Safety Science 181 (2025) 106693

Contents lists available at ScienceDirect

Safety Science

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

From knowledge to action: Assessing the effectiveness of immersive virtual reality training on safety behaviors in confined spaces using the Kirkpatrick model

Alessandro Evangelista [a], Vito M. Manghisi [a,*], Vito De Giglio [a], Raffaele Mariconte [b], Claudia Giliberti [b], Antonio E. Uva [a]

[a] Department of Mechanics Mathematics and Management, Polytechnic University of Bari, Via Edoardo Orabona, 4, 70126 Bari, Italy
[b] Dipartimento Innovazioni Tecnologiche e Sicurezza degli Impianti, Prodotti ed Insediamenti Antropici Inail Rome, Italy

ARTICLE INFO

Keywords:
Virtual reality
Safety training
Confined space
User study
Validation procedure

ABSTRACT

Safety in the industrial sector is paramount, with national legislations establishing specific regulations to reduce accidents. These regulations emphasize risk removal and proper operator training as preventive measures. However, traditional training can be time-consuming and costly, leading companies to meet only the minimum requirements. In particular, confined spaces pose significant dangers, often resulting in fatal, cascading accidents due to a lack of proper safety procedures. Therefore, effective training is crucial to mitigate these risks.

Advanced technologies like Immersive Virtual Reality (IVR) present new opportunities for cost-effective and extensive training campaigns by eliminating the risks associated with real environment training, offering a safe yet realistic experience.

Despite the widespread adoption of IVR training applications in industrial research, there is a notable lack of systematic approaches to evaluate their effectiveness. Conducting such evaluations is crucial to ascertain their impact on essential training outcomes, including participant engagement, knowledge transfer, and enhanced safety behavior during work activities.

To address this gap, we developed an IVR-based training platform for confined space safety procedures and established a systematic validation procedure using the Kirkpatrick model to assess its effectiveness compared to conventional training methods. Results demonstrate that IVR training provides an excellent user experience, better knowledge transfer than the traditional approach, and improved performance in simulated procedures, reducing execution errors and completion times. This dual focus on creating and validating the IVR system underscores its potential as an effective training tool in hazardous environments such as confined spaces, where proper training can prevent tragic fatalities.

1. Introduction

A confined space (CS) is a space with limited or restricted means for entry or exit that, while allowing workers to enter and possibly perform activities, is not designed to be continuously occupied by workers (Spaces and Overview, 2023). Most industrial environments include CSs where cleaning, maintenance, inspection, and other tasks are required. The specific types of CSs can vary based on the industry and context. Examples of CSs in industrial environments include tanks, silos, pits, maintenance holes, tunnels, and utility vaults. Operators working in CSs are more exposed to the risk of accidents than in other industrial scenarios. In CSs, the common risk factors are exacerbated by the environments characterized by no easy escape routes and limited air renewal capabilities.

Despite the effort to define specific safety standards and procedures, the number of accidents, often fatal, remains unacceptably high for working activities in CSs. The Italian National Institute for Insurance against Accidents at Work (INAIL) reported that in Italy, between 2001 and 2019, 103 accidents out of 184 that occurred in CSs were fatal and caused the death of 151 workers (Ambienti confinati e/o sospetti di

* Corresponding author.
E-mail addresses: alessandro.evangelista@poliba.it (A. Evangelista), vitomodesto.manghisi@poliba.it (V.M. Manghisi), v.degiglio1@phd.poliba.it (V. De Giglio), r.mariconte@inail.it (R. Mariconte), c.giliberti@inail.it (C. Giliberti), antonio.uva@poliba.it (A.E. Uva).

https://doi.org/10.1016/j.ssci.2024.106693
Received 7 May 2024; Received in revised form 25 September 2024; Accepted 3 October 2024
Available online 9 October 2024

inquinamento e assimilabili - Prodotti di ricerca dell'istituto - INAIL, , 2023). A significant number of these accidents occurred inside storage tanks containing a poisonous atmosphere. These accidents follow similar patterns in every developed country. Indeed, according to a fact sheet from the U.S. Bureau of Labor Statistics, between 2011 and 2018, 1,030 workers died as a result of CS-related workplace injuries (Fig. 1); among them, the most common were silos and tanks for water and oil (Fatal occupational injuries involving confined spaces : U.S. Bureau of Labor Statistics, 2023). Despite the complexity of accident dynamics, data collected by regulatory agencies point to two factors as major causes: i) underestimation of risk and ii) lack of adequate worker information and training. Therefore, these agencies recommend practices for safety and health based on programs that include the removal of risk exposure (Management and Hazard Prevention, 2024) and appropriate training of workers (Management and Education, 2024).

Risk removal is accomplished a priori by analyzing the conditions at the operation site. If complete risk removal is not possible, it must be minimized using appropriate Personal Protective Equipment (PPE) and Collective Protective Equipment (CPE) as recommended by occupational safety procedures. However, it is often impossible to completely predict potential risks, especially those arising from the human-dependent dynamics of the intervention. Therefore, workers' responsiveness and ability to quickly recognize and analyze the severity of the risk are critical to making appropriate decisions in real time.

Most EU countries enacted specific laws and standards to improve CSs' safety by requiring workers to be specifically qualified to work in such environments. For instance, Decree 177/2011 in Italy aims to enhance security in confined and/or suspected pollution spaces (Di Donato et al., 2020). However, the different regulations have a complexity that makes them difficult to understand and still provide fragmented procedures in CSs (Botti et al., 2018).

For this reason, the regulations need to be concretized into a useful education and training course to enable operators to learn the correct ways to operate CS safely. These considerations underline the importance of an adequate education and training process for workers about the risks associated with CSs and the correct safety procedures to be carried out. This process must involve workers operating inside CSs and those supervising operations outside to alarm and/or intervene in rescue procedures.

The conventional education and training process consists of face-to-face or e-lectures (usually in classrooms) and practical training in a controlled real environment by simulating operative and emergency procedures (Lu et al., 2022). Practical training can be challenging due to limitations in time and resources, such as the availability of instructors, equipment, and appropriate training sites that can replicate the conditions of a real CS without exposing trainees to risks. Consequently, companies tend to limit the number of operators involved in the education and training process to the minimum mandatory ones, thus reducing the widespread dissemination of awareness of risks and safety procedures concerning CSs operations. Furthermore, it is worth noting that the regulations do not provide for continuous training (Unico, 2023). Thus, before the qualification to work in CSs expires, several months could elapse since the operators took the general training course. These considerations highlight the need for innovative approaches that, while retaining the effectiveness of traditional ones, make CS safety training more accessible by making it available to more and more workers.

Nowadays, the introduction of the Metaverse (Ning et al., 2023) and the technological readiness achieved by some technologies paves the way for innovative training approaches. For instance, Immersive Virtual Reality (IVR) technologies are emerging as a support tool for a viable alternative to traditional education and training. IVR can easily overcome the drawbacks of the conventional education and training approach thanks to the capability of allowing highly immersive experiences, the affordability of costs, and the ease of portability. Thus, it is reasonable to expect that this technology could significantly improve the education and training process in CSs, allowing workers to learn training procedures to face different dangerous scenarios with the possibility of failure without any risk concerns (Kaplan et al., 2020). Furthermore, the IVR implementation represents a foundational element in the development of next-generation applications that are "shared," "persistent," and "decentralized" in accordance with the Metaverse paradigm (Min and Cai, 2022).

Numerous studies have demonstrated the effectiveness of IVR-based training tools for workers in different industrial scenarios (Patle et al., 2019). However, the literature analysis reveals that such studies have three main limitations: i) only a few studies exploited IVR for CSs education and training, ii) the effectiveness was poorly investigated as most studies only evaluate the User Experience (UX) or the theoretical knowledge transfer, iii) as it is difficult to guarantee identical experimental conditions while measuring quantitative metrics as execution errors and reaction times, most studies neglect to evaluate the effectiveness in terms of positive behavior improvement, that is the capability to practically apply the theoretical knowledge.

This evaluation is of paramount importance for assessing safety training effectiveness; indeed, only converting theoretical knowledge into practical safety training skills can help avoid accidents. Furthermore, the outcomes of studies on the efficacy of IVR training vary according to the field of application, indicating that these results cannot be generalized across specific domains (Radhakrishnan et al., 2021).

The present work aims to bridge this gap by answering the main research question: "Are IVR-based education and training platforms

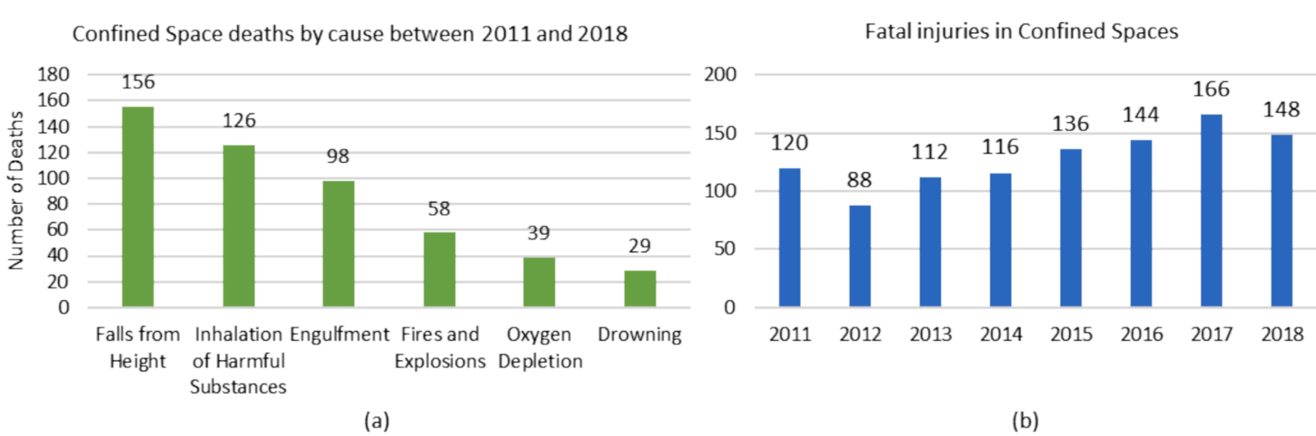

Fig. 1. Deaths by accidents in CSs in the US from 2011 to 2018. (a) Death by cause; (b) deaths by year. Data ()
Source Fatal occupational injuries involving confined spaces : U.S. Bureau of Labor Statistics, https://www.bls.gov/iif/factsheets/fatal-occupational-injuries-confined-spaces-, 2023.

effective tools supporting training for work safety in CSs?".

Supported by experts in the field of work safety, an IVR-based training platform to enable virtual training sessions in synthetic environments that accurately simulate real-world CS interventions was designed and implemented. In addition, a specific validation procedure to evaluate the effectiveness of the IVR platform as an educational and training tool for work safety in CSs was devised and applied in this study.

This study applies a rigorous evaluation and validation of an IVR-based educational platform, evidencing its efficacy and viability for industrial applications. Through a detailed articulation of the design process and subsequent assessment methodologies, the research highlights the considerable potential of IVR technology to enhance current industrial training frameworks. From this study, it is expected that the IVR-based education and training platform will demonstrate its effectiveness in knowledge retention and behavioral improvement and will also be perceived positively by the workers in terms of UX.

2. Background

The use of 3D Virtual Environments (VEs) evidenced numerous advantages for the learning process that can be related to the learning affordances that 3D environments provide. These affordances include facilitating tasks that lead to enhanced spatial knowledge representation, greater opportunities for experiential learning, increased motivation and engagement, improved contextualization of education, and more effective collaborative learning compared to tasks made possible by 2D alternatives (Dalgarno and Lee, 2010). Recently, such a result was confirmed by De Lorenzis et al. (De Lorenzis et al., 2023), who reported an improvement in motivation when learners immersively experience a 3D environment using IVR technologies for training.

Based on such evidence, IVR has been extensively studied, and the literature is replete with applications dedicated to procedural training. IVR-based applications have demonstrated effectiveness for procedural training, outperforming traditional methods such as classroom, paper, and video-based material (Abich et al., 2021).

Many studies report that one of the main advantages of IVR in training is to allow experiencing firsthand real contexts with the possibility of risk-free failure (De Lorenzis et al., 2023) allowing training based on the Productive Failure Theory, that demonstrated effectiveness in this field (Lu et al., 2024).

Furthermore, exploiting IVR technologies saves the time and cost of organizing and executing a real training session. For example, in the field of shipyards, Wahidi et al. (Wahidi et al., 2022) designed an IVR simulator for safety training, and they observed that the theoretical safety knowledge transferred through IVR increased by more than 14 % compared with the traditional one, also resulting in cost savings of around 16 %.

IVR training can be applied to many contexts (Radianti et al., 2020; Pérez et al., 2019; Manghisi et al., 2022; Chittaro and Buttussi, 2015; Buttussi and Chittaro, 2021) and could play a relevant role in industrial safety training. Indeed, there is a need to find new solutions to support education and training to improve occupational health and safety among workers, and using the IVR technology allows for acquiring and practicing the correct behavior in a controlled and safe environment (Grabowski and Jankowski, 2015).

