WP 6: Tools and guidelines for improving/designing a resilient BE assessed through case studies and virtual training

T6.1 - Virtual training development: identification of performance-based features; implementation protocols of innovative solutions within VR/AR environments, accessed on site or remotely (smartphone, tablet…) and targeted on different users' profiles (i.e. technicians, rescuers, users).

DELIVERABLE ID	D6.1.2
Deliverable Title	Demo-lab for virtual training of different users to increase resilience through Resourcefulness
Delivery month	M23
Revision	1.0
Main partner	BA
Additional partners	AN
Authors of the contribution	Fabio Fatiguso (POLIBA); Mariella De Fino (POLIBA), Federica Cassano (POLIBA)
Deliverable type	report
Number of pages	20

Abstract

According to the workplan of the research project, with reference to the Bes2ecure Virtual Reality – Serious Game (VR-SG) prototype presented in D6.1.1, a demo-lab for demonstrating the tool and collecting feedbacks from potential users has been planned that can evaluate knowledge improvement of the training contents, as well as engagement, perceived ease of use and efficacy of the proposed tool, among the others. To this end, based on the review of the state-of-the-art on the topic, a questionnaire has been set, including thematic sessions/questions and reporting/analyzing methods, and a workflow has been scheduled for conducting the survey. The survey is related to three different interaction modes, namely non-immersive game through desktop, immersive game through VR headset and non-interactive recording of videogame exemplary sessions, in view of their pairwise comparison. The questionnaires are meant to cover different users' ages, including at least n.60 people 18-35 years old, n.60 people 36-50 years old, n.50 people 51-60 years old.

Keywords

Virtual Training, Serious Game, Multi-hazard scenarios, Prototype Validation, Feedback questionnaire, Real-word case studies.

Approvals

Role	Name	Partner	Date
Coordinator	Enrico Quagliarini	UNIVPM	-
Task leader	Fatiguso Fabio	POLIBA	-

Revision versions

Revision	Date	Short summary of modifications	Name	Partner

Summary

1. Introduction

Several definitions have been provided for the term Virtual Reality (VR) in the last decades, including *"the illusion of participation in a synthetic environment that relies on three-dimensional, stereoscopic, head-tracked display, hand-body tracking and binaural sound"* and *"a combination of computer and interface devices (goggles, gloves, etc.) that present a user with the illusion of being in a three dimensional world of computer-generated objects"* (Gigante 1993). Conversely, according to (Burbules 2004), VR rather refers to *"a computer-mediated simulation that is three-dimensional, multisensory, and interactive, so that the user's experience is 'as if' inhabiting and acting within an external environment",* where the *'as if'* depends on the immersion, defined by four interrelated factors *"interest, involvement, imagination, and interaction"*. Such a vision is not so much focused on the technology that produces the sense of immersion, e.g. by means of goggles, gloves, and head tracking devices, but rather on the sense of immersion itself. Moreover, it denies the idea of the virtual world as *"synthetic or illusory"* and embraces a *"medial concept, neither real nor imaginary, or better, both real and imaginary".*

In this framework, desktop VR, although classified as being low-immersive, might result in successful learning and training outcomes, if VR features are well addressed, in order to enhance presence, motivation, affective and cognitive benefits, active learning and reflective thinking, as well as to consider different spatial abilities and learning styles (Makransky, G., & Petersen 2019) (Ai-Lim Lee et al. 2010). Nevertheless, headset immersive VR is effective, only by choosing appropriate instructional methods (Klingenberg et al. 2023), since the accredited improvement of presence and motivation compared to other media does not necessarily results in outperforming learning, especially if it implies excessive stimuli and burdens (Makransky et al. 2019).

In line with the above-mentioned issues, it should also be observed that the acceptability of new technologies is greatly based on perceived usefulness and perceived ease of use, which are hypothesized to be fundamental determinants of user acceptance (Davis 1989). These are not necessarily related to the level of immersivity. In fact, desktop non-immersive environments can be easier to use than headset immersive environments because people are already familiar with controlling the computer, and such tools do not subject the user to the physical and psychological stress often associated with immersive environments (Hedberg, J., Harper, B. & Dalgarno 2000).

Furthermore, the assumption that immersive VR enhances representation fidelity and realism does not necessarily implies that it consequently enhances the learning process (Skulmowski et al. 2022)(Skulmowski and Rey 2021). In fact, even though the strong resemblance between the virtual visualization and the real world is acknowledged as effective in retrieving information, in some cases and depending on the education tasks and users, the realism can be detrimental. It can add distracting cognitive load that overburdens learners with complexity, it can negatively affect the transfer capability to different contexts compared to more flexible and less concrete mental schemas, and it can prevent the acquisition and application of abstract knowledge against visual knowledge.

Finally, all the underlined issues might greatly depend on the users in terms of learning context, age, gender and experiences in VR, resulting in a different so-called user experience and usability for the same tested VR technology (Anwar et al. 2018) (Lorenz et al. 2023). Thus, the evaluation of new virtual tools with different levels of immersivity and interactivity for different people is worth specific attention.

