NOTICE: This is a paper that was accepted for publication in the **IEEE Transactions on Learning Technologies**. It is not the final published version. The DOI of the definitive version available in IEEE Explore is https://doi.org/10.1109/TLT.2020.3033766

A Comparison of Procedural Safety Training in Three Conditions: Virtual Reality Headset, Smartphone, and Printed Materials

Fabio Buttussi and Luca Chittaro

***Abstract*—Virtual reality (VR) experiences are receiving increasing attention in education and training. Some VR setups can deliver immersive VR training (e.g., on multiple projected screens), while others can deliver non-immersive VR training (e.g., on standard desktop monitors). Recent, consumer VR headsets make it possible to deliver immersive VR training with six-degrees-of-freedom tracking of trainees' head as well as hand controllers, while most smartphones can deliver non-immersive VR training without the need for additional hardware. Previous studies compared immersive and non-immersive VR setups for training, highlighting effects on performance, learning, presence and engagement, but no study focused on contrasting procedural training with (immersive) VR headsets and (non-immersive) smartphones. This paper conducts a comparison of these two VR setups in the aviation safety domain. The considered training concerned door opening procedures in different aircraft, and included a virtual instructor. In addition, we compared the two VR setups with the traditional printed materials used in the considered domain, i.e., safety cards. Results show that both VR setups allowed gaining and retaining more procedural knowledge than printed materials, and led to higher confidence in performing procedures. However, only the VR headset was considered to be significantly more usable than the printed materials, and presence was higher with the VR headset than the smartphone. The VR headset turned out to be important also for engagement and satisfaction, which were higher with the VR headset than both the printed materials and the smartphone. We discuss the implications of these results.**

***Index Terms*—Virtual reality, procedural training, virtual instructor, immersive VR, non-immersive VR, VR headset, smartphone, mobile devices, user study, aviation safety.**

I. Introduction

VIRTUAL reality (VR) has been employed in education and training for decades and its effectiveness and current limitations in different domains are described by several reviews [1]–[5]. Some VR setups can make trainees feel surrounded by the virtual environment (VE), delivering *immersive* VR training (e.g., a VR headset or multiple projected screens). Other VR setups display the VE on a standard screen (e.g., using a desktop monitor or a smartphone), delivering *non-immersive* VR training [6]. Recent, consumer VR headsets can deliver immersive VR training with six-degrees-of-freedom (6-DOF) tracking of users' head as well as of hand controllers to recognize trainees' actions. However, most smartphones can deliver non-immersive VR training (statistics about mobile graphics support can be found in [7]) without the need for additional hardware and with the advantage of using only a familiar device [6]. Moreover, about 80% of the population in developed countries owns a smartphone [8].

To help make informed choices about training effectiveness of these two VR setups (VR headsets and smartphones) thorough comparisons are needed. Such comparisons should not be limited to the main outcomes of training such as knowledge gain or performance in the considered task, but extend to psychological constructs. For example, Domagk et al. [9] identified three types of engagement that can play a role in training, i.e., cognitive, behavioral, and affective, and all three may be influenced by the VR setup. Plass et al. [10] identified several features of the training experience that could affect cognitive engagement, and some of them (e.g., situatedness, gestures and movement) can vary with different VR setups. In [10], behavioral engagement is linked with motivation and the three questions that shape it, identified by Eccles et al. [11]: "Can I do this?", "Do I want to do this, and why?", and "What do I need to do in order to succeed?". Plass et al. [10] highlighted that affirmative answers to the first question can be fostered by allowing trainees to fail gracefully and by providing them with adaptive help so that they can try again and learn from their mistakes. Such features can be offered through any VR setup. However, making gestures with the tracked controllers of a VR headset or interacting with the touchscreen of a smartphone might have different effects on trainees' understanding of the actions needed to carry out a procedure or their confidence in them. Finally, VR setups might play a role on affective engagement because VR setups may affect presence (i.e., the sense of being in the VE [12]), and presence can play a role on emotions [13]. Previous studies conducted comparisons between different VR setups for education and