One of the common fields where these technologies are involved refers to the construction industry (Scorgie et al., 2024; Stefan and Mortimer, 2023). Sacks et al. (2013) demonstrated the feasibility of utilizing IVR environments, such as power-wall displays, to simulate construction sites for safety training. This approach demonstrated the potential to be more effective than traditional classroom-based methods, as it was found to engage trainees' attention more effectively. Taking into account these considerations, subsequent studies (Nykänen et al., 2020) have evaluated the impact of novel, participatory safety training methods, including IVR-based programs. Results once more showed a significant improvement in workers' safety competencies, particularly in enhancing safety motivation, self-efficacy, and safety-related outcome expectancies when compared to traditional lecture-based training. Joshi et al. (2021) developed an IVR training system for safety protocols in the concrete industry using a VR Head Mounted Display (HMD). Their analysis indicated that IVR training is more engaging and provides a better understanding of safety protocols concerning conventional training methods. Liang et al. (2019) developed an IVR-based system for safety training in underground environments with rock-related risks. The authors observed an efficient and faster safety knowledge acquisition and an improved familiarity with the dangerous environment and operations compared with the traditional video training. Recent research has highlighted the significant advantages of using interactive VEs for fire safety training over traditional methods. In particular, IVR training has been shown to not only improve knowledge acquisition but also increase self-efficacy in learned procedures compared to non-interactive video training. Additionally, the use of 3D virtual worlds for fire safety and emergency evacuation training has been shown to be more engaging and effective, providing an immersive and accessible training experience without the need for advanced equipment (Shiradkar et al., 2021; Lovreglio et al., 2021).

Despite its wide application in many safety training fields, studies regarding the use of IVR for safety training in CSs are relatively scarce and incomplete, as they do not thoroughly investigate the effectiveness of training. Regarding this field of application, Grabowski and Jankowski (Grabowski and Jankowski, 2015) evaluated the effects of IVR-based training among a group of 21 workers in the mining industry, a working scenario where statistics highlight a specific danger for young, inexperienced employees. A pilot study revealed that IVR boosted participants' confidence and knowledge, even after three months, and generated a positive reaction from the supervisors and overall environment. Similarly, Gürer et al. (Gürer et al., 2023) developed an IVR training platform for the mining industry and involved mining engineers and game developers to test it. Despite the positive feedback and high participants' motivation in using these tools for training in occupational health and safety, these studies limited their scope to UX evaluation and did not assess the effectiveness concerning knowledge retention and behavior improvement.

Di Donato et al. (2020) developed an IVR-based system for safety training in CSs. By comparing the UX associated with two different training approaches, they evidenced that short training experiences, based on single tasks of interest, involve increased self-confidence and a lower feeling of pressure than a single multiple-task long experience. Unfortunately, this study does not evaluate the effectiveness of the IVR-based training system in transferring knowledge and correct behavior.

To the best of the author's knowledge, only a few studies evaluated an IVR-based training system in CSs regarding knowledge transfer effectiveness. Lu et al. (2020) developed and tested an IVR-based training system for rescue operations in CSs. In an initial study, they evaluated the acceptability (regarding reality reproduction fidelity, interaction, and storyline quality) and the knowledge transfer effectiveness by administering open questions. They found that the system significantly impacted the participants' understanding of CS rescue operations. In a successive study (Lu et al., 2022), the authors compared the IVR-based training system with an e-lecture-based one by administering these two training approaches to two groups of participants. Subsequently, by asking participants to execute the rescue tasks in a real scenario, the authors observed that the IVR-users group outperformed the e-lecture one. Other studies (Avveduto et al., 2017; Dang et al., 2023) involved IVR training systems specifically to evaluate their effectiveness compared to the traditional learning methods in high-risk environments, as well as in CSs. The authors noted that the IVR approach was more effective in terms of CSs knowledge and practical learning of safety procedures, with a high sense of presence and engagement experienced by the participants.

Safety in industrial environments extends beyond merely understanding proper workplace procedures (Lu et al., 2022); it also involves safeguarding the overall well-being of operators. CSs are often characterized by several disruptive elements, such as noise, which can affect workers' performance and well-being. Recent studies (Evangelista et al., 2024; Manghisi et al., 2022) investigated the effects of industrial noise within CSs on the cognitive abilities of operators using an IVR-based platform. The findings from these studies indicate that exposure to non-speech noise levels below the thresholds established by safety regulations has no effect on participants' cognitive performance or physiological responses. However, the results indicated that workers still perceived the noise as a disruptive element. This perception suggested that while cognitive performance may remain unaffected, the presence of industrial noise can negatively impact workers' perceived well-being.

In the literature, most studies dealing with IVR-based training in CSs predominantly focus on UX evaluation. Indeed, an IVR-based educational platform is evaluated on its ability to generate a virtual scenario that simulates real-world situations (Harris et al., 2020). To achieve this goal, the VR environment must be real enough to allow users to develop new skills that can be applied to real tasks. The necessary conditions for learning in VR environments occur only if users perceive the right sense of presence and immersion (Krassmann et al., 2022; Grassini et al., 2020; Dengel, 2018). For this reason, it is essential to assess the effectiveness of the virtual training environment based on subjective metrics (Graham and McAleer, 2018). Another challenging aspect of evaluating an IVR-based educational platform is its effectiveness in transferring knowledge to learners. Consequently, only a limited number of studies have assessed the effectiveness of theoretical knowledge transfer. Furthermore, at the best of the authors' knowledge, only two studies (Lu et al., 2022; Avveduto et al., 2017) evaluated the platform's ability to positively change behavior. This scarcity of research at the behavioral level may be attributed to the complexity of evaluating user performance in CS operations.

Today, state-of-the-art in IVR technology enables the cost-effective and realistic emulation of real-world environments. Moreover, the existing, albeit limited, studies on the transfer of knowledge and practical skills support the utility of IVR-based educational platforms. With these considerations in mind, an IVR-based educational platform specifically tailored for training in CS was designed and developed. Furthermore, this study exploited the IVR technology to overcome the challenges related to user performance evaluation for operations in CSs. Specifically, an experimental procedure was devised to validate the platform's effectiveness in terms of UX, theoretical knowledge transfer, and behavior improvement. This approach allowed for a comprehensive assessment of the impact of the IVR-based training system.

This study was carried out to investigate the following hypotheses:

H1- Users value the IVR-based educational platform through positive evaluation of UX and system usability;
H2– The IVR-based educational platform performs similarly or better than a traditional approach with respect to the ability to transfer knowledge;
H2– The IVR-based educational platform performs similarly or better than the conventional training approach concerning improving the correctness of the behavior.

3. Methodology

3.1. Platform design and development

The IVR system was designed for VR room-scale experiences optimized for HTC Vive Pro Eye HMD and hand controllers, utilizing Unity 2019.4 and the XR Interaction Toolkit framework. Virtual scenarios were reconstructed using 3D models based on technical drawings or acquired images. Non-player characters were added, including basic animated 3D characters and KinectV2-based motion-captured animations. Real-world audio inputs were incorporated using spatialized 3D audio simulation. Locomotion within the VE combined walking (Fig. 2 (a)) and Point and Teleport techniques (Dengel, 2018); (Bozgeyikli et al., 2016). Interaction and object manipulation were achieved using virtual

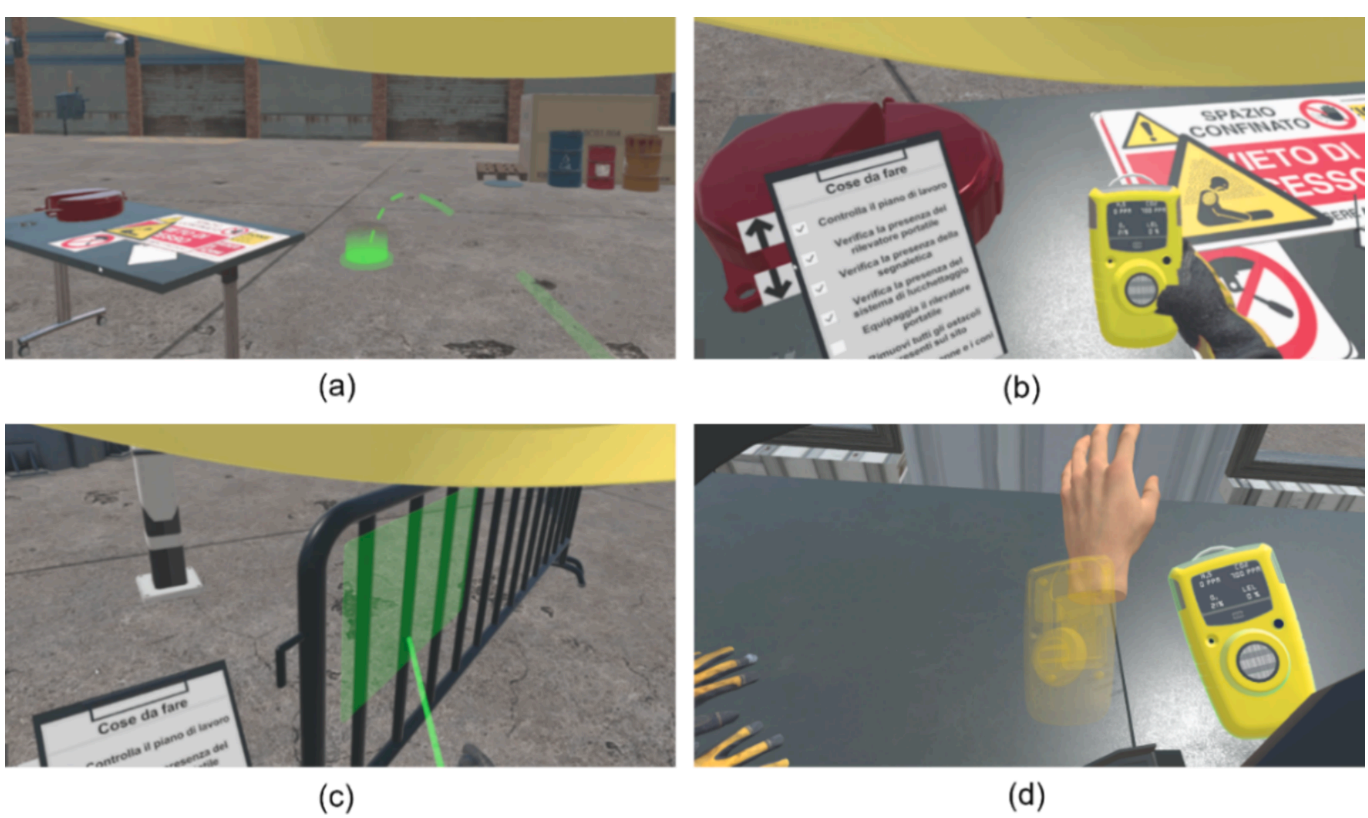

Fig. 2. (a) Point and Teleport, (b) Virtual Hands, (c) Laser-pointer based on Ray-casting, (d) Snap Zone.

hands (Lougiakis et al., 2020) and a laser pointer based on Ray-casting (Lee et al., 2003) (Fig. 2(b) and 2(c), respectively). A Snap Zone method (Saravas et al., 2020) facilitated precise object positioning by rendering transparent snap areas in high-contrast colors (Fig. 2(d)).

In designing the system, special consideration was given to the target end users, i.e., industrial operators often of medium to high age who may be unfamiliar with IVR and its interaction metaphors. To address this, a specific VE has been developed to familiarize users with the interaction metaphors used in the virtual environment for safety training. This module henceforth called the Interaction Training Module (IT module), is the initial VE in which participants are introduced to IVR. It includes detailed explanations of each IVR interaction metaphor, provided through a guiding voice and sample animations (Fig. 3). Furthermore, the module involves repeated practical exercises to ensure users achieve complete mastery of these interaction metaphors.

The list of the exercises to be carried out by the user in the IT module is reported in Table 1. The module provides two use modalities, the Training and Test modes. The former allows training according to the previously described modalities; the latter allows executing a predefined sequence of interaction metaphors without the guiding voice and sample animations, meanwhile recording the sequence completion time.

In addition to the IT module, developed for familiarization with IVR technologies and interactions, two other IVR modules have been developed specifically for the safety training program. The objectives and contents of these IVR modules are described in Sections 3.2, 3.3, and 3.4. Fig. 9 in section 3.7 explains how and when the IT module, and the other two IVR modules are used during the experimental procedure.

3.2. Structure of the training path

The IVR educational platform aims to transfer theoretical knowledge on safety measures and practical behavioral skills in CSs. The designed structure follows the experiential learning theory (Kolb et al., 2000), in which workers can experience different CSs, tasks, and roles, deploying their resources and skills to elaborate or reorganize actions to reduce risk exposure. Experiential learning also enables the subject to face uncertain situations by developing adaptive behavior and improving the ability to manage dangerous situations involving increased psychological stress. Therefore, the virtual experience in CSs becomes the knowledge base of the worker. Furthermore, by allowing workers to experience different risky scenarios with risk-free failure, the IVR system allows them to make mistakes and observe the consequences first-hand. As a result, according to the learning from mistake theory (Chialvo and Bak, 1999), workers increase their awareness of the risks associated with CSs.