In the light of the above-mentioned issues, the present document describes and discusses the procedures and methods that were undertaken, in order to assess the Virtual Reality – Serious Game (VR-SG) prototype presented in D6.1.1. In detail, for both the desktop non-immersive and headset immersive mode, a demo-lab for demonstrating the tools and collecting feedbacks from potential users of different ages has been planned, also in comparison of more traditional media, such as video recording. To this end, based on the review of the State of the Art (section2) on the topic, a questionnaire has been set, including thematic sessions/questions, reporting/analyzing methods and number/typology of interviewees (section 3), and a workflow has been scheduled for conducting the survey (section 4).

2. State of the art

All the VR-based training experiences, which were analyzed for the Bes2ecure VR-SG prototype development, have been herein assessed, in terms of testing methods and phases, where applicable (Table 1). Thus, for the papers that report the testing session, an analysis of the main testing topics and procedures was carried out (Table 2).

In detail, beyond a pre-training collection of data on **participants' demographics**, the majority of the studies is focused first and foremost on the **knowledge assessment**, unless the research is a follow-up of a previous assessment mainly focused on the game improvement (Feng et al. 2022a), it is specifically addressed to the usability (Sukirman et al. 2019) and perception/acceptance of the application (Therón et al. 2020) or it is only meant to assess the users' psychological and psychometric behaviors (Irshad et al. 2021).

The evaluation, related to the efficacy and effectiveness in communicating the training concepts and contents (Table 3) before and after the training, might be based on closed-ended questions (Shiradkar et al. 2021) (Lovreglio et al. 2021), including true or false questions (Feng et al. 2022a) or open-ended questions (Rahouti et al. 2021)(Lovreglio et al. 2022)(Chittaro and Sioni 2015)(Oliva et al. 2019). The former approach is used to speed up the survey and/or guide the answers avoiding misleading or meaningless responses, while the latter approach is used in order to avoid prompting with possible answers and/or limited responses. In some cases, a mixed approach is proposed depending on the types of questions (Irshad et al. 2021) or to make the survey less monotonous for specific target groups such as children (Smith and Ericson 2009). In case of open-ended questions, a pre-set list of right answers is defined by the surveyors, in order to enable a quantitative assessment of the free text answers.

Alternatively, the knowledge acquisition might rely on some direct survey during the game, including behavior-based observations, such as the annotation by the surveyor of actions/decisions taken during the game (Lovreglio et al. 2018) (Oliva et al. 2019) and parameter-based measurements, in terms of time and distance covered by the player to accomplish the required tasks (Cao et al. 2019).

Finally, it was found that in a few cases, the testing session relies on the comparison with traditional methods, such as slide-based training, videos and leaflets (Shiradkar et al. 2021)(Rahouti et al. 2021)(Lovreglio et al. 2021)(Feng et al. 2022a), that always proved that innovative tools are more engaging, effective and pervasive than conventional ones.

Risk	Users	VR solution	Immersivity	Perspective	Motion	Setting	Ref	Testing
Fire	University students	SG	-	TP	-	Floor plan of university building type	(Shiradkar et al. 2021)	Yes
	General public	SG	NM	FP	FL	Indoor rooms of building type	(Yang et al. 2021)	No
	Hospital staff	SG	NM	TP	-	Floor plan of the Vincent Van Gogh Hospital, Belgium	(Rahouti et al. 2021)	Yes
	University students	SG	-	TP	FL	Interiors and exteriors of building at Bangkok University, Thailand	(Sacfung et al. 2014)	No
	General	SG	IM	FP	TP	Interiors of office building type	(Oliva et al. 2019)	Yes
	Museum visitors	SG	NM	-	FL	Interiors of museum type	(Cao et al. 2019)	Yes
	Children	SG	IM	FP	FL	Interiors of house type	(Smith and Ericson 2009)	Yes
	General public	SG	IM	FP	FL	Interiors of airplane cabin, hotel room, control room, kitchen, warehouse and factory types	(Therón et al. 2020)	Yes
	Generic public	SG	IM	FP	-	Warehouse, electrical, office and worksite types	(Lovreglio et al. 2021)	Yes
	Firefighters	SG	IM	FP	FL	Interiors of two-storey building type	(Diez et al. 2016)	No
	Firefighters	EM	NM	-	-	Urban/forest environment type	(Moreno et al. 2014)	No
	Firefighters	SG	IM	TP	FL	Interiors of Jukryeong road tunnel, South Korea	(Cha et al. 2012)	No
Earthquake	University staff	SG	IM	FP	TP	Floor plan of an office building at the University of Auckland, New Zealand	(Feng et al. 2022a)	Yes
	Hospital staff	SG	IM	FP	TP	Floor plan at the Auckland City Hospital, New Zealand	(Feng et al. 2020a)	No
	School staff	SG	IM	FP	TP	Interiors of school and office types	(Feng et al. 2020b)	Yes
	General public	SG	IM	FP	FL	Interiors of house and office types	(Li et al. 2017)	No
	Hospital staff	SG	IM	FP	TP	Floor plan at the Auckland City Hospital, New Zealand	(Lovreglio et al. 2018)	Yes
	University students	SG	IM	TP	FL	Interiors of dormitories at Nankai University, China	(Gong et al. 2015)	No
	Building occupants	SG	IM	-	FL	Interiors/exteriors of house type	(Sukirman et al. 2019)	Yes
Flood	General public	SG	IM	FP	TP FL	Interiors of parking lot and exteriors of town types	(Irshad et al. 2021)	Yes
	General public	SG	NM	FP	FL	Interiors of building and exteriors of a town types	(D'Amico et al. 2022)	Yes
	General public	EM	NM	-	-	Exteriors of the old town of Cosenza, Italy	(Macchione et al. 2019)	No
	General public	EM	IM	TP	FL	Exteriors of town and river types	(Fujimi and Fujimura 2020)	No
Terrorism	University students	SG	IM	FP	TP	Floor plan of university building type	(Lovreglio et al. 2022)	Yes
	General public	SG	IM	FP	TP	Interiors/exteriors of train station type	(Chittaro and Sioni 2015)	Yes