Manuscript received September 13, 2019; revised August 20, 2020 and October 12, 2020; accepted October 22, 2020. Date of publication TBD, date of current version October 22, 2020. This work was supported in part by a grant from the U.S. Federal Aviation Administration (*Corresponding author: Fabio Buttussi.*)

The authors are with the Human-Computer Interaction Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Udine 33100, Italy (e-mail: fabio.buttussi@uniud.it; luca.chittaro@uniud.it).

training showing the impact of the VR setup on performance, knowledge, presence, and engagement [14]–[31]. In particular, some studies compared immersive VR setups, including VR headsets and multiple projected screens (also known as Cave Automatic Virtual Environments or CAVEs), with non-immersive VR setups that display the VE on a standard desktop monitor [14]–[16], [29], [31] or a tablet [28]. However, no study focused on comparing VR headsets and smartphones in *procedural training*, i.e., training that concerns the demonstration and execution of the steps of a procedure.

In this paper, we focus on procedural training in the aviation safety domain, which is representative of many other safety training domains, as described in [29]. The real-world case considered in the study concerns the procedures for opening the overwing exits of different aircraft types in the event of an emergency evacuation. Before takeoff, flight attendants have a few minutes to describe these procedures to the passengers seated on the exit rows. All passengers (including those seated in the exit rows) are invited to look at the safety cards, i.e., printed materials available in the seat pocket with pictorials that illustrate procedures for the specific aircraft. Unfortunately, most passengers do not to pay attention to safety cards [32], and even those who pay attention show an unacceptable level of comprehension of the pictorials [33]. This calls for innovative solutions [32], which should be more engaging and effective than safety cards. Therefore, this paper proposes a VE with a three-dimensional (3D) virtual flight attendant that plays the role of instructor, demonstrates each procedure, and invites trainees to perform it. As trainees try to perform the procedure, the virtual instructor provides feedback after each step, and at the end of the procedure. We developed the VE with the virtual instructor for both the VR headset and the smartphone setups. This paper aims at advancing knowledge in different directions. Our study is the first to compare a VR headset and a smartphone with the aim of exploring which of these VR setups might be more suited for procedural safety training. The study considers a variety of measures including presence, knowledge gain, and confidence in performing procedures. The study also explores the effects of the VR setups on engagement, satisfaction, usability, and retention, which have been previously considered only in a few studies [27]–[29], [31]. In particular, most studies that compared engagement with different VR setups [28], [29] employed serious games [34], i.e., video games that use entertainment for a serious purpose, while our study is the first to compare engagement with two VR setups in training with a virtual instructor. Finally, our study includes a third condition to compare the results of the two VR setups with the traditional printed materials (safety cards) used to instruct about the same procedures, with the aim of assessing if one or both VR setups can overcome the well-known limitations of safety cards in terms of engagement [32] and knowledge gain [33].

The paper is organized as follows. Section 2 introduces previous studies that compared different VR setups in education and training. Section 3 describes the VE with the virtual instructor. Section 4 describes the study, while Section 5 and 6 respectively report and discuss its results. Finally, Section 7 concludes the paper outlining future work.

II. Related Work

Studies comparing different VR setups (Table I) addressed several education and training domains, e.g., natural science [15], [18], [28], [31], history [21], mathematics [23], and medicine [17], [22], [25], [27], [30]. As reported in Table I, the tasks performed by participants varied considerably, and included attending a seminar or a lesson [14], [21], [31] exploring complex 3D structures or datasets [18], [22]–[25], and interacting with different types of simulations [15]–[17], [29], [31]. Different VR setups for procedural training where compared in studies where participants trained in: moving objects from a location to another following a given sequence of actions [19], [20], a medical procedure [27], [30], a visual scanning strategy [26], an evacuation procedure from an aircraft during a specific emergency [29], and biology lab procedures [31]. While in a few studies participants had to interact with characters [17], [29], [31], only one study [31] involved a virtual instructor, which can play an important pedagogical role in a VE as discussed in [35]–[37].

