Safety Science 185 (2025) 106803

Contents lists available at ScienceDirect

Safety Science

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

On the user-based assessments of virtual reality for public safety training in urban open spaces depending on immersion levels☆

Mariella De Fino [a,*], Federica Cassano [a], Gabriele Bernardini [b], Enrico Quagliarini [b], Fabio Fatiguso [a]

[a] Departement of Civil, Environmental, Land, Construction and Chemistry, Politecnico di Bari, Bari, Italy
[b] Department of Construction, Civil Engineering and Architecture, Università Politecnica delle Marche, Ancona, Italy

ARTICLE INFO

Keywords:
Safety training
Virtual reality
Urban squares
User evaluations
Multi-risk
Serious gaming

ABSTRACT

The increasing use of VR (Virtual Reality) training tools in safety-related fields has prompted the scientific community to explore methods for assessing their effectiveness across different levels of immersion. While numerous studies have been conducted in sectors such as healthcare, transportation, agriculture, aviation, mining, firefighting, and construction, one area that remains underexplored is risk training for general public against natural and man-made disasters in both indoor and outdoor built environments. In this context, the paper aims at validating a prototype for VR multi-risk (heat wave and earthquake) training in urban open spaces that was developed according to a Serious Game (SG) approach. To address insights on its extensive adoption, the VR-SG was tested for comparison of three modes: traditional by video recording, non-immersive by desktop, immersive by headset. Outputs from feedback questionnaires on knowledge gain and transfer to expansive contexts, as well as on user experience, suggest that the effectiveness of knowledge acquisition differs significantly between headset and video, as well as between desktop and video, but shows no significant difference between headset and desktop. Additionally, while headset outperforms desktop in terms of engagement, perceived usefulness, and realism, it falls behind in ease of use. Finally, further analyses on training outcome by age, contents of open-ended answers, in-game errors and administration procedures supported the discussion of key aspects in targeting VR for public safety communication.

1. Introduction

The increasing use of VR (Virtual Reality) tools for education and training across various fields has recently led the scientific community to explore not only the innovative technical aspects of design and development but also the methods for evaluating the effectiveness of VR-based learning (Marques et al., 2024; Strojny and Dużmańska-Misiarczyk, 2023). This interest has extended to safety training in sectors such as health services, transportation, agriculture, aviation, mining, fire and construction (Scorgie et al., 2024; Stefan et al., 2023), the latter primarily focused on hazard identification and management on construction sites and related operations (Akindele et al., 2024; Man et al., 2024).

Key aspects in evaluating effectiveness of VR-based safety training include its impact on knowledge acquisition and user experience in comparative terms with other training tools. In many studies, VR is compared with traditional methods, such as video (Liang et al., 2019; Poyade et al., 2021), paper-based material (Kinateder et al., 2013) and lecture-based instruction (Dzeng et al., 2015; Nykänen et al., 2020), with findings suggesting that VR performs as well or better. Other studies compare different levels of immersion and devices, such as headset-based training versus desktop-based training (Buttussi and Chittaro, 2018; Dhalmahapatra et al., 2021; Jung and Ahn, 2018), with some also including traditional methods, evaluating three different modes (Buttussi and Chittaro, 2021; Makransky et al., 2019a). These studies, like the present one, assume that a headset-based system offers greater immersion than a desktop-based system, considering immersion as an "*objective property of a technology supporting natural sensorimotor contingencies for perception, including the response to a perceptual action*" (Slater, 2018), and delivering "*an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant*" (Kisker et al., 2021). Immersion is strictly associated with the technical aspects of a

☆ This article is part of a special issue entitled: 'XR for Safety' published in Safety Science.

* Corresponding author.
E-mail address: mariella.defino@poliba.it (M. De Fino).

https://doi.org/10.1016/j.ssci.2025.106803
Received 25 June 2024; Received in revised form 20 January 2025; Accepted 27 January 2025
Available online 8 February 2025

virtual system that aid the user in experiencing a sense of presence (Wiepke and Heinemann, 2024; Wilkinson et al., 2021), understood as the "*subjective feeling of being in the virtual environment rather than the real space*" (Felton and Jackson, 2022), and generally assessed as being higher in head-mounted displays than in screen applications and traditional methods, as a result of increased sensory stimuli, representation fidelity, spatial realism, embodiment and agency (Buttussi and Chittaro, 2021, 2018; Dhalmahapatra et al., 2021; Grassini et al., 2020; Kim et al., 2020; Makransky et al., 2019b; Pallavicini et al., 2019; Piccione et al., 2019; Szóstak et al., 2024),

In the studies comparing head-mounted, desktop and traditional training, higher levels of immersion are generally associated with improved knowledge acquisition and user experience, although differences are not always statistically significant and, in some cases, the results may even show a reverse trend for certain indicators or target groups.

In fact, as seen in broader VR-based education and training literature, high immersion does not always lead to superior learning outcomes, especially if implying excessive stimuli and burdens (Makransky et al., 2019b), causing physical and psychological stress associated with unused tools (Hedberg et al., 2000), or being unsuitable for users with different spatial abilities and learning styles (Ai-Lim Lee et al., 2010; Makransky et al., 2019). All these issues might also depend on the users' learning context, age, gender and experiences in VR, resulting in a different so-called user experience and usability (Anwar et al., 2018; Lorenz et al., 2023). For instance, in some works, differences in performance and experience between immersion levels are specifically related to age groups, such as seniors showing higher memory using desktop platforms rather than headsets, while expressing comparable acceptance for the two modes (Plechatá et al., 2019) or underperforming in spatial-based performance by VR compared with VR plus a bird-view map, and two-dimensional/2D map presentations (Lin and Yang, 2007). Conversely, in other works the perceived learning outcomes of VR is unaffected by age (Pedram et al., 2022).

Regarding the effectiveness of VR-based safety training in relation to the levels of immersion, an underexplored area is risk training against natural and man-made disasters in indoor and outdoor built environments. In fact, despite numerous studies on the development of VR solutions, many of which follow a Serious Game (SG) approach (Feng et al., 2018; Liu and Liu, 2025), there are no comparative assessments between headset-based, desktop-based, and traditional solutions to the best of the authors' knowledge. However, this type of assessments is highly desirable, particularly in the context of urban open spaces, such as squares, parks and streets, due to their higher vulnerability and exposure compared to indoor settings.

In general terms, open spaces are relevant scenarios for single and multi-risk, because they can attract many users over time (up to possible critical crowding conditions), but, at the same time, they seem to be characterized by a limited implementation of specific structural and non-structural risk mitigation strategies (Coaffee et al., 2009; Pacheco Barzallo et al., 2022; Rosso et al., 2023; Yang et al., 2024). Recent efforts to increase performance of open spaces using VR-based approaches have been provided, e.g. in the context of users' health and by promoting participatory approaches to retrofit strategies design (Dane et al., 2024). On the contrary, risk training actions are still limited to specific kinds of disasters (mainly, tsunami, flood, hurricane) (Fujimi and Fujimura, 2020; Liu and Liu, 2025; Sermet and Demir, 2019). Since promoting proper user behaviours through VR training could boost the safety of open spaces pushing on disaster response and non-structural management solutions, additional extensive efforts should comprise other relevant scenarios and validation of relative VR solutions. Moreover, in this field, assessment is crucial to validate diverse and flexible tools, since different stakeholders (e.g., varying by age or familiarity with the location) and interaction systems (computers, tablets, smartphones, headsets) are involved, depending on the technological infrastructures where the training might take place.

Based on these considerations, the present research has aimed to perform user-based assessments of a framework for VR multi-risk training against earthquake and heat wave in urban open spaces that was previously developed according to a SG approach (De Fino et al., 2023). The user-based assessments have been designed for two complementary objectives: (1) to compare immersive (Head Mounted Display – HMD), non-immersive (PC) and recorded video (video) versions of the VR-SG, in order to demonstrate its efficacy and versatility, while addressing the previously mentioned gap; (2) to validate the training framework in terms of knowledge gain and user experience, focusing on its novel features and related assessment challenges, while drawing general considerations on how to make this and similar tools more effective and pervasive in VR-based public safety communication.

To this end, the paper is structured as follows: Section 2 reviews literature studies involving testing sessions of VR-based risk training in buildings and urban built environments, focusing on natural and man-made disasters; Section 3 describes the methodological workflow for testing the different VR-SG prototype, detailing the training approach, participant sample, feedback questionnaire, testing sessions and criteria for data analysis; Section 4 presents comparative results related to immersion levels, in comparative terms regarding the level of immersion, with sub-analyses on age-based groups, contents of open-ended answers and in-game errors; Section 5 discusses key findings of the analysis, outlining their value for the broader dissemination of the proposed systems.

2. Research background

The user-based assessments of the present study were designed based on key validation criteria and testing procedures found in literature on risk training for natural (e.g. fire, earthquake, flooding) and man-made (e.g. terroristic attack) disasters in indoor and outdoor built environments, as summarized in Table 1. Additionally, the data analysis on immersion levels was structured according to the experiences developed in the general field of safety training, as referenced in the introduction (Buttussi and Chittaro, 2021, 2018; Dhalmahapatra et al., 2021; Jung and Ahn, 2018; Makransky et al., 2019a), as well using standard methods of statistics for behavioural sciences.

Concerning the validation criteria for VR-based risk training in buildings and built environments, these mainly focus on three key areas: (1) *demographics of participants*; (2) *knowledge acquisition/assessment*; (3) *user experience*.

Demographics of participants are assessed before the training session to profile users, by focusing, in particular, on age and gender. The sample group is generally related to the specific context, such as hospital staff (Feng et al., 2020a), university students (Lovreglio et al., 2022), rescue operators (Diez et al., 2016).

Knowledge acquisition/assessment relates to the efficacy and effectiveness in communicating the training concepts and contents before and after the training. It might be based on closed-ended questions (Lovreglio et al., 2021; Shiradkar et al., 2021), including true or false questions (Feng et al., 2022a) or open-ended questions (Chittaro and Sioni, 2015; Lovreglio et al., 2022; Oliva et al., 2019; Rahouti et al., 2021). Closed-ended questions are used to streamline the survey and prevent ambiguous or irrelevant answers, while open-ended questions help avoid prompting with possible answers and/or limited responses. In some cases, a mixed approach is proposed depending on the types of questions (D'Amico et al., 2023; Irshad et al., 2021) or to make the survey less monotonous for specific target groups like children (Smith and Ericson, 2009). In case of open-ended questions, a pre-set list of right answers is defined by the researchers to enable quantitative assessment of the free text answers. Alternatively, *knowledge acquisition* can be evaluated by in-game surveys, including behavioural observations, such as researchers annotating actions/decisions during the game (Lovreglio et al., 2018; Oliva et al., 2019) and parameters –like the time and distance covered by players to complete tasks (Cao et al., 2019). In

Table 1
Validation criteria in literature.

Risk	Reference	(i)	(ii)	(iii)														
		Demographics	Knowledge acquisition	Self-reported engagement	Self-reported efficacy	Recommendation simplicity	Recommendation efficacy	Perceived usefulness	Perceived ease of use	Perceived vulnerability	Severity	Self-reported fear	Realism	Representation fidelity	Attention level	Presence	Simulation sickness	Customization
Fire	(Shiradkar et al., 2021)	x	x	x											x			
	(Rahouti et al., 2021)	x	x	x		x		x	x					x		x		
	(Oliva et al., 2019)	x	x	x	x								x	x				
	(Cao et al., 2019)	x	x	x					x	x		x	x					
	(Smith and Ericson, 2009)	x	x															
	(Therón et al., 2020)	x		x				x	x			x	x				x	
	(Lovreglio et al., 2021)	x	x		x	x								x				
Earthquake	(Feng et al., 2022a)	x	x	x	x				x				x					x
	(Feng et al., 2020b)	x		x					x				x					
	(Lovreglio et al., 2018)	x	x						x				x					
	(Sukirman et al., 2019)	x		x									x					
Flood	(Irshad et al., 2021)	x			x							x						
	(D'Amico et al., 2023)	x	x															
Terrorism	(Lovreglio et al., 2022)	x	x	x	x	x	x	x		x		x	x					
	(Chittaro and Sioni, 2015)	x	x		x	x	x			x	x	x			x			

this general context, comparisons with traditional methods, such as slide-based training, videos and leaflets, demonstrate that innovative tools are more effective and pervasive (Feng et al., 2022a; Lovreglio et al., 2021; Rahouti et al., 2021; Shiradkar et al., 2021).

