Fire Technology (2026) 62:50
https://doi.org/10.1007/s10694-026-01879-2

Comparison of AR Training Effectiveness Based on Directive Instruction and Productive Failure Teaching Theories: A Case Study in Fire Safety

Peizhen Gong[1] · Ying Lu[1] · Xiaer Xiahou[1] · Yunxuan Deng[1] · Ruggiero Lovreglio[2]

Received: 9 April 2025 / Accepted: 27 January 2026

Abstract
Effective fire safety training is critical to mitigating property losses and human injuries and deaths. Augmented Reality (AR) is an emerging solution for safety training, offering a dynamic and interactive learning environment. Despite the emergence of various AR training prototypes demonstrating potential advantages, there is limited research on how teaching theories impact the efficacy of augmented reality. This study aims to integrate and test two teaching theories, directive instruction and productive failure, into AR-based fire safety training. Two novel AR prototypes, directive instruction training and productive failure training, were designed and prototyped in this work. These two prototypes were tested in a controlled experiment involving 68 participants. A comparison was carried out, focusing on knowledge acquisition, knowledge retention, intrinsic motivation, and self-efficacy. The results indicate that productive failure training has better performance in enhancing knowledge acquisition and knowledge retention. However, both prototypes perform equally in the overall learning experience. As such, these findings offer valuable insights into future augmented reality development for safety training.

Keywords Safety training · Augmented reality · Fire safety

✉ Ying Lu
luying_happy@126.com

Peizhen Gong
pzgong@seu.edu.cn

Xiaer Xiahou
xhcmre@seu.edu.cn

Yunxuan Deng
deng.yx@seu.edu.cn

Ruggiero Lovreglio
R.Lovreglio@massey.ac.nz

[1] School of Civil Engineering, Southeast University, 2 Southeast University Rd., Jiangning District, Nanjing, Jiangsu Province 211189, China

[2] School of Built Environment, Massey University, Auckland, New Zealand

Published online: 25 February 2026

1 Introduction

Fires pose significant threats to the built environment, particularly in densely populated residential areas and complex commercial zones. Incidents occurring in these areas can quickly lead to numerous casualties and significant property damage [1, 2]. According to the World Fire Statistics Report published by the Geneva Association in 2014 [3], global fire losses account for about 1% of annual GDP. The direct loss alone reaches tens of billions of US dollars each year. In the United States, for example, the direct fire loss in 2010 was 13 billion US dollars, more than ten times the total economic loss caused by tornadoes in the same year [3, 4]. In 2023, the UK Home Office recorded a total of 178,737 fire incidents, indicating a notable 17% increase from the previous year's 152,639 cases [5]. Given these alarming trends, strengthening fire preparedness through safety training has become an essential strategy for mitigating potential losses. Empirical evidence shows that safety training not only reduces the likelihood of injuries but also helps reduce accident-related economic losses. Dong et al. [6] found that construction workers in Washington State who received safety training were less likely to file injury compensation claims. Yiu et al. [7] reported that adopting safety training in Hong Kong's construction industry reduced both accident frequency and associated costs.

Traditional safety training methods commonly use slide presentations, videos, or text-based materials [8–12]. While these methods have supported knowledge transfer and can contribute to memory retention, especially when well-designed, they also have certain limitations. Pedagogically, these approaches often rely on passive delivery modes that may reduce learner engagement and limit opportunities for experiential learning [11–13]. Reduced interactivity and limited contextualization can make it more difficult for trainees to internalize procedures or apply knowledge under stress [14, 15]. Moreover, traditional training methods frequently lack real-time feedback and adaptability, offering limited capacity to adjust instructional content according to individual learning needs [16, 17].

Augmented Reality (AR) technology, as an emerging educational tool, has the potential to revolutionize safety training and its effectiveness. By augmenting real-life scenarios and seamlessly blending virtual elements with the physical environment, AR technology can provide trainees with an interactive educational experience [18–22]. Several studies have highlighted the crucial role of AR in enhancing training efficacy, demonstrating its potential to significantly improve fire safety training effectiveness [12, 23–25]. For example, Gong et al. [13] found that AR-based metro construction safety training outperformed slide-based training and improved short-term knowledge acquisition of risk identification. Paes et al. [20] found that AR fire safety training enhanced intrinsic motivation and promoted better self-efficacy retention. Shringi et al. [26] found that AR visualizations helped learners retain knowledge more effectively than traditional classroom training. Dallasega et al. [27] found that AR in MEP installation training improved information management and increased adoption compared to conventional visual management methods.

Nevertheless, two significant research gaps persist. Firstly, while existing research has concentrated on advancements in AR technology and system enhancements, there are gaps in the explicit integration of established teaching theories into AR-based safety training [12]. This includes theories of Directive Instruction (DI) and Productive Failure (PF), which are rooted in various disciplines, such as statistics, behavioral sciences, and psychology [28–30]. DI theory stresses the importance of clear instruction and immediate feedback to

expedite trainees' skill acquisition [31, 32]. Conversely, PF theory advocates for creating error-rich scenarios to prompt trainees to contemplate and rectify mistakes, thereby fostering a deeper comprehension and mastery of fundamental principles [33]. Secondly, although the effectiveness of PF and DI has been demonstrated in traditional learning contexts, there remains a dearth of research on their application in AR-based fire safety training to optimize training outcomes. Rigorous research and assessments are crucial to evaluating their potential impact on enhancing training effectiveness. Bridging these research gaps will deepen the understanding of AR applications in fire safety training. It will also shed light on new ways to enhance training methodologies by integrating established teaching theories. As such, this study does not aim to re-establish whether PF is superior to DI, but seeks to fill an existing research gap by examining their applicability in AR-based fire safety training.

The primary aim of this work is to evaluate the pedagogical effectiveness of DI and PF within an AR learning environment. To achieve this, we develop two prototypes of AR fire safety training, DI AR and PF AR, with a focus on fire extinguisher inspection. The effectiveness of these prototypes is compared in terms of knowledge acquisition, knowledge retention, intrinsic motivation, and self-efficacy. As such, this study represents one of the first to compare teaching theories in the context of AR safety training. This comparative analysis reveals potential synergies between established teaching theories and AR technology. It also advances the understanding of using teaching theories to improve training effectiveness and user engagement in safety-critical situations. This work lays the groundwork for future developments in augmented reality within the wider domain of safety training practices.

2 Background

In this section, we review the application of AR in safety training in Sect. 2.1, while reviewing directive instruction and productive failure teaching theories in Sect. 2.2.

2.1 Application of AR in Safety Training

Workplace accidents and disasters often result from unsafe behaviors exhibited by workers, leading to severe outcomes, which can have serious consequences including fatalities and large financial losses. Safety training is critical in addressing this issue, aiming to enhance workers' safety awareness, promote safe behavioral practices, and impart essential safety knowledge [8, 12]. While traditional safety training approaches have been widely used and have proven effective in certain contexts, they may not always fully engage learners or support experiential learning [8, 13]. These methods are typically cost-effective, scalable, and easy to implement, making them suitable for various settings. However, such methods often rely on passive delivery modes, which can limit interactivity and reduce learner motivation and attention, particularly for complex or procedural tasks [9, 10, 20, 34].

AR technology has emerged as a potential tool for safety training, leveraging its ability to seamlessly integrate digital information into real-world environments. AR transcends the limitations of traditional classroom settings by creating interactive learning environments. This enables trainees to directly engage with virtual elements in realistic scenarios [18, 20–22, 35]. By providing trainees with opportunities to practice skills in a secure virtual space,

AR can enhance the adaptability and effectiveness of training while mitigating workplace risks. AR technology has found widespread application in various safety training fields. For example, in the construction industry, Gong et al. [13] introduced an innovative AR safety training prototype for subway construction. Their work demonstrates the superiority of this prototype over traditional methods in quickly acquiring risk identification knowledge and retaining crucial risk identification, assessment, and response skills over time. Wu et al. [9] developed an AR application based on cognitive ergonomics to improve on-site assembly efficiency for construction personnel. This application shows promise in improving workers' skills in tasks such as tying steel bars and enhancing their overall performance. In the manufacturing sector, Aivaliotis et al. [36] created an AR application to facilitate operator interaction with flexible mobile robot assistants, offering a virtual interface with essential information on processes and production status. Chu and Ko [37] focused on spatially constrained AR assembly tasks. Their study confirmed the effectiveness of various supplementary information in AR applications, such as the assembly interface, manual dexterity guidance, and component handling instructions. These examples highlight the remarkable adaptability and practicality of AR technology in meeting safety training needs across diverse industries.

However, alongside its advantages, AR-based training also faces several practical and pedagogical challenges that merit attention. First, the translation of AR-acquired knowledge into real-world emergency behaviour remains underexplored, especially in high-stakes and time-sensitive situations. While virtual environments offer realism, they may not fully replicate the stress, unpredictability, and physical demands of actual emergencies [38, 39]. In addition, AR training often requires significant development resources, including custom content design and device calibration, which may hinder scalability and adoption in resource-limited settings [40, 41].

Existing research confirms the promise of AR for safety training, but gaps remain in its application to the fire safety domain. Currently, only a limited number of AR applications in fire safety training address topics such as safety sign recognition, decision-making on safe behaviors, and preparation for evacuations [20, 42, 43]. These applications primarily focus on short-term training outcome assessments, such as knowledge acquisition, while neglecting essential aspects of long-term knowledge retention. To address this gap, this study focuses on AR-based fire extinguisher inspection training within the fire safety sector. The study aims to evaluate rapid knowledge acquisition post-training and knowledge retention over a four-week period. This study seeks to illuminate the potential of AR technology in enhancing the efficacy of fire safety training practices.

2.2 Directive Instruction and Productive Failure Teaching Theories

DI theory and PF theory were selected because they represent contrasting pedagogical approaches—DI as a traditional, structured method and PF as an innovative, constructivist approach [28, 29, 32]. This contrast allows for a meaningful comparison of how different teaching methods affect training effectiveness in the context of AR-based fire safety training. Additionally, both theories are well-established in education and have been previously applied in skill-based and technical domains, making them suitable for comparison in the context of procedural learning tasks such as fire extinguisher inspection.

DI theory embodies a teaching approach that equips learners with clear directives on actions to take and how to perform them. This method entails offering detailed, step-by-step guidance and unambiguous instructions, leaving little room for interpretation or autonomous decision-making [31, 32]. In DI instruction, educators take on a proactive role by providing explicit guidance, demonstrations, and explanations, thus reducing ambiguity and enhancing task accuracy. DI is often employed when learners are unfamiliar with a subject or skill, offering the necessary explicit direction for accurate knowledge acquisition or task execution. One of the primary advantages of DI lies in its capacity to deliver a straightforward learning experience, thereby diminishing cognitive load and augmenting learning efficiency [44, 45]. By delineating a well-defined path for learners, DI streamlines the learning process, enabling learners to comprehend intricate concepts and execute tasks with greater precision. Moreover, the structured nature of DI cultivates a sense of security and confidence among learners, as they receive unambiguous guidance that minimizes confusion and uncertainty [31, 32, 46]. Through the provision of clear instructions, educators empower students to methodically approach challenges, thereby enhancing their problem-solving skills and task performance. Despite its advantages, DI may potentially limit learners' autonomy and problem-solving skills by emphasizing adherence to instructions over fostering independent thinking or decision-making abilities [32, 47]. As a result, DI is frequently categorized as a passive instructional strategy.

The PF theory presents a novel teaching approach that prioritizes active learning and problem-solving ahead of explicit instruction [28–30, 33]. In the implementation of PF, learners are initially tasked with addressing intricate challenges independently, without immediate explicit guidance. These initial phases of exploration and problem-solving aim to ignite learners' cognitive engagement and hone their critical thinking abilities [48, 49]. A key strength of PF theory lies in its capacity to foster profound learning through the experiential process of encountering and surmounting obstacles [50–52]. By empowering learners to confront challenges, PF cultivates resilience in the presence of ambiguity and intricacy. This method urges learners to actively explore diverse solutions through an iterative problem-solving loop, scrutinizing outcomes and honing their strategies. Consequently, learners cultivate a deeper comprehension of fundamental concepts and principles, facilitating enduring retention and the application of knowledge in novel contexts. Moreover, PF theory nurtures metacognitive skills by prompting learners to reflect on their problem-solving methodology, recognize misconceptions, and enhance strategies based on feedback and self-evaluation [53, 54]. This metacognitive consciousness augments learners' capability to monitor and regulate their learning, nurturing a sense of self-assurance and independence in their educational odyssey. PF has found successful implementation in various educational settings. For example, Song and Kapur [51] compared the PF flipped classroom model with the traditional model in a two-week multi-course unit at a secondary school in Hong Kong. The results indicated that students in the PF group showed a strong conceptual understanding. Palominos et al. [55] examined nursing students' perceptions of PF simulations, emphasizing their effectiveness in enhancing knowledge and skills. Kerrigan et al. [50] noted that engaging in PF processes within a flipped mathematics classroom sparked students' intellectual curiosity as they tackled invention tasks.