Table I lists the different conditions in the studies, highlighting the features within a VR setup or the types of VR setups they compared. For example, some studies focused on specific visualization features of VR setups, such as field of view (FOV) [20], [26], field of regard (FOR, i.e., the total size of the visual field surrounding the user [38], which depends on FOV and on the possibility to rotate the head) [20], [22], [24], [25], and/or stereoscopy [22], [24], [25], typically enabling or disabling different features on the same VR setup. Some studies also considered interaction features, e.g. enabling or disabling 6-DOF head tracking [22], [24], [25] or room-scale tracking (i.e., the possibility to physically move in a room to move in the VE, instead of being physically seated and move in the VE using a controller) [27]. Other studies considered instead different types of VR setups, typically comparing immersive ones based on VR headsets [14]–[17], [28], [29], [31], CAVEs [18], [23], or two projection displays [19] with non-immersive ones based on a desktop monitor [14]–[16], [29], [31], a laptop monitor [19], a single projected screen [17], [18], [23], or a tablet [28]. In these studies, the different VR setups offered different visualization features, and in some cases different interaction features, e.g., providing different DOF values of head-tracking ranging from 0 to 6 [16], [17], [23], [29], and/or using different input devices such as gamepad or keyboard [16]. A study also compared two VR setups based on a tablet, using touchscreen interaction or a physical endovascular tool [30]. Some studies that compared different VR setups included in the comparison also a traditional learning method, e.g., attending a seminar in a real classroom [14], listening to an audio recording [14], or watching a video podcast [30]. Considering assessed measures, knowledge gain or performance (time or correctness in completing the given task) were included in all studies except [28]. Presence was assessed in [14], [15], [17], [29], [31], while participants' confidence in their answers or performance, self-efficacy, or other self-evaluation ratings were included in [14], [17], [22], [25], [29]–[31]. Participants' engagement and/or satisfaction were measured in [27]–[29], [31], while usability only in [27]. Retention was assessed only in [29].

VR setups offering more immersive visualization features had a positive effect on knowledge gain in [18] and in one of the two topics considered in [15]. Similarly, VR setups offering more immersive visualization features and, in some cases, more natural interaction features, had a positive effect on performance in [16] (but only for one of the two performance measures), [19] (but only for complex procedures), [20], [22]–[24], [26]. On the contrary, no difference in knowledge gain was found in [14], [27], [29], and in one of the two topics considered in [15]. No effect of the VR setup was found on one of the two performance measures in [16], and on performance measures in [19] for simple procedures, while the effect in [30] was limited to two steps of the procedure. In [31], a VR headset led to less knowledge gain than a desktop monitor. No difference in knowledge retention was found in [29]. Linear regression analysis showed that presence predicted knowledge gain in [15], but differences in presence did not correspond to differences in knowledge gain in [14] and [31]. In [29], presence as well as engagement were higher with the VR setup that provided better visualization and interaction features. Engagement (only in a subscale related to immersion) and satisfaction (only in a subscale related to overall reactions) were higher with room-scale tracking enabled than disabled [27]. Higher engagement with a VR headset than a tablet was claimed in [28], but without reporting statistical significance. Self-efficacy or participants' self-evaluation ratings about some aspects of their performance or learning were higher with VR setups offering more immersive visualization and/or more natural interaction features in [17] and [30], but not in [29] and [31]. Participants' confidence in their answers was lower with a VR headset than a desktop monitor in [14].

Overall, the effects of VR setups in education and training are not clear, because results vary with different tasks, procedures or topics, as also shown in [25], where different combinations of the considered features of the VR setup led to different results for different tasks. This calls for further studies comparing different VR setups on different types of training. In particular, more studies are needed to assess the effects of VR setups on procedural training, because existing studies about procedural training led to the above-described different outcomes. Such new studies of VR setups should include knowledge retention, a fundamental measure of the effectiveness of training over time, which has been considered by only one study [29].