Table 1a. Review of papers from D6.1.1 including testing sessions

Risk	Ref	Participants*	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. 2022)	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			

Table 2. Overview of topics covered by testing sessions in literature

Risk	Ref	Knowledge	Feedback Questionnaire (FQ) versus Direct Survey (DS)	Type of questions (FQ)	Type of observations (DS)	Knowledge comparison
Fire	(Shiradkar et al. 2021)	x	FQ	close-ended		Slide-based training
	(Rahouti et al. 2021)	x	FQ	open-ended		Slide-based training
	(Oliva et al. 2019)	x	FQ & DS	open-ended	annotations by the surveyor of performed actions/decision according to a pre-set list	
	(Cao et al. 2019)	x	DS		distance and time covered during the game	
	(Smith and Ericson 2009)	x	FQ	fill in the blank + true or false + open-ended		
	(Lovreglio et al. 2021)	x	FQ	open-ended		Video
Earthquake	(Feng et al. 2022a)	x	FQ	true or false		Leaflet
	(Lovreglio et al. 2018)	x	DS		annotations by the surveyor of performed actions/decision according to a pre-set list	
Flood	(D'Amico et al. 2022)	x	FQ	open-ended and close – ended (Likert scale)		
Terroris	(Lovreglio et al. 2022)	x	FQ	open-ended		
	(Chittaro and Sioni 2015)	x	FQ	open-ended		

Table 3. Criteria in knowledge assessment in literature

Beyond **participants' demographics** and **knowledge assesssment**, all the studies also refer to the **evaluation of the tool** itself (Table 2), to be carried out only after training. Several indicators might be taken into account. Among them:

- The **self-reported engagement** is related to the interest, involvement, motivation and enjoyment during the experience.
- The **recommendation simplicity** is related to the ease, in terms of time and resources, of applying the suggested actions in real life. It is sometimes combined with the **recommendation efficacy,** which is related to understanding how the suggested actions are beneficial in reducing severe consequences from the threat considered by the simulation.
- The **perceived usefulness** is related to the degree to which a person believes that using a particular system would enhance his/her performance.
- The **perceived ease of use** is defined as the degree to which a person believes that using a particular system would be free of effort.

- The **severity** refers to how serious the individual believes that the threat would be to his/her own life. The **perceived vulnerability** refers to how personally susceptible an individual feels to the threat. The severity and vulnerability are sometimes combined with **fear** that stands between perceptions of severity and vulnerability and the level of the appraised threat.
- The **realism** is related to the plausibility of the virtual environment and the **representation fidelity** to the correspondence between the virtual and the real settings.

Some other testing topics are related to specific applications and/or they are not relevant for the present prototype. For instance:

- The **attention level** (Shiradkar et al. 2021) is assessed by Quantitative electroencephalogram (qEEG) analysis of neural waveform anomalies and, as psychological arousal (Chittaro and Sioni 2015), it is measured through electrodermal activity (EDA) and hearth rate (HR) sensors on the player's hand.
- The **presence** (Rahouti et al. 2021) is the player's sense of being a part of the virtual environment
- The **simulator sickness** (Therón et al. 2020) is a syndrome similar to motion sickness and can be experienced as a side effect during and after exposure to different virtual reality environments.
- The **customization** (Feng et al. 2022a) refers to the capability of changing the game and the storyline according to the player's choices.

All the above-mentioned indicators are scored according to a Likert 5 or 7 scale.