Several previous studies [29, 30, 49, 54] have shown that PF theory can assist learners in acquiring and retaining knowledge more effectively after facing initial challenges, in contrast to DI theory. Lu et al. [56] reported similar findings in the context of Virtual Reality

(VR). However, research on integrating DI and PF theories into AR-based safety training remains limited, and several challenges persist. For instance, implementing PF theory in AR safety training requires a well-structured framework to ensure that failure serves as a learning opportunity rather than a source of discouragement. Conversely, applying DI theory in AR poses the challenge of maintaining learner engagement, as overly prescriptive instructions may diminish the exploratory benefits of AR-based training. Furthermore, the comparative effectiveness of these two instructional approaches in AR-based learning environments remains unclear and needs further investigation. This study endeavors to develop two AR prototypes based on DI theory and PF theory, subsequently evaluating their effectiveness.

3 Materials and Methods

This study aims to develop and evaluate prototypes for AR-based safety training systems tailored for fire extinguisher inspection. Fire extinguisher inspection was chosen due to its critical role in fire safety and the frequent evidence of inadequate maintenance [57]. In many buildings, the use of a fire extinguisher is the main strategy to contain a fire when it is still small enough. Therefore, it is paramount that all these extinguishers are correctly maintained. This task is operationally relevant because errors in inspection can directly compromise fire response effectiveness. In addition, the procedure has a relatively high error occurrence in practice, making it a suitable target for training interventions [58]. The structured, step-by-step nature of extinguisher inspection makes it ideal for procedural learning, allowing trainees to focus on mastering technical details in a low-risk environment [59].

The target audience for this training system is professionals with expertise in fire safety. The primary objective is to equip them with an understanding of the correct procedures for inspecting fire extinguishers. This skill is known to be challenging to master due to the detailed and technical nature of inspection steps. The training consists of a series of tasks designed to promote effective knowledge acquisition and retention. By harnessing the capabilities of AR technology, these systems create dynamic and interactive virtual environments that enable trainees to actively participate in realistic inspection scenarios. A controlled between-subjects experiment was conducted to evaluate their effectiveness. While this research is situated within the broader context of fire safety, it focuses specifically on fire extinguisher inspection, representing a critical component of comprehensive fire safety training.

3.1 Experimental Design

This study compares the efficacy of two different training methods, namely DI AR training and PF AR training, using a between-subjects experimental approach. Each participant receives individual training and is randomized to either the PF AR or DI AR group.

The training mode, which leads to the creation of the two experimental circumstances, is the independent variable in our study. Both groups receive training on fire extinguisher inspection using the HoloLens 2, with differences solely in teaching methods as outlined in Sect. 3.2. The dependent variables focus on participants' learning performance, including knowledge, intrinsic motivation, and self-efficacy levels. Data are collected at three specific

time points: pre-training, immediately after training, and 4 weeks after training. This study calculates the change in knowledge levels from pre-training to immediately after training to evaluate knowledge acquisition. Similarly, comparing the knowledge levels immediately after training with those measured four weeks later allows for the evaluation of knowledge retention. The evaluation of self-efficacy and intrinsic motivation follows the same method. Detailed information regarding the questionnaire items across the three time points is provided in Sect. 3.3. It is worth noting that similar studies [8, 13, 20, 60] typically employ a four-week interval to assess retention, hence the selection of a 4-week duration in this study. The experimental design is displayed in Fig. 1.

3.2 AR Prototype Design

The AR prototype was built on a Windows computer utilizing a Microsoft HoloLens 2 AR head-mounted display and the Unity 3D game engine version 2019.4.40f1c1. The HoloLens 2 provides a suite of features that enhance the augmented reality experience. These include a transparent holographic lens for visualization, a 2 K 3:2 light engine that enables high-resolution rendering, and a holographic density exceeding 2.5 K radiant, which reflects the angular resolution of holograms. Noteworthy functionalities of the HMD encompass eye-based rendering, which optimizes 3D display based on the user's eye position [61].

The AR system was developed on a Windows computer equipped with specifications capable of meeting the requirements of AR development. The setup featured an NVIDIA GeForce GTX 4060 GPU, an Intel Core i7-14700HX 5.5 GHz CPU with 20 cores and 28 threads, 16 GB of DDR5 RAM, and the Windows 10 operating system. For prototyping and user interface (UI) design, the widely adopted Microsoft Mixed Reality Toolkit (MRTK) was utilized. Unity's long-term support version, Unity 2019.4 LTS, served as the primary development platform throughout the research, ensuring a stable and reliable environment for creating AR system prototypes.

The implemented AR system was evaluated in a controlled setting within a soundproof conference room. To maintain consistent training conditions, the room's brightness was set between 300 and 500 lx to ensure adequate lighting, while the temperature was regulated to fall within the range of 20 to 22 ℃. The AR system training sessions were tailored to guide participants in correctly conducting fire extinguisher inspections. The training included

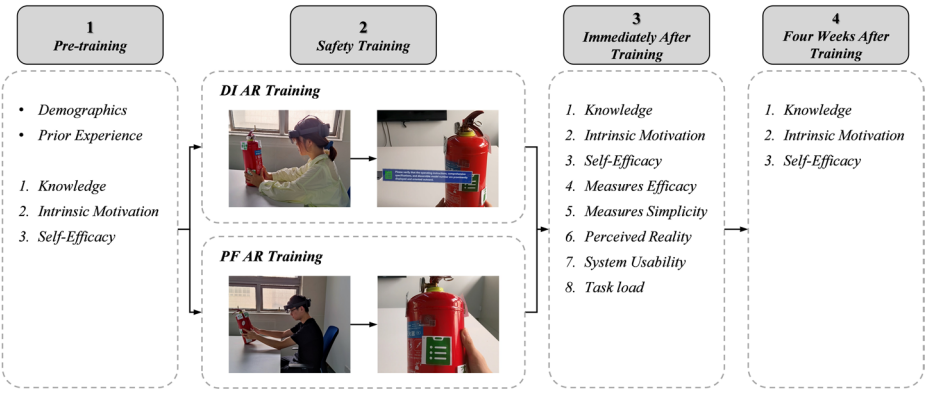

Fig. 1 The experimental design

eight operations carefully chosen to meet fire safety goals. These operations aimed to provide participants with practical knowledge and skills for inspecting fire extinguishers:

- Step 1: Verify that the operating instructions, comprehensive specifications, and discernible model number are prominently displayed and oriented outward;
- Step 2: Examine the fire extinguisher for physical damage, corrosion, or leakage;
- Step 3: Ascertain whether the fire extinguisher has surpassed its designated expiration date;
- Step 4: Observe the position of the needle on the gauge and ascertain whether it resides within the designated green area;
- Step 5: Verify the proper positioning of the safety pin, ensuring that it is securely locked;
- Step 6: Inspect the hose connected to the fire extinguisher, meticulously assessing its condition for any indications of breakage, looseness, blockage, or leakage;
- Step 7: Gently invert the fire extinguisher to facilitate the loosening of the dry powder contained within;
- Step 8: Attach the signature and the current date to the inspection tag, which should be securely fastened to the fire extinguisher.

Figures 2 and 3 depict the design of the DI AR and PF AR prototypes. The developments of the DI AR and PF AR prototypes were guided by DI and PF theories. These prototypes differed in two key aspects: (1) the timing of instructions and feedback and (2) the integration of iterative failures in PF AR. At the start of the AR experience, participants encountered a non-player character (NPC) in the form of a firefighter alongside a UI. This firefighter NPC was designed with a gender-neutral appearance to avoid reinforcing gender stereotypes and to promote inclusivity in the training environment. It provided verbal feedback and directed the participant towards the fire extinguisher on the table. The fire extinguisher featured eight pictures, each representing one of the eight inspection actions that need to be carried out. Specifically:

(1) In the DI AR prototype, only one instructional stage was included. Participants used the HoloLens 2 to view eight UI instructional message boxes pinned to corresponding fire extinguisher pictures. These instructions guided participants in performing the required inspection actions. Subsequently, participants were directed to a specific area where they could access a UI summary message box outlining the correct steps for inspecting the fire extinguisher.
(2) In the PF AR prototype, both an exploration phase and an instructional phase were included. This design was intentionally aligned with the core principles of PF theory. In the exploration phase, participants viewed images of the fire extinguisher inspection process through the HoloLens 2. No instructional message boxes were provided. Participants had to complete the inspection tasks using only their prior knowledge. This phase was designed to create cognitive conflict by placing learners in an ill-structured and ambiguous situation. They were required to make decisions under uncertainty, which is a key mechanism in PF theory. After the inspection, participants entered the instructional phase. They moved to a designated area and accessed the UI self-check message box. This interface aimed to promote reflection. It presented eight inspection-related questions, such as “Have you verified that the operating instructions, comprehensive

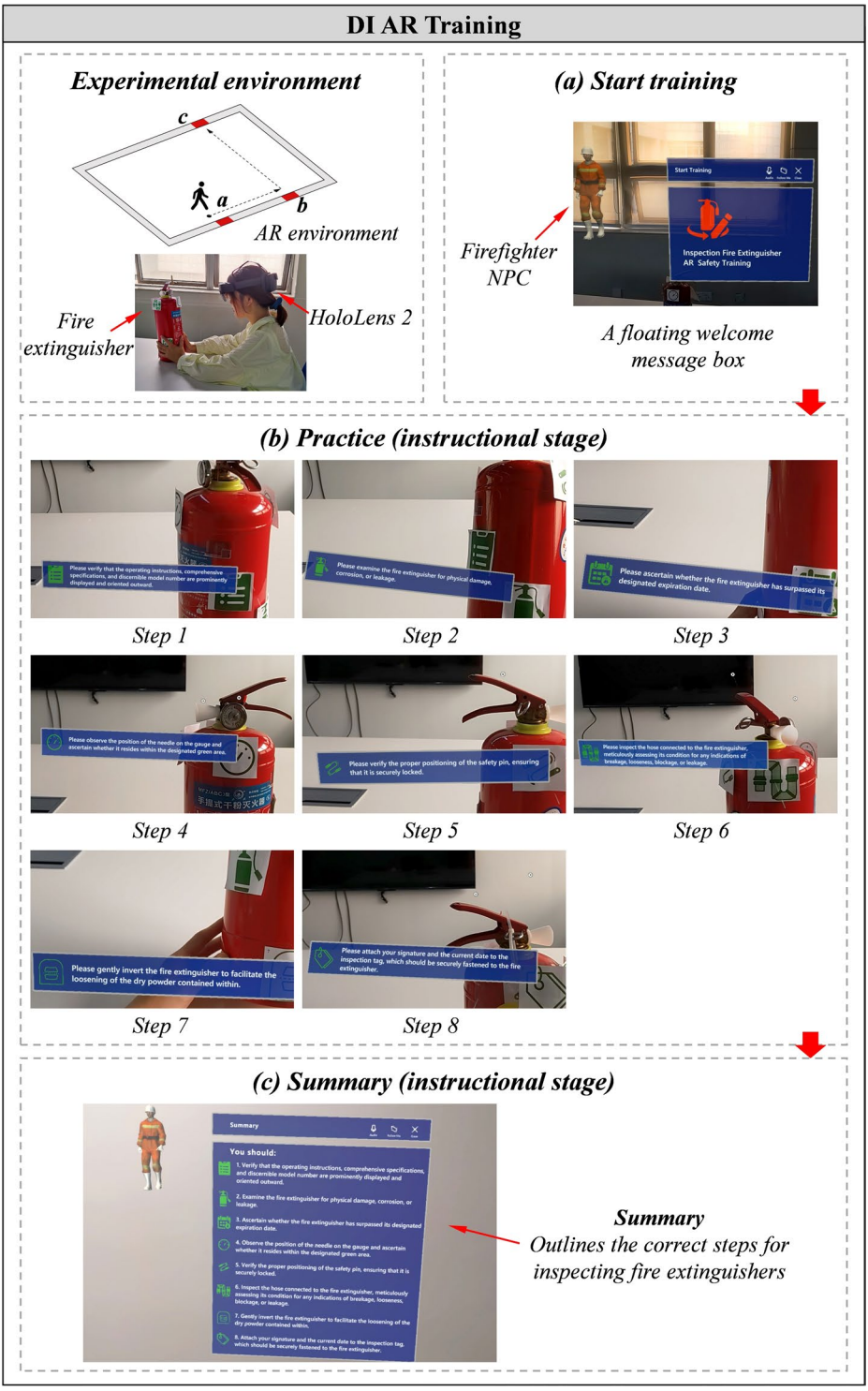

Fig. 2 DI AR prototype design: **a** start training; **b** practice; **c** summary