The training path was structured into two main IVR modules. The first module provides information and education through virtually simulated traditional class lessons describing CSs-related risks and introducing PPE and CPE functionalities and employment. The second module provides practical training by virtually simulating working activities in real CS scenarios. By doing so, the requirements provided by the Italian Legislative Decree 81/08 were also fulfilled, which prescribes three specific components for CSs activities training – specific information, education, and training on risk factors (Unico, 2023). The contents of the first module and the scenarios and storyboard implemented in the second module were defined in collaboration with the Service for Prevention and Safety in the Workplace (SPESAL) staff of the [removed due to double-anonymized review].

Table 1
The IT module exercises in the Training and Test modes.

Interaction metaphor	Exercise	Number of repetitions Training mode	Test mode
Bipedal walking	Reach the destination point indicated by the placeholder	ND*	2
Point and teleport	Teleport sequentially to the three points indicated by the placeholders	ND*	5
Laser pointer interaction	Interact with four different target objects	ND*	2
Virtual hands	Grab four objects and place them in the designated places	ND*	2

*Until the user feels confident with it.

The IVR system is designed to transfer procedural knowledge to reduce risk during CS activities. The specific learning outcomes of the IVR training system consist of:

- Safety Procedures and Protocols: Participants will receive training on industry-standard safety procedures and protocols for working in CSs. This includes entry and exit protocols, atmospheric testing, ventilation requirements, and emergency response procedures.
- Risk Assessment and Mitigation Strategies: Participants will learn how to conduct risk assessments for CSs and implement effective risk mitigation strategies to minimize the risk of accidents and injuries.
- Decision-Making Under Pressure: Participants will simulate high-stress situations commonly encountered in CSs and learn how to make quick, informed decisions under pressure to ensure their safety and the safety of their team.
- Regulatory Compliance Awareness: Participants will gain an understanding of the regulatory requirements and compliance standards related to CS.
- Behavioral Awareness and Hazard Recognition: Participants will gain an understanding of the human factors and behavioral aspects that are relevant to working in CSs. This includes the significance of situational awareness, hazard recognition, and maintaining focus and concentration in high-risk environments.

3.3. The first IVR module – CSs-related risks and PPE/CPE presentation

This module delivers the information and education components in a virtual office container furnished with tables, stands for PPE/CPE, and a virtual screen (Fig. 4) according to the following sequence of actions. First, the user attends a lecture on a virtual screen inside the container describing the risks related to CSs. Then, the presentation of PPE/CPE takes place with a synthetic voice comprehensively describing each

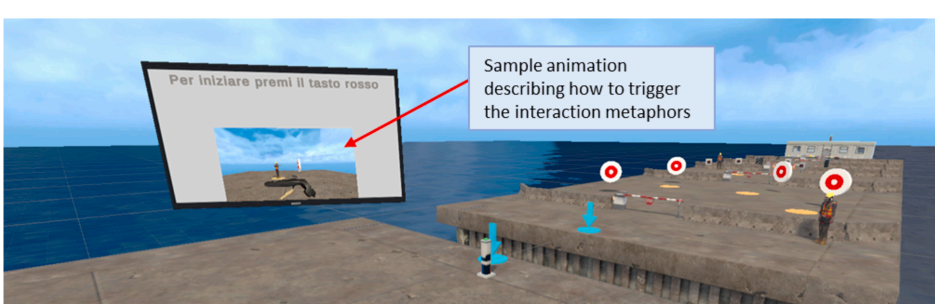

Fig. 3. The VE of the IT module with a virtual display for the interaction metaphors' sample animations and the targets for the interaction training tasks.

Fig. 4. The virtual office container furnished with tables and stands for PPE/CPE.

PPE/CPE by showing its 3D model and main features on the virtual screen. After the presentation of each piece of equipment, the system waits for the user to select it by grabbing it, and when the selection is successful, the system continues with the presentation of the next object. This scenario was implemented following the learning-by-mistakes approach. Therefore, the system does not provide the user with any visual indication of which object to select; instead, it only sends a text message on the virtual screen inviting the user to grab the object. The user's choice is based only on the information received from the synthetic voice and the virtual screen. If the user selects the correct object, the system provides the description of the next object; otherwise, it provides an error message that is shown on the virtual screen. The first IVR module ends when all the PPE and CPE are presented.

3.4. The second IVR module –Virtual training in CS scenarios

This module covers training in three different CS scenarios and allows two IVR experience modes: the Training mode, aimed at training, and the Test mode, aimed at assessing the knowledge acquired about CSs.

3.4.1. The CS scenarios and tasks

Two main factors guided the choice of scenarios: (1) cover the three CS risk-level classes (respectively Class A – high risk, Class B – medium risk, and Class C – low risk) as classified in the document: "Criteria for a Recommended Standard: Working in Confined Spaces" DHHS (NIOSH) Publication Number 80–106 (NIOSH-Issued Publications: Working in Confined Spaces (80-106) | NIOSH | CDC, https://www.cdc.gov/niosh/docs/80-106/default.html, last accessed, 2022); (2) identify scenarios commonly occurring in industrial facilities in the local area.

As a result, the following scenarios were identified:

- Underground utility compartment (Fig. 5) –Class A environment-: this CS presents a high and immediate risk to the health and life of the worker, including a lack of oxygen, flammable or explosive atmospheres, and high concentrations of toxic materials.
- Silo for the winemaking (Fig. 6) –Class B environment-: this CS can lead to injury and illness if preventive measures are not taken, but it is not immediately dangerous to life and health. High levels of carbon dioxide could be trapped inside the silo.
- Food oil storage silo (Fig. 7) –Class C environment-: this CS requires no modification of work procedures due to the negligible risk that does not affect the normal course of work.

Since short training sessions were shown to be more effective than long ones (Longo et al., 2021), the IVR-based experience in each CS scenario was split into three main use cases: i) site preparation, ii) PPE wearing and intervention, and iii) rescue action. A detailed description of use cases is reported in Appendix A.

3.4.2. The training and the test modes

The Training mode provides the worker with vocal guidance and visual hints for task execution. A synthetic voice guides the worker in the execution of each task (e.g., checking the atmosphere inside the CS through the gas detector). If the worker makes a mistake or the operation is not carried out within a predefined timeout (i.e., two minutes), the voice repeats the last instruction and resets the timeout. Given the difficulty of identifying specific objects within a VE densely populated with assets, the proposed system offers visual cues by highlighting the PPE or CPE for interaction. Furthermore, as the task execution order is fundamental for the procedure's safety, the Training mode does not allow interaction with objects not required by the task currently provided. In addition, text messages explicitly flag incorrect interaction attempts as errors.

Fig. 5. The entry space of the underground utility compartment –Class A environment.

Fig. 6. The silo for the winemaking – Class B environment.

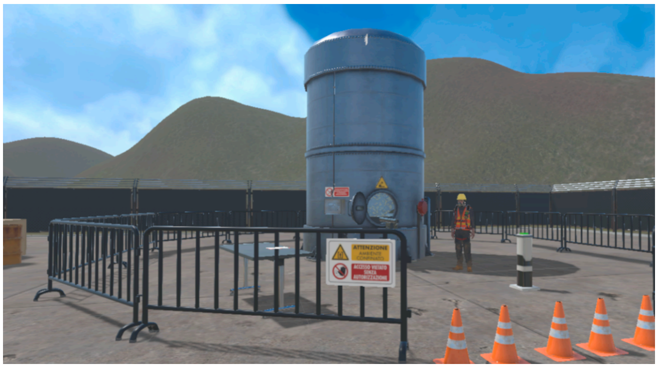

Fig. 7. The food oil storage silo −Class C environment.

The Test mode allows the assessment of the worker's performance. In this mode, there are no voice instructions or visual hints. The worker can freely interact with the virtual scene elements to replicate the correct safety procedure summarized on a virtual checklist mimicking the real operating conditions (Fig. 2(b)). When the test session ends, the system generates a digital report on the worker's performance based on the number and type of errors, as well as the time taken to complete the virtual scenario.

3.5. Validation procedure

Systematic validation of an education and training system requires an analysis that thoroughly investigates the different levels of learning and knowledge transfer (Makransky et al., 2019). In particular, it is crucial to assess not only the ability to transfer theoretical knowledge but also the capacity to positively modify behaviors in accordance with that knowledge. The approach based on Kirkpatrick's model (Smidt et al., 2009) is widely employed in the literature to this end (Alsalamah and Callinan, 2021).

Kirkpatrick's model outlines a specific evaluation approach structured around a 4-level model, which assesses the learning experience and outcomes based on:

1. The "reaction level": measuring participants' interest, motivation, and attention levels through post-course evaluations;
2. The "learning level": measuring what participants have learned in terms of knowledge and/or skills;
3. The "behavior or performance level": assessing the learners' ability to use the acquired knowledge or skills in the workplace;
4. The "The results level": measuring the overall impact of the training program, including financial or morale impacts.

The validation procedure was designed to evaluate the IVR-based training platform in terms of short-term outcomes with respect to the first three levels of Kirkpatrick's model. The hypotheses H1, H2, and H3 were investigated by evaluating the reaction, the learning, and the behavior levels, respectively. These evaluations involved comparing the outcomes of two different groups: the IVR group and the Control group, as detailed in section 3.7. The analysis of the results level was not carried out as its evaluation typically involves tracking longer-term outcomes, such as accident rates or productivity improvements, which require a long follow-up period. Furthermore, this study aimed to evaluate specific aspects of the VR training program, such as its effectiveness in conveying knowledge and in positively improving trainee's behavior, rather than its broader impact on organizational outcomes. Consequently, Level 4 evaluation fell outside the scope of the study.

3.6. Metrics

Specific metrics related to IVR applications have been identified in the literature for each level of Kirkpatrick's model (Fig. 8).

The Reaction level of the IVR Group was assessed in terms of UX and perceived usability. UX was evaluated by administering three questionnaires −the VR Sickness Questionnaire (VRSQ) (Kim et al., 2018), the User Experience Questionnaire (UEQ) (Schrepp et al., 2017), and the System Usability Scale (SUS) questionnaire (Lewis, 2018).

The VRSQ questionnaire assesses the presence of VR sickness-related symptoms by returning a score ranging from zero to 100 (the higher the score, the higher the symptoms). As the presence of such symptoms could be a condition of a participant before the IVR experience (i.e., not IVR-related), it is important to consider the preexisting condition and assess only whether and how much the IVR experience induced VR sickness-related symptoms. For this reason, the questionnaire was administered before and after the IVR experience. The difference in the VRSQ score (pre- and post-IVR experience) enables the assessment of VR sickness-related symptom onset resulting from the IVR experience.

The UEQ allows measuring both pragmatic and hedonic quality aspects by considering six scales: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. Regarding the use of the UEQ, the consistency of the results was analyzed by assessing the corresponding Alpha-Coefficient (Cronbach, 1951).

The SUS questionnaire evaluates perceived usability and returns a score that is compared to threshold values established in the literature (Lewis, 2018; Bangor et al., 2008; Sauro, 2012). Furthermore, users were asked to leave free comments concerning their experience with the IVR-based platform.

The effectiveness concerning the learning level was assessed by comparing the learning outcomes of the two groups. The learning outcomes were assessed as changes in correct answers by administering a custom Knowledge Questionnaire (KQ) in a multiple-choice format (ten questions with four alternative answers) before and after the corresponding training experience were evaluated. Thus, the KQs were administered before and after the training phase to measure the knowledge gain. The relevance of the observed effects was analyzed through a post-hoc power analysis to assess the effect size for the main effects by calculating the rank-biserial correlation and the power (1-β error probability) (Lakens, 2013; Cureton, 1956; Zhang et al., 2019).

The effectiveness concerning the behavior level was evaluated by analyzing users' performance (i.e., completion times and execution errors) in accomplishing operating procedures in CSs. The relevance of the observed effects was analyzed through a post-hoc power analysis to assess the effect size for the main effects by calculating the rank-biserial correlation and the power (1-β error probability).

3.7. Experimental procedure

The procedure follows a between-subjects design by comparing two groups. The first group, from now on the IVR group, undergoes the full IVR training course (i.e., first and second IVR modules). The second one, from now on, the Control group, undergoes a training course consisting of watching an e-lecture on a PC monitor (Lu et al., 2022) (i.e., a slideshow with a voiceover and videos describing CS-related risks, PPE and CPE, and safety procedures by giving practical examples). Professional trainers in the field of confined environment safety provided the e-lecture.

The ethics committee of the [removed due to double-anonymized review] approved the experimental protocol. The procedure consists of three phases. Beginning with the participants' enrollment, the training phase branches into two paths, one for each group. Finally, the test phase is common to both groups. Fig. 9 illustrates the procedure and its phases for the two groups.