User experience is also a key factor to evaluate the tool/prototype itself thanks to post-training surveys. Several indicators are commonly used. Among them, the most recurring ones are: *self-reported engagement*, namely interest, involvement, motivation and enjoyment during the experience, e.g. (Shiradkar et al., 2021); *perceived usefulness* as the degree to which a person believes that using a particular system would enhance his/her performance, e.g. (Rahouti et al., 2021); *perceived ease of use* as the degree to which a person believes that using a particular system would be free of effort, e.g. (Feng et al., 2020a); *recommendation simplicity* in applying the suggested actions in real life, often combined with *recommendation efficacy*, assessing how the suggested actions are beneficial in in mitigating risks from the simulated threat (Lovreglio et al., 2021); *realism* related to the plausibility of the virtual environment often combined with *representation fidelity,* which measures how closely the virtual setting matches the real one, e.g. (Cao et al., 2019); *severity* referring to how serious the individual believes that the threat would be to his/her own life, closely connected with *perceived vulnerability* denoting how personally susceptible an individual feels to the threat and *self-reported fear,* standing between perceptions of severity and vulnerability and the level of the appraised threat, e.g (Irshad et al., 2021)*; self-efficacy* as the individual's beliefs about whether he/she is able to perform the recommended coping response, e.g. (Chittaro and Sioni, 2015); *presence* as the player's sense of being a part of the virtual environment, e.g. (Rahouti et al., 2021); and *simulator sickness* and *motion sickness* that can be experienced as a side effect during and after exposure to different virtual reality environments, e.g. (Therón et al., 2020).

Finally, beyond the previous three main topics, recent works (Feng et al., 2022b) also suggest that research perspective should be linked with the *knowledge transfer* or *skill transfer* to expansive context. These concepts, related to the capability of shifting from an artificial environment to a real one that is similar or identical to the one in which the knowledge and skills will be actually used in practice, are widely applied in related VR training fields (Strojny and Dużmańska-Misiarczyk, 2023). This issue seems to be particularly interesting in view of overcoming a case-oriented approach in the research area. Similarly, the possibility to investigate the role of *repetitions following errors* during the game is pointed out as beneficial by other studies for strengthening the learning process assuming that "*mistakes are necessary steps in retracing information rather than an undesirable outcome*" (Chittaro and Buttussi, 2022).

3. User study

The methodological workflow has been developed to address the research objectives.

In detail, the comparison between the three versions of the VR-SG framework (HMD, PC, video) has been organized into four main phases: identification of the training approaches (including VR-SG prototypes); (section 3.1); selection of topics and indicators to be measured by a feedback questionnaire (section 3.2); definition of criteria for data analysis (section 3.3); and, the design and implementation of the testing procedure, including sample of participants (section 3.4).

3.1. Materials: Training framework

The training framework (Annex 1) is developed within the research project BE^{2}SECURE (De Fino et al., 2023) according to an innovative approach:

- It is founded on a multi-risk vision, coupling SLow (SLOD) and SUdden (SUOD) Onset Disasters. The specific risks (heat wave as SLOD and earthquake as SUOD) were selected as the most relevant, based on the cross-assessment of available databases on critical events in Italian cities, and were analyzed according to a SLOD-to-SUOD approach: the SLOD resulting in higher temperatures in certain areas, acts as the primary hazard and alters the environmental conditions in the built environment, potentially affecting the distribution of people, who may gather in cooler areas.
- It is set within a typological environment of urban square, with representative features, based on the cluster analysis of real case studies (D'Amico et al., 2021). The scene includes schematic volumes, simple decorations and plain surfaces to balance realism and computational load.
- It displays maps and patterns, resulting from rigorous analytical simulations on surface temperature distribution and falling debris, as well as from agent-based models of crowd position and motion (Bernardini et al., 2023) to animate Non-Playable Characters (NPCs).

Furthermore, in terms of operational functionalities, the framework is arranged in training sessions (Heat Wave – HW, Earthquake-E and Post-Earthquake-PE), which are further organized in training modules. Each module is associated with a training item, positioned at a specific location of the scene. Players are teleported to these locations and prompted to answer multiple-choice questions. They receive immediate feedback on the correct solution and, if the first attempt is wrong, they have the option to retry their selection. The game allows players to earn or lose points based on the accuracy of their answers and to assign a final score at the end. In more detail, the previously developed training modules/items used in this work (De Fino et al., 2023) concerns the recognition of: three temperature mitigating elements (trees, fountain, shading building) during the HW session; four safe behaviours during the E session (avoiding closeness/use for: buildings that could collapse, glazed surfaces/objects that could break, vehicles that could hinder rescue operations, and electric devices that could catch fire or emit sparks); and three characteristics of the safe areas in the PE session (centre of the square, pre-designated, far from buildings and falling objects). All the objectives/items are derived from national institutional recommendations.

The above-mentioned framework has been developed in both immersive mode on PC and non-immersive mode for HMD. Within Epic Unreal Engine 4.6® simulative environment, different user interfaces have been developed: in the PC mode, the mouse and the keyboard are used to control the 360° game view and behavioural choices, respectively; in the HMD mode, tested on a VR Meta Quest 2 headset, visual orientation relies on gyroscopic sensors and laser tracking, while gesture control is enabled by hand-based oculus touch controllers.

In addition, the non-interactive video of a game exemplary session was recorded to provide a "traditional" training mode. This version can be compared with the VR-SG prototypes and potentially validated for users who prefer conventional tools or in situations where technological resources are limited.

3.2. Measures: Feedback questionnaire

In line with all the studies involving testing sessions in Section 2, the feedback questionnaire has been divided into three parts (Annex 2).

The first part, to be completed only before the training, collects demographic information, such as age, gender, education level, previous experiences of training and VR (Annex 2.1).

The second part is related to knowledge and is arranged in six questions (three open-ended and three closed-ended). Participants were requested to respond to the same questions of this part before and after the training session. The three open-ended questions align with the three modules/items within the training framework (compare Section 3.1). Participants are asked to freely list: (i) the elements that are risk decreasing for the heat wave, (ii) the elements that are risk increasing

for the earthquake, and (iii) the areas that should be designated as safe places in a post-earthquake scenario. Then, the answers are scored (Annex 2.2) taking into account that the training modules/items concerns the recognition of three temperature mitigating elements (vegetation, water sources, areas in shadow) during the HW session; four safe behaviours during the E session (avoiding closeness/use for: buildings that could collapse, glazed surfaces/objects that could break, vehicles that could hinder rescue operations, and electric devices that could catch fire or emit sparks) and three characteristics of the safe areas in the PE session (centre of the square, pre-designated, far from buildings and falling objects). The structure of the questions and the criteria for scoring the answers follow previous studies (Chittaro and Sioni, 2015; D'Amico et al., 2023; Feng et al., 2020a; Lovreglio et al., 2022) and the results are analyzed based on the number of right answers. The three closed-ended questions are based on the recognition of the training items within the Virtual Tour (VT) of 360° panoramas of a real case study, embedded in the same Google form used for administering the questionnaire (Fig. 1A). This approach tests the incorporation of repeated exercises in expansive contexts that could contribute to the transferring of knowledge and applying of skills to new settings. The selected real case study is Piazza dei Priori in Narni, Centre Italy, one of the urban squares considered for the identification of the VR-SG typological environment and one of the three pilot sites of the research project BE^{2}SECURE for the validation of assessment methods and mitigation strategies for multi-risk scenarios of earthquake and heat wave. Three spherical pictures of Piazza dei Priori were edited with hotspots, marked by capital letters (Annex 2.3), corresponding to several items (Fig. 1B) similar to those displayed in the prototype (Fig. 1C), so that the trainee is asked to point out the relevant ones (Annex 2.4) through closed-ended questions and their responses are scored accordingly (Annex 2.5).

Finally, the third part is related to the user experience. The indicators were selected among the most commonly used in testing VR-based risk training, as reported in Section 2, namely self-reported engagement, perceived usefulness, perceived ease of use, recommendation simplicity and efficacy, and realism (Annex 2.6). Moreover, the questions were adapted from the available literature to validate the main functional and operational choices of the training framework, concerning the outdoor settings (QE2.3), the multi-risk approach (QE2.5), the presence of crowds (QE2.6), and the realism of the typological environment (QE2.4 and QE5.2). All the answers follow a 7-point Likert scale, as proposed in other studies – e.g. (Feng et al., 2022a; Lovreglio et al., 2022, 2018), ranging from 1 (strongly disagreed) to 7 (strongly agreed).

In addition to the above-mentioned indicators, in-game errors are tracked by a log file, automatically generated at the end of the game both for PC and HMD modes. This log records he sequence of "true" and "false" answers for each question, in view of investigating whether or not errors provide useful feedback to participants, pinpointing where their knowledge needs improvement, and prompting refinement of their mental model.

3.3. Methods: Data analysis

Data analysis has been developed according to the following phases and using different statistical methods depending on the assumptions and purposes (Gravetter et al., 2021):

- Assessment of statistical variability in knowledge, depending on the immersion level, considering three modes (video, PC, HMD). This phase comprises:
  - Evaluating whether the participants exhibit comparable pre-training knowledge levels between the modes, by running a one-way ANOVA test (or non-parametric equivalent if the normal distribution is not satisfied) for three samples and confirming the null-hypothesis of no significant differences (ρ>0.05).
  - Evaluating whether the participants achieve relevant knowledge gain through training in one or more mode, by running a two-way mixed-design ANOVA test or non-parametric equivalent (pre-post training as the within-subjects factor, mode as the between-subjects factor). If the ANOVA test highlights relevant interactions rejecting