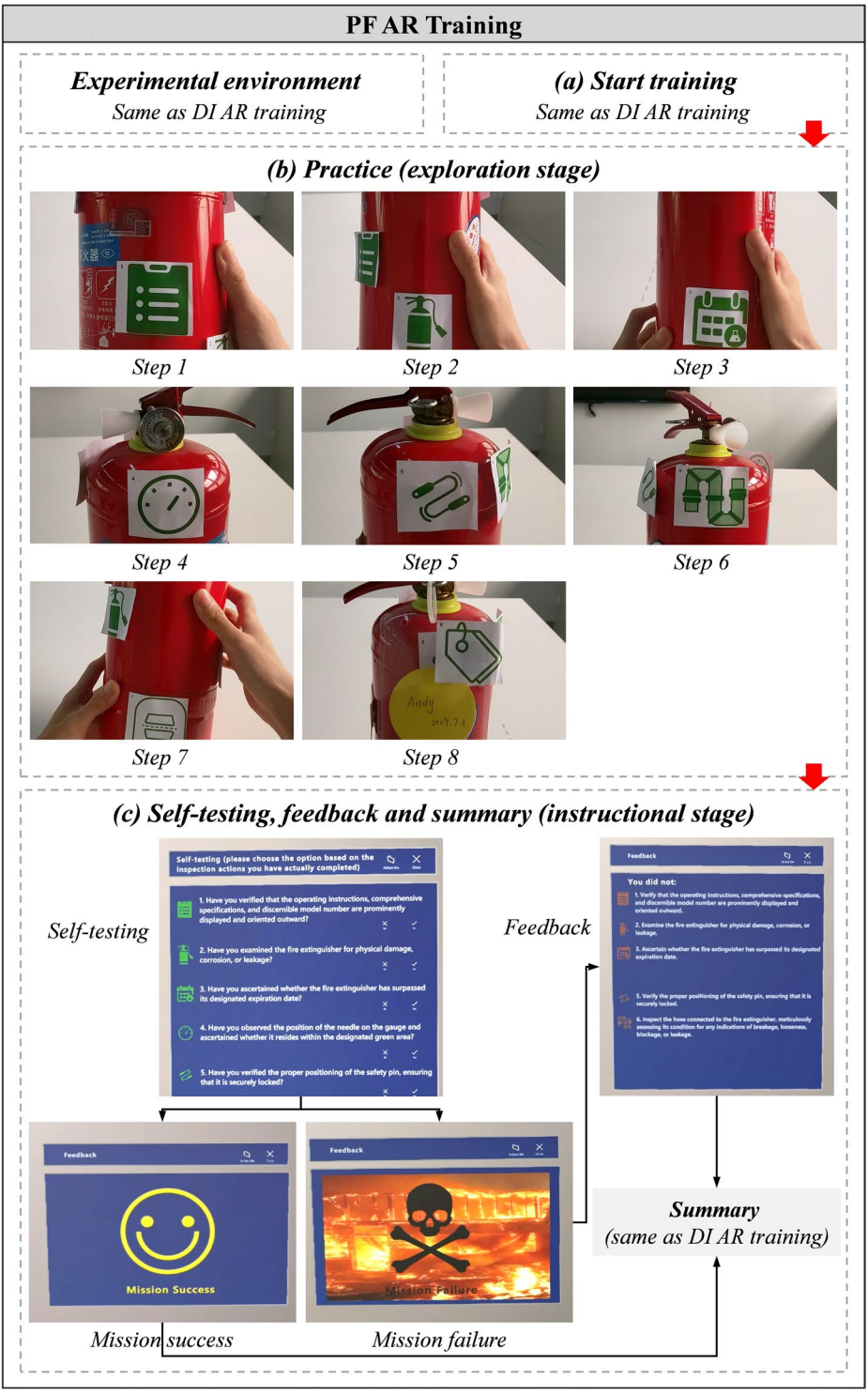

Fig. 3 PF AR prototype design: **a** start training; **b** practice; **c** summary

specifications, and discernible model number are prominently displayed and oriented outward?" Participants responded using check marks (√) for completed tasks or crosses (×) for missed ones. These prompts reflected real-world inspection standards and required participants to think critically about their actions. After the self-check, participants received feedback through a UI message box. If all tasks were correct, a success message appeared. If any tasks were incorrect or missed, a failure message listed the specific errors. This error-driven feedback helped participants identify gaps between their assumptions and standard procedures. It supported the process of revising and rebuilding their understanding. Next, participants entered a UI summary message box. This step showed the correct inspection steps, like the DI AR prototype. However, in the PF design, this occurred only after learners had experienced challenge and feedback. Participants' answers in the self-check had to match their actual inspection behavior. This reinforced accountability and discouraged superficial engagement. The overall design of the PF AR prototype was grounded in PF theory. Exploration before instruction, delayed feedback, self-assessment, and realistic failure scenarios were not random features. They were essential elements aimed at increasing cognitive effort, encouraging active learning, and simulating real workplace tasks. Together, these features supported deeper understanding and improved decision-making in safety-critical contexts. In both AR prototypes, participants could use gestures to choose and execute what they believed to be the correct actions.

Vuforia image tracking technology was utilized to generate anchors, termed image targets, in the research setting. Vuforia image tracking technology, developed by PTC Inc., is a leading AR software development kit that has been widely adopted for AR applications [62]. It primarily leverages advanced computer vision algorithms to detect and track planar images in real-time. The technology analyzes unique visual features such as edges, corners, and textures within these targets to accurately estimate their spatial position and orientation relative to the AR device's camera. This capability enables the stable and precise overlay of virtual 3D content onto corresponding physical surfaces, creating an immersive and interactive user experience. As shown in Fig. 4, these image targets consisted of eight operational images, about 5 cm in length and width, attached to the fire extinguisher. Each operational image was strategically positioned to correspond with a specific task, ensuring a seamless integration of the AR experience with the physical surroundings. While wearing a HoloLens 2 and looking at the image target, the AR application identified the image target and

Fig. 4 Examples of anchors. **a** Anchor 1, **b** Anchor 2

automatically generated a UI instructional message box next to it for the user. As the user's viewpoint changed, new images and UI instructional message boxes dynamically appeared. For example, an image target was situated near the operational instructions, indicating the necessity to verify that the operating instructions, comprehensive specifications, and discernible model number are prominently displayed and oriented outward. Figures 2 and 3 present a series of screenshots illustrating the AR environment.

Prior to initiating experiments, calibration was essential to ensure the accurate alignment of various elements within the AR environment, including image targets, floating UI message boxes, and a firefighter NPC. The eight images were sequentially attached to the operational instructions, appearance, production date, pressure gauge, safety pin, hose, bottom, and inspection label of the fire extinguisher. It is important to note that the calibration process allows for customization and adjustments to meet specific requirements. Figure 5 outlines the operational framework of the AR prototypes.

To offer customization, the AR prototypes integrated a versatile interaction mode manager designed to facilitate a range of user engagements. This manager controlled the playback of audio clips in each scene of the application and managed the rendering and closure of UI message boxes. The initial scene served as a configuration platform, allowing participants to adjust various parameters to suit their preferences. These parameters included the dimensions, placement, and associated configurations of the floating UI message box text, as well as the validation of the Vuforia state. In addition to visual enhancements, audio elements played a significant role in enhancing the AR experience. Carefully recorded audio clips corresponding to the various UI message boxes have been seamlessly integrated into the prototypes. These audio cues could be played at the proper moment, providing authentic sound effects.

3.3 Data Collection Instruments

Three questionnaires were given out at three distinct times during the data collection process: pre-training, immediately after training, and 4 weeks after training. Every questionnaire is

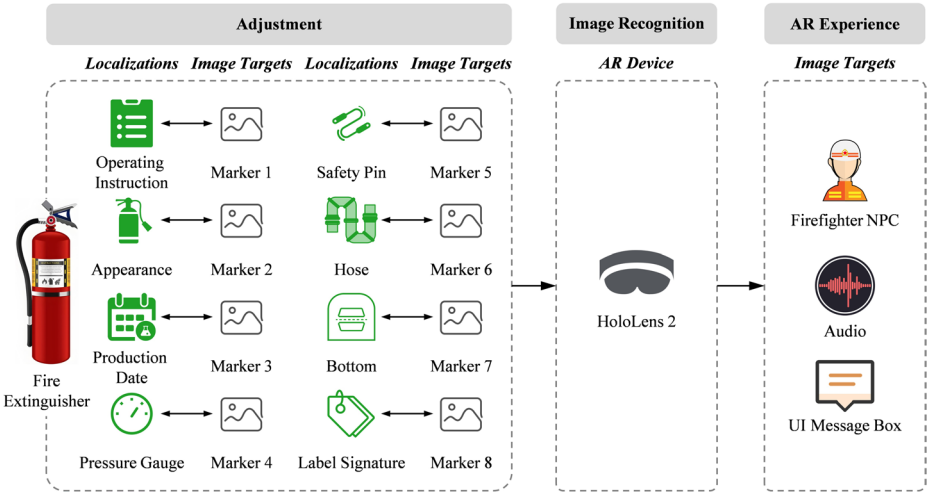

Fig. 5 AR prototype

carefully constructed to collect certain data and replies from participants. The items in these questionnaires are structured to gather data on the following aspects: (1) demographics, (2) prior experience, (3) knowledge, (4) self-efficacy, (5) intrinsic motivation, (6) measures efficacy, (7) measures simplicity, (8) perceived reality, (9) system usability, and (10) task load. These items encompass details regarding the participants' personal backgrounds, levels of knowledge, and user experiences. Gathering this data can provide a more comprehensive understanding of how various experimental conditions affect the participants. Several studies [8, 13, 20, 34, 63, 64] have utilized these specific items for data collection purposes. The pre-training questionnaire covers groups 1 to 5. The questionnaire administered just after the training includes groups 3 to 10. The questionnaire distributed 4 weeks after training involves groups 3 to 5. This approach is crucial in gaining a thorough understanding of the research variables and their implications [13, 20, 65]. The items in the questionnaires are detailed in Table 1.

It should be noted that, to evaluate participants' knowledge, we included an open-ended question: "If assigned to inspect a fire extinguisher, what steps would you take?" Using open-ended questions is a well-established procedure [11, 20, 66]. This solution is preferred over closed-ended questions as it reduces the chance of guessing and helps uncover misconceptions that may otherwise remain hidden. Responses were scored on a scale from 0 to 8, based on the number of accurate procedural steps provided, corresponding to the steps outlined in Sect. 3.2. Each correctly mentioned step earned one point. For example, a response such as "Check for damage, look at the pressure, make sure it's not expired, and invert it" received 4 points, while "Read the label, check for damage and leaks, confirm expiration, check the pin, inspect the hose, and invert the extinguisher" received 6 points. Table 2 illustrates an example of the scoring procedure used in the pre-training questionnaire.

All responses were assessed by a rater with expertise in fire extinguisher inspection. Moreover, a second rater with professional experience independently evaluated 90 responses. The inter-rater Pearson correlation coefficient was 0.9816, and the Cohen's kappa was 0.8787. These values indicate a very high level of agreement, confirming the reliability of the measurement method.

3.4 Participants and Experiment Session

For this study, a sample of 68 participants was chosen based on predefined eligibility criteria. Participants had to be at least eighteen years old and have at least a high school education to meet the inclusion requirements. Using the G*Power software, an a priori power analysis was performed to establish the required sample size. The findings from the analysis revealed that a minimum of 26 participants per group would be needed, considering a large effect size of 0.8, a statistical power of 0.8, and a significance level of 0.05 when using an independent sample t-test. The sample size for this study is consistent with those employed in previous between-subject research in the field. Noteworthy sample sizes include 30 participants in Hong et al. [65], separate experiments with 20 and 25 participants in Sacks et al. [8], 50 participants in Paes et al. [20], and 60 participants in Abbas et al. [77].

Participant consent and ethical approval were obtained before the commencement of training. Participants were directed not to drink alcohol the day before the experiment to optimize engagement and ensure safety during training. Two experimental conditions were randomly assigned to a total of 68 participants, with 34 individuals assigned to each group.

Random assignment was employed to mitigate biases and ensure an equitable distribution of characteristics across groups. In each experimental condition, participants received individual 15 min training to avoid any potential interference or interaction among participants. To ensure the integrity of the study results, participants were specifically directed not to engage in any other fire extinguisher inspection-related safety training activities for four weeks post-training.

3.5 Participant Demographics

For this study, 68 participants were recruited. A chi-square test was used to evaluate the distinction between the PF AR group and the DI AR group, as shown in Table 3. The chi-square test verified that there were no statistically significant differences in the two groups concerning gender, age, education level, previous fire safety training experience, and video game experience ($p > 0.05$).

While these results do not confirm equivalence, they suggest that the two groups were reasonably comparable with respect to these characteristics. Furthermore, the results indicated that most participants either possessed a high school diploma or a college degree. They had also taken part in one or more fire safety trainings and were familiar with video gaming.

4 Results

The purpose of the data analysis is to examine the following: (a) whether knowledge, intrinsic motivation, and self-efficacy at three distinct time points differ significantly between the two experimental conditions (as described in Sect. 4.1, 4.2, and 4.3); and (b) whether there are significant differences regarding measures efficacy, measures simplicity, perceived reality, system usability, and task load (as detailed in Sect. 4.4). No traditional-method control group was included, so the study does not directly compare AR with conventional training.

A homogeneity of variance test and a normality test were performed to accomplish these goals. The independent sample t-test was employed for significance analysis if the p-values for both tests were more than 0.05, indicating adherence to the homogeneity and normality assumptions. Conversely, the Mann-Whitney U test was utilized if the p-value was 0.05 or below, indicating the violation of the assumptions. In both cases, a p-value less than 0.05 was considered to indicate a statistically significant difference. A p-value equal to or greater than 0.05 was interpreted as not providing sufficient evidence to conclude a significant difference. To account for multiple comparisons, Bonferroni correction was applied. Given a total of seven independent tests (three between-group comparisons and four within-group comparisons), the adjusted significance threshold was set at $\alpha = 0.05/7 = 0.0071$. In addition, Cohen's delta parameter (d) was calculated for each comparison to check the effect size of differences between groups. In education, an effect size of 0.4 or above is generally regarded as sufficient to justify the implementation of a new educational policy. Moreover, the figures in Sect. 4.1 to 4.4 utilized rectangles to represent mean scores and dots to illustrate individual data points.