3. Methodology

The testing procedure for the Bes2ecure VR-SG prototype concerns three different interaction modes, namely non-immersive game through desktop, immersive game through VR headsets and non-interactive recording of videogame exemplary sessions, in view of their pairwise comparison. In particular, the non-interactive recording of videogame exemplary sessions acts as "traditional" mode, in line with previous studies (Lovreglio et al. 2021). However, its validation is interesting per se to prove the versatility of the prototype, targeting users with low digital maturity, eventually depending on age, education and social levels and/or fitting demonstration venues with limited technological equipment, in terms of computers and headsets. Thus, the three modes should be assessed according to a common workflow, involving the same questionnaire and comparable target groups as described below.

3.1 Questionnaire

In line with all the previous studies involving testing sessions, the first part of the Bes2ecure questionnaire, to be completed only before the training, is related to demographic information, such as age, gender, education level, previous experiences of training and virtual reality (Tab.4).

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

Tab.4 Questions and answers of the first section - PARTECIPANTS

Thus, the second part, to be completed before and after the training, is related to prior and acquired knowledge. Due to the approach of the prototype, arranged in training modules/items, three questions in the second part are open-ended, where the interviewees are asked to freely list the elements that are risk decreasing for the heat wave, the elements that are risk increasing for the earthquake, and the areas that should be designated as safe places in a post-earthquake scenario. In line with the literature, open-ended are chosen to avoid influencing the response. Moreover, the structure of the questions and the criteria for scoring the answers are adapted from previous studies (Chittaro and Sioni 2015) (Feng et al. 2020a) (D'Amico et al. 2022) (Lovreglio et al. 2022), so that the results are analysed in terms of number of right answers (Tab.5).

Questions/Answers	Scores	Assessment
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; (i) water sources; (iii) 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
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	
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	

Tab.5 Open-ended questions and scores of the second section - KNOWLEDGE

Furthermore, the knowledge assessment also relies on the direct evaluation of the answers during the game. The value can be deduced from a log file, automatically generated after the game session, both non-immersive and immersive, is completed. In this case, the results are analysed in terms of wrong answers (Tab.6). Such supplementary data is considered interesting, in view of assessing whether or not errors are serving an informative function for the learner, pinpointing where his/her knowledge needs improvement, and prompting refinement of his/her mental models (Chittaro and Buttussi 2022).
Furthermore, a customized set of questions for knowledge assessment is added, based on the recognition of the training items in virtual tours of 360° panoramas of real case studies, corresponding to the typological environment. This choice comes from the need to test the incorporation of repeated exercises in expansive contexts that could contribute to the transferring of knowledge and applying of skills to new settings (Engle et al. 2012). To this end, three spherical pictures (Fig. 1) of one of the BeS2ecure case study, Piazza dei Priori in Narni (TR), were edited with hotspots corresponding to several items (Tab.7), so that the trainee is asked to point out the relevant items through closed-ended questions and scored accordingly (Tab.8).

Questions/Answers	Scores	Assessment
How many times the wrong answer was selected during the HW training?	3 points for no wrong answers out of three; 2 points for one wrong answer; 1 point for two wrong answers; 0 points for three wrong answers	Sum of scores
How many time the wrong answer was selected during the E training?	4 points for no wrong answers out of four; 3 points for one wrong answer; 2 points for two wrong answers; 1 point for three wrong answers; 0 points for four (or more) wrong answers	
How many time the wrong answer was selected during the PE training?	4 points for no wrong answers out of four; 3 points for one wrong answer; 2 points for two wrong answers; 1 point for three wrong answers; 0 points for four (or more) wrong answers	

Tab.6 Log file based scores of the second section - KNOWLEDGE

Figure 1 Panoramas of Piazza dei Priori, Narni (TN) with hotspots

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

Tab.7 Hotspots within the virtual tour

It is worth mention that, following the approach of the VR-SG prototype and in order to highlight some correspondences within the multi-hazard vision, some items of the virtual tour are both risk decreasing for one hazard (e.g. buildings casting shadow that mitigates the temperature) and risk increasing for the other one (e.g. buildings collapsing due to seismic shocks).

Questions	Scores	Assessment
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
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	
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	

Tab.8 Closed-ended questions and scores of the second section - KNOWLEDGE

The questionnaire includes a third part related to the assessment of the tool itself, to be completed only after training. In detail, the testing topics for the present application were selected based on the most representative found in the literature (Table 2), including:

- **self-reported engagement**
- **perceived ease of use**
- **realism**

Moreover, although very frequent in previous studies, **self-reported efficacy** was not included and replaced by the less common **recommendation simplicity and efficacy** and **perceived usefulness** that somehow are related to it. This is because these indicators seemed to better assess some specific issues of the prototype, such as the modular arrangement chosen to make the information easy and effective, as well as the integration of multi-hazard data, typological settings and crowd simulation, as it will be discussed right after.