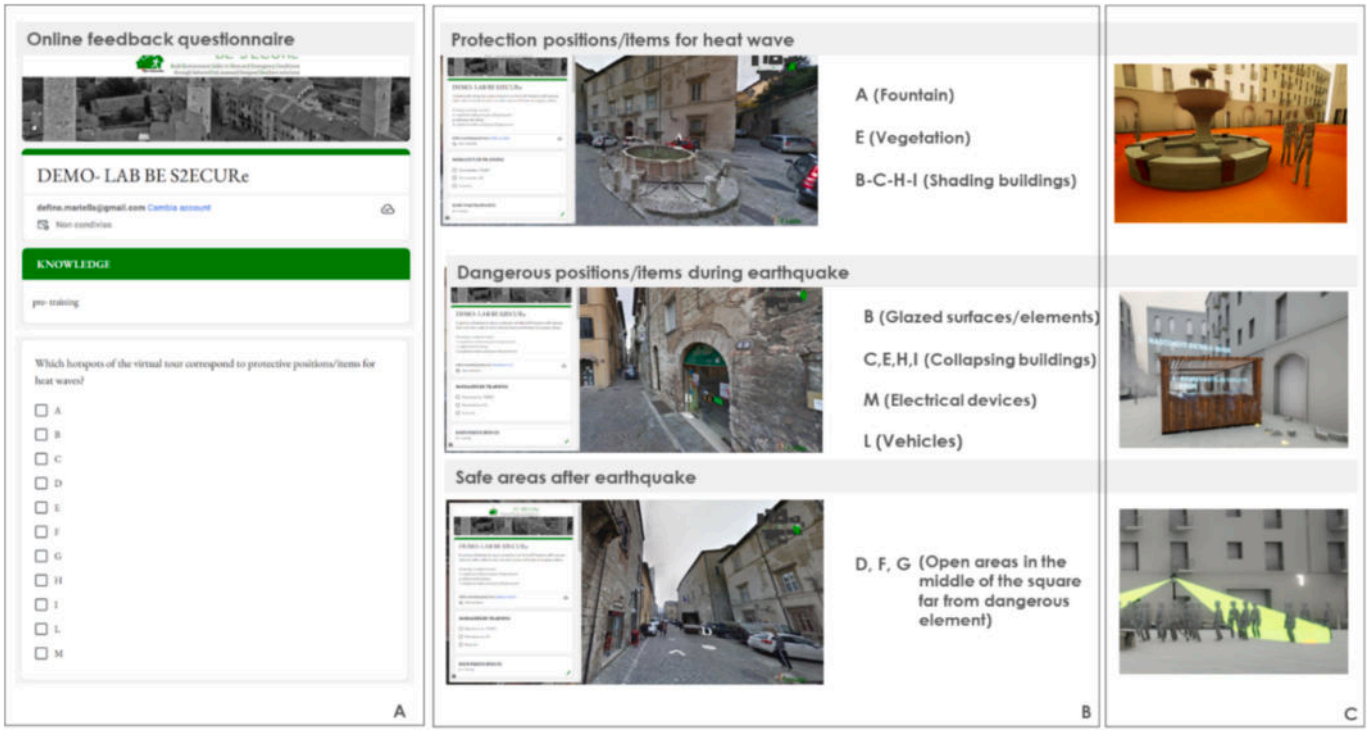


Fig. 1. Questionnaire overview for closed-ended questions, based on VT visualization: A- Example of Google Form questionnaire page, B- Example of Virtual Tour views and hotspots within the Google Form, C- Example of training items in the VR-SG.

the null-hypothesis ($\rho<0.05$), three pairwise comparisons between pre-post training knowledge for each mode are performed by *t*-test for dependent samples or non-parametric equivalent. The results are then analysed using post-hoc (e.g. Bonferroni) correction on ρ, to avoid Type I errors in multiple comparisons ($\rho_{corrected} = 0.05/\text{number of tests} = 0.05/3 = 0.167$) for rejecting or confirming the null hypothesis.

  - Evaluating whether the participants exhibit different post-training knowledge depending on the immersion level, by running a one-way ANOVA test or non-parametric equivalent for three modes. If the ANOVA test highlights statistical differences ($\rho<0.05$), three pairwise comparisons of post-training knowledge between modes are run by t-test for independent samples or non-parametric equivalent. The results are then analysed using post-hoc correction on $\rho = 0.167$, similarly to the previous point.

- Assessment of statistical variability in user experience, depending on the immersion level, considering two modes (PC, HMD). This phase comprises:
  - Checking for each indicator (engagement, perceived usefulness, ease of use, simplicity/efficacy of recommendations, realism) that the average Likert-scale scores are statistically different for the two VR testing apparatus, by running a t-test for independent samples (or non-parametric equivalent if the normal distribution is not met).

Furthermore, some data sub-analyses are addressed to investigate further issues, such as:

- Assessment of the distribution of post-knowledge and user-experience average scores depending on age-based groups, in order to withdraw some general insights on the correlation age-immersion although on small sub-samples of users.
- Assessment by content analysis of the knowledge open-ended questions, calculating the percentage of participants mentioning each item in their free lists before and after the training, to understand which concepts were previously known and which ones were acquired by the training.
- Assessment of the in-game errors and repetitions, through the log files automatically generated during the game. The errors have been pairwise compared between the VR training modes following the statistical procedure, previously mentioned for the user experience indicators. Moreover, they are evaluated against the corresponding average knowledge to investigate whether mistakes strengthen the information delivery, as well as with experiences in VR and videogames to evaluate if errors are potentially affected by difficulties in managing the apparatus.

3.4. Procedure: Participants and experimental set-up

The sample size has been determined by power analysis with the statistical software G*Power (Franz et al., 2007), according to previous works (Latini et al., 2023). In particular, to enable the comparison of the three training modes by two-way ANOVA test, a sample of 54 people per training mode was selected. This number was based on an effect size equal to 0.32, as ex-post estimated by (Buttussi and Chittaro, 2021) in a similar study, with a first type error (α) equal to 0.05. The given sample size ensures a statistical power equal to 98 %.

The 162 participants were recruited through email invitations sent to all teaching and administrative staff from the university departments of the authors, as well as during events organized for high school student orientation, technical seminar for engineers and architects and presentation of the BE^{2}SECURE project results to the public (Fig. 2). The decision to conduct testing at various university campuses, professional associations, and project case studies, while actively engaging participants from diverse age groups, genders, and backgrounds, aimed to enhance the external validity of the questionnaire and ensure its generalizability beyond the study sample.

The participants voluntarily participated to the tests. According to national regulations and university rules, no sensitive information was collected and data were anonymized. Before starting the training, participants were informed about the training purpose and content and signed a consent form to allow the research team to store and use the collected data for scientific purposes.

The tests were conducted on standard PCs, both desktops and laptops, for video playback and the non-immersive game, while the immersive game used Meta Quest 2 headsets. During public events, when multiple demonstrations were held simultaneously, testing stations were spaced out to avoid interference from sounds and movements. Differently, for individual sessions with participants responding the invitation email, a dedicated laboratory room was set up. In both cases, interactions among participants who had completed and those waiting to begin were properly managed to avoid information exchange that could affect the results. Each participant was supported during the training and testing by an individual assistant, who was instructed to provide only general pre-training instructions and solve in-training technical problems, without interfering in any way with the user's experience. Each assistant was also asked to note any comments or observations from participants, particularly regarding discomfort during the immersive game. In this regard, the trainees testing the HMD mode were specifically informed about the potential motion sickness and asked to report any related problem and feel free to quit the session.

Finally, the feedback questionnaires were administered, before and after the training, through the Google form, incorporating the questions and the VT.

Fig. 2. Testing sessions in the laboratory (left) and during public event in Narni (right).

4. Results

4.1. Demographic of participants

As far as the first part of the feedback questionnaire on demographic information is concerned (Fig. 3), an overall gender balance was achieved (52 % women and 48 % men in total) with a predominance of medium–high levels of education (67 % at least graduated), mainly due to the academic and professional contexts of the demonstration sessions. Age ranges from 18 to 60 years old. Limited previous experiences in

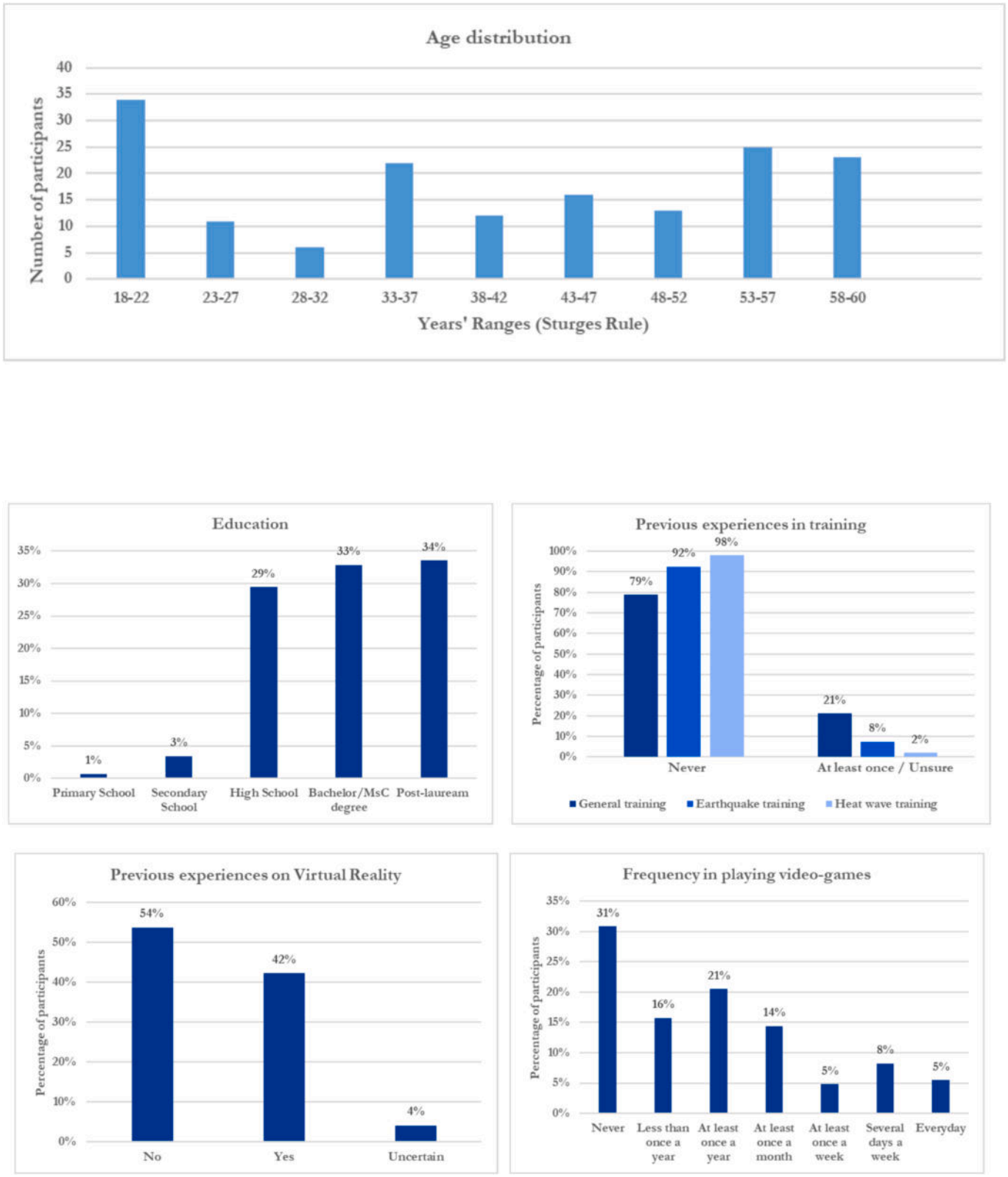


Fig. 3. Overview of demographic information.

training were declared, including very high percentages of interviewees with no experiences in training for earthquake (92 %) and heat wave (98 %). Moreover, about half of the sample declared no previous experiences in VR (54 %) and one third expressed low frequency in playing video-games (68 % in total answering "never", "less than once a year", "at least once a year").

When cross-observing the most significant demographic sub-indicators against the training modes, no relevant differences were found in terms of gender (48 %, 54 % and 54 % of men, respectively for video, PC and HMD). Moreover, among the participants testing the VR modes, about half confirmed previous VR experiences in both cases (50 % and 54 %, respectively for PC and HMD) and comparable shares (28 % and 26 % respectively for PC and HMD) declared medium–high frequency in playing video-games (answering "at least once a month" o greater occurrence).

4.2. Knowledge and user experience assessment

According to section 3.3 assumptions, the analyses included a preliminary verification that the pre-training knowledge was statistically comparable in all the three training modes. To this end, since the data were found normally distributed according to the Kolmogorov-Smirnoff test, the one-way ANOVA test for three independent samples has been used, resulting in $F(2, 159) = 1.21$ and $\rho = 0.3 > 0.05$ that have confirmed the samples to be statistically similar.