Table 1 Item properties

No.	Items	Descriptions	Scores range	Calculation method	Source
1	Demographics	This section includes two items collecting participants' characteristics, specifically gender, age, and education level.	–	–	[13, 20]
2	Prior experience	This section consists of two items designed to gather data on participants' experiences with fire extinguisher safety training and video games.	–	–	[13, 20]
3	Knowledge	This section includes an open-ended question to assess participants' understanding of the steps required for inspecting a fire extinguisher. Participants were presented with the question, "If assigned to inspect a fire extinguisher, what steps would you take?"	Scores range from 0 to 8, depending on the number of unique actions in the participant's response (Steps 1 to 8 in Sect. 3.2).	Summation.	[20, 66]
4	Self-efficacy	This section includes five items that assess participants' confidence levels in understanding basic and complex fire safety concepts, performing well in fire safety training, expectations for performance in fire safety training, and their belief that they had acquired the necessary skills from prior fire safety courses.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[67–70]
5	Intrinsic motivation	This section includes five items that assess participants' enjoyment, interest, potential boredom, concentration, and overall engagement in fire safety training activities.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[70–72]
6	Measures efficacy	This section has two items, evaluating how participants perceived the usefulness of the fire extinguisher inspection measures taught and their confidence in these measures guiding correct procedures in real-world scenarios.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[20, 59]
7	Measures simplicity	This section includes three items designed to capture participants' perceptions of the ease of learning, remembering, and applying the training measures.	Participants rated their responses on a 7-point Likert scale, with −3 denoting "very difficult" and +3 denoting "very easy."	Mean.	[20, 59]
8	Perceived reality	This section consists of three items aimed at capturing participants' perceptions of the realism of specific elements within the AR training environment. These items assessed how participants view the virtual firefighter NPC, sound effects, and floating message box effects.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[20, 73]
9	System usability	Ten items make up this section, which aims to gauge participants' impressions of their proficiency using and interacting with the AR prototype. These items assess participants' perceptions of various usability aspects.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[13, 20, 74]

Table 1 (continued)

No.	Items	Descriptions	Scores range	Calculation method	Source
10	Task load	This segment incorporates six items drawn from the NASA Task Load Index (TLX), a reliable and validated tool for assessing participants' perceived task load.	Participants' responses were rated on a 7-point Likert scale, with −3 denoting "strongly disagree" and +3 denoting "strongly agree."	Mean.	[13, 75, 76]

Table 2 Example of scoring procedure

Participant	#1	#2	#3	#4	#5	#6	…	#68
Step 1	1							1
Step 2								
Step 3		1						
Step 4		1						
Step 5				1				1
Step 6				1				1
Step 7	1				1			1
Step 8								1
Total	2	2	0	2	1	0	…	5

Table 3 Participant demographics

Parameter	DI AR group ($n=34$)	PF AR group ($n=34$)	Significance test
Gender			
Woman	15	16	$\chi^2=0.059$
Man	19	18	$p=0.808$
Age			
Under 25 years old	8	7	$\chi^2=0.697$
26–35 years old	18	21	$p=0.874$
36–45 years old	6	4	
Over 45 years old	2	2	
Education level			
High school degree	15	13	$\chi^2=0.250$
Undergraduate degree	12	13	$p=0.883$
Master degree	7	8	
Previous fire safety training experience			
4 times and more	1	1	$\chi^2=1.367$
3 times	2	1	$p=0.897$
2 times	7	9	
1 time	18	19	
0	6	4	
Experience with video games using a smartphone, game console, or computer			
More than once a day	9	10	$\chi^2=1.905$
Once a day	7	8	$p=0.910$
Once a week	5	6	
Once a month	6	5	
Once a year	5	2	
Never	2	3	

The participant count remained consistent throughout the study, ensuring uniformity across all phases of data collection without any dropouts. Using the two AR prototypes, none of the participants expressed being uncomfortable or having motion sickness.

4.1 Analysis of Knowledge

The purpose of this section is to assess the knowledge scores that the participants of the PF AR and DI AR groups received at three different time points. Given the non-normal distribution of the data ($p<0.05$), group differences were analyzed using the Mann–Whitney U test along with effect size d.

Upon analyzing the data across the different time points, it was found that when comparing the knowledge scores prior to and immediately after training, there was a significant improvement ($p<0.001$) in each group. However, the larger effect size (d = −3.015) for the PF AR group further supports its stronger impact on learning outcomes. When comparing scores obtained immediately after training to those measured 4 weeks later, a significant decline was observed in PF AR group. These findings are illustrated in Fig. 6; Table 4.

At the pre-training stage, no statistically significant difference was observed between the DI AR and PF AR groups ($p=0.707$). However, the PF AR group demonstrated significantly higher knowledge scores both immediately ($p<0.001$, d = −0.886) and 4 weeks after training ($p<0.001$, d = −0.924) compared to the DI AR group. These effect sizes indicate that the PF AR training had a practically meaningful impact on knowledge acquisition and retention.

4.2 Analysis of Self-efficacy

Figure 7; Table 5 present the self-efficacy scores of the participants. A normality test indicated that the scores did not conform to the assumption of a normal distribution, with a

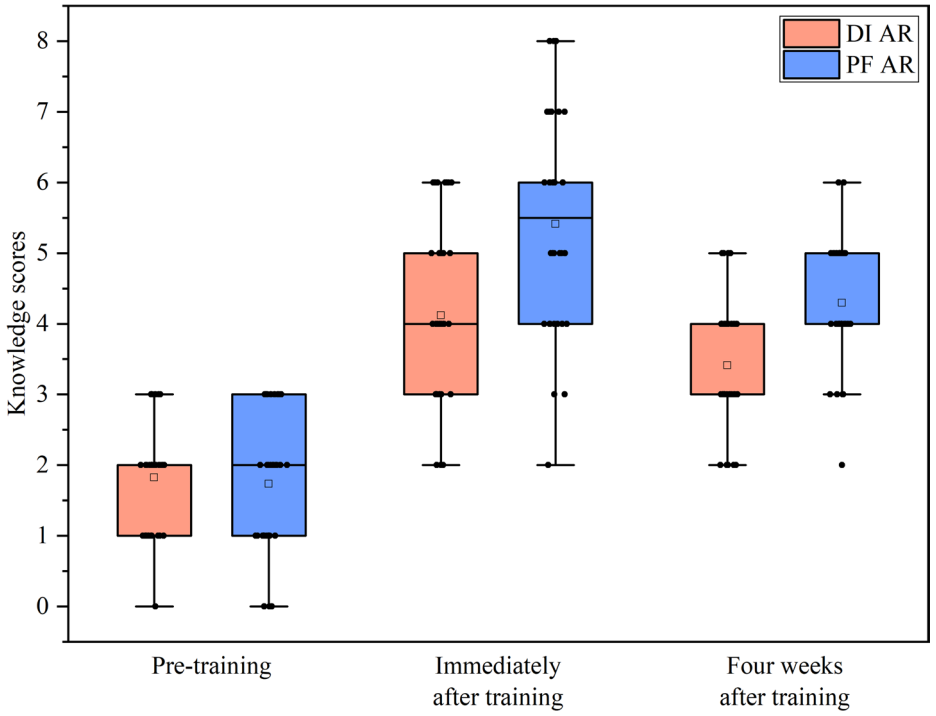

Fig. 6 Knowledge scores

Table 4 Comparison of knowledge scores in DI AR and PF AR training settings

Time points	Parameter	DI AR	PF AR	DI AR vs. PF AR
Pre-training	N	34	34	Mann-Whitney U = 549.000
	M	1.824	1.735	Z = −0.375
	SD	0.797	0.963	d = −0.091
				p = 0.707
Immediately after training	N	34	34	Mann-Whitney U = 310.500
	M	4.118	5.412	Z = −3.341
	SD	1.343	1.500	d = −0.886
				p < 0.001***
4 weeks after training	N	34	34	Mann-Whitney U = 307.000
	M	3.412	4.294	Z = −3.459
	SD	0.925	0.970	d = −0.924
				p < 0.001***
Pre-training vs. Immediately after training	Mann-Whitney U	89.500	23.000	–
	Z	−6.122	−6.872	–
	d	−2.216	−3.015	–
	p	< 0.001***	< 0.001***	–
Immediately after training vs. 4 weeks after training	Mann-Whitney U	406.000	317.000	–
	Z	−2.175	−3.286	–
	d	−0.547	−0.869	–
	p	0.030	0.001**	–

p-value below 0.05. Consequently, to identify any significant differences between the experimental settings, the Mann-Whitney U test and effect size d were employed.

Following the trainings, there was a significant rise (PF AR: $p < 0.001$, d = −1.431; DI AR: $p < 0.001$, d = −1.795) in self-efficacy scores for both groups, indicating that the trainings were successful in enhancing participants' self-efficacy. The effect sizes demonstrate the substantial impact of training on self-efficacy regardless of modality.

When comparing self-efficacy scores immediately after training and 4 weeks after training, no statistically significant decrease (DI AR: $p = 0.101$; PF AR: $p = 0.058$) was observed in either the DI AR group or the PF AR group. As shown in Table 5, the pre-training self-efficacy scores of the two groups did not differ significantly ($p = 0.541$). Additionally, no statistically significant differences were found between the groups immediately after training ($p = 0.140$) or 4 weeks after training ($p = 0.119$).

Figure 8 shows the relationship between self-efficacy and knowledge. In the DI AR group, there was a positive association ($\rho = 0.31$, $p = 0.002$). The PF AR group also demonstrated a positive relationship, which was slightly stronger ($\rho = 0.46$, $p < 0.001$).

4.3 Analysis of Intrinsic Motivation

Figure 9; Table 6 present the intrinsic motivation scores collected throughout the study. The experiment groups were compared for significant differences using the Mann-Whitney U test and effect size d due to non-compliance with the tests of homogeneity of variance and normality ($p < 0.05$).

Both groups showed a significant rise (DI AR: $p < 0.001$, d = −1.707; PF AR: $p < 0.001$, d = −1.745) in intrinsic motivation scores after the trainings. These large effect sizes indicate

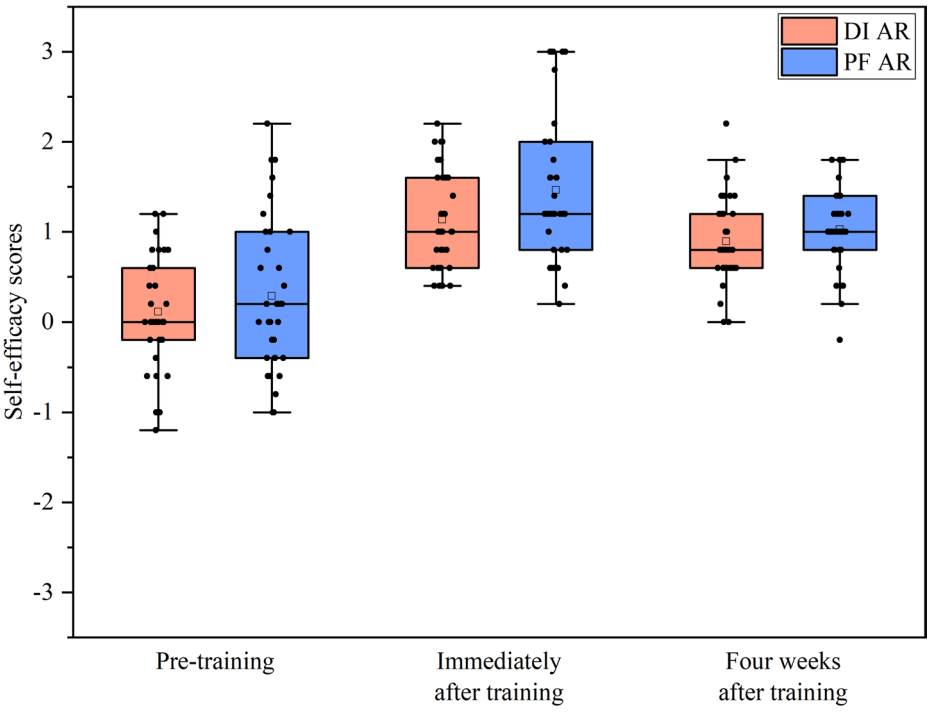

Fig. 7 Self-efficacy scores

Table 5 Comparison of self-efficacy scores in DI AR and PF AR training settings

Time points	Parameter	DI AR	PF AR	DI AR vs. PF AR
Pre-training	N	34	34	Mann-Whitney U = 528.500 Z = −0.611 d = −0.149 *p* = 0.541
	M	0.112	0.288	
	SD	0.617	0.846	
Immediately after training	N	34	34	Mann-Whitney U = 458.500 Z = −1.475 d = −0.364 *p* = 0.140
	M	1.135	1.465	
	SD	0.552	0.844	
4 weeks after training	N	34	34	Mann-Whitney U = 452.000 Z = −1.561 d = −0.386 *p* = 0.119
	M	0.894	1.029	
	SD	0.483	0.452	
Pre-training vs. Immediately after training	Mann-Whitney U	131.000	188.000	–
	Z	−5.508	−4.799	–
	d	−1.795	−1.431	–
	p	< 0.001***	< 0.001***	–
Immediately after training vs. 4 weeks after training	Mann-Whitney U	445.500	424.500	–
	Z	−1.639	−1.898	–
	d	−0.406	−0.473	–
	p	0.101	0.058	–