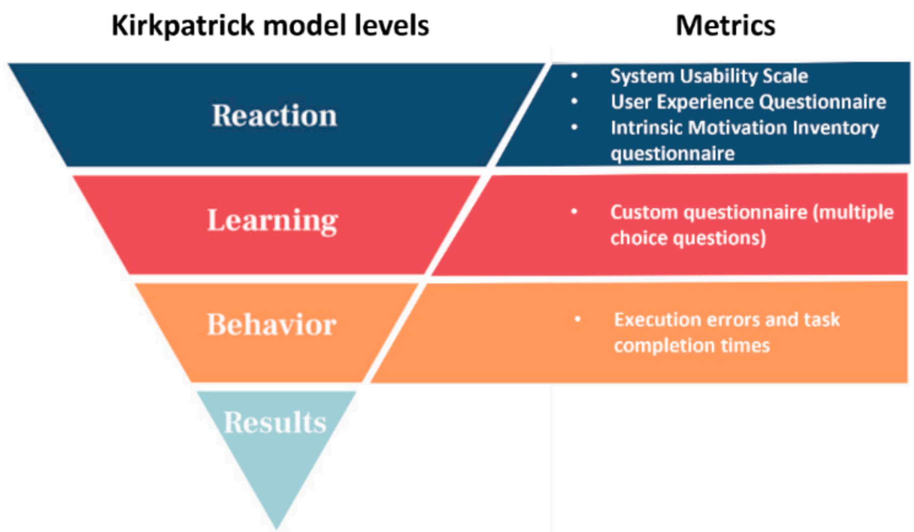

Fig. 8. The metrics used to assess the first three levels of Kirkpatrick's model.

3.7.1. Participants enrollment

This study aims to validate the platform with actual operators who previously attended a formal training course on CSs safety but do not routinely work in such spaces. For this reason, the recruitment process required a significant effort to achieve a population size that aligns with similar studies in the literature (Sacks et al., 2013; Rahouti et al., 2021; Jung and Ahn, 2018; Checa and Bustillo, 2020). Fifty participants (all males, ages ranging from 19 to 57 years, mean 30, SD 13.04) were recruited from four local industrial companies. No users reported suffering from visual impairment or phobias that would make them unable to participate in the experiment. The experiments were conducted within the spaces used by the companies as training classrooms.

After signing the informed consent, four questionnaires were administered to participants to divide them into two homogeneous groups based on demographics (Demographic Questionnaire, DQ), self-perceived familiarity with IVR (IVR questionnaire, IVRQ), self-perceived familiarity with operations in CSs (CS Questionnaire, CSQ), and prior Knowledge Questionnaire (KQ) on CSs.

Both the second and the third questionnaires (i.e., IVRQ and CSQ) consisted of one question asking to evaluate participants' familiarity according to a 7-point Likert scale with anchors on "not familiar at all" and "extremely familiar." Fig. 10 reports the collected data (Petrillo et al., 2011).

After answering the KQ questionnaire, participants were randomly split into two groups. It was positively checked that there were no statistically significant differences between the two groups regarding mean age, the number of correct answers, and familiarity with VR and operations in CSs (Table 2 reports the statistical tests' results).

3.7.2. The IVR group training phase (Fig. 9 – right column)

Before beginning the IVR experience, each participant of the IVR group completed the VRSQ. Subsequently, they started the IVR training experience by wearing the HMD (Fig. 11).

To become familiar with the VE interaction metaphors, each participant started the experience by practicing the IT Module in the Training mode. Subsequently, they experimented with the first IVR module. After completing it and an additional resting period of 15 min, the participant started experiencing the second IVR module in the Training mode. The IVR training was carried out using three use cases—site preparation, PPE application and intervention, and rescue—within the Class B scenario, which involved a silo used for winemaking. This scenario was chosen as it represents one of the most common CS accident environments in local industries. The IVR group training phase (i.e., the first and second IVR modules) lasted about 120 min. At the end of the IVR training phase, each participant completed the KQ questionnaires.

3.7.3. The Control group training phase (Fig. 9 – left column)

The Control group components attended the e-lecture on a 15.6-inch monitor of a notebook PC. Each participant could navigate the slideshow and its multimedia contents backward and forward according to their needs. Furthermore, an experienced CSs safety trainer supervised the e-lecture administration. The information provided during the e-lecture was consistent with that provided within the two IVR training phases. In particular, both Control and IVR groups received the same learning content regarding CS-related risks, PPE/CPE presentation, and correct operating procedures in the Class B environment scenario. At the end of the e-lecture, each participant completed the KQ questionnaires. The Control group training phase lasted about 120 min.

Since the aim is to exploit IVR technology for user performance assessment, it is important to ensure that users' confidence in using the technology does not affect the experiment's outcome. Therefore, after a 20-minute resting period and completing the VRSQ questionnaire, each Control group participant was asked to experience the IT module in the Training mode to become familiar with the IVR interaction metaphors. The participants completed the IT module in Training mode once they could easily accomplish each interaction task.

3.7.4. The test phase

To verify that there were no differences between the two groups concerning the ability and dexterity to use the IVR system, at the beginning of this phase, each participant performed a session of using the IT module in the Test mode. During this session, the execution times of a predetermined sequence of interaction tasks were recorded. Samples were positively checked for normality and homoscedasticity of distributions. The paired samples *t*-test did not evidence any statistically significant difference between the two groups (Table 3).

After a 60-minute rest break, evaluation of behavior level begins by using the second IVR module in the Test mode. At this stage, each participant in the two groups must complete the segregation procedure for another scenario, the Class A scenario. A different scenario was chosen to avoid performance being influenced by a learning effect related to the procedure sequence. All participants were encouraged to perform the procedure to the best of their abilities, as the top three

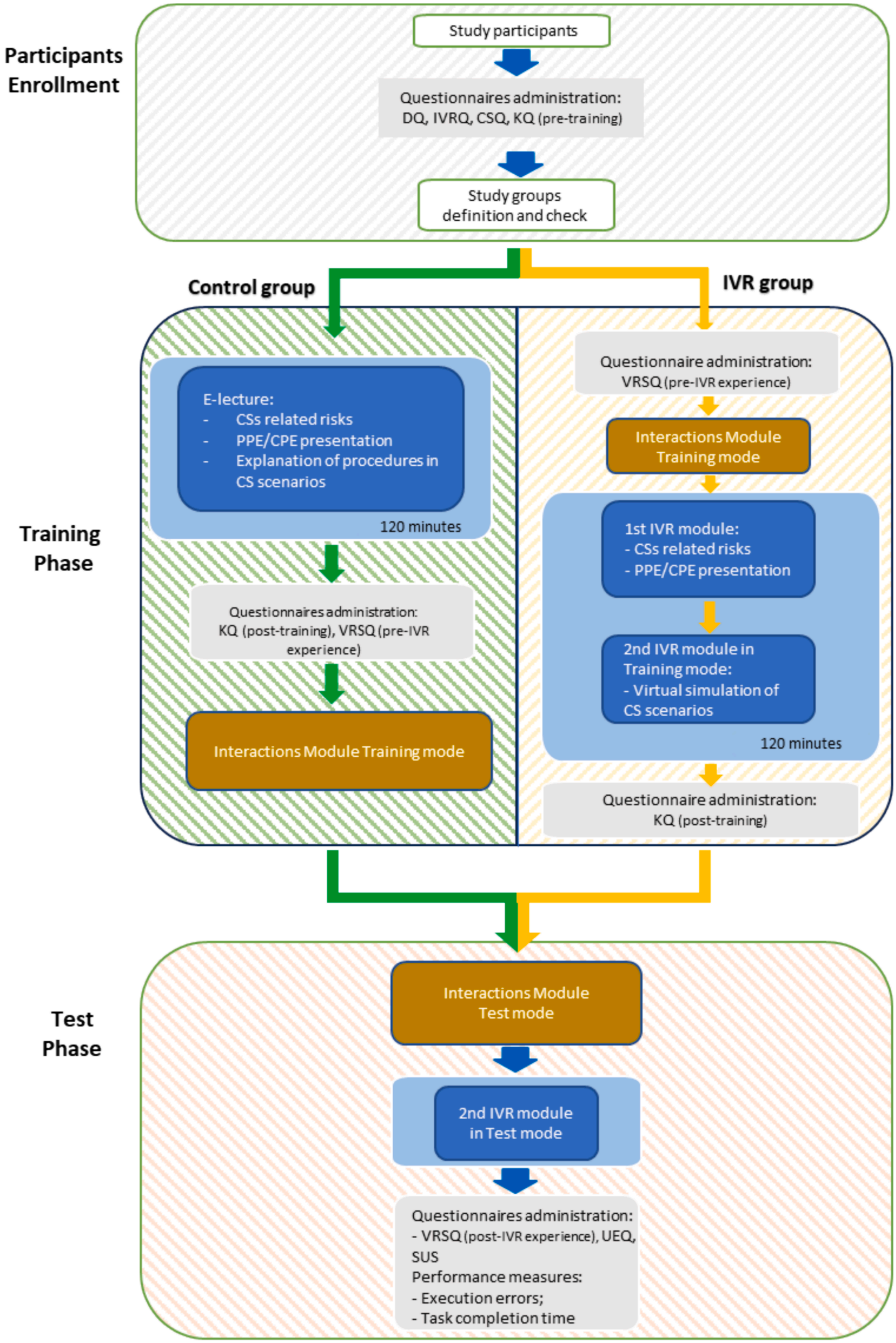

Fig. 9. The experimental procedure.

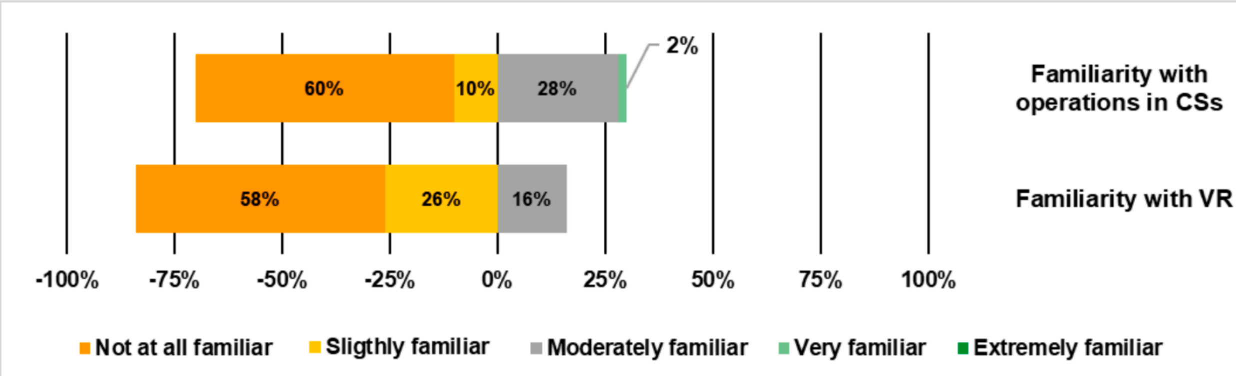

Fig. 10. Results from the IVRQ and the CSQ.

Table 2
Statistical analysis of the study groups' homogeneity.

Variable	Test	Statistics	p-value
Age	Mann-Whitney *U* test	U = 264.50	0.327
KQ correct answers	Mann-Whitney *U* test	U = 289.0	0.637
VR familiarity	Mann-Whitney *U* test	U = 236.0	0.094
CS familiarity	Mann-Whitney *U* test	U = 285.0	0.545

Fig. 11. An operator experiences the IVR platform during the experimental campaign.

scorers would be awarded a shopping voucher at the end of the testing campaign. During execution, the IVR platform recorded procedure completion times and execution errors. At the end of the procedure, all participants completed the VRSQ, the UEQ, and the SUS questionnaires. Furthermore, the experimenter asked to leave free feedback and comments on the IVR experience.

Table 3
Completion time comparison between the two groups for the IT module execution in Test mode.

Variable	Mean	Std. dev.	Test statistics		
IVR group Execution times	148.6	11.5	Ind. samples t-Test	t(48) = 0.918	p = 0.363
Control group Execution times	145.3	13.7			

4. Results

4.1. Reaction level

The VRSQ scores collected during the pre- and post- IVR experience for each group were compared. The Wilcoxon signed-rank test for paired samples did not allow the rejection of the null hypothesis (the median of the differences of the population before and after the IVR experience is zero) (Table 4). Thus, the two groups did not show any statistically significant difference before and after the IVR experience concerning the onset of IVR-related VR sickness symptoms.

Table 5 reports the scores obtained on the six evaluation scales of the UEQ (Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, And Novelty) and the corresponding Cronbach's Alpha, the Average Variance Extracted (AVE), and the Composite Reliability (CR) analysis results.

A UEQ analysis tool was used to compare the UEQ scores with a benchmark data set containing data from 21,175 persons from 468 studies concerning different products (Questionnaire, 2023). The comparison allowed for an estimation of the quality of the perceived UX. Fig. 12 reports the results of the comparison.

Regarding perceived usability, the SUS scores had a mean value of 82.20 and a Standard Deviation of 13.43.

4.2. Learning level

The correct answers (KQCAs) given by the participants to the KQ were collected. Thus, the Knowledge Improvement Index (KII) for each participant was calculated using the following formula:

$$KII = \frac{Post\text{-training } KQCAs - Pre\text{-training } KQCAs}{Pre\text{-training } KQCAs} \times 100$$

The IVR group achieved a mean KII of 38 %, while the Control group achieved a mean KII of 23 %. Fig. 13 shows a boxplot comparing the KII for both groups.