Thus, the verification that meaningful knowledge improvement occurred in one or more training modes has relied on the two-way mixed-design ANOVA test (pre-post training as the within-subjects factor, mode as the between-subjects factor) resulting in significant interaction effect $(2,159) = 7.2$ and $\rho = 0.0008 < 0.05$. The following pairwise comparisons (pre-post for each training mode) with t-test applying the Bonferroni correction ($\rho_{corrected} = 0.167$) has always resulted in significant statistical difference between previous and acquired knowledge regardless the apparatus (for video: $t(53) = -4.21$, $\rho = 9.83\ E^{-5}$; for PC: $t(53) = -10.15$, $\rho = 5\ E_{-14}$; for HMD: $t(53) = -13.09$, $\rho = 2.99\ E^{-18}$). Consequently, all the training modes have resulted in effective communication of risk-related information content.

Based on the above verifications, the statistical analyses of post-training scores, supported by one-way ANOVA test for three independent samples, has been carried out, showing some statistical differences, with $F(2,59) = 10.52$ and $\rho = 5.1\ E^{-5}$.

Looking at the following pairwise comparisons by *t*-test for independent samples with Bonferroni correction ($\rho_{corrected} = 0.167$), the differences are significant between video and PC ($t(106) = -2.92$, $\rho = 0.04$) and between video and HMD ($t(106) = -3.57$, $\rho = 0.0005$) and not significant between PC and HMD ($t(106) = -0.64$, $\rho = 0.519$). In terms of indexes of variability, the knowledge increases with the immersion (Fig. 4).

In the case of knowledge after training through close ended questions following the visualization of the virtual tour, the results are consistent with the previous analysis (for video-PC: $t(106) = -4.15$, $\rho = 6.59\ E^{-05}$; for video-HMD: $t(106) = -3.59$, $\rho = 0.0005$; for PC-HMD: $t(106) = 0.49$, $\rho = 0.62$) and the related indexes of variability show a comparable trend to those from overall knowledge (Fig. 4).

Concerning the user experience, a Cronbach's alpha was preliminarily calculated to assess the internal consistency of the questionnaire's questions for each indicator, resulting in at least acceptable values (0.767 engagement, 0.723 perceived usefulness, 0.698 ease of use, 0.856 simplicity/efficacy of recommendations, 0.755 realism).

Moreover, looking at the results, all the indicators were scored positively (score 5 or higher in the 7-point Likert scale) for both VR modes – PC and HMD – by at least two third of the interviewees (Fig. 5), except for the ease of use of the HMD mode. Particularly, the highest shares of positive scores are related to engagement (81 % PC, 96 % HMD), perceived usefulness (83 % PC, 98 % HMD) and efficacy/simplicity of recommendations (93 % both PC and HMD), compared to ease of use (67 % PC, 39 % HMD) and realism (68 % PC, 91 % HMD). Table 2 provides an overview of the results for each question.

Nevertheless, the pairwise comparisons between the VR modes, supported by the t test for independent samples given their normal distribution, has showed some statistical differences ($\rho < 0.05$) between the modes for engagement ($t(106) = -3.56$; $\rho = 0.0005$), perceived usefulness ($t(106) = -2.71$; $\rho = 0.007$), ease of use ($t(106) = 4.00$, $\rho = 0.0001$) and realism ($t(106) = -3.72$; $\rho = 0.003$), differently from simplicity and efficacy of recommendations ($t(106) = -1.1.4$; $\rho = 0.25$). Looking at the indexes of variability, it can be observed that engagement, perceived usefulness and realism scores increase with the level of immersion, while PC is found easier to use than HMD (Fig. 6).

4.3. Age-based groups

A sub-analysis of the results has been conducted for groups categorized by age, selected as the most significant demographic indicator. This is due to the greater variability in the data (compared to factors, like education, training experiences, and VR), and to its relevance to specific target groups in training applications (e.g., in schools, organized tourist groups). Moreover, age was found the most recurring demographic indicator in related studies in literature (Lin and Yang, 2007; Pedram et al., 2022; Plechatá et al., 2019).

Three age groups were preliminarily identified, each with a balanced distribution per mode (18 participants): 18–35 years old (youth), 36–49 years old (adults), 50–60 years old (seniors).

In terms of knowledge, a summary of the post-training and VT post-training values is shown in Table 3. The results indicate that, for both

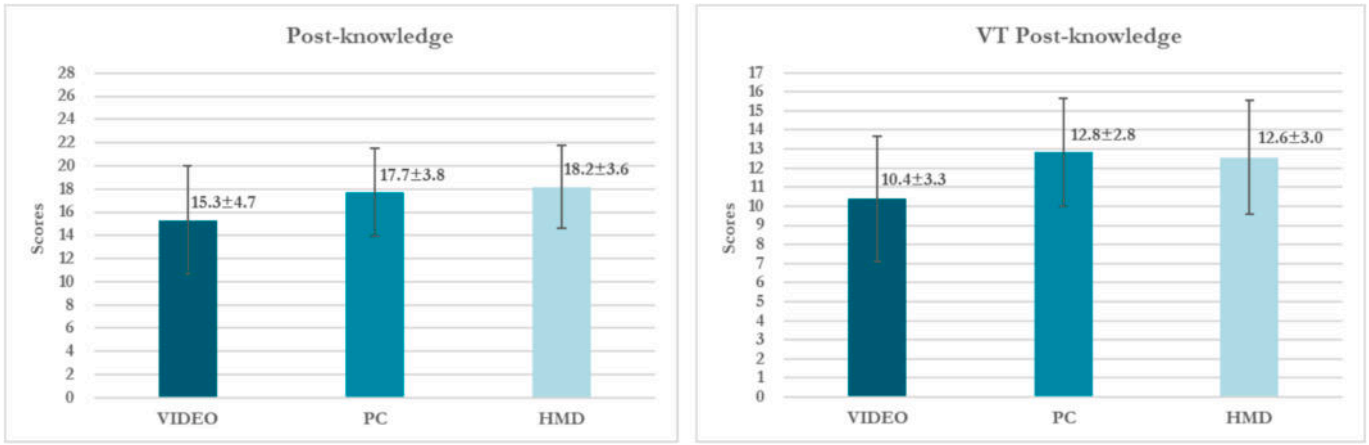


Fig. 4. Mean and standard deviation from post-training knowledge scores (maximum 28) and Virtual Tour knowledge scores (maximum 17) by mode.

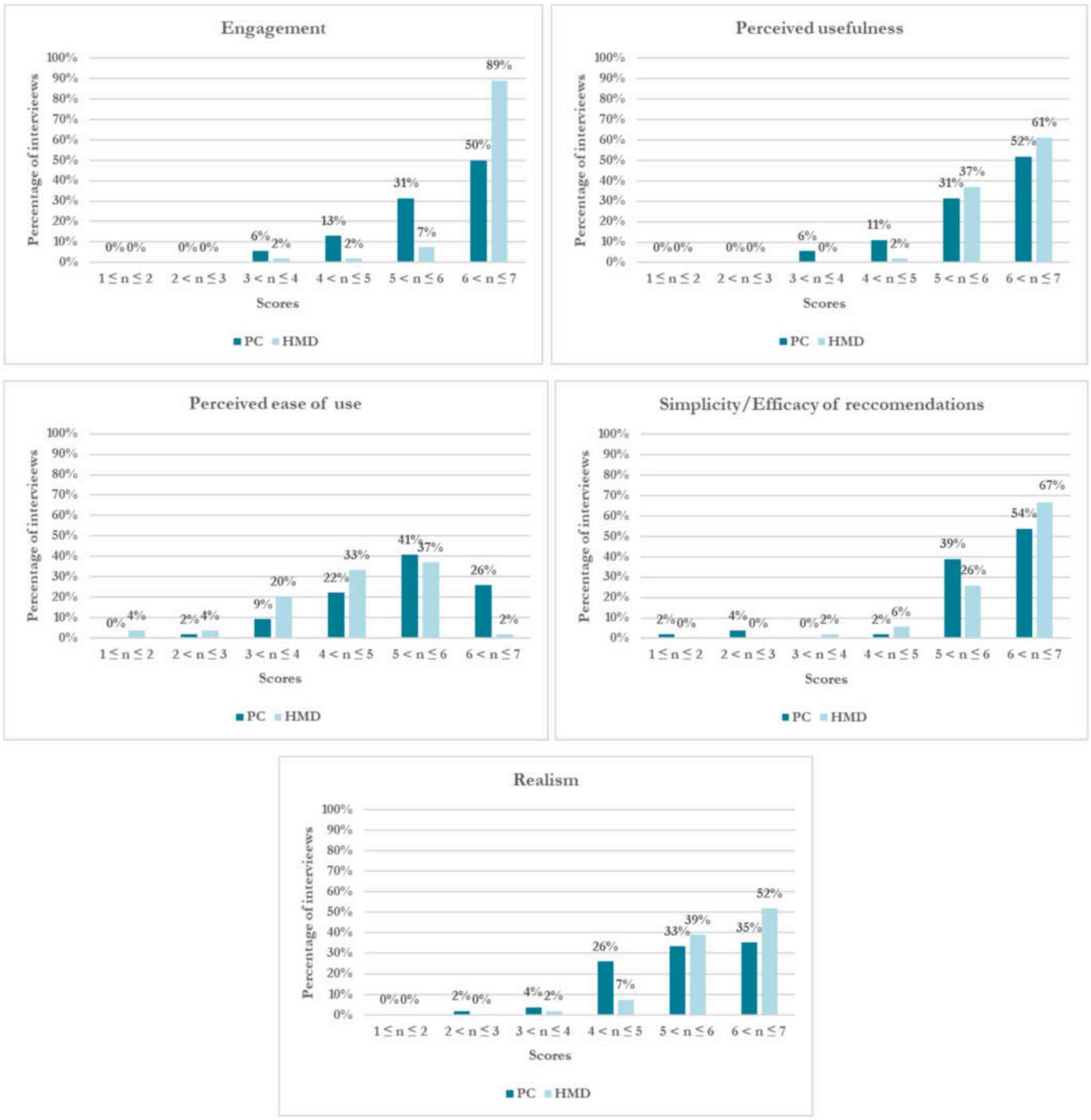


Fig. 5. Percentage of 7-point Likert scale scores for the indicators of user experience.

categories, the first two groups (youth and adults) exhibit increased knowledge gains with increased immersion, while in the third group (seniors) PC outperforms both video and HMD.

Similarly, a summary of the mean and standard deviation values on the 7-point Likert scale for the user experience indicators and questions is presented in Table 4. The results show consistent trends across age groups, with the HMD receiving higher ratings for QE1 (engagement), QE2 (perceived usefulness) and QE5 (realism), while the PC mode was rated higher for QE3 (ease of use). It is also noteworthy that the youth group generally gave higher scores across all categories, including ease of use. Responses from the adult and senior groups were largely similar, except for ease of use, which seniors rated the lowest across all indicators and modes.

4.4. *Content analysis*

The open-ended answers, related to lists of items, were further analysed, with specific focus on the Heat Wave and Earthquake game sessions (Fig. 7). A simple content analysis on the recurrence of the right items has showed that, for the Heat Wave, participants mostly referred to buildings casting shadow before the training (about 57 % of the questionnaires), while water sources were rarely mentioned (about 6 %). After the training, the answers were more balanced and the knowledge gaps were filled by exceeding 50 % for all the items.

Table 2
Results from user experience scores by questions and mode.