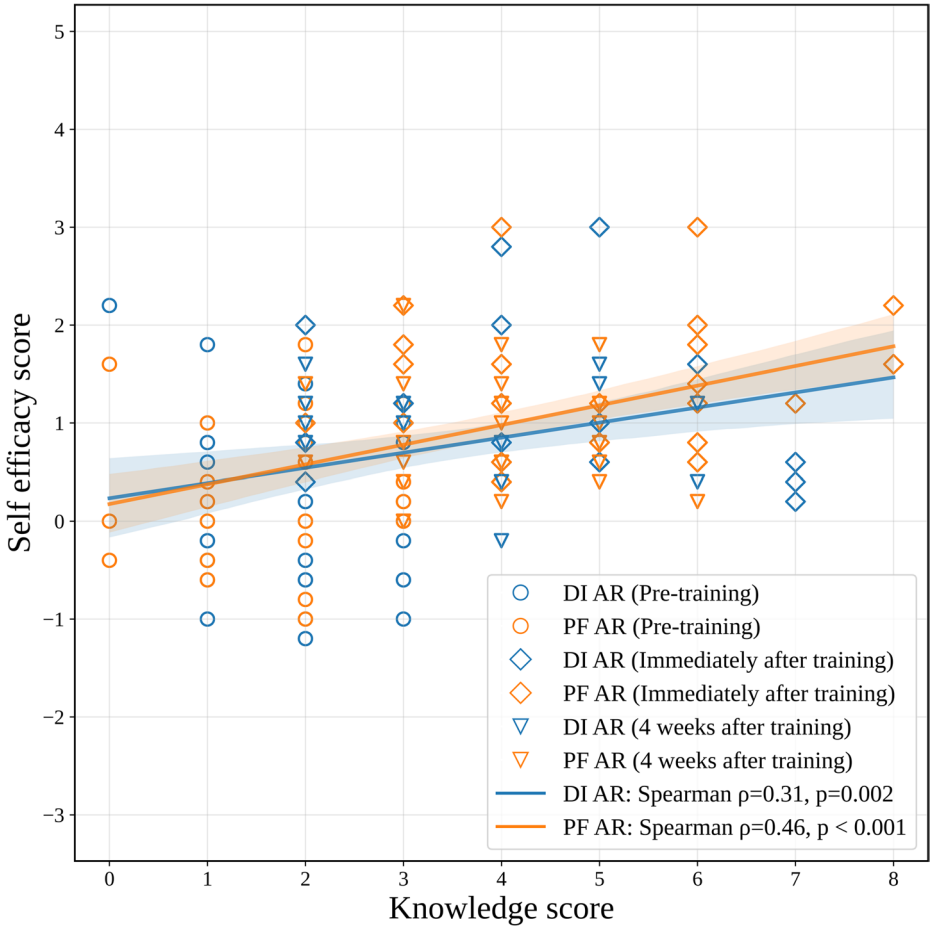

Fig. 8 Trends of self-efficacy and knowledge

that both training methods were highly effective in increasing intrinsic motivation immediately after training. Furthermore, when comparing the DI AR and PF AR groups' intrinsic motivation levels immediately following training and four weeks later, no statistically significant drop (DI AR: $p=0.226$; PF AR: $p=0.085$) was observed. As shown in Table 5, there was no statistically significant difference ($p=0.777$) in intrinsic motivation scores between the two groups before training. Similarly, no statistically significant differences were found between the groups immediately after training ($p=0.054$) or 4 weeks after training ($p=0.057$).

Figure 10 depicts the trends of intrinsic motivation and knowledge. For the DI AR group, intrinsic motivation and knowledge were positively correlated ($\rho=0.45$, $p<0.001$). A similar positive correlation was found in the PF AR group, with a slightly higher ($\rho=0.47$, $p<0.001$).

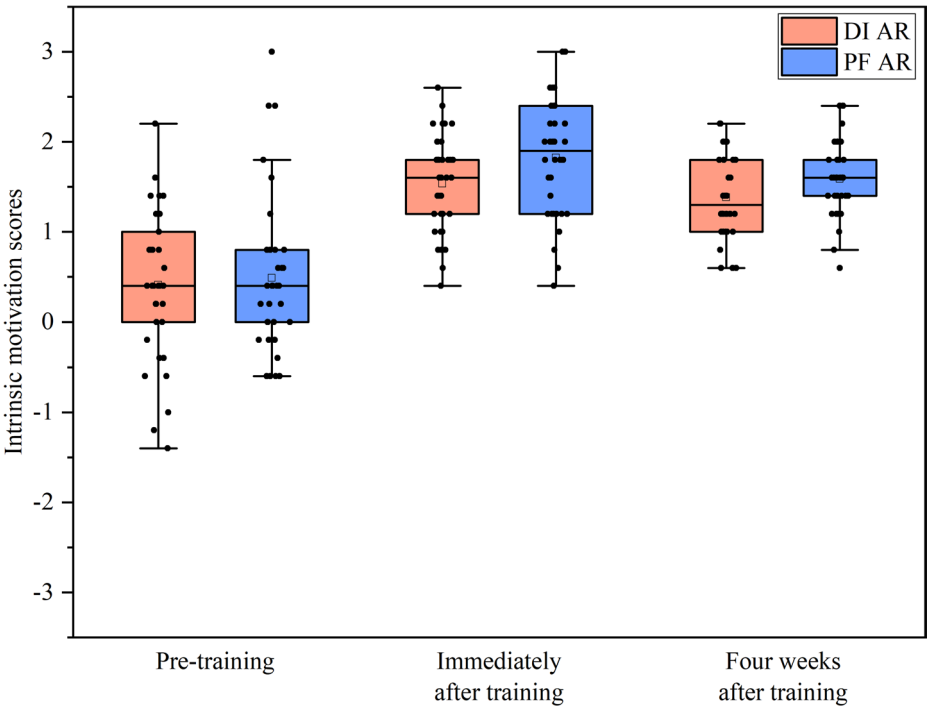

Fig. 9 Intrinsic motivation scores

Table 6 Comparison of intrinsic motivation scores in DI AR and PF AR training settings

Time points	Parameter	DI AR	PF AR	DI AR vs. PF AR
Pre-training	N	34	34	Mann-Whitney U = 555.000 Z = −0.284 d = −0.069 *p* = 0.777
	M	0.412	0.488	
	SD	0.825	0.887	
Immediately after training	N	34	34	Mann-Whitney U = 422.000 Z = −1.924 d = −0.480 *p* = 0.054
	M	1.535	1.824	
	SD	0.543	0.666	
4 weeks after training	N	34	34	Mann-Whitney U = 424.000 Z = −1.906 d = −0.475 *p* = 0.057
	M	1.382	1.588	
	SD	0.460	0.415	
Pre-training vs. Immediately after training	Mann-Whitney U	143.000	137.000	–
	Z	−5.353	−5.422	–
	d	−1.707	−1.745	–
	p	< 0.001***	< 0.001***	–
Immediately after training vs. 4 weeks after training	Mann-Whitney U	480.000	438.500	–
	Z	−1.211	−1.723	–
	d	−0.297	−0.427	–
	p	0.226	0.085	–

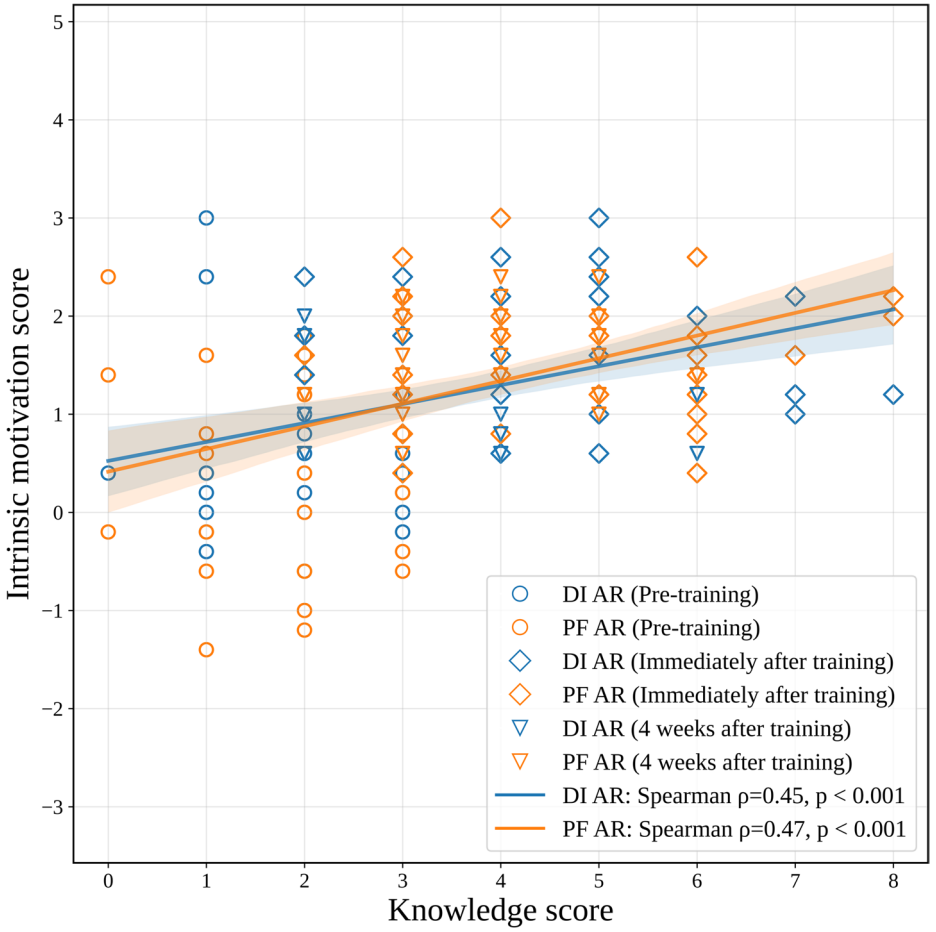

Fig. 10 Trends of intrinsic motivation and knowledge

4.4 Analysis of Measures Efficacy, Measures Simplicity, Perceived Reality, System Usability, and Task Load

A normality test conducted on the measures efficacy, measures simplicity, perceived reality, system usability, and task load scores revealed that they did not conform to the assumption of a normal distribution ($p<0.05$). To ascertain if there were any significant differences between the two experimental groups, the Mann-Whitney U test and effect size d were utilized. Figures 11 and 12, as well as Tables 7 and 8, illustrate the results of the measures efficacy, measures simplicity, perceived reality, system usability, and task load scores. No statistically significant differences ($p>0.05$) were found between the DI AR and PF AR groups across these five aspects.

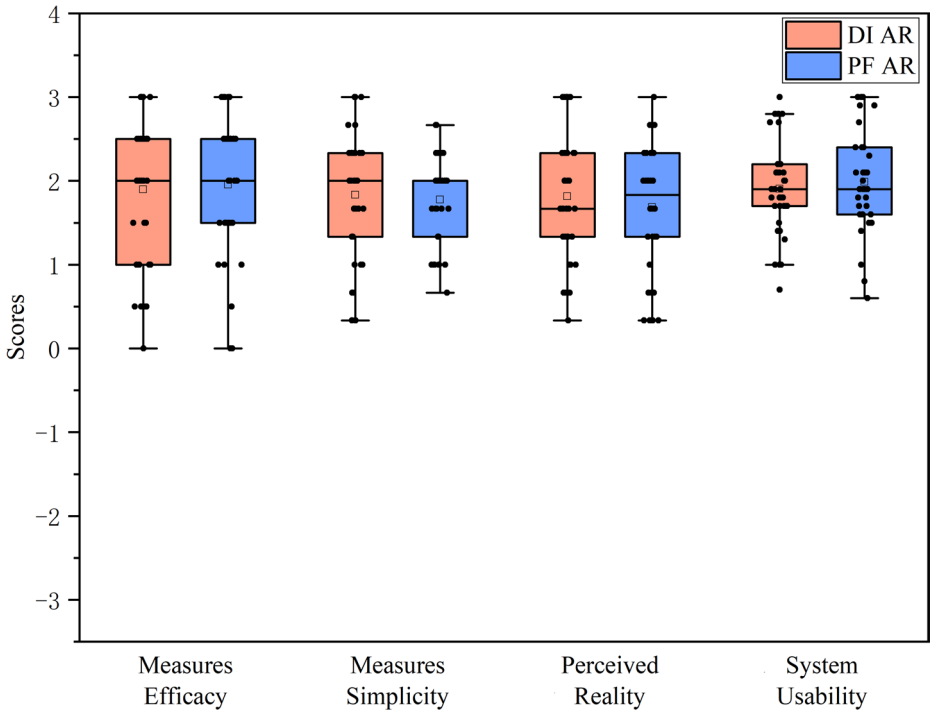

Fig. 11 Measures efficacy, measures simplicity, perceived reality, and system usability scores

5 Discussions

This study first introduces teaching theories into AR safety training. We aim to compare the effectiveness of two different AR safety trainings, DI AR and PF AR. Although prior educational research has shown the advantages of PF over DI, such evidence has rarely been examined within AR safety training contexts. By embedding established learning theories into AR system design, this study extends PF research from conventional settings to applied fire safety training. The contribution, therefore, lies not in reaffirming PF superiority but in demonstrating how PF theory can be operationalized and evaluated within AR-based procedural safety training. A range of critical indicators were applied, including knowledge, intrinsic motivation, self-efficacy, measures efficacy, measures simplicity, perceived reality, system usability, and task load. In addition, the study sought to compare the short- and long-term impacts of the trainings through two follow-up evaluations conducted four weeks apart. No control group using traditional methods was included, so comparisons with conventional approaches were not directly tested. The results provide evidence to support the efficacy of augmented reality in conveying fire extinguisher training knowledge.