Topic	Questions	Scores	Assessment
Self-reported engagement	The training experience was fun and enjoyable (Shiradkar et al. 2021)	1 (strongly disagreed) – 7 (strongly agreed)	Mean value and standard deviation of the Likert scale scores
	Safety training activities are boring (Lovreglio et al. 2022)		
	I would describe safety training as very interesting (Lovreglio et al. 2022)		
	Safety training does not hold my attention at all (Lovreglio et al. 2022)		
	It was easy for me to concentrate on my learning (Feng et al. 2022a)		
Perceived usefulness	Using this type of virtual reality simulation as an educational tool will enhance my learning (Rahouti et al. 2021)	1 (strongly disagreed) – 7 (strongly agreed)	Mean value and standard deviation of the Likert scale scores
	This type of simulation is useful as a learning supplement (Davis 1989)		
	This type of simulation is as useful as simulation of indoor spaces		
	This type of simulation is useful for behaving properly in real case, too		
	The combination of two risks (heat wave + earthquake) is effective because it simulates real conditions		
	The simulation of the crowd helped me in taking the right decisions		
Perceived ease of use	This simulation tool is rigid and inflexible to interact with (Davis 1989)	1 (strongly disagreed) – 7 (strongly agreed)	Mean value and standard deviation of the Likert scale scores
	I think this training tool is easy to use (Davis 1989)(Rahouti et al. 2021)		
Recommendation simplicity and efficacy	I could easily remember the recommendations provided in the virtual experience (Chittaro and Sioni 2015)(Lovreglio et al. 2021)(Rahouti et al. 2021)	1 (strongly disagreed) – 7 (strongly agreed)	Mean value and standard deviation of the Likert scale scores
	The recommendations provided in the training experience are useful for my safety (Chittaro and Sioni 2015)(Lovreglio et al. 2021)(Rahouti et al. 2021)		
Realism	The built environment was realistic (Feng et al. 2022a)	1 (strongly disagreed) – 7 (strongly agreed)	Mean value and standard deviation of the Likert scale scores
	The built environment reminded me of a familiar place		
	The VR experience was realistic (Feng et al. 2022a)		
	The realism of the virtual world motivates me to learn (Rahouti et al. 2021)		
	The virtual world makes learning more interesting (Dalgarno et al. 2002)		

Tab.9 Likert scale questions and scores of the third section - TOOL

Tab. 9 summarizes the questions and the references that they were adapted from. In detail, beyond the references in Tab.1, some general studies on testing virtual tools were also taken into account (Ai-Lim Lee et al. 2010)(Davis 1989)(Dalgarno et al. 2002).
The questions for these criteria will have answers rated according to a Likert scale ranging from 1 (strongly disagreed) to 7 (strongly agreed), so that the results will be analysed in terms of mean value and standard deviation (Feng et al. 2022a)(Lovreglio et al. 2018)(Lovreglio et al. 2018).

As anticipated, it is worth mention that some questions were specifically defined for the present application, in order to test some peculiar contents of the VR-SG prototype, particularly in terms of perceived usefulness. In fact, the questionnaire includes opinions on the meaningfulness of the simulation, as set in an outdoor space (*This type of simulation is as useful as simulation of indoor spaces*), referred to a typological context representative of real case studies *(This type of simulation is useful for behaving properly in real case, too),* based on a multi-hazard approach *(The combination of two risks (heat wave + earthquake) is effective because it simulates real conditions),* including the agent-based crowd simulation (*The simulation of the crowd helped me in taking the right decisions).* The above-mentioned aspects are also the most innovative compared to the state-of-the-art as thoroughly discussed in D.6.1.1.

3.2 Target groups

The Bes2ecure VR-SG prototype is meant to boost urban community resilience to multi-hazard scenarios in open spaces. Thus, it foresees wide scale application for large player populations. For this reasons, a comprehensive set of target groups covering different ages, such 18-35 years old, 36-49 years old, 50-60 years old, is foreseen corresponding to representative categories of the Italian population with comparable number of citizens (about 10 million people), based on the available statistical data (ISTAT 2023). This approach is different from all the studies in Table 1-2-3, in which the age distribution derives from the main selection of the recipient type (e.g. hospital staff, university students, museum visitors). However, it seems more appropriate for the presented prototype in view of understanding whether and to what extent it fits the general public of urban citizens. For the same reasons, the trainees within a target group should test all the foreseen training methods (i.e. non-interactive visualization of video-recording, non-immersive game on computer/tablet, and immersive game with VR headset).
In order to identify a suitable sample size, a power analysis with the statistical software G*Power v.3.1.9.7 was run. In particular, in order to enable a pairwise comparison between training modes, the sample size should be at least 53 persons for each training mode, considering a medium effect size (d= 0.5), as estimated by (Cohen 1988), a first type error (α=0.05) and a power of 80%, as proposed by (Feng et al. 2020b). This setting would allow a Mann Whitney U test, which was chosen because the knowledge, as well as all the categories used for assessing the tool, are not normally distributed. Mann–Whitney u-test is suitable to assess whether there are statistically significant differences between two independent modes and at different test steps, i.e. knowledge before and after training for the same training mode (Rahouti et al. 2021)(Lovreglio et al. 2021)(Feng et al. 2020b).
Thus, the following scheme of testing administration is foreseen, with 60 questionnaires for each training mode and 60 questionnaires for each training groups, so that 20 people of each target group should test one of the three testing modes. Consequently 180 questionnaires are the minimum required standard for the testing phase (Fig.2).