Questions	*PC*	*HMD*
QE1.1. The training experience was fun and enjoyable	5.9 ± 1.2	6.6 ± 0.7
*QE1.2. Safety training activities are boring**	6.1 ± 1.4	6.8 ± 0.5
QE1.3. I would describe safety training as very interesting	5.9 ± 1.2	6.4 ± 0.9
*QE1.4. Safety training does not hold my attention at all**	6.3 ± 1.2	6.5 ± 0.9
QE1.5. It was easy for me to concentrate on my learning	5.6 ± 1.5	6.1 ± 1.1
QE2.1. Using this type of simulation as an educational tool will enhance my learning	5.9 ± 1.3	5.9 ± 0.8
QE2.2. This type of simulation is as useful as indoor simulations	5.6 ± 1.7	6.3 ± 1.0
QE2.3. This type of simulation is useful for behaving properly in real case. too	6.2 ± 1.2	6.6 ± 0.8
QE2.4. This type of simulation is useful as a learning supplement	5.9 ± 1.1	5.8 ± 0.8
QE2.5. The combination of two risks is effective because it simulates real conditions	6.3 ± 1.0	6.7 ± 0.4
QE2.6. The simulation of the crowd helped me make the right decisions	5.2 ± 1.7	6.0 ± 1.2
*QE3.1. This simulation tool is rigid and inflexible to interact with**	5.4 ± 1.3	4.7 ± 1.6
QE3.2. I think this training tool is easy to use	5.7 ± 1.0	4.8 ± 1.1
QE4.1. I could easily remember the recommendations provided in the virtual experience	6.2 ± 1.3	6.4 ± 0.8
QE4.2. The recommendations provided in the training experience are useful for my safety	6.3 ± 1.2	6.5 ± 0.7
QE5.1. The built environment was realistic	6.1 ± 1.0	5.9 ± 1.0
QE5.2. The built environment reminded me of a familiar place	5.3 ± 1.5	5.6 ± 1.0
QE5.3. The VR experience was realistic	5.0 ± 1.5	6.1 ± 0.9
QE5.4. The realism of the virtual world motivates me to learn	5.4 ± 1.5	6.3 ± 0.9
QE5.5. The virtual world makes learning more interesting	5.9 ± 1.4	6.7 ± 0.6

Similarly, for the Earthquake, before the training, the recommendation to avoid closeness to buildings was widely known (about 75 % of the questionnaires), while the other proper behaviours were less familiar. However, for the latter ones, a significant knowledge gain (delta knowledge data in Fig. 7), i.e. difference between post- and pre-training, was assessed after the game testing.

4.5. In-game errors

Concerning the assessment of in-game "false" answers and repetitions, the pairwise comparison of the number of errors between the VR training modes was run by the non-parametric Mann-Whitney u-test, since the data were found not normally distributed. No statistical differences were found ($U_{value} = 1420$; z-score = 0.23; $\rho = 0.8181 > 0.05$).

Furthermore, for the Heat Wave and Earthquake game sessions, the number of errors was associated to the difference of post- and pre-training scores from the open-ended answers, corresponding to the lists of items. Fig. 8 shows that, regardless the number of errors, the knowledge gain is quite consistent across the sample. Thus, the repetitions have somewhat brought the players who made more mistakes achieve a similar improvement in knowledge to those who made few or no mistakes. Moreover, the number of errors somewhat corresponds to

Table 3
Mean and standard deviation for the scores of the second part by mode.

Indicator	*Group*	*VIDEO*	*PC*	*HMD*
Post-Knowledge	***18–35***	16.4 ± 4.1	18.1 ± 3.6	18.7 ± 3.0
	36–49	13.6 ± 4.9	16.5 ± 2.5	18.4 ± 3.5
	50–60	16.0 ± 4.8	18.6 ± 4.8	17.4 ± 4,2
Post-Knowledge VT	***18–35***	10.9 ± 2.4	12.9 ± 2.6	13.3 ± 2.2
	36–49	9.1 ± 3.7	12.5 ± 2.4	12.5 ± 3.2
	50–60	11.2 ± 3.4	13.1 ± 3.4	11.9 ± 3.4

Table 4
Mean and standard deviation for all the indicators of the third part by mode.

Indicator	*Group*	*PC*	*HMD*
QE1. Self-reported engagement	***18–35***	6.1 ± 0.8	6.4 ± 0.5
	36–49	5.9 ± 1.1	6.4 ± 0.7
	50–60	5.8 ± 0.9	6.6 ± 0.5
QE2. Perceived usefulness	***18–35***	6.1 ± 0.7	6.4 ± 0.7
	36–49	5.7 ± 1.0	6.2 ± 0.4
	50–60	5.8 ± 0.8	6.1 ± 0.4
QE3. Perceived ease of use	***18–35***	5.9 ± 0.9	4.9 ± 0.8
	36–49	5.6 ± 1.2	4.8 ± 1.5
	50–60	5.2 ± 0.8	4.5 ± 1.0
QE4. Recommendation simplicity and efficacy	***18–35***	6.7 ± 0.5	6.4 ± 0.6
	36–49	5.9 ± 1.6	6.4 ± 0.7
	50–60	6.1 ± 1.0	6.4 ± 0.9
QE5. Realism	***18–35***	5.8 ± 0.7	5.8 ± 0.8
	36–49	5.4 ± 1.2	6.3 ± 0.6
	50–60	5.4 ± 0.8	6.2 ± 0.6

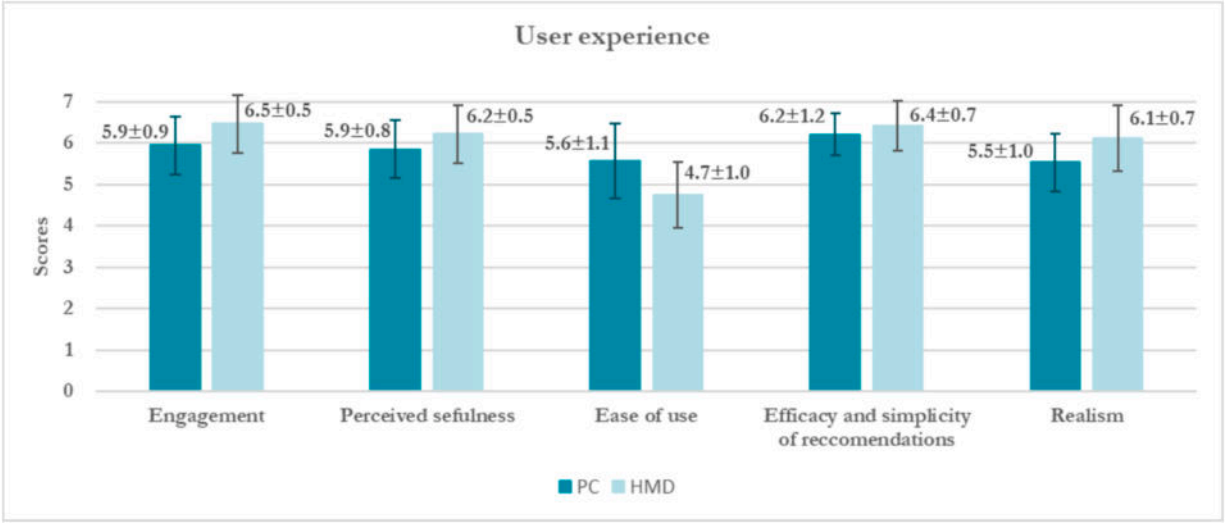


Fig. 6. Results from user experience scores by indicator and mode.

Fig. 7. Knowledge gain versus training items.

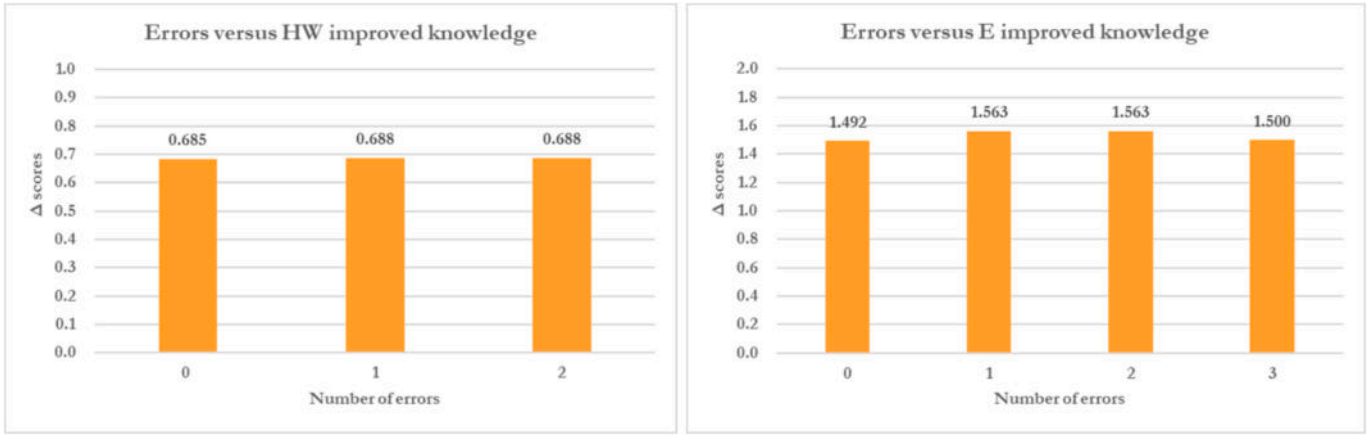


Fig. 8. In-game errors versus knowledge gain.

the VR familiarity (median 1, maximum 4, mode 1 for users declaring no previous experience and median 2, maximum 5 and mode 2 for users declaring it), and to the frequency of playing videogames (median 1, maximum 3, mode 0 for users replying "at least once a month" or more frequent and median 1, maximum 5, mode 1 for users replying "at least once a year" or less frequent).

5. Discussion

In general terms, the results showed that the effectiveness in gaining knowledge is statistically different for both HMD compared to video and PC compared to video, but statistically similar between HMD and PC. Moreover, HMD outperforms PC for engagement, perceived usefulness and realism, while underperforming for ease of use. Additionally, the sub-analyses have shown that: senior groups achieve the highest mean knowledge scores using the PC, while reporting the lowest mean scores for ease of use with the HMD; most training concepts/items related to risk and safety were poorly known before the training, but widely recognized after; repeated exercises are beneficial in strengthening the informative content; and, higher reliability on attention and accountability by participants was achieved during group public events rather than individual sessions.

The following sub-sections provide an outlook of some advantages due to the adoption of the BE S^2ECURe VR-SG framework, as well as some recommendations on future development and testing of virtual training tools for risk communication within the safety management of the built environment.

5.1. Immersion level

Concerning the testing outcome against the level of immersion, the significant knowledge gain and transfer after training in all the three modes – video, PC and HMD – demonstrates the efficacy and versatility of the tool. Moreover, since both HMD and PC modes outperform the video one, but are statistically similar in pairwise comparison, the efficacy and acceptability of non-immersive training solutions are confirmed in terms of successful learning and subject assessment, particularly compared with traditional approaches, as suggested by other authors (D'Amico et al., 2023; Rahouti et al., 2021). Thus, the possibility to develop multiple modes for the same game is recommended, especially in case of great variability of the training receivers, e.g. by age. In fact, as noted for the senior group, the PC mode can be more effective in knowledge compared to both video and HMD (Table 3). The explanation does not seem to lie in the low attractiveness of the immersive tool for senior users, who scored the HMD mode very highly for engagement, consistently with other studies (Tseng and Giau, 2022). Instead, the issue seems to lie in the ease of use that received lower scores for the HMD mode, both compared to the PC mode within the same group and compared to the same mode for the other groups (Table 4). Thus, specific attention should be paid to senior users in wide replication of the training, by ensuring that the users are aware of the options to lower the level of immersion by selecting the PC mode or they are fully aware on how the HMD apparatus, relative controllers and interaction functionalities work, even based on pre-training demonstration. This is paramount to prevent the player from being disoriented by the environment and distracted from the knowledge delivery.

