Gonçalves, H. Coelho, P. Monteiro, M. Melo, Maximino Bessa, Systematic review of comparative studies of the impact of realism in immersive virtual experiences, ACM Comput. Surv. 55 (6) (2022), https://doi.org/10.1145/3533377. Article 115 (June 2023), 36 pages.

[49] L. Chittaro, R. Sioni, Serious games for emergency preparedness: evaluation of an interactive vs. a non-interactive simulation of a terror attack, Comput. Hum. Behav. 50 (2015) 508–519, https://doi.org/10.1016/j.chb.2015.03.074.

[50] M.C. Lam, M.J. Sadik, N.F. Elias, The effect of paper-based manual and stereoscopic-based mobile augmented reality systems on knowledge retention, Virtual Reality 25 (2021) 217–232, https://doi.org/10.1007/s10055-020-00451-9.

[51] G. Makransky, G.B. Petersen, Investigating the process of learning with desktop virtual reality: a structural equation modeling approach, Comput. Educ. 134 (2019) 15–30, https://doi.org/10.1016/j.compedu.2019.02.002.

[52] W. Tarng, Y.-S. Lin, C.-P. Lin, K.-L. Ou, Development of a lunar-phase observation system based on augmented reality and mobile learning technologies, Mob. Inf. Syst. vol. 2016 (2016), https://doi.org/10.1155/2016/8352791. Article ID 8352791, 12 pages.

[53] N. Xu, Y. Li, J. Lin, L. Yu, H.-N. Liang, User retention of Mobile augmented reality for cultural heritage learning, in: 2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), Singapore, Singapore, 2022, pp. 447–452, https://doi.org/10.1109/ISMAR-Adjunct57072.2022.00095.

[54] J. Yip, S.H. Wong, K.L. Yick, K. Chan, K.H. Wong, Improving quality of teaching and learning in classes by using augmented reality video, Comput. Educ. 128 (2019) 88–101, https://doi.org/10.1016/j.compedu.2018.09.014.

[55] R.W. Rogers, Cognitive and physiological processes in fear appeals and attitude change: A revised theory of protection motivation, in: J. Cacioppo, R. Petty (Eds.), Social Psychophysiology, Guilford Press, New York, 1983, pp. 153–177.

[56] Z. Feng, V.A. González, R. Amor, M. Spearpoint, J. Thomas, R. Sacks, R. Lovreglio, G. Cabrera-Guerrero, An immersive virtual reality serious game to enhance earthquake behavioral responses and post-earthquake evacuation preparedness in buildings, Adv. Eng. Inform. 45 (2020) 101118, https://doi.org/10.1016/j.aei.2020.101118.

[57] R. Lovreglio, D. Ngassa, A. Rahouti, D. Paes, Z. Feng, A. Shipman, Prototyping and testing a virtual reality counterterrorism serious game for active shooting, Int. J. Disaster Risk Red. 82 (2022) 103283, https://doi.org/10.1016/j.ijdrr.2022.103283.

[58] R. Schwarzer, M. Jerusalem, Generalized self-efficacy scale, in: In J. Weinman, S. Wright, M. Johnston (Eds.), Measures in Health Psychology: A user's Portfolio. Causal and Control Beliefs, NFER-NELSON, Windsor, UK, 1995, pp. 35–37.

[59] A. Bandura, Self-Efficacy: The Exercise of Control, W H Freeman/Times Books/ Henry Holt & Co., 1997.

[60] E.L. Deci, H. Eghrari, B.C. Patrick, D.R. Leone, Facilitating internalization: the self-determination theory perspective, J. Pers. 62 (1) (1994) 119–142, https://doi.org/10.1111/j.1467-6494.1994.tb00797.x.

[61] R.M. Ryan, E.L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, Am. Psychol. 55 (1) (2000) 68–78, https://doi.org/10.1037/0003-066X.55.1.68.

[62] J.M. Sáez-López, R. Cózar-Gutiérrez, J.A. González-Calero, C.J. Gómez Carrasco, Augmented reality in higher education: an evaluation program in initial teacher training, Educ. Sci. 10 (2) (2020) 26, https://doi.org/10.3390/educsci10020026.

[63] F. De Crescenzio, M. Fantini, F. Persiani, L. Di Stefano, P. Azzari, S. Salti, Augmented reality for aircraft maintenance training and operations support, IEEE Comput. Graph. Appl. 31 (1) (2011) 96–101, https://doi.org/10.1109/MCG.2011.4.

[64] D.H. Arjoni, I. de Souza Rehder, J.M. Pereira Figueira, E. Villani, Augmented reality for training formation flights: an analysis of human factors, Heliyon 9 (3) (2023) e14181, https://doi.org/10.1016/j.heliyon.2023.e14181.

[65] Q.T. Le, A. Pedro, C.R. Lim, H.T. Park, C.S. Park, H.K. Kim, A framework for using mobile based virtual reality and augmented reality for experiential construction safety education, Int. J. Eng. Educ. 31 (3) (2015) 713–725.

[66] T. Chandrasekera, The Effect of Augmented Reality and Virtual Reality Interfaces on Epistemic Actions and the Creative Process, Ph.D. thesis. University of Missouri–Columbia, 2015, https://doi.org/10.32469/10355/57754.