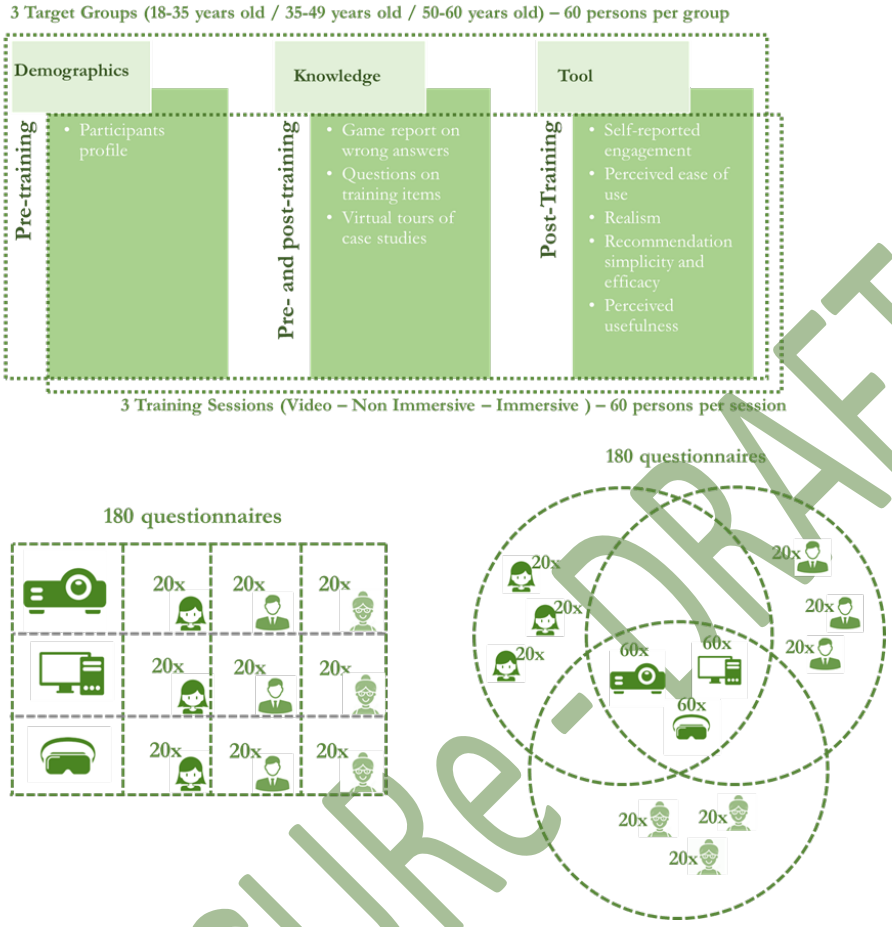

Figure 2. Testing scheme

3.3. Virtual demo-lab set-up

In order to set-up the virtual demo-lab sessions, a Google form was prepared in Italian and English (see Annex 1-2). Within the Google form, the virtual tour of Piazza dei Priori in Narni (TR), has been integrated in order to facilitate the answers of the closed-ended questions related to the hotspots (Figure 3).

Moreover, in order to avoid the potential risks of privacy violation (Skulmowski 2023), a consent form, by which the users give their consent to take the trial and for their data to be collected anonymously, has been prepared in Italian and English, submitted and approved by the Ethical Committee of the Polytechnic of Bari (see Annex 3-4).

The invitation to take part to the testing phase has been carried out by emails to the staff of the Department DICATECh (Polytechnic of Bari), asking to extend it to their relatives, in order to have a wide age distribution. The selection of the training mode is chosen by the user itself, after presenting the three types of media/environments. In particular, the potential risks of reduction of autonomy and health problems (Skulmowski 2023) are clearly stressed in the invitation, so that the trainees should have the option to lower the level of immersivity by selecting the non-immersive and non-interactive modes.

Moreover, specific testing sessions were scheduled in the Bachelor and Master Courses in Building Engineering of the Polytechnic of Bari, as well as in the city of Narni (TR), within a seminar in the Post-graduate Course in "Architectural design for the recovery of historic buildings and public spaces" of the University La Sapienza of Rome.