5.1 Effects on Knowledge Acquisition and Knowledge Retention

In terms of knowledge acquisition, this study found that the PF AR group demonstrated significantly better training results than the DI AR group. These results are in line with the

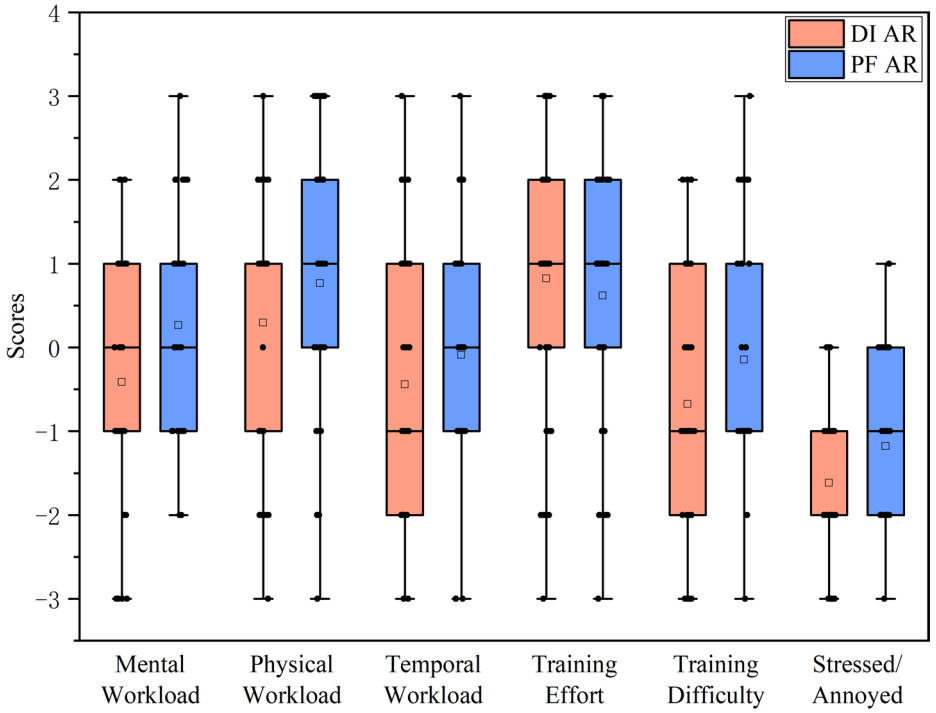

Fig. 12 Task load scores

Table 7 Comparison of measures efficacy, measures simplicity, perceived reality, and system usability scores for DI AR and PF AR training settings

DI AR vs. PF AR	Measures efficacy	Measures simplicity	Perceived reality	System usability
Mann-Whitney U	558.500	530.000	544.500	529.500
Z	−0.244	−0.598	−0.415	−0.597
d	−0.059	−0.145	−0.101	−0.145
p	0.807	0.550	0.678	0.551

Table 8 Comparison of task load scores for DI AR and PF AR training settings

DI AR vs. PF AR	Mental workload	Physical workload	Temporal workload	Training effort	Training difficulty	Stressed/ annoyed
Mann-Whitney U	450.000	485.000	490.000	530.000	474.500	441.000
Z	−1.608	−1.161	−1.099	−0.601	−1.340	−1.788
d	−0.398	−0.284	−0.269	−0.146	−0.329	−0.444
p	0.108	0.246	0.272	0.548	0.180	0.074

research by Lu et al. [56] in the realm of VR. Several factors may explain these results. Firstly, in the PF AR group, initial exposure to challenges and failures can facilitate deeper learning and improve problem-solving skills [78–81]. Trainees are encouraged to explore and experiment by engaging them in AR environments with open-ended tasks. During the

PF AR exploration phase, participants commonly made two main types of procedural errors. The first type was omission errors, such as skipping essential inspection steps. The second type was misidentification errors, in which participants confused different components of the fire extinguisher. These errors were not random. They often reflected participants' underlying misconceptions about the function of the equipment. Importantly, many participants who initially made these errors were able to self-correct during the AR exploration phase or in the post-test. For example, approximately one-third of participants initially skipped steps such as checking the safety pin or pressure gauge. However, after reviewing task feedback or re-experiencing the AR simulation, they recognized their omissions and incorporated these steps into their procedural understanding. Similarly, misidentification errors often occurred with components that were visually similar or functionally abstract. Most participants corrected these errors after interacting with the AR interface, indicating improved component recognition. These observations suggest that the AR-based PF approach not only exposes learners to specific errors but also actively facilitates error detection and correction. This process reinforces procedural knowledge and promotes deeper conceptual understanding. By confronting errors and receiving immediate or delayed cues from the AR environment, participants developed a more accurate understanding of task logic, which contributed to superior post-training performance. This differs from the DI AR group, which tends to follow a rigid structure, limiting autonomy and hindering critical thinking and adaptability [46, 56, 82].

Secondly, AR technology enhances the effectiveness of the PF theory by providing an interactive learning experience. AR overlays digital content onto the real world, enabling trainees to interact with virtual elements [13, 20, 83]. This multimodal engagement enriches cognitive processes, spatial understanding, and contextualization of knowledge. Trainees can actively manipulate objects, simulate scenarios, and engage in trial-and-error decision-making, aligning with the active learning principles of the PF theory [33, 84–87]. Furthermore, the integration of PF theory and AR aligns well with the complexity of real-world problems. Trainees are encouraged to embrace uncertainty, engage in iterative attempts, and develop adaptive thinking and resilience for ambiguous situations [88]. This nurtures an active learning mindset and prepares them for challenges that lack clear instructions [89–91]. In contrast, DI methods often prioritize memorization, limiting the application of knowledge in dynamic settings. It is important to note that this study does not suggest the inefficiency of DI AR. Both groups demonstrated significant knowledge improvements post-training. The conclusion drawn solely underscores the superiority of PF AR over DI AR.

Nevertheless, the mean score immediately after training reached 5.4 in the PF AR group and 4.1 in the DI AR group, corresponding to an accuracy rate of 67.5% and 51.25%, respectively. Although these gains are statistically significant compared to pre-training performance, they show that participants did not achieve near-complete procedural mastery after a single training session. This outcome may be due to the cognitively demanding nature of the assessment, which required accurate recall of multiple inspection steps in an open-ended format, as well as the limited time participants had for the training. From a practical safety perspective, this level of performance should be seen as the acquisition of foundational procedural knowledge rather than full operational readiness. Accordingly, the developed AR prototypes are better positioned as an introductory or supplementary training tool that sup-

ports early-stage learning and conceptual understanding, rather than as a standalone solution sufficient to ensure real-world fire safety performance.

Regarding knowledge retention, the PF AR group experienced a significant decline after four weeks. This decrease aligns with prior findings [20, 56], potentially linked to the training's complexity. Nonetheless, the PF AR group maintained significantly higher knowledge levels than the DI AR group, demonstrating the efficacy of PF AR in retaining knowledge. While the absolute mean differences in scores were relatively small (1.294 points immediately after training and 0.882 points at retention), the corresponding effect sizes were $d = -0.886$ immediately after training and $d = -0.924$ at four weeks. Based on Cohen's and Kraft's benchmarks ($d = 0.2$, 0.5, and 0.8 for small, medium, and large effects, respectively), these effect sizes can be regarded as large and practically meaningful [92, 93]. In fact, within the field of education, an effect size of 0.40 or higher is often viewed as a meaningful threshold for justifying the adoption of new policies [94]. In fire extinguisher use, even modest gains in procedural knowledge can enhance response confidence and reduce hesitation during emergencies. Therefore, the superior performance of the PF AR group may indicate a greater likelihood of applying learned knowledge under pressure, which is critical in occupational health and safety scenarios. However, it is important to acknowledge that the observed differences between the PF AR and DI AR groups may not be solely attributable to instructional sequencing. The AR prototypes differed not only in pedagogical structure but also in several system design elements. For instance, the PF AR prototype incorporated a self-check interface, iterative failure-feedback loops, and verbal guidance from an NPC. Importantly, these features were not arbitrarily included but were deliberately designed to embody the principles of productive failure—such as promoting metacognitive reflection, encouraging exploration, and supporting learning from errors. While they operationalize key aspects of the PF theory, these elements may have exerted their own influence on learning outcomes. Future studies could address this limitation by employing a factorial design that systematically manipulates both instructional theory and interface features. This would help disentangle the individual and combined contributions of pedagogy and AR design elements.

5.2 Effects on Motivation, Perceived Usability, and Cognitive Load

Participants in both groups demonstrated significant increases in self-efficacy and intrinsic motivation scores after training. There was no significant decrease even after four weeks, which highlights the effectiveness of the training intervention. Furthermore, there were no statistically significant variations in the two groups' self-efficacy and intrinsic motivation scores at any of the three time periods, indicating equal success in enhancing these vital factors. This shared effectiveness was reinforced by no significant differences in measures efficacy, measures simplicity, perceived reality, and system usability, with both groups achieving high scores. These results validate the efficiency and user-friendliness of the two AR training prototypes developed in this study. The two prototypes offer an interactive experience that engages participants' interest and motivates them to explore this technology for educational purposes. Regarding task load, both experimental conditions presented relatively high demands in terms of physical workload and training effort. Nevertheless, participants found both trainings equally manageable to navigate. It is worth noting that no challenges, stress, or frustrations were reported during the trainings. The smooth progress

of the training likely contributed to the similar level of knowledge acquisition observed in both trainings, suggesting that participants effectively interacted with the presented AR environment.

5.3 Theoretical and Practical Contributions

This study provides several important contributions. In terms of theoretical contributions, firstly, it advances theoretical understanding by integrating teaching theories into AR safety training methodologies. By combining DI AR and PF AR methods, this study expands the application of teaching theories in the dynamic realm of computer-assisted learning within AR environments. This fusion enriches the theoretical framework for AR applications. It also establishes the foundation for a detailed examination of the effectiveness of various teaching methods in safety training environments. Secondly, by meticulously comparing the efficacy of DI AR and PF AR, the study delves into the nuances of distinct teaching methodologies in AR safety training scenarios, revealing the comparative advantages and limitations of these approaches. Building upon the works of Kennedy-Clark et al. [95] and Lu et al. [56], this paper clarifies optimal teaching strategies for enhancing learning results within safety training environments. It lays a robust groundwork for comprehending the diverse effects of teaching theories on educational practices employing augmented reality technology. Furthermore, this study enriches the theoretical understanding of educational practices in AR environments by evaluating the effectiveness of different teaching strategies in AR safety training. It also contributes to a broader discussion on teaching design and technology integration in educational frameworks. The insights gained offer valuable guidance on utilizing teaching theories to improve learning outcomes in safety training scenarios. This opens up new possibilities for advanced and effective teaching practices in AR-enhanced learning environments. This study will also contribute to future efforts of running a meta-analysis by integrating findings across studies that compare two training conditions.

In terms of practical contributions, this study has developed two AR safety training prototypes specifically tailored for fire extinguisher inspections. Their effectiveness and usability were rigorously evaluated through comparative experiments. These prototypes not only streamline the supervision process, potentially curbing the costs associated with manual oversight, but also enhance the feasibility of delivering training. This cost-efficient approach underscores a pragmatic strategy for enhancing safety training outcomes. Moreover, a user-centered design approach was emphasized in the development of these AR safety training prototypes. From the perspective of end-users, this study enriches the comprehension of the practicality and effectiveness of AR-based safety training for fire extinguisher inspections. By enhancing trainees' awareness, safety comprehension, and task proficiency, these AR-based trainings play a crucial role in reinforcing safety standards and enhancing survival probabilities in emergency scenarios. Future research and development efforts aimed at creating specialized, efficient AR training systems that satisfy users' unique demands and safety management criteria might build on the conclusions and suggestions of this study.

5.4 Limitations and Future Research

This study also has several limitations. Firstly, the reliance on a limited sample size may hinder a comprehensive understanding of the benefits that AR training can offer to diverse

user groups. To address this limitation, future research could focus on enlarging the sample size and improving participant diversity.

Secondly, the target audience for the developed training system primarily consisted of professionals with expertise in fire safety. However, to broaden participant inclusion and ensure feasibility, individuals from the general adult population were recruited for this study. This difference between the intended users and actual participants may affect the generalizability of the findings to professional training settings. Future studies could conduct more targeted surveys and recruit participants with relevant backgrounds to better align with intended users.