TABLE I
PREVIOUS STUDIES ABOUT THE EFFECTS OF DIFFERENT VR SETUPS IN EDUCATION AND TRAINING DOMAINS

Reference	Task	Conditions	Measures	Main Findings
Mania & Chalmers (2001) [14]	Attend a 15-minutes seminar on a non-science topic	Seminar presentation: real classroom / 3D virtual classroom on 21" desktop monitor / 3D virtual classroom on VR headset without head-tracking / audio only	Presence, knowledge gain, participants' confidence in their answers	Higher presence only in real classroom; some differences in presence did not correspond to differences in knowledge gain; confidence lower for VR headset than desktop monitor
Winn et al. (2002) [15]	Interact with a simulation about water movement and salinity	VR setup: VR headset / desktop monitor	Presence, knowledge gain	Presence predicted knowledge gain; higher knowledge gain about water movement with VR headset; no effect about salinity
Aoki et al. (2008) [16]	Point at targets and navigate in a simulated space station	VR setup: 3-DOF head-tracked VR headset with gamepad / 17" desktop monitor with gamepad / 17" desktop monitor with keyboard and integrated touchpad	Time to point from destination to start and from start to different destinations, egress time	Lower time to point with VR headset; no effect on egress time
Johnsen & Lok (2008) [17]	Interact with two virtual patients to train in communication skills	VR setup: 6-DOF head-tracked VR headset/ large-screen projection display	Quality of the simulated patient, presence, self-evaluation ratings, performance grades given by an expert	Higher self-evaluation ratings about use of empathy with VR headset; correlation between performance grades and self-evaluation ratings about use of empathy only with large-screen projected display
Limniou et al. (2008) [18]	Study molecule structure and chemical reactions	VR setup: 4-screen CAVE / single projected screen	Knowledge gain	Higher knowledge gain with CAVE
Sowndara-rajan et al. (2008) [19]	Train in a simple and a complex procedure (moving virtual objects from a location to another)	VR setup: laptop monitor / 2-screen projection display	Procedure completion time, number of errors	Less time and less errors with 2-screen projection display on complex procedure; no effect on simple procedure
Ragan et al. (2010) [20]	Train in a procedure (moving virtual objects from a location to another), recall procedure in the assessment environment	Software FOV: matched / unmatched; FOV: 60°/ 180°; FOR: 1 screen / 3 screens; assessment environment: virtual / physical	Time to complete procedure in assessment environment, number of errors in assessment environment	No effect of assessment environment; less time and errors for matched software FOV, high FOV, and high FOR
Fassbender et al. (2012) [21]	Attend a virtual history lesson	VR setup: 3-monitor setting / larger curved screen; background music: yes / no	Knowledge gain	Overall higher knowledge gain with 3-monitor setting; interaction between VR setup and background music; higher knowledge gain with larger curved screen if there was music (only in the second half of the lesson)

TABLE I
(CONTINUED)