5.2. Realism and knowledge transfer

The prototype was developed by displaying a typological scene (repetitive patterns, schematic volumes, simple decorations, plain surfaces), as representative of a wide set of case studies. Nonetheless, the realism is scored quite positively (Fig. 6 and Table 2), for both the built environment (see *QE5.1. The built environment was realistic*) and the experience (see *QE5.3. The VR experience was realistic*), even acknowledging the resemblance with familiar places (see *QE5.2. The built environment reminded me of a familiar place)*. Moreover, it was recognised as valuable in the learning process (see *QE5.4. The realism of the virtual world motivates me to learn*). In this regard, the HMD mode emerged as the most effective, likely due to the enhanced fidelity by immersion. Moreover, the post-training knowledge measured by the close ended questions, viewed the VT of a real case study, results in very good scores for all the interviewees (Fig. 4). This confirms the benefit of integrating the VT within the training itself in order to boost the overall consistency with real life scenarios though the skill transfer. This ability is seen as highly beneficial in all fields of virtual training and it was not previously investigated for risk training in buildings and built environments through virtual tools (Strojny and Dużmańska-Misiarczyk, 2023). Finally, while the perception of realism is scored relatively lower than other user experience indicators, such as engagement, this aspect presents an opportunity for improvement. To this end, in view of keeping the computational load manageable, it is recommended to identify and enhance only the most influencing visual and sound effects, e.g. smoke, alarms, falling debris, (Lu et al., 2020). Additionally, incorporating sensors, possibly within wearable devices, could simulate smells, vibrations and heat that might foster the credibility of the generated virtual world, in line with recommendations from other studies (Gagliardi et al., 2023). Moreover, the application on further VR devices, including CAVEs (Cave Automatic Virtual Environments) should be assessed to understand how additional VR solutions can impact final training effects on users. In particular, CAVEs could be explored since previous works seem to address a possible high-level visuals and large-scale immersion, along with possible integration with real-world elements and advanced lighting, shadow, and reflection effects, and group experiences, too (de Back et al., 2020; Liu and Liu, 2025; Scorgie et al., 2024).

5.3. Simplicity and usefulness

All the participants rated he simplicity and efficacy of recommendations very highly (Fig. 6 and Table 2), confirming that the modularity of the prototype effectively delivers information. The results are not dependent from the immersion level. Concerning the perceived usefulness, the scores are medium–high and high for PC and HMD modes respectively, including the aspects related to outdoor training (see *QE2.2. This type of simulation is as useful as indoor* simulations) and multi-risk approach (see *QE3.5. The combination of two risks is effective because it simulates real conditions*). These features distinctively feature the prototype compared to previous applications. However, one question received relatively lower scores regarding the role of the crowd (see *QE5.6. The simulation of the crowd helped me make the right decisions*), resulting from the agent-based simulation to estimate the time and paths covered by Non Playable Characters (NPCs) to reach a safe area. The result might point out that the crowd motion does not play a prominent role in leading/encouraging individual actions throughout the game, as informally reported by several interviewees. who found it negligible. A solution could be to utilize agent-based data for animating Playable Characters (PCs) that actively instruct players, as tested in other studies (Feng et al., 2020a, 2020c; Serafini and Chittaro, 2023; Tucker et al., 2018).

Furthermore, allowing participants to receive explanations and retry after providing "false" answer is confirmed to be beneficial for the learning process. Players who failed several times to provide the right answer achieved a comparable level of knowledge gain to those who made few or no mistakes. This error/repetition approach seems to be particularly useful for compensating low familiarity with the tools, in terms of experience with VR and videogames and consequent difficulties in concentrating during the training. In this regard, to the best of the authors' knowledge, this research introduces a novel aspect of investigation, although further studies are needed to validate these insights.

5.4. Participants' awareness

Although the analysis focused on simple keywords to verify correct answers examining the most recurring terms in open-ended questions, both from pre- and post- training questionnaire sections, was found quite interesting and useful. It allowed for a general assessment of concepts that require strengthening in risk communication (Fig. 7). This kind of content analysis, which could be empowered by customized design of questions and more complex methods of data mining (e.g. by natural language processing algorithm), has been successfully used in other fields combining VR and AI (Antel et al., 2022) and might be beneficial in risk training, too. Moreover, whenever the tool is arranged in subsequent modules and is based on teleporting, thus limiting freedom of observable actions and movements of the player, this approach may be more worthwhile than others a (e.g. behavioural analysis. sensor-based measurements) with different training layouts and motion modes.

Finally, concerning the participants' awareness, the questionnaires were administrated in targeted events whit several users involved in demonstration sessions or individually in a laboratory. It was qualitatively observed that individual testing was less reliable than group testing, in terms of user concentration and the reliability of the outcome. In some cases, results had to be discarded for evident inconsistencies (high Likert scale scores for both positive and negative questions, no answers to the post-training questions, out of topic open answers). This is a shortcoming in view of large-scale dissemination, particularly for the non-immersive mode, that could be deliverable remotely by web-based smartphone and tablet applications (Chittaro and Buttussi, 2022), where the users should show autonomy and trustworthiness in developing the training. Furthermore, the above-mentioned issues suggest that the social dimension in learning and training is crucial. The development of team gaming solutions should be addressed, as collective experiences in a virtually mediated space, where there is a shared goal/problem which learners must attend to collaboratively (Paulsen et al., 2024). To this end, synchronous game sessions, by CAVEs or multiple HMDs, could better fulfil the purpose.

5.5. Limitations of the work

The study presents certain limitations that may suggest future research. The sample size did not allow for robust analyses of the results based on target groups across different demographic indicators. However, the sub-analysis on age groups indicated that this topic warrants rigorous evaluation and is particularly relevant to risk training for a broad spectrum of participants. To this end, a greater diversification of possible responses to the questions in the first part (e.g., experience with VR) could support a more accurate assessment of the impact on training

outcomes. Additionally, the evaluation of knowledge and user experience was based on the most common indicators reported in the literature for risk training against natural and human-made hazards in the built environment, excluding certain areas of investigation (e.g., situational awareness, risk perception, or physiological/psychological responses), which could provide more comprehensive insights into users' responses to VR-based training. Finally, a noteworthy aspect deserving attention is the assessment of the impact of microclimatic conditions – e.g., lighting, temperature – on user performance that was neglected in this study.

6. Conclusion

The study has meant to validate the HMD immersive and PC non-immersive prototypes of a VR-SG framework, developed for multi-risk training (heatwave and earthquake) in a typological urban open space. Particularly, in view of broad application, feedback questionnaires were administered toward the identification of relevant differences in knowledge gain and transfer to real cases studies, as well as in perceived user experience, based on the level of immersion.

The outcome has shown the overall efficacy and usability of the prototypes, suggesting that HMD and PC modes do not show statistical differences in knowledge gain and transfer. They are also rated very highly for engagement, perceived usefulness, ease of use, realism and simplicity/efficacy of recommendations. However, HMD outperforms PC for engagement, perceived usefulness and realism, while PC outperforms HMD for ease of use.

Moreover, the validation of several features of the VR-SG, including the typological outdoor setting, the modular structure and the multi-risk approach was achieved, along with the suggestion of further improvements related to the realism of the scene and the role of NPCs. The overall outcome has confirmed the need of accurate and rigorous user-based empirical assessments of this kind of solutions.

All the above-mentioned issues are particularly relevant to the adoption of VR in the broader field of risk and safety communication, including safety awareness increase, risk perception assessment, and strategic consultation on mitigation plans. This is especially important in the context of inclusive community engagement in developing processes and solutions for managing hazards, emergencies, and the effects of disasters.

In particular, VR tools can be conveniently used by local administrators to convey specific risk and safety instructions to citizens and visitors through tailored dissemination events. These events can focus on potential dangers and/or protective measures within specific open spaces, such as public parks, squares, or high-risk areas. By simulating emergency scenarios in a virtual environment, users can become familiar with safety procedures in a controlled and interactive setting, significantly improving their understanding and preparedness for real-life dangers. For example, VR simulations can help people practice evacuations from crowded areas or train them on how to respond to disasters, thus enhancing public safety readiness.

Moreover, the SG approach can be explored more broadly, independent of the specific context, such as in schools and associations, to raise general awareness and assess behavioural choices and attitudes. Through gamification, participants can simulate critical decision-making processes, develop practical risk response skills, and evaluate the consequences of their actions in simulated environments. This educational experience is especially useful in preventing risky behaviours and fostering self-protection, by teaching best safety practices and reducing exposure to hazards.

Finally, from an urban planning and infrastructure design perspective, the structure underlying the digital environment and the game narrative in modules, questions, answers, and explanations can be easily adapted for visualizing future mitigation solutions. This provides significant benefits to designers and planners allowing them to demonstrate the compatibility of strategies with physical spaces and integrate these strategies effectively into the built environment. At the same time, it enables immediate evaluation of the perception and acceptability of these strategies by expert (financiers, regulatory bodies) and non-expert users (urban community and touristic groups). This type of visualization and simulation also helps to identify and address safety concerns, making the planning process more inclusive, secure, and transparent for all stakeholders involved.

Ethical approval

The ethical issues included the necessity to properly and fully inform volunteers about the experiment procedures and the analyzed data. No sensitive information was collected. and data were anonymized. The consent form signed by the participants was preliminary approved by the University Ethics Committee during the meeting held on 4th April 2023.

CRediT authorship contribution statement

Mariella De Fino: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Federica Cassano:** Writing – review & editing, Methodology, Investigation, Formal analysis, Data curation. **Gabriele Bernardini:** Writing – review & editing, Methodology, Formal analysis. **Enrico Quagliarini:** Validation, Supervision, Funding acquisition. **Fabio Fatiguso:** Validation, Supervision, Funding acquisition, Conceptualization.

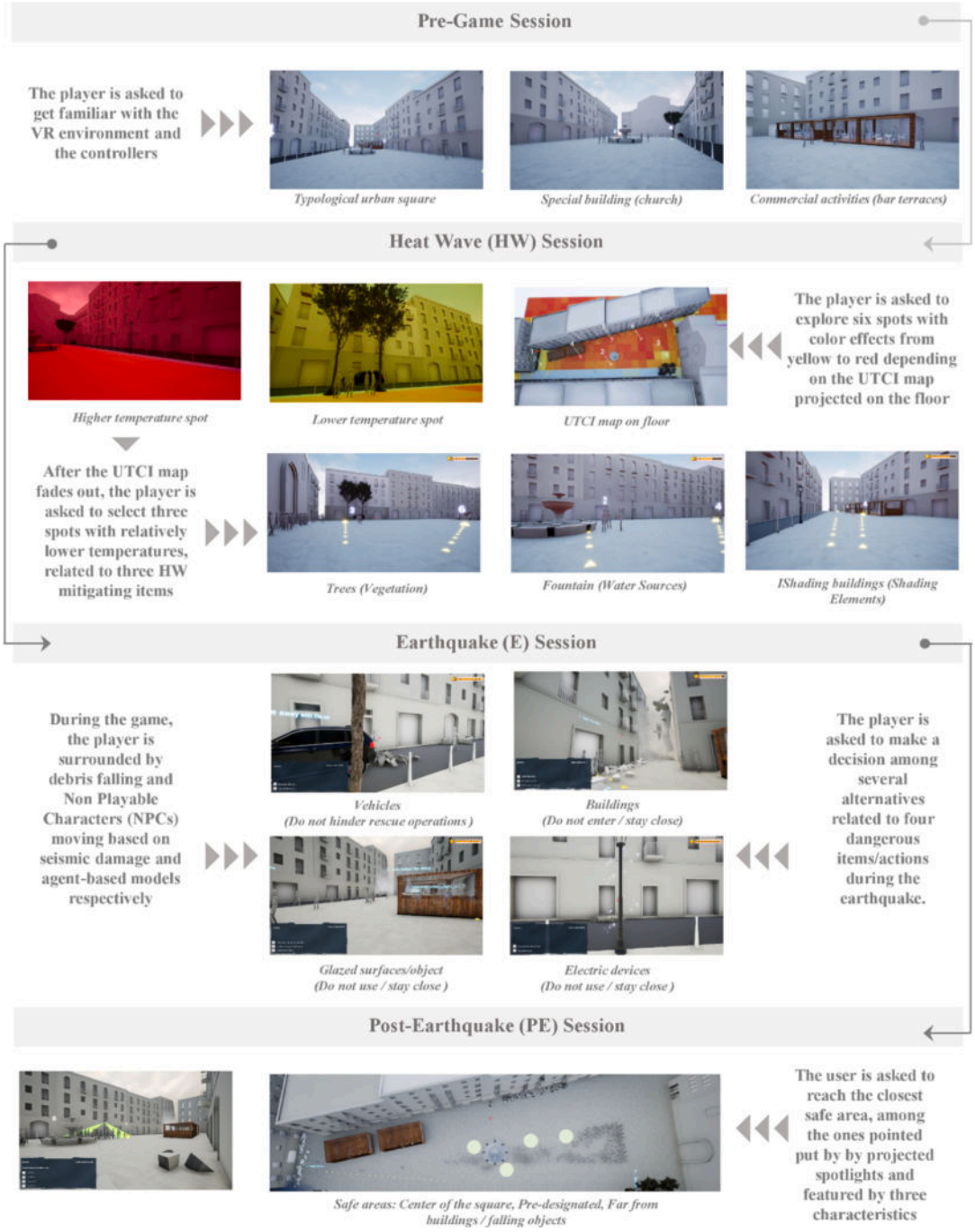
Declaration of competing interest

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

Acknowledgment

This research was funded by MIUR (Italian Ministry of Education. University. and Research) Project BE S^2ECURe—(make) Built Environment Safer in Slow and Emergency Conditions through behavioural assessed/designed Resilient solutions (grant number: 2017LR75XK).