The normality of data for both groups was checked using the Shapiro-Wilk test. The results indicated that the Control group values were not normally distributed ($p < 0.05$), while the samples of the other

Table 4
VRSQ scores and statistical analysis.

VRSQ scores							Test statistics		
	Pre- IVR experience			Post- IVR experience					
	Mean	Median	SD	Mean	Median	SD			
Control group	1.83	0.00	3.20	3.03	0.00	3.76	Wilcoxon signed-rank	Z = -1.786	p = 0.074
IVR group	3.57	0.00	7.67	2.83	0.00	3.52	Wilcoxon signed-rank	Z = -0.465	p = 0.642

Table 5
UEQ ratings.

Scale	Mean	SD	Cronbach's Alpha	AVE	CR
Attractiveness	2.41	0.43	0.86	0.53	0.61
Perspicuity	2.14	0.67	0.74	0.54	0.59
Efficiency	1.90	0.46	0.55	0.30	0.35
Dependability	1.71	0.63	0.42	0.25	0.39
Stimulation	2.36	0.41	0.67	0.46	0.56
Novelty	2.46	0.57	0.71	0.52	0.59

group followed a normal distribution.

A Mann-Whitney *U* test was conducted to assess the differences between the two independent groups. The test indicated a statistically significant difference ($U = 209.0$, $Z = -2.015$, $p = 0.044$), suggesting a meaningful divergence between the groups. The IVR group had a mean rank of 29.64, whereas the Control group had a mean rank of 21.36. However, the effect size, calculated using the rank-biserial correlation ($rb = -0.285$) (Cureton, 1956), reflects a small effect. A post hoc power analysis indicated that the achieved power of the test was $P(1 - \beta) = 0.521$, which is considerably below the standard threshold of 0.80. This suggests that the study had only a 52.1 % chance of detecting an effect of this magnitude, highlighting a potential risk of Type II error. According to Cohen's guidelines, this suggests that although the result is statistically significant, the magnitude of the effect is limited (Lakens, 2013).

4.3. Behavior level

Regarding completion times, Fig. 14 reports the collected data for the two groups.

As samples were not normally distributed, a Mann-Whitney *U* test was performed to compare two independent groups. The IVR group had a mean score of 428.40, while the Control group had a higher mean score of 788.60. The analysis revealed a highly significant difference between the two groups ($U = 28.0$, $Z = -5.521$, $p < 0.001$), indicating a strong divergence in the distributions of the two groups' ranks.

The effect size, calculated using both the rank-biserial correlation ($rb = -0.781$) and Cohen's $d = -2.499$, suggests a large negative effect. The negative sign of the rank-biserial correlation indicates that although the Control group has a higher mean score, the IVR group tends to rank higher relative to the Control group. This suggests a substantial and meaningful difference between the groups, with the IVR group consistently outperforming the Control group in terms of rankings.

Additionally, a post hoc power analysis revealed that the achieved power was $P(1 - \beta) = 1$, indicating that the study had sufficient power (100 %) to detect the observed effect. This means there is essentially no risk of a Type II error, providing high confidence in the reliability of the significant result.

In regard to the execution errors, Fig. 15 reports the collected data for the two groups.

As samples were not normally distributed, a Mann-Whitney *U* test was conducted to compare the execution errors between two groups: the IVR group (mean 1.16, SD 0.99) and the Control group (mean 2.44, SD 1.58). The test revealed a statistically significant difference between the two groups ($U = 157.0$, $Z = -3.103$, $p = 0.002$), suggesting that the IVR group made significantly fewer execution errors compared to the Control group.

The effect size, calculated using the rank-biserial correlation ($rb = -0.439$) and Cohen's $d = -0.977$, suggests a large negative effect. The negative sign of the rank-biserial correlation indicates that the IVR group, which had fewer execution errors, consistently ranked better (lower number of errors) compared to the Control group. This large effect size indicates a strong and meaningful practical difference in performance between the two groups.

Additionally, a post hoc power analysis revealed that the achieved power of the test was $P(1 - \beta) = 0.9105$, indicating that the study had a 91.05 % chance of detecting an effect of this magnitude. This level of power is above the commonly accepted threshold of 0.80, suggesting

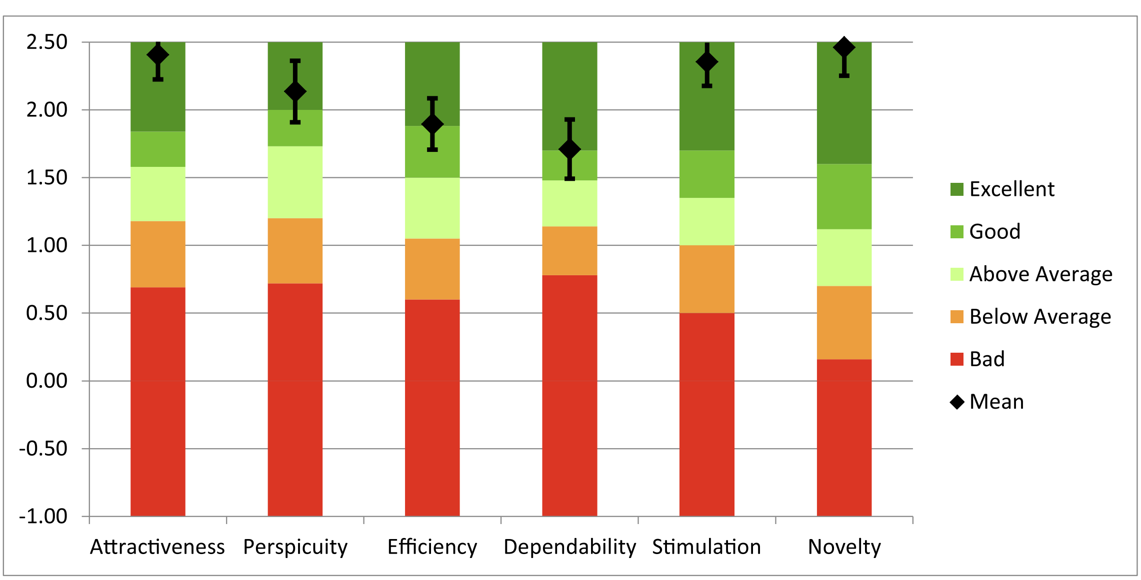

Fig. 12. UEQ ranking according to the UEQ analysis tool.

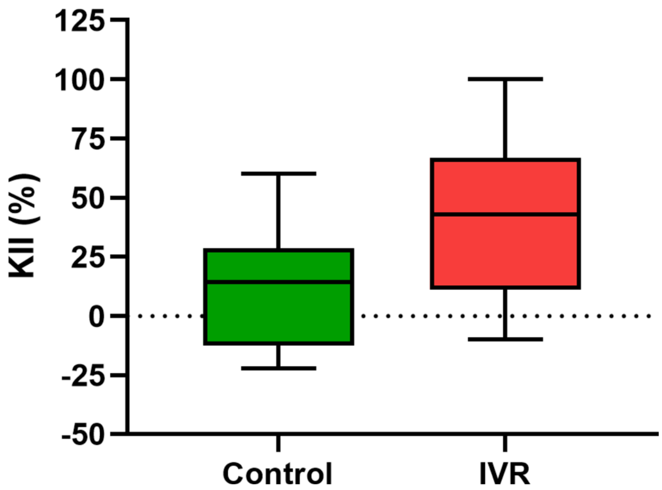

Fig. 13. Box plot for the comparison of the KII for the two groups.

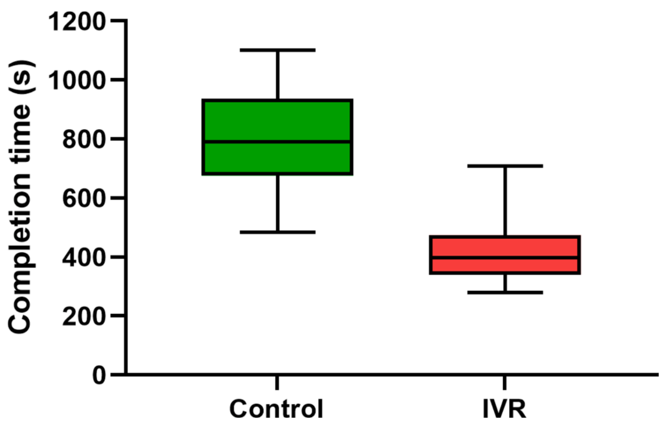

Fig. 14. The completion times (in seconds) for the procedure execution in Test mode.

that the study was sufficiently powered to detect the observed effect with a low risk of a Type II error.

5. Discussion

The results obtained from the experimental validation procedure concerning the three levels of Kirkpatrick's model (Reaction level, Learning level, and Behavior level) suggest the adoption of the IVR-based platform as a tool for supporting training in this specific field.

5.1. Reaction level

The obtained results support hypothesis *H1, suggesting that users highly value the IVR-based educational platform due to its positive evaluation of UX and system usability.*

The VRSQ scores collected to assess the potential occurrence of simulator sickness do not provide statistically significant evidence of such a problem in the tested population. This result is coherent with the one obtained by Nykänen et al. (Nykänen et al., 2020) and Lu et al. (Lu et al., 2022) and contradicts the findings of Joshi et al. (Joshi et al., 2021) and Martirosov et al. (Martirosov et al., 2022). It can be asserted that, in this deployment scenario, IVR technology does not induce issues of this nature. The different outcomes obtained by Martirosov et al. could be attributed to the platform's design choices and the technological maturity achieved by IVR over the years. Despite users' extended use of IVR, it is reasonable to believe that the choice of interaction metaphors, such as bipedal walking and teleportation, has mitigated spatial disorientation, which is a contributing factor to cybersickness (Bhandari and MacNeilage, 2018). Similarly, achieving high frame rates (over 120 fps) helps mitigate symptoms caused by latency of visual information perceived in IVR (Wang et al., 2023).

The UEQ analysis tool indicates a highly positive self-evaluation of the perceived UX. The scores across all scales are 'excellent,' placing within the top 10 % of the best results in the database. This indicates that only 10 % of the results in the database are better, while 75 % are worse.

Specifically, the Attractiveness scale obtained an average score of 2.41, with 'excellent' user ratings. The positive score obtained on this scale could be attributed to the high quality of the assets used to generate the VEs, which includes their proper arrangement within the scenarios. Additionally, particular attention was given to recreating the scenarios to be as consistent as possible with reality. This meticulous approach to design and attention to detail likely contributed to the platform's ability to attract and engage users, leading to an excellent evaluation of the Attractiveness scale.

Regarding Perspicuity, the IVR platform obtained an average score of 2.14, with an 'excellent' evaluation. The Perspicuity scale measures clarity and ease of understanding of the system (how easy it is for users to familiarize themselves with it). The positive score on this scale could be justified by the presence of the IT module, which allowed users to master the interaction and locomotion metaphors. This combination of well-established metaphors and the opportunity for users to become proficient in their use likely enhanced the overall clarity and ease of use of the IVR platform.

The IVR platform obtained an average score of 1.90 on the Efficiency scale, corresponding to an 'excellent' evaluation. This dimension measures how fast and effective the interaction with the system is in achieving users' goals. It includes system responsiveness, interaction efficiency, and ease of task completion. In this case, the score obtained could be attributed to the implemented interaction metaphors that, despite the complexity of the implemented use cases, enabled the user to interact with the VEs in a natural and efficient manner. Completing a use case requires significant effort from the user, as it involves numerous tasks and sub-tasks that must be completed in precise order as dictated by safety procedures in CSs.

Regarding the Dependability scale, although users reported a limited or no familiarity with VR before training through the IT module, at the end of the test, users mastered the interaction and locomotion system, as evidenced by the positive results of the IT module test. Thus, the IVR platform instilled an excellent sense of trust and reliability in users who tested it, with an average score of 1.71.

The IVR platform achieved an 'excellent' rating of 2.36 on the Stimulation scale, which evaluates the emotional aspect of the UX, assessing the system's ability to evoke positive emotions like excitement,

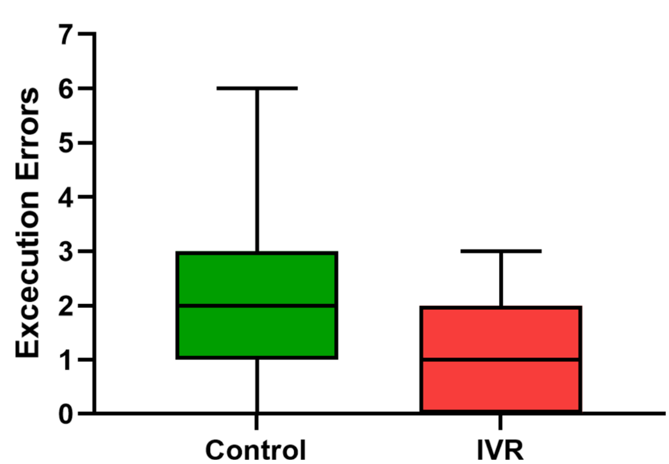

Fig. 15. The execution errors in Test mode.

interest, and entertainment. This remarkable result can be attributed to the distinctive features of IVR. Most users were new to such an experience, never having encountered the sensations it provides. The complete sensory immersion, including visual and auditory stimuli, likely provided the users with a noteworthy sense of presence. This, along with the lifelike and vibrant virtual scenarios, may have stirred intense and genuine emotions in the users. Lastly, the Novelty scale, which measures users' perceived originality and innovation of the IVR platform, received an excellent evaluation with an average score of 2.46. Since all operators involved had previously completed traditional training and education programs, the use of IVR might have been perceived as something novel, arousing curiosity and stimulating the users' interest. Discovering this new technology as an engaging and rewarding tool for training and education is considered a novelty by the users, in contrast to traditional frontal lessons, which could be perceived as boring and unstimulating.