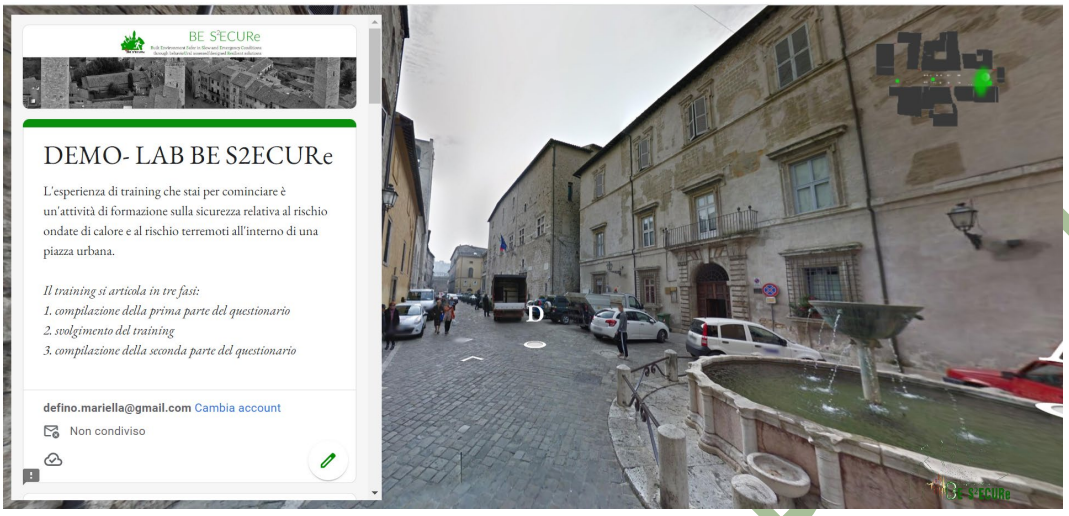

Figure 3. Virtual tour integrated in the Google form

4. Conclusions and remarks

The virtual demo-lab, including the development of the feedback questionnaires, the identification of the target groups and the arrangement of documents and facilities for running the testing sessions, is meant to collect relevant insights on the effectiveness and usability of the VR-SG Bes2ecure prototype. In particular, the assessment concerns the knowledge improvements on heat wave protection and earthquake response, as well as the engagement, ease of use, usefulness, simplicity, efficacy and realism of the proposed tools. Covering different interaction modes (i-non-interactive video recording; ii-desktop non-immersive; iii-headset immersive) and different users (i-18-35 years old; ii-36-50 years old; iii-51-60 years old) should enable a cross understanding on which modes are potentially more suitable for certain target groups, also based on their familiarity with training and virtual reality experiences. Nonetheless, the questionnaire has been developed to assess some specific and innovative aspects of the prototype, as set in an outdoor space, referred to a typological context representative of real case studies, based on a multi-hazard approach and including agent-based crowd simulations. Finally, specific attention will be given to the exercises in expansive contexts through the navigation of the virtual tour of panoramic scenes for a real case study, that could contribute to the transferring of knowledge and applying of skills to new settings. In fact, if successful in the testing phase, virtual tours of real case studies might be considered for integration in the training phase itself. To this end, the demonstration on the typological environment could be followed by visualization of photorealistic 360° images and videos, in order to assess whether or not the player is able to recognize similar danger sources and safe conditions, when the representation fidelity and the complexity of the scene are enhanced. This would also avoid using more complex three-dimensional modelling of the actual open space, in view of large scale application(Chittaro and Buttussi 2022) (Feng et al. 2022b)(Anwar et al. 2018).

References

Ai-Lim Lee E, Wong KW, Fung CC (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

Anwar MS, Wang J, Ullah A, et al (2018) User Profile Analysis for Enhancing QoE of 360 Panoramic Video in Virtual Reality Environment. Proc - 8th Int Conf Virtual Real Vis ICVRV 2018 106–111. https://doi.org/10.1109/ICVRV.2018.00030

Burbules NC (2004) Rethinking the Virtual. E-Learning Digit Media 1:162–183. https://doi.org/10.2304/elea.2004.1.2.2

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

Cha M, Han S, Lee J, Choi B (2012) A virtual reality based fire training simulator integrated with fire dynamics data. Fire Saf J 50:12–24. https://doi.org/10.1016/j.firesaf.2012.01.004

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

Cohen J (1988) Statistical Power Analysis for the Behavioral Sciences Second Edition

D'Amico A, Bernardini G, Lovreglio R, Quagliarini E (2022) A Non-Immersive Virtual Reality Serious Game Application for Flood Safety Training. SSRN Electron J. https://doi.org/10.2139/ssrn.4110990

Dalgarno B, Hedberg J, Harper B (2002) the Contribution of 3D Environments To Conceptual Understanding. Learning 1:44–54

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

Diez H V., Garcia S, Mujika A, et al (2016) Virtual training of fire wardens through immersive 3D Environments. Proc 21st Int Conf Web3D Technol Web3D 2016 43–50. https://doi.org/10.1145/2945292.2945296

Engle RA, Lam DP, Meyer XS, Nix SE (2012) How Does Expansive Framing Promote Transfer? Several Proposed Explanations and a Research Agenda for Investigating Them. Educ Psychol 47:215–231. https://doi.org/10.1080/00461520.2012.695678