Thirdly, this study focused on a highly structured and procedural task, fire extinguisher inspection, which represents one of the most constrained and predictable components of fire safety training. Although specific procedures may vary across equipment types, the task remains rule-based and sequential in nature. Accordingly, the instructional effects of DI and PF observed in this study should be interpreted within this narrow task boundary. They should not be generalized to fire safety training contexts that involve situational judgment, dynamic risk assessment, or non-routine decision-making. Extending this comparison to more complex and ill-structured emergency scenarios remains an important direction for future research. Task difficulty was not explicitly quantified in this study. Future research could describe baseline performance ranges or use difficulty ratings to better capture task complexity. The training duration was limited to a single 15-minute session, and learning outcomes were assessed only once, four weeks after training. While this controlled setup ensures internal validity, it may limit ecological validity. Real fire safety situations involve more complex tasks such as situational awareness, evacuation coordination, and emergency communication, which unfold over longer time periods and under dynamic conditions. Moreover, the current assessment focused primarily on recall of procedural steps through an open-ended question. The same open-ended question was used at pre-training, immediately after training, and four weeks after training. This repeated use introduces the possibility of a test-retest effect due to item familiarity, which may have partially contributed to the observed learning gains. Furthermore, the time spent on each PF phase, including exploration, self-testing, feedback, and summary, was not recorded. This limits the ability to examine how procedural knowledge translates into actual performance. Future research should incorporate phase-specific timing, objective behavioral measures, and varied or counterbalanced test items to enhance the validity of the findings. These measures would help evaluate the relationship between engagement in each phase of the PF, knowledge gain, error reduction, and real-world skill transfer. Specifying minimum and maximum time limits per phase could also help ensure fairness when comparing PF and DI. In addition, memory retention and behavioural change may develop non-linearly over extended periods, which cannot be captured by a single follow-up point. Future studies could address this limitation by including diverse emergency scenarios and conducting longitudinal assessments at multiple time points over several months. Distributed practice, extended exposure, and adaptive feedback strategies may also improve the depth and sustainability of training outcomes.

Furthermore, the current AR training relies on image markers attached to fire extinguishers to support procedural learning. This dependency may limit direct transfer to unmarked extinguishers. Future research should explore optimized marker configurations or marker-free AR solutions, and further test and mark fewer solutions using 3-D anchor techniques to minimize the reliance on visual markers. Quantitative metrics such as occlusion rate,

recognition failures, and task-time impact could be recorded. Future experiments should also examine whether participants trained with AR are more capable of performing accurate and complete fire extinguisher inspections, thereby assessing real-world skill transfer. In addition, this study compared only two theory-based instructional approaches under a fixed training setting. While this comparison offers useful insights into immersive instructional design, it did not examine how these dimensions interact with one another. Future studies could use multivariate designs or mediation analyses to explore their interactions and combined effects on learning outcomes. Moreover, this study did not include a control group based on traditional training methods, such as reading materials or slide-based presentations. Future studies should consider adding traditional training conditions to better assess the added value of theory-based AR training.

Lastly, this study focused on developing fire safety training prototypes using optical see-through (OST) AR display technology, specifically utilizing HoloLens 2 in both training methods. Currently, head-mounted AR devices use two main display technologies: optical see-through (OST) and video see-through (VST). With the use of an optical display module, OST allows users to view virtual images and real-world elements concurrently through a transparent display or glasses [96]. In contrast, VST enables users to engage with augmented reality information by superimposing virtual graphics onto a live video stream of the physical surroundings via various display devices [96, 97]. Noteworthy devices include Google Glass, HoloLens 2 for OST, and Apple Vision Pro for VST. Future research could conduct comparative analyses of these two display technologies to highlight their strengths and weaknesses. This effort would provide essential insights for refining AR training techniques and choosing suitable devices in educational environments [12].

6 Conclusion

This study integrated two teaching theories into AR safety training, comparing the effectiveness of directive instruction training and productive failure training. Two AR prototypes were designed and prototyped for fire safety training. In a controlled experiment, 68 participants were randomly allocated to two training groups.

The findings show that productive failure training is more effective in terms of knowledge acquisition and knowledge retention. Nevertheless, both trainings demonstrate similar effectiveness in providing a thorough and varied learning experience. This highlights the potential advantages of AR training based on productive failure theory to improve trainee safety performance. These results offer valuable insights for the future advancement of augmented reality applications in safety training.

Acknowledgements This study was supported by the Ministry of Education of Humanities and Social Science Project of China (Grant No. 23YJA630069), the Jiangsu Province Construction System Science and Technology Project (Grant No. 2023JH04004), SEU Innovation Capability Enhancement Plan for Doctoral Students (CXJH_SEU 25088), the National Natural Science Foundation of China (Grant No. 72101054), the Marsden Fast Start (MAU2204) and the Rutherford Discovery Fellowship (RDF-MAU2201).

Author Contributions Peizhen Gong—Funding acquisition, Data curation, Writing-original draft, Visualization, Conceptualization, Methodology. Ying Lu—Funding acquisition, Conceptualization, Methodology, Supervision. Xiaer Xiahou—Funding acquisition, Resources, Project administration. Yunxuan Deng—Project administration, Visualization. Ruggiero Lovreglio—Funding acquisition, Conceptualization, Writing-review and editing, Supervision.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability All data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

References

1. Vu M, Lin S-Y (2024) Empirical assessment of fire safety in high-rise residential buildings in Vietnam and residents' knowledge and awareness regarding fire safety. Fire Saf J 146:104162. https://doi.org/10.1016/j.firesaf.2024.104162
2. Chen H, Hou L, Zhang G, Moon S (2021) Development of BIM, IoT and AR/VR technologies for fire safety and upskilling. Autom Constr 125:103631. https://doi.org/10.1016/j.autcon.2021.103631
3. Association G (2014) Fire and climate risk, The Geneva Associate Bulletin World Fire Statistics
4. Wikipedia Tof (2010) Available online: https://en.wikipedia.org/wiki/Tornadoes_of_2010, Accessed 7 Oct 2025
5. UK-Home-Office (2023) Fire and rescue incident statistics: England, year ending March Available online: https://www.gov.uk/government/statistics/fire-and-rescue-incident-statistics-england-year-ending-march-2023, Accessed 31 July 2024
6. Dong X, Entzel P, Men Y, Chowdhury R, Schneider S (2004) Effects of safety and health training on work-related injury among construction laborers. J Occup Environ Med 46:1222–1228. https://doi.org/10.1097/01.jom.0000147268.42094.de
7. Yiu NSN, Sze NN, Chan DWM (2018) Implementation of safety management systems in Hong Kong construction industry – a safety practitioner's perspective. J Saf Res 64:1–9. https://doi.org/10.1016/j.jsr.2017.12.011
8. Sacks R, Perlman A, Barak R (2013) Construction safety training using immersive virtual reality. Taylor & Francis, pp 1005–1017
9. Wu S, Hou L, Chen H, Zhang G, Zou Y, Tushar Q (2023) Cognitive ergonomics-based Augmented Reality application for construction performance. Autom Constr 149:104802. https://doi.org/10.1016/j.autcon.2023.104802
10. Feng Z, Gonzalez VA, Amor R, Lovreglio R, Cabrera-Guerrero G (2018) Immersive virtual reality serious games for evacuation training and research: a systematic literature review. Comput Educ 127:252–266. https://doi.org/10.1016/j.compedu.2018.09.002
11. Scorgie D, Feng Z, Paes D, Parisi F, Yiu TW, Lovreglio R (2024) Virtual reality for safety training: a systematic literature review and meta-analysis. Saf Sci 171:106372. https://doi.org/10.1016/j.ssci.2023.106372
12. Gong P, Lu Y, Lovreglio R, Lv X, Chi Z (2024) Applications and effectiveness of augmented reality in safety training: a systematic literature review and meta-analysis. Saf Sci 178:106624. https://doi.org/10.1016/j.ssci.2024.106624
13. Gong P, Lu Y, Lovreglio R, Yang X, Deng Y (2024) Comparing the effectiveness of AR training and slide-based training: the case study of metro construction safety. Saf Sci 176:106561. https://doi.org/10.1016/j.ssci.2024.106561

14. Cherrett T, Wills G, Price J, Maynard S, Dror IE (2009) Making training more cognitively effective: making videos interactive. Br J Educ Technol 40:1124–1134. https://doi.org/10.1111/j.1467-8535.2009.00985.x
15. Gwynne SMV, Kuligowski ED, Boyce KE, Nilsson D, Robbins AP, Lovreglio R, Thomas JR (2019) Roy-poirier, enhancing egress drills: preparation and assessment of evacuee performance. Fire Mater 43:613–631. https://doi.org/10.1002/fam.2448
16. Ipinnaiye O, Risquez A (2024) Exploring adaptive learning, learner-content interaction and student performance in undergraduate economics classes. Computers Educ 215:105047. https://doi.org/10.1016/j.compedu.2024.105047
17. Raeisinafchi R, Bhandari S, Perry L, Hallowell MR, Albert A, Correll J (2025) Comparing training delivery methods: impact on learning outcomes and engagement among construction workers. Saf Sci 187:106870. https://doi.org/10.1016/j.ssci.2025.106870
18. Alkan IB, Basaga HB (2023) Augmented reality technologies in construction project assembly phases. Autom Constr 156:105107. https://doi.org/10.1016/j.autcon.2023.105107
19. Fazel A, Adel A (2024) Enhancing construction accuracy, productivity, and safety with augmented reality for timber fastening. Autom Constr 166:105596. https://doi.org/10.1016/j.autcon.2024.105596
20. Paes D, Feng Z, King M, Khorrami Shad H, Sasikumar P, Pujoni D, Lovreglio R (2024) Optical see-through augmented reality fire safety training for building occupants. Autom Constr 162:105371. https://doi.org/10.1016/j.autcon.2024.105371
21. Tan Y, Xu W, Chen P, Zhang S (2024) Building defect inspection and data management using computer vision, augmented reality, and BIM technology. Autom Constr 160:105318. https://doi.org/10.1016/j.autcon.2024.105318
22. Wu S, Chen H, Hou L, Zhang G, Li C-Q (2024) Using eye-tracking to measure worker situation awareness in augmented reality. Autom Constr 165:105582. https://doi.org/10.1016/j.autcon.2024.105582
23. Kolaei AZ, Hedayati E, Khanzadi M, Amiri GG (2022) Challenges and opportunities of augmented reality during the construction phase. Autom Constr. https://doi.org/10.1016/j.autcon.2022.104586
24. Alkan IB, Basaga HB (2023) Augmented reality technologies in construction project assembly phases. Autom Constr. https://doi.org/10.1016/j.autcon.2023.105107
25. Kim K, Kim H, Kim H (2017) Image-based construction hazard avoidance system using augmented reality in wearable device. Autom Constr 83:390–403. https://doi.org/10.1016/j.autcon.2017.06.014
26. Shringi A, Arashpour M, Golafshani EM, Dwyer T, Kalutara P (2023) Enhancing safety training performance using extended reality: a hybrid Delphi–AHP multi-attribute analysis in a type-2 fuzzy environment. Buildings 13:625. https://doi.org/10.3390/buildings13030625
27. Dallasega P, Schulze F, Revolti A (2023) Augmented reality to overcome visual management implementation barriers in construction: a MEP case study. Constr Manage Econ 41:232–255. https://doi.org/10.1080/01446193.2022.2135748
28. Hartmann C, van Gog T, Rummel N (2022) Productive versus vicarious failure: do students need to fail themselves in order to learn? Appl Cogn Psychol 36:1219–1233. https://doi.org/10.1002/acp.4004
29. Kapur M, Saba J, Roll I (2023) Prior math achievement and inventive production predict learning from productive failure. NPJ Sci Learn. https://doi.org/10.1038/s41539-023-00165-y
30. Steenhof N, Woods NN, Mylopoulos M (2020) Exploring why we learn from productive failure: insights from the cognitive and learning sciences. Adv Health Sci Educ 25:1099–1106. https://doi.org/10.1007/s10459-020-10013-y
31. Beach TAC, Stankovic T, Carnegie DR, Micay R, Frost DM (2018) Using verbal instructions to influence lifting mechanics - Does the directive lift with your legs, not your back attenuate spinal flexion? J Electromyogr Kinesiol 38:1–6. https://doi.org/10.1016/j.jelekin.2017.10.008
32. Goebel K, Helmke A (2010) Intercultural learning in English as foreign language instruction: the importance of teachers' intercultural experience and the usefulness of precise instructional directives. Teach Teach Educ 26:1571–1582. https://doi.org/10.1016/j.tate.2010.05.008
33. Chowrira SG, Smith KM, Dubois PJ, Roll I (2019) DIY productive failure: boosting performance in a large undergraduate biology course. NPJ Sci Learn. https://doi.org/10.1038/s41539-019-0040-6
34. Perlman A, Sacks R, Barak R (2014) Hazard recognition and risk perception in construction. Saf Sci 64:22–31. https://doi.org/10.1016/j.ssci.2013.11.019
35. Chen Y-J, Lai Y-S, Lin Y-H (2020) BIM-based augmented reality inspection and maintenance of fire safety equipment. Autom Constr 110:103041. https://doi.org/10.1016/j.autcon.2019.103041
36. Aivaliotis S, Lotsaris K, Gkournelos C, Fourtakas N, Koukas S, Kousi N, Makris S (2023) An augmented reality software suite enabling seamless human robot interaction. Int J Comput Integr Manuf 36:3–29. https://doi.org/10.1080/0951192X.2022.2104459
37. Chu C-H, Ko C-H (2021) An experimental study on augmented reality assisted manual assembly with occluded components. J Manuf Syst 61:685–695. https://doi.org/10.1016/j.jmsy.2021.04.003