However, the results need to be carefully considered by evaluating the consistency of the scales. To assess this, the Alpha-Coefficient (Cronbach, 1951) is a useful tool that measures the scale's consistency. Although many authors suggest that a scale should have an alpha value greater than 0.7 to be considered sufficiently consistent, these are more like rules of thumb and not based on rigorous statistical evidence (Schmitt, 1996). In this context, the values for the Attractiveness, Perspicuity, Novelty, and Stimulation scales are above or across the 0.7 threshold and can be considered consistent. However, the Efficiency and Dependability scales fall below the threshold value, and their interpretation requires careful handling. The analysis of Cronbach's alpha yielded results that were consistent with those from AVE and CR. AVE is a measure used to assess convergent validity (Netemeyer et al., 2003). Convergent validity measures the level of correlation between multiple indicators of the same construct that are in agreement. An AVE value of 0.5 or higher confirms convergent validity. CR is a measure of internal consistency in scale items. Values of composite reliability between 0.6 and 0.7 are acceptable (Shrestha, 2021).

The deviation from the target values may indicate that some items in these scales were interpreted in an unexpected way by some participants. It is also possible that, due to the novelty of the technology being used, some users may have encountered difficulties in understanding the concepts of efficiency (i.e., the users' capability to solve tasks without unnecessary effort) and dependability (i.e., the control and predictability of the interaction). As this aspect remains unclear, further investigation by conducting another experiment with the IVR experience will be planned, followed by interviews focused on this specific topic, guided by the obtained current results.

Nevertheless, the evaluations provided by the UEQ are consistent with the scores obtained from the System Usability Scale (SUS), which assesses perceived usability. In this case, the average score achieved is 82.20, surpassing the threshold value of 68, which is considered the minimum acceptable level of usability according to the literature. Specifically, the score of 82.20 corresponds to an 'excellent' rating on the scale developed by Bangor et al. (Bangor et al., 2008). This outcome is particularly meaningful considering the cohort compositions. Gürer et al. (Gürer et al., 2023) obtained scores of 85 and 87 for groups of mining engineers and game developers, respectively, while this study was carried out with actual industrial operators, of whom only 16 % declared to be familiar with VR technologies. Additionally, the score observed in the current research is significantly higher than that of Joshi et al. (Joshi et al., 2021). The free comments left by participants after the IVR experience further support the positive evaluation of the UX and involvement. The most relevant ones were the following:

- "By using the system, I was able to see up close things that had only been described to me in theory";
- "By using the system, I remembered things that I was forgetting after a long time";
- "It feels like really living the experience. I was completely caught up in it".

5.2. Learning level

Results support the hypothesis *H2-The IVR-based educational platform performs similarly or better than a conventional approach concerning the ability to transfer knowledge.*

The improved performance of the IVR group, as measured by the KII, could be attributed to the fact that IVR allows visualization and firsthand interaction through a more effective process of learning by doing and learning by mistakes, as evidenced in (Dalgarno and Lee, 2010; De Lorenzis et al., 2023).

This result extends the experimental validation of the effectiveness of this type of platform for training in this field of application and is consistent with previous findings by Lovreglio et al. (Lovreglio et al., 2021) for fire extinguisher training, by Saks et al. (Sacks et al., 2013), by Nykänen et al. (Nykänen et al., 2020) and Dang et al. (Dang, H., Serne, J., Tafazzoli, M.: Virtual Reality Safety Training Assessment in Construction Management and Safety and Health Management Programs. Computing in Civil Engineering, 2023) for construction workers, and Lu et al. for CS rescue training (Lu et al., 2022). These findings are in contrast with other studies reported in the literature. Specifically, Joshi et al. (Joshi et al., 2021) identify a dependency of knowledge transfer on the type of training in the precast concrete industry but fail to detect any statistically significant differences. Similarly, Shiradkar et al. (Shiradkar, S., Rabelo, L., Alasim, F., Nagadi, K.: Virtual World as an Interactive Safety Training Platform. Information, 2021), in their study on emergency and evacuation training, and Avveduto et al. (Avveduto et al., 2017), in the field of inspection and maintenance, did not report statistically significant outcomes.

While the results indicate a statistically significant difference between the groups, the moderate negative effect size ($rb = -0.285$) indicates that the difference in ranks is not large. The IVR group has a higher mean score (KII = 38 %) compared to the Control group (KII = 23 %), but the rank-biserial correlation suggests that the Control group is still performing relatively better in terms of rankings. The low power of the study further suggests that the sample size was insufficient to conclusively detect a robust effect.

Given the low statistical power and the conflicting results in the literature, it can be hypothesized that the effectiveness of VR-based training in knowledge transfer may not differ significantly from that of traditional training methods. Furthermore, it should be kept in mind that the results refer to a cohort of participants who had previously attended a specific conventional training course. Although it was verified that the level of prior knowledge was homogeneous between the two groups, the effectiveness of the IVR platform concerning the learning level was evaluated only regarding its use as a tool to support training. Thus, it would be appropriate to integrate the investigation explained in this work by carrying out a specific study involving participants at the beginning of their training path. By doing so, it would be possible to investigate the effectiveness of an IVR-based platform not only as a training support tool but also as a possible alternative training tool.

5.3. Behavior level

The improved completion times and the reduced number of execution errors in the IVR group, compared to the Control group, suggest that the IVR-based platform can effectively leverage the benefits of experiential learning (KOLB, D.: The Process of Experiential Learning. Strategic Learning in a Knowledge Economy. 313–331, 2000). The findings support hypothesis *H3– The IVR-based educational platform performs similarly or better than the conventional training approach concerning improving the correctness of the behavior.*

VR appears to be suitable for this type of learning, as it enables users to participate in hands-on experiences within a safe and controlled environment (Chen et al., 2024). As previously noted, acquiring this level of practical experience is often difficult in traditional training due to safety concerns and cost limitations.

Regarding execution errors, the present findings align with those of Avveduto et al. (Avveduto et al., 2017) and Lu et al. (Lu et al., 2022), who identified a similar pattern of reduced errors in participants trained with a VR simulator compared to those trained using traditional methods. However, execution times could not be compared, as their studies did not include such data.

The statistically significant result ($p = 0.002$) indicates a real difference between the IVR and Control groups in terms of execution errors, with the IVR group performing substantially better. The large effect size ($rb = -0.439$, $d = -0.977$) suggests that the difference in performance is both statistically significant and practically meaningful. The IVR group, on average, made significantly fewer execution errors than the Control group, and this difference is reflected strongly in the ranks, with the IVR group consistently outperforming the Control group in terms of error reduction. Furthermore, the high post hoc power ($P(1-\beta) = 0.9105$) indicates that the study was well-designed in terms of sample size and had more than adequate power to detect the observed effect. This substantially reduces the risk of a Type II error, ensuring confidence that the significant difference is not attributable to chance or an inadequate sample size.

Certainly, given the limitations in terms of the haptic feedback that can be returned to the user, the IVR training system cannot currently fully replace traditional practical training in physical simulations of real environments and tools. It is believed that traditional practical training remains paramount for improving operators' dexterity with specific working tools such as PPE and CPE. Regardless, the reduction in execution errors suggests that IVR training can facilitate the effective acquisition of behaviors that comply with the execution steps outlined in safety procedures. Consequently, IVR-based training could be used as both a preparatory and a reinforcing tool to enhance knowledge acquisition, retention, and recall, as suggested by (Sacks et al., 2013; Lovreglio et al., 2021). Results encourage the organization of a focused study to further validate the proposed procedure and assess its alignment with outcomes from evaluations conducted in real-world environments.

The difference in effect size between the learning and behavioral levels, along with the findings from other studies in the literature, suggests that training using IVR platforms is particularly effective in improving practical behavior. In contrast, the effectiveness of theoretical knowledge transfer appears comparable to that of traditional systems. This finding could be useful for planning the use of these tools in training programs.

5.4. Limitations

Among the objectives of the present work was to conduct a study that is ecologically valid in terms of experimental conditions and the population involved in the experiment. Achieving this goal required close collaboration with the local Service for Prevention and Safety in the Workplace to define the VEs and the training path structure. Additionally, considerable effort was devoted to recruiting the test cohort, which had to consist of actual industrial operators who had previously undergone safety training for CSs.

As a result, this study presents some limitations. First, the final sample size consisted of 50 participants. Although a larger population was initially screened to maximize the experimental cohort, the final selection was constrained by rigorous inclusion criteria to ensure the chosen sample's ecological validity (industrial operators who had completed formal training in CSs safety but did not routinely work in such environments). Additionally, it was not possible to achieve gender balance within the sample, as none of the companies involved had female personnel who met the selection criteria. Nonetheless, the sample size is comparable to those reported in similar studies in the literature (Scorgie et al., 2024; Sacks et al., 2013; Rahouti et al., 2021; Jung and Ahn, 2018; Checa and Bustillo, 2020).

Although the experimental phase yielded promising results, it is crucial to validate the effectiveness of the IVR platform in a real CS environment. This validation is necessary because the transferability of skills and knowledge from a virtual training environment to high-risk real-world scenarios is not guaranteed. However, as noted in section 3.5, such an analysis, in line with the Kirkpatrick model's outcome level, requires extended follow-up periods. Therefore, this aspect is beyond the scope of the current study and will be addressed in future research.

Although the present analysis identified a statistically significant difference between the groups ($p < 0.05$) regarding the KII, the moderate effect size and low statistical power constrain the robustness of these findings. To gain a more definitive understanding of the relationship between these variables and to minimize the risk of Type II error, future research should incorporate a larger sample size.

When interpreting the results, it is crucial to acknowledge that the experimental cohort solely consists of male individuals. This limitation arises from the absence of females qualified for CS operations within the collaborating companies. Consequently, to generalize the findings of this work, it would be necessary to conduct additional inclusive, experimental campaigns involving female participants.

6. Conclusions

The adoption of IVR technology for industrial training is becoming increasingly prominent in applied research projects, especially in high-risk environments like CSs. The technology's success is largely attributed to its ability to reduce costs, control experimental conditions, and facilitate risk-free simulations. These simulations are particularly beneficial for training in potentially hazardous environments and procedures. Empirical evidence has consistently demonstrated IVR technology's effectiveness in user acceptance and knowledge transfer. However, there is still limited validation of its potential to positively influence user behavior due to the complexities and expenses associated with such evaluations.

In this study, the challenges in evaluating user performance during CS operations were addressed by harnessing the capabilities of IVR technology. A simulated CS using IVR was created, allowing users to experience these operations without any actual risk. This approach enabled a comprehensive assessment of user behavior and performance in a realistic yet safe manner, thereby significantly enhancing the evidence supporting the effectiveness of IVR-based training systems in improving work safety procedures in CSs. By combining the IVR platform with the application of Kirkpatrick's model, valuable insights into the platform's impact on users' reactions, knowledge transfer, and behavioral changes were gained. However, it is important to emphasize that the validation was conducted with actual industrial operators, and this approach required limiting the number of participants. This comprehensive approach contributed to a deeper understanding of the potential of IVR in CS training. Specifically, the investigation explained in this study revealed the following key findings:

- Users expressed positive evaluations regarding their experience with the system, considering hedonism, pragmatism, and usability aspects.
- Under the defined experimental conditions, the use of IVR did not result in any side effects related to simulation sickness.
- The IVR-based training approach led to a comprehensive enhancement of behavioral skills.
- Engaging in "hands-on" training within the IVR environment further contributed to performance improvement.
- Training in IVR resulted in a reduction in errors and execution time.

These findings distinguish this study from previous IVR research by offering a focused evaluation within the context of CS operations and highlighting the specific benefits of IVR on safety behavior in this high-risk environment. Moreover, the study suggests potential fields of application for IVR training, such as large-scale basic training aimed at improving "risk and safety literacy," serving as a supportive tool for conducting risk-free practical simulations and acting as a "reinforcement" tool for individuals who have previously completed traditional training courses.

The present study demonstrates the significant potential of IVR technology for industrial training in CSs. Its unique contributions lie in the focused evaluation of IVR within the context of CS operations and the provision of empirical evidence supporting its effectiveness. The results of this study broaden the experimental evidence and contribute to a deeper understanding within the scientific community of how IVR technology can be effectively developed and implemented in training programs.

However, further research and validation are necessary to deepen understanding in this field and ensure successful implementation across various industries.