Feng Z, González VA, Amor R, et al (2020a) An immersive virtual reality serious game to enhance earthquake behavioral responses and post-earthquake evacuation preparedness in buildings. Adv Eng Informatics 45:. https://doi.org/10.1016/j.aei.2020.101118

Feng Z, González VA, Mutch C, et al (2022a) Exploring spiral narratives with immediate feedback in immersive virtual reality serious games for earthquake emergency training. Multimed Tools Appl.

https://doi.org/10.1007/s11042-022-13306-z

Feng Z, González VA, Mutch C, et al (2020b) Towards a customizable immersive virtual reality serious game for earthquake emergency training. Adv Eng Informatics 46:. https://doi.org/10.1016/j.aei.2020.101134

Feng Z, Liu C, Gonzalez VA, et al (2022b) Prototyping an immersive virtual reality training system for urban-scale evacuation using 360-degree panoramas. IOP Conf Ser Earth Environ Sci 1101:. https://doi.org/10.1088/1755-1315/1101/2/022037

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

Gigante MA (1993) 1 - Virtual Reality: Definitions, History and Applications. In: Earnshaw RA, Gigante MA, Jones HBT-VRS (eds). Academic Press, Boston, pp 3–14

Gong X, Liu Y, Jiao Y, et al (2015) A novel earthquake education system based on virtual reality. IEICE Trans Inf Syst E98D:2242–2249. https://doi.org/10.1587/transinf.2015EDP7165

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

ISTAT (2023) Population and families (in Italian). In: Popul. Fam. https://www.istat.it/it/popolazione-e-famiglie?dati

Klingenberg S, Fischer R, Zettler I, Makransky G (2023) Facilitating learning in immersive virtual reality: Segmentation, summarizing, both or none? J Comput Assist Learn 39:218–230. https://doi.org/10.1111/jcal.12741

Li C, Liang W, Quigley C, et al (2017) Earthquake Safety Training through Virtual Drills. IEEE Trans Vis Comput Graph 23:1275–1284. https://doi.org/10.1109/TVCG.2017.2656958

Lorenz M, Brade J, Klimant P, et al (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, Daphney-Chloe N, Rahouti A, et al (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

Lovreglio R, Duan X, Rahouti A, et al (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, Gonzalez V, Feng Z, et al (2018) Prototyping virtual reality serious games for building earthquake preparedness: The Auckland City Hospital case study. Adv Eng Informatics 38:670–682. https://doi.org/10.1016/j.aei.2018.08.018

Macchione F, Costabile P, Costanzo C, De Santis R (2019) Moving to 3-D flood hazard maps for enhancing

risk communication. Environ Model Softw 111:510–522. https://doi.org/10.1016/j.envsoft.2018.11.005

Makransky, G., & Petersen GB (2019) Investigating the Process of Learning with Desktop Virtual Reality: A Structural Equation Modeling Approach. Comput Educ 134:15–30. https://doi.org/https://doi.org/10.1016/j.compedu.2019.02.002

Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn Instr 60:225–236. https://doi.org/https://doi.org/10.1016/j.learninstruc.2017.12.007

Moreno A, Posada J, Segura Á, et al (2014) Interactive fire spread simulations with extinguishment support for Virtual Reality training tools. Fire Saf J 64:48–60. https://doi.org/10.1016/j.firesaf.2014.01.005

Oliva D, Somerkoski B, Tarkkanen K, et al (2019) Virtual reality as a communication tool for fire safety – Experiences from the VirPa project. CEUR Workshop Proc 2359:241–252

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

Sacfung A, Sookhanaphibarn K, Choensawat W (2014) Serious game for fire safety evacuation plan. Adv Mater Res 931–932:583–587. https://doi.org/10.4028/www.scientific.net/AMR.931-932.583

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

Skulmowski A (2023) Computers & Education : X Reality Ethical issues of educational virtual reality. Comput Educ X Real 2:100023. https://doi.org/10.1016/j.cexr.2023.100023

Skulmowski A, Nebel S, Remmele M, Rey GD (2022) Is a Preference for Realism Really Naive After All? A Cognitive Model of Learning with Realistic Visualizations. Educ Psychol Rev 34:649–675. https://doi.org/10.1007/s10648-021-09638-1

Skulmowski A, Rey GD (2021) Realism as a retrieval cue: Evidence for concreteness-specific effects of realistic, schematic, and verbal components of visualizations on learning and testing. Hum Behav Emerg Technol 3:283–295. https://doi.org/10.1002/hbe2.209

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

Sukirman, Wibisono RA, Sujalwo (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

Therón R, Coelho AF, Tan B, et al (2020) The Evaluation of Virtual Reality Fire Extinguisher Training. Front Psychol. https://doi.org/10.3389/fpsyg.2020.593466

Yang Y, Xu Z, Wu Y, et al (2021) Virtual fire evacuation drills through a web-based serious game. Appl Sci 11:. https://doi.org/10.3390/app112311284