Annex 1. Overview of the BE²SECURE VR-SG prototype



Annex 2. Feedback questionnaire

2.1 Questions and answers of the first part – PARTECIPANTS

Questions	Answers
Gender	male / female
Age	(free text)
Educational level	primary school/ secondary school/ high school/ bachelor-msc degree/ post graduate specialization
Previous experience in training	never / once / twice / more than twice / unsure
Previous experience in earthquake training	never / once / twice / more than twice / unsure
Previous experience in heat wave training	never / once / twice / more than twice / unsure
Frequencies of playing videogames	never / less than once a year / at least once a year/ at least once a month / at least once a wee / several days a week / everyday
Experience with VR	no / yes / unsure

2.2 Open-ended questions and scores of the second part – Knowledge

Questions/Answers	Scores	Assessment
QK1.Where would you feel less exposed to high temperatures in an open space? (Free text)	3 points for knowing 3 out of 3 among the following items or similar: (i) area in shadow (e.g. building); (i) water sources (e.g. fountain); (iii) vegetation (e.g. trees); 2 points for 2 out of 3 items; 1 point for 1 out of 3 items; 0 point for knowing nothing	Sum of scores
QK2.What would you avoid doing during an earthquake in an open space? (Free text)	4 points for knowing 4 out of 4 among the following items or similar: (i) stay close to buildings; (ii) stay close to glazed elements; (iii) stay close to electric devices; (iv) use vehicles; 3 points for 3 out of 4 items; 2 points for 2 out of 4 items 1 point for 1 out of 4 items; 0 point for knowing nothing	
QK3.Where would you go after an earthquake in an open space? (Free text)	3 points for knowing 3 out of 3 among the following items or similar: (i) reach the center of the square; (ii) reach pre-set designated areas; (iii) reach areas free of buildings and falling objects 2 points for knowing 2 out of 3 items 1 point for knowing 1 out of 3 items 0 points for knowing nothing	

2.3 Hotspots within the virtual tour

Heat Wave – Protective elements	Earthquake – Dangerous elements	Post-Earthquake
Closeness to shading building: B.,C, H, I Closeness to trees/vegetation: E Closeness to fountain/water: A	Buildings potentially collapsing: C, E, H, I Glazed surfaces/elements: B Electric devices: M Vehicles: L	Open areas in the middle of the square far from dangerous elements: D. F. G

2.4 Panoramas of Piazza dei Priori in Narni with hotspots marked by capital letters from A to M

2.5 Closed-ended questions and scores of the second part – KNOWLEDGE

Questions	Scores	Assessment
QK4.Which hotspots of the virtual tour correspond to protective positions/items for heat waves?	1 point for each right item (A.B.C.E.H.I) up to 6 points	Sum of scores
QK5.Which hotspots of the virtual tour correspond to dangerous positions/items during an earthquake?	1 point for each right item (B.C.E.H.I.L.M) up to 7 points	
QK6.Which hotspots of the virtual tour correspond to safe area positions after an earthquake?	1 point for each right item (A.D.F.G) up to 4 points	

2.6 Likert scale questions and scores of the third section – USER EXPERIENCE

Indicators	Questions	Scores	Assessment
QE1. Self-reported engagement	QE1.1.The training experience was fun and enjoyable (Shiradkar et al., 2021) QE1.2.Safety training activities are boring (Lovreglio et al., 2022) QE1.3.I would describe safety training as very interesting (Lovreglio et al., 2022) QE1.4.Safety training does not hold my attention at all (Lovreglio et al., 2022) QE1.5.It was easy for me to concentrate on my learning (Feng et al., 2022a)	1 (strongly disagreed) –7 (strongly agreed)	Average score
QE2. Perceived usefulness	QE2.1.Using this type of virtual reality simulation as an educational tool will enhance my learning (Rahouti et al., 2021) QE2.2.This type of simulation is useful as a learning supplement (Davis, 1989) QE2.3.This type of simulation is as useful as simulation of indoor spaces QE2.4. This type of simulation is useful for behaving properly in real cases. too QE2.5.The combination of two risks (heat wave + earthquake) is effective because it simulates real conditions QE2.6.The simulation of the crowd helped me in taking the right decisions	1 (strongly disagreed) –7 (strongly agreed)	Average score
QE3. Perceived ease of use	QE3.1.This simulation tool is rigid and inflexible to interact with (Davis, 1989) QE3.2.I think this training tool is easy to use (Davis, 1989; Rahouti et al., 2021)	1 (strongly disagreed) –7 (strongly agreed)	Average score

(*continued on next page*)

(*continued*)

Indicators	Questions	Scores	Assessment
QE4. Recommendation simplicity and efficacy	QE4.1.I could easily remember the recommendations provided in the virtual experience (Chittaro and Sioni, 2015; Lovreglio et al., 2021; Rahouti et al., 2021) QE4.2.The recommendations provided in the training experience are useful for my safety (Chittaro and Sioni, 2015; Lovreglio et al., 2021; Rahouti et al., 2021)	1 (strongly disagreed) –7 (strongly agreed)	Average score
QE5. Realism	QE5.1.The built environment was realistic (Feng et al., 2022a) QE5.2.The built environment reminded me of a familiar place QE5.3.The VR experience was realistic (Feng et al., 2022a) QE5.4.The realism of the virtual world motivates me to learn (Rahouti et al., 2021) QE5.5.The virtual world makes learning more interesting (Dalgarno et al., 2002)	1 (strongly disagreed) –7 (strongly agreed)	Average score

Data availability

Data will be made available on request.

References

Ai-Lim Lee, E., Wong, K.W., Fung, C.C., 2010. How does desktop virtual reality enhance learning outcomes? a structural equation modeling approach. *Comput. Educ.* 55, 1424–1442. https://doi.org/10.1016/J.COMPEDU.2010.06.006.

Akindele, N., Taiwo, R., Sarvari, H., Oluleye, B.I., Awodele, I.A., Olaniran, T.O., 2024. A state-of-the-art analysis of virtual reality applications in construction health and safety. *Res. Eng.* 23, 102382. https://doi.org/10.1016/j.rineng.2024.102382.

Antel, R., Abbasgholizadeh-Rahimi, S., Guadagno, E., Harley, J.M., Poenaru, D., 2022. The use of artificial intelligence and virtual reality in doctor-patient risk communication: a scoping review. *Patient Educ. Couns.* 105, 3038–3050. https://doi.org/10.1016/j.pec.2022.06.006.

Anwar, M.S., Wang, J., Ullah, A., Khan, W., Li, Z., Ahmad, S., 2018. User profile analysis for enhancing QoE of 360 panoramic video in virtual reality environment. *Proc. Int Conf. Virtual Real. vis. ICVRV* 2018, 106–111. https://doi.org/10.1109/ICVRV.2018.00030.

Bernardini, G., Quagliarini, E., Orazio, M.D., 2023. Historic Public Open Spaces, users and multihazard : using be- havioural-design to assess and mitigate heatwaves and terrorist act in typological scenarios. *Heritage* 1–15.

Buttussi, F., Chittaro, L., 2018. Effects of different types of virtual reality display on presence and learning in a safety training scenario. IEEE, pp. 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117.

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.

Cao, L., Lin, J., Li, N., 2019. A virtual reality based study of indoor fire evacuation after active or passive spatial exploration. *Comput. Human Behav.* 90, 37–45. https://doi.org/10.1016/j.chb.2018.08.041.

Chittaro, L., Buttussi, F., 2022. Learning safety through public serious games: a study of "prepare for impact" on a very large, international sample of players. IEEE Trans. Vis. & Comput. Graph. 28, 1573–1584. https://doi.org/10.1109/TVCG.2020.3022340.

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

Coaffee, J., O'Hare, P., Hawkesworth, M., 2009. The visibility of (in) security: The aesthetics of planning urban defences against terrorism. *Secur. Dialogue* 40, 489–511.

D'Amico, A., Russo, M., Angelosanti, M., Bernardini, G., Vicari, D., Quagliarini, E., Currà, E., 2021. Built environment typologies prone to risk: a cluster analysis of open spaces in Italian cities. *Sustain.* 13, 1–32. https://doi.org/10.3390/su13169457.

D'Amico, A., Bernardini, G., Lovreglio, R., Quagliarini, E., 2023. A non-immersive virtual reality serious game application for flood safety training. *Int. J. Disaster Risk Reduct.* 96, 103940. https://doi.org/10.1016/j.ijdrr.2023.103940.

Dalgarno, B., Hedberg, J., Harper, B., 2002. The contribution of 3D environments to conceptual understanding. *Learning* 1, 44–54.

Dane, G., Evers, S., van den Berg, P., Klippel, A., Verduijn, T., Wallgrün, J.O., Arentze, T., 2024. Experiencing the future: evaluating a new framework for the participatory co-design of healthy public spaces using immersive virtual reality. *Comput. Environ. Urban Syst.* 114, 102194. https://doi.org/10.1016/j.compenvurbsys.2024.102194.

Davis, F.D., 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 319–340. https://doi.org/10.2307/249008.

de Back, T.T., Tinga, A.M., Nguyen, P., Louwerse, M.M., 2020. Benefits of immersive collaborative learning in CAVE-based virtual reality. *Int. J. Educ. Technol. High. Educ.* 17, 51. https://doi.org/10.1186/s41239-020-00228-9.

De Fino, M., Tavolare, R., Bernardini, G., Quagliarini, E., Fatiguso, F., 2023. Boosting urban community resilience to multi-hazard scenarios in open spaces: a virtual reality – serious game training prototype for heat wave protection and earthquake response. *Sustain. Cities Soc.* 99, 104847. https://doi.org/10.1016/j.scs.2023.104847.

Dhalmahapatra, K., Maiti, J., Krishna, O.B., 2021. Assessment of virtual reality based safety training simulator for electric overhead crane operations. *Saf. Sci.* 139, 105241. https://doi.org/10.1016/j.ssci.2021.105241.

Diez, H.V., Garcia, S., Mujika, A., Moreno, A., Oyarzun, D., 2016. Virtual training of fire wardens through immersive 3D Environments. *Proc. Int. Conf. Web3D Technol.* 43–50. https://doi.org/10.1145/2945292.2945296.

Dzeng, R.-J., Hsueh, H.-H., Chang, R.-N., 2015. 3D game-based training system for hazard identification on construction site, *in*: 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD). pp. 2453–2458. DOI: 10.1109/FSKD.2015.7382339.