The proposed validation procedure effectively facilitates the identification of distinctions between the two groups and could also be applicable for validating any training applications based on IVR technology. It eliminates the risks associated with evaluating behavior through the execution of procedures in hazardous real environments. However, an additional validation step is required to assess its consistency with the results obtained from such evaluations. The next steps in this research aim to broaden the scope of validation by integrating the IVR training system into the Metaverse. The properties of sharing, persistence, and decentralization will be tested for their potential positive or negative impact on the effectiveness of the training program.

CRediT authorship contribution statement

Alessandro Evangelista: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Vito M. Manghisi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Vito De Giglio:** Writing – original draft, Visualization, Investigation, Data curation. **Raffaele Mariconte:** Supervision. **Claudia Giliberti:** Supervision. **Antonio E. Uva:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antonio Emmanuele Uva reports financial support was provided by National Institute for Insurance against Accidents at Work.

Acknowledgments

This work is part of the SPIRiT project (Secure Personalized Immersive Reality innovative Trainer) that was supported by the 2019 BRIC grant funded by the National Institute for Insurance against Accidents at Work (INAIL) of Italy.

Appendix A. — use cases

The three use cases, common to the three CS scenarios, cover the different phases of work according to safety best practices in CSs. Each use case involves user tasks that differ according to the characteristics of the CS. In Table A1, the tasks related to the different use cases of the identified CSs are reported.

Table A1
The use cases and the related user tasks for the CSs.

Class A Environment Use case	**User Task**	**Task description**
Site Preparation	1. Removal of bulky objects and environment confinement; 2. Valve/electrical switch closure with safety systems; 3. Access opening, room ventilation, and indoor atmosphere check; 4. Installation of rescue anti-fall tripod and descent ladder; 5. Posting caution and warning signs.	• Removal of objects obstructing the passage and positioning of safety barriers and safety cones; • Valve switching, locking, and padlocking; • Site opening and ventilation, O_2 levels, explosiveness, and pollutants checks; • Positioning of anti-fall device and ladder for descent into the site; • Caution and warning signs posting.
PPEs wearing and intervention	1. Wear PPEs; 2. Verification of safety conditions outside the site; 3. 3. Entering the site, pressure gauge check, and painting.	• Displaying the list of PPEs, after wearing/checking them for function check-list is updated; • Control of operators' presence, communication system, atmospheric conditions, safety rope attachment, and painting gun holding; • Check that the pressure gauge reads 1 atm and start the painting procedure.
Rescue action	1. Site monitoring by the external operator and verification of internal operator conditions; 2. Activation of public emergency service and site evacuation.	• The external operator oversees and monitors the site, and the internal operator with the communication system; • If necessary, the external operator activates the public emergency service and recovers/rescues the internal operator by evacuating the site.
Class B Environment Use case	**User Task**	**Task description**
Site Preparation	1. Removal of bulky objects and environment confinement; 2. Valve/electrical switch closure with safety systems; 3. Access opening, room ventilation; 4. Installation of rescue anti-fall tripod and indoor atmosphere check; 5. Posting caution and warning signs.	• Removal of objects obstructing the passage and positioning of safety barriers and safety cones; • Valve/switching locking and padlocking; • Site opening and ventilation; • Positioning of anti-fall device and verification of O_2 levels, explosiveness, and pollutants; • Caution and warning signs posting.
PPEs wearing and intervention	1. Wear PPEs; 2. Verification of safety conditions outside the silo; 3. 3. Entering the silo, inspection, and water-jet cleaning;	• Displaying the list of PPEs, after wearing/checking them for function check-list is updated; • Control of operators' presence, communication system, atmospheric conditions, and safety rope attachment; • Start of cleaning procedure and periodic atmosphere breathability check.

(*continued on next page*)

Table A1 (*continued*)

Class A Environment Use case	User Task	Task description
Rescue action	1. Site monitoring by the external operator and verification of internal operator conditions; 2. Activation of public emergency service and site evacuation.	• The external operator oversees and monitors the site, and the internal operator with the communication system; • If necessary, the external operator activates the local emergency service and recovers/rescues the internal operator by evacuating the site.
Class C Environment Use case	**User Task**	**Task description**
Site Preparation	1. Removal of bulky objects and environment confinement; 2. Valve/electrical switch closure with safety systems; 3. Access opening, room ventilation; 4. Indoor atmosphere check and installation of a side-entry rescue system; 5. Posting caution and warning signs.	• Removal of objects obstructing the passage and positioning of safety barriers and safety cones; • Valve/switching locking and padlocking; • Site opening and ventilation; • Verification of O_2 levels and pollutants, positioning of the life-saving system; • Caution and warning signs posting.
PPEs wearing and intervention	1. Wear PPEs; 2. Verification of safety conditions outside the silo; 3. 3. Entering the silo, inspection, and water-jet cleaning;	• Displaying the list of PPEs, after wearing/checking them for function, the check-list is updated; • Control of operators' presence, communication system, atmospheric conditions, safety rope attachment, and high-pressure washer lance holding; • Start of cleaning procedure and periodic atmosphere breathability check.

Appendix B. — CSs scenarios and use cases implementation

The IVR system provides an interface allowing the users to choose both the type of confined environment and the use case to be experienced. Through a monitor interface, an instructor selects the scenario and use case while the worker, who has already worn the HMD, waits for the start in a virtual office container. Thus, when the instructor starts the scenario, the worker is teleported into a VE that faithfully simulates the chosen CS.

As stated previously, for each CS scenario, the developed system provides three different use cases (site preparation, PPEs wearing and intervention, and rescue) in which every single task must be performed in a specific order using the action-driven method.

The site preparation, designed to educate and train in securing the area surrounding the CS entrance, begins with the user activating a specific 3D button. A synthetic voice describes the CS in the Training mode, presenting its main characteristics and potential risks. Next, the entry clearance document and a checklist for the operating procedure are displayed. The procedure checklist is always available to the worker, and a synthetic voice guides the user during the performance of tasks (for the tasks list, see Table 1). Once all tasks are completed, the user receives feedback based on a synthetic voice message on completing the site preparation procedure. If the task is not executed correctly or the correct execution order is not followed, the system reports the error with a text message displayed on the screen. The system does not allow the user to proceed to the next task if the previous one is not correctly executed.

The PPE-wearing and intervention procedure is designed to educate and train the operators to carry out the intervention within a CS, which is the most dangerous phase of the work. This scenario was implemented into two different parts: PPE-wearing and intervention. The PPE-wearing process occurs inside the office container, where PPE and other equipment are placed on or around a table. The wearing procedure starts when a synthetic voice describes the CS presenting its main characteristics and potential dangers. Depending on the type of CS, the system requests the user through a synthetic voice to wear the PPE necessary for the intervention, also showing them on a checklist.

Furthermore, the system will visually indicate the PPE to wear, highlighting it with a green halo. For some PPEs, a specific order of wearing is required. If the wearing order is incorrect, the system reports the error with a text message. The system does not allow the user to proceed to wear the next PPE if the previous one is not correctly worn. The user can access the intervention procedure inside the CS at the end of the wearing procedure. The intervention procedure requires the performance of user tasks that strictly depend on the type of CS (Class A, Class B, or Class C) experienced. This procedure is carried out with the same modalities of the site preparation procedure, i.e., with a synthetic voice and checklist that guides the user during the performance of tasks. However, particular attention is paid to implementing negative consequences (explosion, poisoning, and subsequent death). In fact, in the case of incorrect execution of the tasks, the error will no longer be notified on the screen via a text message but through a scripted scene which will lead to the death of the operator and the end of the scenario simulation. Furthermore, during the intervention in the CS, the system allows for generating unforeseen risks that the worker must face in a precise time frame.

The rescue procedure, implemented for Class A and Class B environments, is designed to educate and train on procedures for rescuing the unconscious operator from the CS. In this training scenario, the operator must monitor, by continuous visual contact, the operator within the CS. In case of loss of consciousness of the operator within the CS (simulated by an NPC), the worker will be required to implement procedures to extract the unconscious operator. In addition, the worker will be required to alert the public emergency. If the operations are not carried out correctly, timely, and in the correct order of execution, the system will show negative consequences (death of the rescued operator).

References

Abich, J., Parker, J., Murphy, J.S., Eudy, M., 2021. A review of the evidence for training effectiveness with virtual reality technology. Virtual Real. 25, 919–933. https://doi.org/10.1007/S10055-020-00498-8/TABLES/2.

Alsalamah, A., Callinan, C., 2021. The Kirkpatrick model for training evaluation: bibliometric analysis after 60 years (1959–2020). Ind. Commer. Train. 54, 36–63. https://doi.org/10.1108/ICT-12-2020-0115/FULL/XML.

Ambienti confinati e/o sospetti di inquinamento e assimilabili - Prodotti di ricerca dell'istituto - INAIL, https://www.inail.it/cs/internet/comunicazione/pubblicazioni/catalogo-generale/pubbl-ambienti-confinati-prodotti-di-ricerca-istituto.html, last accessed 2023/07/21.

Avveduto, G., Tanca, C., Lorenzini, C., Tecchia, F., Carrozzino, M., Bergamasco, M.: Safety Training Using Virtual Reality: A Comparative Approach. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 10324 LNCS, 148–163 (2017). https://doi.org/10.1007/978-3-319-60922-5_11.

Bangor, A., Kortum, P.T., Miller, J.T., 2008. An empirical evaluation of the system usability scale. Int. J. Hum. Comput. Interact. 24, 574–594. https://doi.org/10.1080/10447310802205776.