Reference	Tasks	Conditions	Measures	Main Findings
Laha et al. (2012) [22]	Analyze volume data visualizations of tomography datasets (search and description tasks of different complexity)	Head tracking: 6-DOF on / off; stereoscopy: on / off; FOR: 4 screens / 1 screen	Performance (task correctness, time to complete tasks), subjective measures (perceived difficulty, confidence in task correctness)	Overall better performance with 6-DOF on, stereoscopy on, and 4 screens; different effects of different combinations of features for different tasks
Bacim et al. (2013) [23]	Search for intersections, follow paths, identify connections, compare lengths in mathematical graphs	VR setup: 4-screen CAVE with stereoscopy and 6-DOF head tracking / one screen with no stereoscopy and no head tracking	Performance (task correctness, time to complete task)	Better performance using 4-screen CAVE with stereoscopy and 6-DOF head tracking
Ragan et al. (2013) [24]	Count interconnecting tubes in virtual underground cave structures	Head tracking: 6-DOF on / off; stereoscopy: on / off; FOR: 1 screen / 4 screens	Number of errors, time to complete task	Less errors with higher FOR or head tracking on; less time with stereoscopy and head tracking on
Laha et al. (2014) [25]	Search for features, recognize patterns, judge positions and orientations, estimate properties, describe shapes in volume data visualizations of tomography scans	Head tracking: 6-DOF on / off; stereoscopy: on / off; FOR: 1 screen / 4 screens	Performance (task correctness, time to complete tasks), subjective measures (perceived difficulty, confidence in task correctness)	Different effects of different combinations of features for different tasks
Ragan et al. (2015) [26]	Search for targets in virtual city streets following a given scanning strategy	FOV: 35.81° / 80.44° / 120.41°; scenario complexity: three levels with different number of objects and different detail of textures	Detection performance (percentage of correct identifications, percentage of false-positive identifications) in training and in assessment, adherence to given scanning strategy (scores given by three raters following given rules)	Better detection performance in training for higher FOV; no effect in assessment
Shewaga et al. (2017) [27]	Train in the preparation for performing an epidural procedure	Room-scale tracking: on / off	Knowledge, satisfaction, usability, engagement	Higher satisfaction (only for overall reaction subscale) and higher engagement (only for immersion subscale) with room-scale tracking
Silva et al. (2017) [28]	Learn about human cell biology by playing a game	VR setup: VR headset / tablet	Engagement	Results suggest higher engagement with the VR headset, but no statistical analysis was carried out
Buttussi & Chittaro (2018) [29]	Learn how to evacuate from an aircraft by playing a game	VR setup: desktop monitor / VR headset with narrow FOV and 3-DOF tracking / VR headset with wide FOV and 6-DOF tracking	Knowledge gain, retention, self-efficacy, engagement, presence	Higher engagement and presence with VR headset with wide FOV and 6-DOF tracking; no effect on knowledge gain, retention, and self-efficacy
Aeckersberg et al. (2019) [30]	Train in an endovascular procedure	Training method: conventional learning through a video podcast / VR setup with tablet visualization and touchscreen interaction / VR setup with tablet visualization and physical endovascular tool interaction	Skill performance for each step of the procedure, interest in taught topics, self-efficacy	Increased interest in taught topics when using tablet visualization and physical endovascular tool interaction; higher self-efficacy with tablet visualization and physical endovascular tool interaction rather than the other methods; differences in skill performance limited to two steps
Makransky et al. (2019) [31]	Attend a virtual biology lesson including a lab simulation with a virtual instructor	VR setup: VR headset / desktop monitor	Knowledge gain, presence, self-evaluation ratings about learning, satisfaction, cognitive load	Higher presence and cognitive load with VR headset, higher knowledge gain with desktop monitor

III. Training with the Virtual Instructor

The VE used in the study is meant for trainees who want to learn some aircraft emergency procedures that can be performed by flight attendants as well as passengers. More specifically, the VE teaches trainees how to open six different types of overwing exits in the event of an emergency evacuation, which can be initiated after an accident (e.g., a water landing) or incident (e.g., a fault causing smoke in the cabin). The six exits represent those in use in the following aircraft types: Airbus A320, Boeing 737-800, Bombardier CRJ200, Bombardier CRJ900, Embraer ERJ145, and Embraer E190. The position of the trainees in the VE is on the aircraft seat closest to the considered exit (Fig. 1–3). Since flight attendants are responsible of instructing passengers seated on the exit rows of commercial flights about how to open overwing exits, we have dressed the virtual instructor in our VE in a flight attendant uniform. In a previous study that compared different VR setups for aviation safety training [29], we presented trainees with a VR reproduction of a real emergency scenario. While such approach was effective for learning [29], it might be scary for some trainees in the general public, so the present

Fig. 1. Demonstration phase. The virtual instructor is demonstrating the procedure on an overwing exit. Trainees are sitting in front of the trainer and there is a second overwing exit of the same type on their right.