Felton, W.M., Jackson, R.E., 2022. Presence: A Review. *Int. J. Human–computer Interact.* 38, 1–18. https://doi.org/10.1080/10447318.2021.1921368.

Feng, Z., Liu, C., González, V., Lovreglio, R., 2022b. Prototyping an immersive virtual reality training system for urban-scale evacuation using 360-degree panoramas, *in*: CIB 2022. RMIT, Melbourne, Australia.

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

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

Feng, Z., González, V.A., Mutch, C., Amor, R., Rahouti, A., Baghouz, A., Li, N., Cabrera-Guerrero, G., 2020b. Towards a customizable immersive virtual reality serious game for earthquake emergency training. *Adv. Eng. Inform.* 46. https://doi.org/10.1016/j.aei.2020.101134.

Feng, Z., González, V.A., Trotter, M., Spearpoint, M., Thomas, J., Ellis, D., Lovreglio, R., 2020c. How people make decisions during earthquakes and post-earthquake evacuation: using verbal protocol analysis in immersive virtual reality. *Saf. Sci.* 129, 104837. https://doi.org/10.1016/j.ssci.2020.104837.

Feng, Z., González, V.A., Mutch, C., Amor, R., Cabrera-Guerrero, G., 2022a. Exploring spiral narratives with immediate feedback in immersive virtual reality serious games for earthquake emergency training. Tools Appl. Multimed. https://doi.org/10.1007/s11042-022-13306-z.

Franz, F., Edgar, E., Albert-Georg, L., Axel, B., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191.

Fujimi, T., Fujimura, K., 2020. Testing public interventions for flash flood evacuation through environmental and social cues: the merit of virtual reality experiments. *Int. J. Disaster Risk Reduct.* 50, 101690. https://doi.org/10.1016/j.ijdrr.2020.101690.

Gagliardi, E., Bernardini, G., Quagliarini, E., Schumacher, M., Calvaresi, D., 2023. Characterization and future perspectives of virtual reality evacuation drills for safe built environments: a systematic literature review. *Saf. Sci.* 163, 106141. https://doi.org/10.1016/j.ssci.2023.106141.

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. https://doi.org/10.3389/fpsyg.2020.01743.

Gravetter, F.J., Wallnau Larry, B., Lori-Ann, B.F., Witnauer, J.E., 2021. Essentials of statistics for the behavioral sciences. Cengan.

Hedberg, J., Harper, B. & Dalgarno, B., 2000. The contribution of 3D environments to conceptual understanding, *in*: Winds of Change in the Sea of Learning: Proceedings of the 19th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education. pp. 149–158.

Irshad, S., Perkis, A., Azam, W., 2021. Wayfinding in virtual reality serious game: an exploratory study in the context of user perceived experiences. *Appl. Sci.* 11. https://doi.org/10.3390/app11177822.

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, e1812.

Kim, J., Ahn, S., Han, S., 2020. Impact of VR Settings on Immersiveness in VR-Based Construction Skills Training: Case Study of Scaffolding Work, *in*: Construction Research Congress 2020: Computer Applications.

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

Kisker, J., Gruber, T., Schöne, B., 2021. Behavioral realism and lifelike psychophysiological responses in virtual reality by the example of a height exposure. *Psychol. Res.* 85, 68–81. https://doi.org/10.1007/s00426-019-01244-9.

Latini, A., Di Giuseppe, E., D'Orazio, M., 2023. Immersive virtual vs real office environments: a validation study for productivity, comfort and behavioural research. *Build. Environ.* 230, 109996. https://doi.org/10.1016/j.buildenv.2023.109996.

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.

Lin, D.-Y.-M., Yang, P.-Y.-D., 2007. The Use of Virtual Reality to Train Older Adults in Processing of Spatial Information BT - Universal Access in Human-Computer Interaction. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 451–459.

Liu, Q., Liu, R., 2025. Virtual reality for indoor emergency evacuation studies: design, development, and implementation review. *Saf. Sci.* 181, 106678. https://doi.org/10.1016/j.ssci.2024.106678.

Lorenz, M., Brade, J., Klimant, P., Heyde, C.E., Hammer, N., 2023. Age and gender effects on presence, user experience and usability in virtual environments-first insights. *PLoS One* 18, e0283565. https://doi.org/10.1371/journal.pone.0283565.

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

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.

Lovreglio, R., Daphney-Chloe, N., Rahouti, A., Paes, D., Feng, Z., Shipman, A., 2022. Prototyping and testing a virtual reality counterterrorism serious game for active shooting. *Int. J. Disaster Risk Reduct.* 82, 103283. https://doi.org/10.1016/j.ijdrr.2022.103283.

Lu, X., Yang, Z., Xu, Z., Xiong, C., 2020. Scenario simulation of indoor post-earthquake fire rescue based on building information model and virtual reality. *Adv. Eng. Softw.* 143, 102792. https://doi.org/10.1016/j.advengsoft.2020.102792.

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

Makransky, G., Borre-Gude, S., Mayer, R.E., 2019a. 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.

Makransky, G., Terkildsen, T.S., Mayer, R.E., 2019b. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learn. Instr.* 60, 225–236. https://doi.org/10.1016/j.learninstruc.2017.12.007.

Man, S.S., Wen, H., So, B.C.L., 2024. Are virtual reality applications effective for construction safety training and education? a systematic review and meta-analysis. *J. Safety Res.* 88, 230–243. https://doi.org/10.1016/j.jsr.2023.11.011.

Marques, B., Santos, B.S., Dias, P., 2024. Ten years of immersive education: overview of a virtual and augmented reality course at postgraduate level. *Comput. Graph.* 124, 104088. https://doi.org/10.1016/j.cag.2024.104088.

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. Safety Res.* 75, 205–221. https://doi.org/10.1016/j.jsr.2020.09.015.

Oliva, D., Somerkoski, B., Tarkkanen, K., Lehto, A., Luimula, M., 2019. Virtual reality as a communication tool for fire safety – experiences from the VirPa project. *CEUR Workshop Proc.* 2359, 241–252.

Pacheco Barzallo, A., Fariña, J., Álvarez de Andrés, E., 2022. Public open spaces: enabling or impeding inclusive evacuation during disasters. *J. Public Sp.* 7, 79–92. https://doi.org/10.32891/jps.v7i2.1474.

Pallavicini, F., Pepe, A., Minissi, M.E., 2019. Gaming in virtual reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games? *Simul. Gaming* 50, 136–159. https://doi.org/10.1177/1046878119831420.

Paulsen, L., Dau, S., Davidsen, J., 2024. Designing for collaborative learning in immersive virtual reality: a systematic literature review. *Virtual Real.* 28. https://doi.org/10.1007/s10055-024-00975-4.

Pedram, S., Palmisano, S., Miellet, S., Farrelly, M., Perez, P., 2022. Influence of age and industry experience on learning experiences and outcomes in virtual reality mines rescue training. *Front. Virtual Real.* 3, 1–12. https://doi.org/10.3389/frvir.2022.941225.

Piccione, J., Collett, J., De Foe, A., 2019. Virtual skills training: the role of presence and agency. *Heliyon* 5, e02583. https://doi.org/10.1016/j.heliyon.2019.e02583.

Plechatá, A., Sahula, V., Fayette, D., Fajnerová, I., 2019. Age-related differences with immersive and non-immersive virtual reality in memory assessment. *Front. Psychol.* 10, 1–12. https://doi.org/10.3389/fpsyg.2019.01330.

Poyade, M., Eaglesham, C., Trench, J., Reid, M., 2021. A transferable psychological evaluation of virtual reality applied to safety training in chemical manufacturing. *ACS Chem. Heal. Saf.* 28, 55–65. https://doi.org/10.1021/acs.chas.0c00105.

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.

Rosso, F., Bernabei, L., Bernardini, G., Cadena, J.D.B., Russo, M., D'Amico, A., Salvalai, G., Currà, E., Quagliarini, E., Mochi, G., 2023. Mitigating Multi-risks in the Historical Built Environment: A Multi-strategy Adaptive Approach, *in*: Littlewood, J., Howlett, R.J., Jain, L.C. (Eds.), Sustainability in Energy and Buildings 2022 . SEB 2022. Smart Innovation, Systems and Technologies, Vol 336. Springer, Singapore, pp. 197–207. DOI: 10.1007/978-981-19-8769-4_19.

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.

Serafini, M., Chittaro, L., 2023. Perception of Virtual Agents as Communicators in Virtual vs. Augmented Reality by a Male Sample, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). DOI: 10.1007/978-3-031-30933-5_3.

Sermet, Y., Demir, I., 2019. Flood action VR, *in*: ACM SIGGRAPH 2019 Posters. ACM, New York, NY, USA, pp. 1–2. DOI: 10.1145/3306214.3338550.

Shiradkar, S., Rabelo, L., Alasim, F., Nagadi, K., 2021. Virtual world as an interactive safety training platform. *Inf.* 12, 1–19. https://doi.org/10.3390/info12060219.

Slater, M., 2018. Immersion and the illusion of presence in virtual reality. *Br. J. Psychol.* 109, 431–433. https://doi.org/10.1111/bjop.12305.

Smith, S., Ericson, E., 2009. Using immersive game-based virtual reality to teach fire-safety skills to children. *Virtual Real.* 13, 87–99. https://doi.org/10.1007/s10055-009-0113-6.

Stefan, H., Mortimer, M., Horan, B., 2023. Evaluating the effectiveness of virtual reality for safety-relevant training: a systematic review. *Virtual Real.* 27, 2839–2869. https://doi.org/10.1007/s10055-023-00843-7.

Strojny, P., Dużmańska-Misiarczyk, N., 2023. Measuring the effectiveness of virtual training: a systematic review. *Comput. Educ. X Real.* 2, 100006. https://doi.org/10.1016/j.cexr.2022.100006.

Sukirman, W., Sujalwo, R.A., 2019. Self-evacuation drills by mobile virtual reality application to enhance earthquake preparedness. *Procedia Comput. Sci.* 157, 247–254. https://doi.org/10.1016/j.procs.2019.08.164.

Szóstak, M., Mahamadu, A.-M., Prabhakaran, A., Caparros Pérez, D., Agyekum, K., 2024. Development and testing of immersive virtual reality environment for safe unmanned aerial vehicle usage in construction scenarios. *Saf. Sci.* 176. https://doi.org/10.1016/j.ssci.2024.106547.

Therón, R., Coelho, A.F., Tan, B., Saghafian, M., No, M.S., Laumann, K., Akhtar, R.S., Skogstad, M.R., 2020. The evaluation of virtual reality fire extinguisher training. *Front. Psychol.* https://doi.org/10.3389/fpsyg.2020.593466.

Tseng, K.C., Giau, D.T.N., 2022. A feasibility study of using virtual reality as a pre-occupancy evaluation tool for the elderly. *Autom. Constr.* 134, 104037. https://doi.org/10.1016/j.autcon.2021.104037.

Tucker, A., Marsh, K.L., Gifford, T., Lu, X., Luh, P.B., Astur, R.S., 2018. The effects of information and hazard on evacuee behavior in virtual reality. *Fire Saf. J.* 99, 1–11. https://doi.org/10.1016/j.firesaf.2018.04.011.

Wiepke, A., Heinemann, B., 2024. A systematic literature review on user factors to support the sense of presence in virtual reality learning environments. *Comput. Educ. X Real.* 4, 100064. https://doi.org/10.1016/j.cexr.2024.100064.

Wilkinson, M., Brantley, S., Feng, J., 2021. A mini review of presence and immersion in virtual reality. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* 65, 1099–1103. https://doi.org/10.1177/1071181321651148.

Yang, S., Dane, G., Arentze, T.A., 2024. Modeling pedestrians' activity time-use choices in a virtual urban public space: the influences of the environment and affective experience. *Cities* 154, 105399. https://doi.org/10.1016/j.cities.2024.105399.