Bhandari, J., MacNeilage, P., interface, E.F.-G., 2018, undefined: Teleportation without Spatial Disorientation Using Optical Flow Cues. graphicsinterface.org. (2018).
Botti, L., Duraccio, V., Gnoni, M.G., Mora, C., 2018. An integrated holistic approach to health and safety in confined spaces. J. Loss Prev. Process Ind. 55, 25–35. https://doi.org/10.1016/J.JLP.2018.05.013.
Bozgeyikli, E., Raij, A., Katkoori, S., Dubey, R., 2016. Point & Teleport locomotion technique for virtual reality. In: CHI PLAY 2016 - Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play, pp. 205–216. https://doi.org/10.1145/2967934.2968105.
Buttussi, F., Chittaro, L., 2021. A comparison of procedural safety training in three conditions: virtual reality headset, smartphone, and printed materials. IEEE Trans. Learn. Technol. 14, 1–15. https://doi.org/10.1109/TLT.2020.3033766.
Checa, D., Bustillo, A., 2020. A review of immersive virtual reality serious games to enhance learning and training. Multimed Tools Appl. 79, 5501–5527. https://doi.org/10.1007/S11042-019-08348-9/FIGURES/13.
Chen, P.H., Ho, H.W., Chen, H.C., Tam, K.W., Liu, J.C., Lin, L.F., 2024. Virtual reality experiential learning improved undergraduate students' knowledge and evaluation skills relating to assistive technology for older adults and individuals with disabilities. BMC Med. Educ. 24, 1–12. https://doi.org/10.1186/S12909-024-05085-Y/TABLES/5.
Chialvo, D.R., Bak, P., 1999. Learning from mistakes. Neuroscience 90, 1137–1148. https://doi.org/10.1016/S0306-4522(98)00472-2.
Chittaro, L., Buttussi, F., 2015. Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety. IEEE Trans vis Comput Graph. 21, 529–538. https://doi.org/10.1109/TVCG.2015.2391853.
Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555/METRICS.
Cureton, E.E., 1956. Rank-biserial correlation. Psychometrika 21, 287–290. https://doi.org/10.1007/BF02289138/METRICS.
Dalgarno, B., Lee, M.J.W., 2010. What are the learning affordances of 3-D virtual environments? Br. J. Educ. Technol. 41, 10–32. https://doi.org/10.1111/J.1467-8535.2009.01038.X.
Dang, H., Serne, J., Tafazzoli, M.: Virtual Reality Safety Training Assessment in Construction Management and Safety and Health Management Programs. Computing in Civil Engineering 2023: Visualization, Information Modeling, and Simulation - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2023. 28–35 (2024). https://doi.org/10.1061/9780784485231.004.
De Lorenzis, F., Pratticò, F.G., Repetto, M., Pons, E., Lamberti, F., 2023. Immersive Virtual Reality for procedural training: Comparing traditional and learning by teaching approaches. Comput Ind. 144, 103785. https://doi.org/10.1016/J.COMPIND.2022.103785.
Dengel, A., Mazdefrau, J.: Immersive Learning Explored: Subjective and Objective Factors Influencing Learning Outcomes in Immersive Educational Virtual Environments. Proceedings of 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering, TALE 2018. 608–615 (2018). https://doi.org/10.1109/TALE.2018.8615281.
Di Donato, L., Longo, F., Ferraro, A., Pirozzi, M., 2020. An advanced solutions for operators' training working in confined and /or pollution suspected space. Procedia Manuf. 42, 254–258. https://doi.org/10.1016/J.PROMFG.2020.02.080.
Evangelista, A., Manghisi, V.M., De Giglio, V., Martellotta, F., Giliberti, C., Mariconte, R., Uva, A.E., 2024. Exploiting Immersive Virtual Reality for Investigating the Effects of Industrial Noise on Cognitive Performance and Perceived Workload. Lecture Notes in Mechanical Engineering. 311–318. https://doi.org/10.1007/978-3-031-52075-4_35.
Fatal occupational injuries involving confined spaces : U.S. Bureau of Labor Statistics, https://www.bls.gov/iif/factsheets/fatal-occupational-injuries-confined-spaces-2011-19.htm, last accessed 2023/07/21.
Grabowski, A., Jankowski, J., 2015. Virtual Reality-based pilot training for underground coal miners. Saf Sci. 72, 310–314. https://doi.org/10.1016/J.SSCI.2014.09.017.
Graham, A.C., McAleer, S., 2018. An overview of realist evaluation for simulation-based education. Adv. Simul. 3, 1–8. https://doi.org/10.1186/S41077-018-0073-6/TABLES/1.
Grassini, S., Laumann, K., Rasmussen Skogstad, M., 2020. The Use of Virtual Reality Alone Does Not Promote Training Performance (but Sense of Presence Does). Front. Psychol. 11, 553693. https://doi.org/10.3389/FPSYG.2020.01743/BIBTEX.
Gürer, S., Surer, E., Erkayaoğlu, M., 2023. MINING-VIRTUAL: A comprehensive virtual reality-based serious game for occupational health and safety training in underground mines. Saf Sci. 166, 106226. https://doi.org/10.1016/J.SSCI.2023.106226.
Harris, D.J., Bird, J.M., Smart, P.A., Wilson, M.R., Vine, S.J., 2020. A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. Front. Psychol. 11, 524804. https://doi.org/10.3389/FPSYG.2020.00605/BIBTEX.
Joshi, S., Hamilton, M., Warren, R., Faucett, D., Tian, W., Wang, Y., Ma, J., 2021. Implementing Virtual Reality technology for safety training in the precast/prestressed concrete industry. Appl Ergon. 90, 103286. https://doi.org/10.1016/J.APERGO.2020.103286.
Jung, J., Ahn, Y.J., 2018. Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. Comput. Animat. Virtual Worlds. 29. https://doi.org/10.1002/CAV.1812.
Kaplan, A.D., Cruit, J., Endsley, M., Beers, S.M., Sawyer, B.D., Hancock, P.A., 2020. The Effects of Virtual Reality, Augmented Reality, and Mixed Reality as Training Enhancement Methods: A Meta-Analysis. Human Factors: the Journal of the Human Factors and Ergonomics Society. 63, 706–726. https://doi.org/10.1177/0018720820904229.
Kim, H.K., Park, J., Choi, Y., Choe, M., 2018. Virtual reality sickness questionnaire (VRSQ): Motion sickness measurement index in a virtual reality environment. Appl Ergon. 69, 66–73. https://doi.org/10.1016/J.APERGO.2017.12.016.
KOLB, D.: The Process of Experiential Learning. Strategic Learning in a Knowledge Economy. 313–331 (2000). https://doi.org/10.1016/B978-0-7506-7223-8.50017-4.
Krassmann, A.L., Melo, M., Pinto, D., Peixoto, B., Bessa, M., Bercht, M., 2022. What Is the Relationship between the Sense of Presence and Learning in Virtual Reality? A 24-Year Systematic Literature Review. Presence Teleop. Virt. 28, 247–265. https://doi.org/10.1162/PRES_A_00350/109997/WHAT-IS-THE-RELATIONSHIP-BETWEEN-THE-SENSE-OF.
Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front. Psychol. 4, 62627. https://doi.org/10.3389/FPSYG.2013.00863/ABSTRACT.
Lee, S., Seo, J., Kim, G.J., Park, C.-M.: Evaluation of pointing techniques for ray casting selection in virtual environments. In: Proceedings Volume 4756, Third International Conference on Virtual Reality and Its Application in Industry. pp. 38–44. SPIE (2003). https://doi.org/10.1117/12.497665.
Lewis, J.R., 2018. The system usability scale: past, present, and future. Int. J. Hum. Comput. Interact. 34, 577–590. https://doi.org/10.1080/10447318.2018.1455307.
Liang, Z., Zhou, K., Gao, K., 2019. Development of Virtual Reality Serious Game for Underground Rock-Related Hazards Safety Training. IEEE Access 7, 118639–118649. https://doi.org/10.1109/ACCESS.2019.2934990.
Longo, F., Padovano, A., Gazzaneo, L., Mirabelli, G., Ferraro, A., Pirozzi, M., Donato, L. D., 2021. Integrating physical and virtual game-based simulation for operators' training to enhance learning effectiveness: An application in hazardous industrial spaces. Int. J. Simul. Process Model. 16, 130–146. https://doi.org/10.1504/IJSPM.2021.115866.
Lougiakis, C., Katifori, A., Roussou, M., Ioannidis, I.-P.: Effects of Virtual Hand Representation on Interaction and Embodiment in HMD-based Virtual Environments Using Controllers. 510–518 (2020). https://doi.org/10.1109/VR46266.2020.00072.
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/FIGURES/10.
Lu, S., Xu, W., Wang, F., Li, X., Yang, J., 2020. Serious Game: Confined Space Rescue Based on Virtual Reality Technology. ACM International Conference Proceeding Series. 66–73. https://doi.org/10.1145/3442705.3442716.
Lu, S., Wang, F., Li, X., Shen, Q., 2022. Development and validation of a confined space rescue training prototype based on an immersive virtual reality serious game. Adv. Eng. Inf. 51, 101520. https://doi.org/10.1016/J.AEI.2021.101520.
Lu, S., Feng, Z., Lovreglio, R., Wang, F., Yuan, X., 2024. Comparing the productive failure and directive instruction for declarative safety knowledge training using virtual reality. J Comput Assist Learn. 40, 1040–1051. https://doi.org/10.1111/jcal.12937.
Lu, Y., Zheng, H., Chand, S., Xia, W., Liu, Z., Xu, X., Wang, L., Qin, Z., Bao, J., 2022. Outlook on human-centric manufacturing towards Industry 5.0. J Manuf Syst. 62, 612–627. https://doi.org/10.1016/J.JMSY.2022.02.001.
Makransky, G., Borre-Gude, S., Mayer, R.E., 2019. Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. J Comput Assist Learn. 35, 691–707. https://doi.org/10.1111/JCAL.12375.
Safety Management - Education and Training | Occupational Safety and Health Administration, https://www.osha.gov/safety-management/education-training, last accessed 2024/04/12.
Safety Management - Hazard Prevention and Control | Occupational Safety and Health Administration, https://www.osha.gov/safety-management/hazard-prevention, last accessed 2024/04/12.
Manghisi, V.M., Martellotta, F., Evangelista, A., Giliberti, C., Mariconte, R., Diano, M., Galasso, V., Uva, A.E.: Investigating the Effects on User Performance and Perceived Workload of Environmental Noise in Immersive Virtual Reality. 2022 IEEE International Workshop on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering, MetroXRAINE 2022 - Proceedings. 46–51 (2022). https://doi.org/10.1109/MetroXRAINE54828.2022.9967541.
Manghisi, V.M., Evangelista, A., Semisa, D., Latorre, V., Uva, A.E., 2022. Evaluating the Acceptance of Cinematic Virtual Reality-Based Applications for Rehabilitative Interventions in Schizophrenia. Games Health J. 11, 385–392. https://doi.org/10.1089/G4H.2021.0223.
Martirosov, S., Bureš, M., Zítka, T., 2022. Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. Virtual Real. 26, 15–32. https://doi.org/10.1007/S10055-021-00507-4/FIGURES/14.
Min, T., Cai, W., 2022. Portrait of decentralized application users: an overview based on large-scale Ethereum data. CCF Transactions on Pervasive Computing and Interaction. 4, 124–141. https://doi.org/10.1007/S42486-022-00094-6/FIGURES/13.
Netemeyer, G., Bearden, W., Sharma, S., 2003. Scaling procedures. Scaling Procedures. https://doi.org/10.4135/9781412985772.
Ning, H., Wang, H., Lin, Y., Wang, W., Dhelim, S., Farha, F., Ding, J., Daneshmand, M., 2023. A Survey on the Metaverse: The State-of-the-Art, Technologies, Applications, and Challenges. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2023.3278329.
NIOSH-Issued Publications: Working in Confined Spaces (80-106) | NIOSH | CDC, https://www.cdc.gov/niosh/docs/80-106/default.html, last accessed 2022/01/25.
Nykänen, M., Puro, V., Tiikkaja, M., Kannisto, H., Lantto, E., Simpura, F., Uusitalo, J., Lukander, K., Räsänen, T., Heikkilä, T., Teperi, A.M., 2020. Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. J. Saf. Res. 75, 205–221. https://doi.org/10.1016/J.JSR.2020.09.015.

Patle, D.S., Manca, D., Nazir, S., Sharma, S., 2019. Operator training simulators in virtual reality environment for process operators: a review. Virtual Real. 23, 293–311. https://doi.org/10.1007/S10055-018-0354-3/FIGURES/4.

Pérez, L., Diez, E., Usamentiaga, R., García, D.F., 2019. Industrial robot control and operator training using virtual reality interfaces. Comput Ind. 109, 114–120. https://doi.org/10.1016/J.COMPIND.2019.05.001.

Petrillo, F., Spritzer, A.S., Dal, C., Freitas, S., Pimenta, M.: Interactive Analysis of Likert Scale Data using a Multichart Visualization Tool. (2011). https://doi.org/10.5555/2254436.2254496.

User Experience Questionnaire (UEQ), https://www.ueq-online.org/, last accessed 2023/07/21.

Radhakrishnan, U., Koumaditis, K., Chinello, F., 2021. A systematic review of immersive virtual reality for industrial skills training. Behav. Inform. Technol. 40, 1310–1339. https://doi.org/10.1080/0144929X.2021.1954693.

Radianti, J., Majchrzak, T.A., Fromm, J., Wohlgenannt, I., 2020. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Comput Educ. 147, 103778. https://doi.org/10.1016/J.COMPEDU.2019.103778.

Rahouti, A., Lovreglio, R., Datoussaïd, 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/FIGURES/15.

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

Saravas, V., Pahalchuk, E., Molchan, A., Kampov, Y.: The Development of Educational Application with Virtual Reality Placing Objects System Using Snap Zone Technology. Proceedings - 2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies, DESSERT 2020. 416–421 (2020). https://doi.org/10.1109/DESSERT50317.2020.9125023.

Sauro, J., Lewis, J.R.: Quantifying the User Experience: Practical Statistics for User Research.

Schmitt, N., 1996. Uses and abuses of coefficient alpha. Psychol Assess. 8, 350–353. https://doi.org/10.1037/1040-3590.8.4.350.

Schrepp, M., Thomaschewski, J., Hinderks, A., 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). International Journal of Interactive Multimedia and Artificial Intelligence. 4, 40–44. https://doi.org/10.9781/IJIMAI.2017.445.

Scorgie, D., Feng, Z., Paes, D., Parisi, F., Yiu, T.W., 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.

Shiradkar, S., Rabelo, L., Alasim, F., Nagadi, K.: Virtual World as an Interactive Safety Training Platform. Information 2021, Vol. 12, Page 219. 12, 219 (2021). https://doi.org/10.3390/INFO12060219.

Shrestha, N., 2021. Factor analysis as a tool for survey analysis. Am. J. Appl. Math Stat. 9, 4–11. https://doi.org/10.12691/ajams-9-1-2.

Smidt, A., Balandin, S., Sigafoos, J., Reed, V.A., 2009. The Kirkpatrick model: A useful tool for evaluating training outcomes. J. Intellect. Dev. Disabil. 34, 266–274. https://doi.org/10.1080/13668250903093125.

Confined Spaces - Overview | Occupational Safety and Health Administration, https://www.osha.gov/confined-spaces, last accessed 2023/07/21.

Stefan, H., Mortimer, M., Horan, B.: Evaluating the effectiveness of virtual reality for safety-relevant training: a systematic review. Virtual Reality 2023 27:4. 27, 2839–2869 (2023). https://doi.org/10.1007/S10055-023-00843-7.

Testo Unico sulla salute e sicurezza sul lavoro | Ministero del Lavoro e delle Politiche Sociali, https://www.lavoro.gov.it/priorita/pagine/testo-unico-sulla-salute-e-sicurezza-sul-lavoro, last accessed 2023/07/21.

Wahidi, S.I., Pribadi, T.W., Rajasa, W.S., Arif, M.S., 2022. Virtual Reality Based Application for Safety Training at Shipyards. IOP Conf. Ser. Earth. Environ. Sci. 972, 012025. https://doi.org/10.1088/1755-1315/972/1/012025.

Wang, J., Shi, R., Zheng, W., Xie, W., Kao, D., Liang, H.N., 2023. Effect of Frame Rate on User Experience, Performance, and Simulator Sickness in Virtual Reality. IEEE Trans vis Comput Graph. 29, 2478–2488. https://doi.org/10.1109/TVCG.2023.3247057.

Zhang, Y., Hedo, R., Rivera, A., Rull, R., Richardson, S., Tu, X.M., 2019. Post hoc power analysis: is it an informative and meaningful analysis? G E NPsychiatr. 32, 100069. https://doi.org/10.1136/GPSYCH-2019-100069.