[67] D. Paes, J. Irizarry, A Usability Study of an Immersive Virtual Reality Platform for Building Design Review: Considerations on Human Factors and User Interface. CRC 2018 – Construction Research Congress, Louisiana State University, New Orleans, LA, United States, 2018, https://doi.org/10.1061/9780784481264.041. April 2-5.

[68] K.M. Stanney, M. Mollaghasemi, L. Reeves, R. Breaux, D.A. Graeber, Usability engineering of virtual environments (VEs): identifying multiple criteria that drive effective VE system design, Int. J. Human-Comput. Stud. 58 (4) (2003) 447–481, https://doi.org/10.1016/S1071-5819(03)00015-6.

[69] T.J. Lin, H.B.L. Duh, N. Li, H.Y. Wang, C.C. Tsai, An investigation of learners' collaborative knowledge construction performances and behavior patterns in an augmented reality simulation system, Comput. Educ. 68 (2013) 314–321, https://doi.org/10.1016/j.compedu.2013.05.011.

[70] B. Zhang, N. Robb, A comparison of the effects of augmented reality N-Back training and traditional two-dimensional N-Back training on working memory, SAGE Open 11 (2) (2021), https://doi.org/10.1177/21582440211014507.

[71] M. Wang, Generalized estimating equations in longitudinal data analysis: a review and recent developments, Adv. Stat. 303728 (2014), https://doi.org/10.1155/2014/303728.

[72] J. Cohen, Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Lawrence Erlbaum Associates, Hillsdale, NJ, 1988.

[73] J. Garzón, J. Acevedo, Meta-analysis of the impact of AR on students' learning gains, Educ. Res. Rev. 27 (1) (2019) 244–260, https://doi.org/10.1016/j.edurev.2019.04.001.

[74] M. Thees, S. Kapp, M.P. Strzys, F. Beil, P. Lukowicz, J. Kuhn, Effects of augmented reality on learning and cognitive load in university physics laboratory courses, Comput. Hum. Behav. 108 (2020), https://doi.org/10.1016/j.chb.2020.106316. Article 106316.

[75] J. Wolf, V. Wolfer, M. Halbe, F. Maisano, Q. Lohmeyer, M. Meboldt, Comparing the effectiveness of augmented reality-based and conventional instructions during single ECMO cannulation training, Int. J. Comput. Assist. Radiol. Surg. 16 (7) (2021) 1171–1180, https://doi.org/10.1007/s11548-021-02408-y.

[76] J. Buchner, K. Buntins, M. Kerres, The impact of augmented reality on cognitive load and performance: a systematic review, J. Comput. Assist. Learn. (2021), https://doi.org/10.1111/jcal.12617.

[77] S. Gargrish, D.P. Kaur, A. Mantri, G. Singh, B. Sharma, Measuring effectiveness of augmented reality-based geometry learning assistant on memory retention abilities of the students in 3D geometry, Comput. Appl. Eng. Educ. 29 (2021) 1811–1824, https://doi.org/10.1002/cae.22424.

[78] S. Brunot, R.B. Sanitioso, Motivational influence on the quality of memories: recall of general autobiographical memories related to desired attributes, Eur. J. Soc. Psychol. 34 (5) (2004) 627–635, https://doi.org/10.1002/ejsp.220.

[79] E. Krokos, C. Plaisant, A. Varshney, Virtual memory palaces: immersion aids recall, Virtual Reality 23 (2019) 1–15, https://doi.org/10.1007/s10055-018-0346-3.

[80] D.C. Park, J. Lodi-Smith, L. Drew, S. Haber, A. Hebrank, G.N. Bischof, W. Aamodt, The impact of sustained engagement on cognitive function in older adults: the synapse project, Psychol. Sci. 25 (1) (2014) 103–112, https://doi.org/10.1177/0956797613499592.

[81] R.B. Sanitioso, Motivated self and recall: visual perspectives in remembering past behaviors, Eur. J. Soc. Psychol. 38 (3) (2008) 566–575, https://doi.org/10.1002/ejsp.456.

[82] D. Paes, J. Irizarry, M. Billinghurst, D. Pujoni, Investigating the relationship between three-dimensional perception and presence in virtual reality-reconstructed architecture, Appl. Ergon. 109 (2023) 103953, https://doi.org/10.1016/j.apergo.2022.103953.

[83] N. Alexandre, M. Coluci, Content validity in the development and adaptation processes of measurement instruments, Cien. Saude Colet. 16 (7) (2011) 3061–3068, https://doi.org/10.1590/S1413-81232011000800006.

[84] M. Kinateder, P. Pauli, M. Müller, J. Krieger, F. Heimbecher, I. Rönnau, U. Bergerhausen, G. Vollmann, P. Vogt, A. Mühlberger, Human behaviour in severe tunnel accidents: effects of information and behavioural training, Transport. Res. F: Traffic Psychol. Behav. 17 (2013) 20–32, https://doi.org/10.1016/j.trf.2012.09.001.

[85] D. Paes, E. Arantes, J. Irizarry, Immersive environment for improving the understanding of architectural 3D models: comparing user spatial perception between immersive and traditional virtual reality systems, Autom. Constr. 84 (2017) 292–303, https://doi.org/10.1016/j.autcon.2017.09.016.

[86] D. Paes, J. Irizarry, D. Pujoni, An evidence of cognitive benefits from immersive design review: comparing three-dimensional perception and presence between immersive and non-immersive virtual environments, Autom. Constr. 130 (2021), https://doi.org/10.1016/j.autcon.2021.103849.