Fig. 2. Practice phase. Trainees perform the procedure on the overwing exit on their right. This screenshot is taken from the VR headset version in which trainees used tracked controllers.

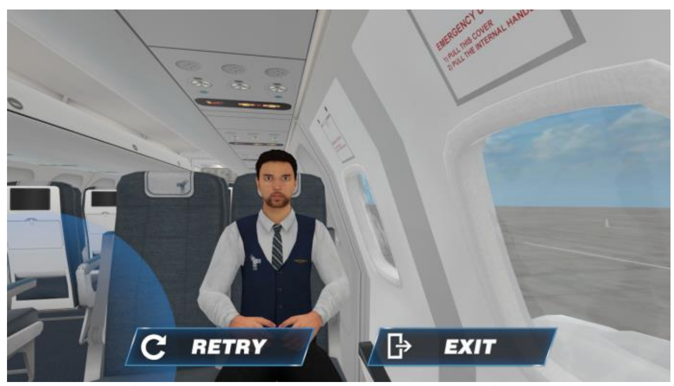

Fig. 3. Final feedback phase. This screenshot is taken from the smartphone version: trainees slide a finger on the semi-transparent blue area at the bottom left to rotate the camera, and touch the buttons to select options. The user interface used the participants' native language (Italian). The figure translates it into English for reader's convenience.

study evaluates a different approach. More precisely, the VE in this study reproduces a safe and calm training scenario, where the virtual instructor is on a seat placed in front of the trainee (Fig. 1–3), demonstrates the procedure on an overwing exit, and provides trainees with feedback as described in the following.

The virtual instructor is able to teach six lessons, one for each overwing exit type. Each lesson is organized in four phases:

1) *Introduction*. The virtual instructor tells the trainees to which aircraft type the overwing exit belongs, and that they should be able to open it in case of evacuation.
2) *Demonstration*. The virtual instructor verbally describes the sequence of steps to open the overwing exit as it demonstrates them on the exit on its left (Fig. 1). For five of the six aircraft types, the procedure is made of six steps. Although the details vary with the aircraft type, the six steps follow this sequence: i) check the external conditions by looking through the exit window to be sure there is no fire or other danger outside, ii) remove the cover that protects the top handle (the position and type of the top handle as well as the type of the cover varies with the aircraft), iii) firmly hold the bottom handle (the position and type of the bottom handle varies with the aircraft), iv) push or pull the top handle to unlock the exit (the push or pull direction varies with the aircraft), v) lift and pull the hatch inside the cabin to open the exit, vi) toss the hatch outside the aircraft. For example, in the case of the Airbus A320, shown in Fig. 1, the top handle must be pulled down to unlock the exit. Only in the case of the Boeing 737-800, the procedure is made of three steps: i) check the external conditions by looking through the exit window to be sure there is no fire or other danger outside, ii) remove the cover that protects the handle near the top of the exit, iii) pull down the handle so that the exit automatically opens. At the end of the demonstration, trainees can choose to watch it again or proceed to the practice phase.
3) *Practice*. The virtual instructor invites trainees to perform the steps of the procedure on the overwing exit at their right (Fig. 2), and monitors their actions (movements of tracked controllers in the VR headset version or interaction with the touchscreen in the smartphone version). The training experience focuses on procedural knowledge rather than mastering motor skills (which would require using a physical door), so the virtual instructor evaluates the correctness of each step considering if trainees interacted with the correct part of the exit and remembered the correct direction of movement. Following Plass et al. [10], the virtual trainer supports trainees' motivation by allowing them to fail gracefully and by providing them with adaptive help, so that they can try again and learn from their mistakes. More precisely, after each attempt to complete a step, the virtual instructor provides trainees with feedback adapted to the number of attempts for that step. If trainees complete a step correctly, the virtual instructor provides them with positive feedback (e.g., saying "Good!"). If trainees fail at their first attempt, the virtual instructor says only that they are not interacting with the current door part correctly (e.g., "you are not acting on the handle as you should"), so that they can try again to perform the correct action on their own. If trainees fail at the second attempt, the virtual trainer recalls the correct action to perform (e.g., "you should unlock the exit by pulling the top handle"). If trainees fail at subsequent attempts, the virtual trainer describes in detail the action to perform (e.g., "pull the top handle towards the bottom to unlock the exit"). In addition, if the trainees remain inactive for 30 seconds after the beginning of a new step, the virtual trainer explains them how to interact with the VE in that step (e.g., "drag the handle in the correct direction") without providing hints on the specific exit part or direction of movement.