38. Castillo-Rodríguez JM, Gómez-Urquiza JL, García-Oliva S, Suleiman-Martos N (2025) Effectiveness of virtual and augmented reality for emergency healthcare training: a randomized controlled trial. Healthcare 13:1034. https://doi.org/10.3390/healthcare13091034
39. Zhu Y, Li N (2021) Virtual and augmented reality technologies for emergency management in the built environments: a state-of-the-art review. J Saf Sci Resil 2:1–10. https://doi.org/10.1016/j.jnlssr.2020.11.004
40. Tan Y, Xu W, Li S, Chen K (2022) Augmented and virtual reality (AR/VR) for education and training in the AEC industry: a systematic review of research and applications. Buildings 12:1529. https://doi.org/10.3390/buildings12101529
41. Mondal H, Mondal S (2025) Adopting augmented reality and virtual reality in medical education in resource-limited settings: constraints and the way forward. American Physiological Society Rockville, MD, pp 503–507
42. Papakostas C, Troussas C, Krouska A, Sgouropoulou C (2021) Measuring user experience, usability and interactivity of a personalized mobile augmented reality training system. Sensors. https://doi.org/10.3390/s21113888
43. Somerkoski B, Tarkkanen K, Oliva D, Lehto A, Luimula M (2022) Pedagogic solutions and results in designing a mobile game for fire safety teaching, CEUR Workshop Proceedings, pp. 44–53
44. Akama F, Keenan J (2023) Attitudes towards staff mentoring by senior leaders of a college of education in Ghana. J High Educ Policy Manag 45:84–95. https://doi.org/10.1080/1360080X.2022.2140749
45. Landy MSH, Vezer E, Bance S, Loskot T, Ip J, Zeifman APP, Mutschler C, Thomas FCC, McShane K, Monson CMM, Stirman SW (2023) Elucidating the elements of clinical Case consultation in cognitive processing therapy elucider les elements de La consultation de Cas cliniques En therapie du processus cognitif. Couns Psychol 51:626–654. https://doi.org/10.1177/00110000231166103
46. Lim K, Kang M, Park SY (2016) Structural relationships of environments, individuals, and learning outcomes in Korean online university settings. Int Rev Res Open Distrib Learn 17:315–330
47. Leria Dulcic FJ, Acosta Pena RN, Sasso Orellana PE (2021) Collao Jofre, do instructions overwhelm the preschool classroom? Early childhood educators' use of instructional vs regulative directive commands. Suvrem Lingvist 47:247–265. https://doi.org/10.22210/suvlin.2021.092.06
48. Kapur M (2008) Productive failure. Cogn Instr 26:379–425. https://doi.org/10.1080/07370000802212669
49. Kapur M, Bielaczyc K (2012) Designing for productive failure. J Learn Sci 21:45–83. https://doi.org/10.1080/10508406.2011.591717
50. Kerrigan J, Weber K, Chinn C (2021) Effective collaboration in the productive failure process. J Math Behav. https://doi.org/10.1016/j.jmathb.2021.100892
51. Song Y, Kapur M (2017) How to flip the classroom - Productive failure or traditional flipped classroom pedagogical design? Educ Technol Soc 20:292–305. https://doi.org/10.3929/ethz-b-000128354
52. Ziegler E, Trninic D, Kapur M (2021) Micro productive failure and the acquisition of algebraic procedural knowledge. Instr Sci 49:313–336. https://doi.org/10.1007/s11251-021-09544-7
53. Palominos E, Levett-Jones T, Power T, Alcorn N, Martinez-Maldonado R (2021) Measuring the impact of productive failure on nursing students? learning in healthcare simulation: a quasi-experimental study. Nurse Educ Today. https://doi.org/10.1016/j.nedt.2021.104871
54. Steenhof N, Woods NN, Van Gerven PWM, Mylopoulos M (2019) Productive failure as an instructional approach to promote future learning. Adv Health Sci Educ Theory Pract 24:739–749. https://doi.org/10.1007/s10459-019-09895-4
55. Palominos E, Levett-Jones T, Power T, Martinez-Maldonado R (2022) We learn from our mistakes': nursing students' perceptions of a productive failure simulation. Collegian 29:708–712. https://doi.org/10.1016/j.colegn.2022.02.006
56. Lu S, Feng Z, Lovreglio R, Wang F, Yuan X (2023) Comparing the productive failure and directive instruction for declarative safety knowledge training using virtual reality. J Comput Assist Learn. https://doi.org/10.1111/jcal.12937
57. Kodur V, Kumar P, Rafi MM (2019) Fire hazard in buildings: review, assessment and strategies for improving fire safety. PSU Res Rev 4:1–23. https://doi.org/10.1108/prr-12-2018-0033
58. Saghafian M, Laumann K, Akhtar RS, Skogstad MR (2020) The evaluation of virtual reality fire extinguisher training. Front Psychol 11:593466
59. Lovreglio R, Duan X, Rahouti A, Phipps R, Nilsson D (2021) Comparing the effectiveness of fire extinguisher virtual reality and video training. Virtual Real 25:133–145. https://doi.org/10.1007/s10055-020-00447-5
60. Feng Z, Lovreglio R, Yiu TW, Acosta DM, Sun B, Li N (2023) Immersive virtual reality training for excavation safety and hazard identification. Smart Sustain Built Environ. https://doi.org/10.1108/SASBE-10-2022-0235
61. Microsoft (2024) Learn about HoloLens 2 features and review technical specs, Available online: https://www.microsoft.com/en-us/hololens/hardware#document-experiences, Accessed 15 July 2024

62. Inc. P (2025) Vuforia Engine Library, Available online: https://library.vuforia.com/, Accessed 15 July 2024
63. Alshiar M, Holtkamp B, Biediger D, Wilson M, Yun C, Kim K (2019) Ieee, SMACK: subjective measure of applied contextual knowledge, 2019 IEEE Games Entertain. Media Conference
64. Daling LM, Tenbrock M, Isenhardt I, Schlittmeier SJ (2023) Assemble it like this! – Is AR- or VR-based training an effective alternative to video-based training in manual assembly? Appl Ergon 110:104021. https://doi.org/10.1016/j.apergo.2023.104021
65. Hong J, Choi J, Lee J, Cho S, Hong T, Han S, Park HS, Lee D-E (2023) Virtual reality-based analysis of the effect of construction noise exposure on masonry work productivity. Autom Constr 150:104844. https://doi.org/10.1016/j.autcon.2023.104844
66. Chittaro L, Sioni R (2015) Serious games for emergency preparedness: evaluation of an interactive vs. a non-interactive simulation of a terror attack. Comput Hum Behav 50:508–519. https://doi.org/10.1016/j.chb.2015.03.074
67. Wang H-Y, Lin V, Hwang G-J, Liu G-Z (2019) Context-aware language-learning application in the green technology building: which group can benefit the most? J Comput Assist Learn 35:359–377. https://doi.org/10.1111/jcal.12336
68. Sun PP, Luo X (2024) Understanding English-as-a-foreign-language university teachers' synchronous online teaching satisfaction: a Chinese perspective. J Comput Assist Learn 40:685–696. https://doi.org/10.1111/jcal.12891
69. Endriulaitiene A, Seibokaite L, Marksaityte R, Slavinskiene J, Arlauskiene R (2020) Changes in beliefs during driver training and their association with risky driving. Accid Anal Prev. https://doi.org/10.1016/j.aap.2020.105583
70. Rahouti A, Lovreglio R, Datoussaid S, Descamps T (2021) Prototyping and validating a non-immersive virtual reality serious game for healthcare fire safety training. Fire Technol 57:3041–3078. https://doi.org/10.1007/s10694-021-01098-x
71. Lim HW, Li N, Fang D, Wu C (2018) Impact of safety climate on types of safety motivation and performance: multigroup invariance analysis. J Manage Eng. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000595
72. Bruhn AB, Lindahl C, Andersson I-M, Rosen G (2023) Motivational factors for occupational safety and health improvements: a mixed-method study within the Swedish equine sector. Saf Sci. https://doi.org/10.1016/j.ssci.2022.106035
73. Le QT, Pedro A, Lim CR, Park HT, Park CS, Kim HK (2015) A framework for using mobile based virtual reality and augmented reality for experiential construction safety education. Int J Eng Educ 31:713–725
74. Gao M, Kortum P, Oswald FL (2020) Multi-language toolkit for the system usability scale. Int J Hum Comput Interact 36:1883–1901. https://doi.org/10.1080/10447318.2020.1801173
75. Hart SG (2006) Nasa-task load index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet 50:904–908. https://doi.org/10.1177/154193120605000909
76. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Advances in Psychology. North-Holland, Elsevier, pp 139–183
77. Abbas A, Seo J, Ahn S, Luo Y, Wyllie Mitchell J, Lee G, Billinghurst M (2023) How immersive virtual reality safety training system features impact learning outcomes: an experimental study of forklift training. J Manage Eng 39:04022068. https://doi.org/10.1061/(ASCE)ME.1943-5479.0001101
78. Perros T, Bisaga I, Broad O, Macon L, Fennell PJ, Tomei J (2022) Lost learnings: breaking the silence of failure in the energy and development sector. Energy Res Soc Sci. https://doi.org/10.1016/j.erss.2022.102804
79. Song Y (2018) Improving primary students' collaborative problem solving competency in project-based science learning with productive failure instructional design in a seamless learning environment. ETR&D 66:979–1008. https://doi.org/10.1007/s11423-018-9600-3
80. Kapur M, Hattie J, Grossman I, Sinha T (2022) Fail, flip, fix, and feed - rethinking flipped learning: a review of meta-analyses and a subsequent meta-analysis. Front Educ. https://doi.org/10.3389/feduc.2022.956416
81. Thorgeirsson S, Sinha T, Friedrich F, Su Z (2022) Does deliberately failing improve learning in introductory computer science? Educating for a new future: making sense of technology-enhanced learning adoption, EC-TEL 2022, pp. 608–614
82. Nervi H (2012) Training for work with the network or remain entangled: that is the question, 5th International Conference of Education, Research and Innovation (ICERI 2012), pp. 476–481
83. Lovreglio R, Kinateder M (2020) Augmented reality for pedestrian evacuation research: promises and limitations. Saf Sci. https://doi.org/10.1016/j.ssci.2020.104750

84. Beeler N, Ziegler E, Navarini AA, Kapur M (2023) Active before passive tasks improve long-term visual learning in difficult-to-classify skin lesions. Learn Instr. https://doi.org/10.1016/j.learninstruc.2023.101821
85. Decaro MS, Isaacs RA, Bego CR, Chastain RJ (2023) Bringing exploratory learning online: problem-solving before instruction improves remote undergraduate physics learning. Front Educ. https://doi.org/10.3389/feduc.2023.1215975
86. Hromkovi J, Staub J (2021) The problem with debugging in current block-based programming environments. Bull Eur Assoc Theor Comput Sci
87. Kapur M, Hattie J, Grossman I, Sinha T (2022) Fail, flip, fix, and feed - Rethinking flipped learning: a review of meta-analyses and a subsequent meta-analysis. Front Educ 7:956416. https://doi.org/10.3389/feduc.2022.1098967
88. Pusey M (2018) Acm, The Effect of Puzzle Video Games on High School Students' Problem-Solving Skills and Academic Resilience, Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY 2018), pp. 63–69
89. Roll I, Holmes NG, Day J, Bonn D (2012) Evaluating metacognitive scaffolding in guided invention activities. Instr Sci 40:691–710. https://doi.org/10.1007/s11251-012-9208-7
90. Savelson ZM, Muldner K (2023) How do students feel and collaborate during programming activities in the productive failure paradigm? Comput Sci Educ. https://doi.org/10.1080/08993408.2023.2237365
91. Vedder-Weiss D, Ehrenfeld N, Ram-Menashe M, Pollak I (2018) Productive framing of pedagogical failure: how teacher framings can facilitate or impede learning from problems of practice. Think Skills Creat 30:31–41. https://doi.org/10.1016/j.tsc.2018.01.002
92. Cohen J (1977) Statistical power analysis for the behavioral sciences. Academic Press, New York
93. Kraft MA (2020) Interpreting effect sizes of education interventions. Educ Res 49:241–253. https://doi.org/10.3102/0013189X20912798
94. Hattie J (2008) Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Routledge
95. Kennedy-Clark S, Jacobson MJ, Reimann P (2010) Scenario-based multi-user virtual environments: productive failure and the impact of structure on learning, sustaining TEL: From Innovation to Learning and Practice, pp. 402–407
96. Carbone M, Domeneghetti D, Cutolo F, D'Amato R, Cigna E, Parchi PD, Gesi M, Morelli L, Ferrari M, Ferrari V (2021) Can liquid lenses increase depth of field in head mounted video see-through devices? J Imaging. https://doi.org/10.3390/jimaging7080138
97. Gattullo M, Uva AE, Fiorentino M, Monno G (2015) Effect of text outline and contrast polarity on AR text readability in industrial lighting. IEEE Trans Vis Comput Graph 21:638–651. https://doi.org/10.1109/TVCG.2014.2385056