4) *Final feedback*. Once the practice phase concludes, the virtual instructor gives a verbal feedback about the entire phase. If trainees have made errors, the virtual instructor tells them that they have made some errors and suggests them to practice more. If trainees have made no errors, the final feedback takes into account how fast they performed the procedure, because in some aircraft evacuations the cabin can become quickly unsurvivable due to smoke, fire or water. For each procedure, we set a time threshold considering the time required to perform all the steps correctly and without hesitations. If trainees have taken more time than the time threshold for the considered procedure, the virtual trainer tells them that they should open the exit faster and invites them to practice more. If trainees have completed the practice phase correctly and within the time threshold, the virtual trainer tells them that they successfully completed practice. Regardless of the outcome, trainees can choose to try the procedure again or conclude the lesson (Fig. 3).

The spoken sentences were recorded in the native language of the study participants (Italian). The English translations in this paper are provided for reader's convenience. We built two versions of the VE interface, one for the VR headset and one for the smartphone. The VE was thus the same except for the way users interacted with it. More precisely, in the VR headset version, trainees control the viewpoint by moving their head (the VR headset had 6-DOF head tracking), while in the smartphone version, trainees rotate the viewpoint by sliding their finger inside a semi-transparent blue area at the bottom left of the interface (Fig. 3). Considering the practice phase, in the VR headset version, trainees perform the required actions by moving their head in the first step and the tracked controllers in all the other steps (Fig. 2). Grasping (e.g., to hold or drag a handle) is activated by pressing the trigger button on the controller. In the smartphone version, trainees perform the actions by touching the object on the screen (e.g., to hold the handle) or sliding their finger on it (e.g., to remove the cover). Finally, when trainees need to select among options, in the VR headset version, they can press virtual buttons on the controller trackpad, while in the smartphone version they can touch buttons on the screen (Fig. 3). The virtual trainer behaves identically in the two versions of the VE, and uses the same sentences except for those that explain how to interact with the VE (e.g., "move the head to check for danger" in the VR headset version and "touch the part that allows you to check for danger" in the smartphone version). To make trainees familiar with the interaction before using the VE with the virtual instructor, we added an initial tutorial VE (see Section IV D).

IV. User Study

We conducted a between-groups study with three conditions: the VR Headset (VH) group tried the VE using a VR headset and its tracked hand controllers (immersive VR), the Smartphone (SP) group tried the VE using a smartphone and its touchscreen (non-immersive VR), and the Printed Materials (PM) group used instead the safety cards. The first two conditions represent two VR setups increasingly used for training, the third is the traditional method provided to all passengers in commercial aviation.

We formulated the following hypotheses for the study:

1) Engagement should be higher with the two VR setups than printed materials. This would be consistent with studies that compared either VR headsets or smartphones with safety cards in training [39]–[41]. Moreover, engagement might be higher in the VH group than SP group as suggested by studies that compared immersive and non-immersive VR setups [29], [42]. It is worth noting that both previous studies [29], [42] concerned games and did not consider a VR setup based on a smartphone. Our study thus aims at possibly extending results to lessons with a virtual instructor, and comparing a VR headset with a smartphone.
2) Presence in the VE should be higher in the VH group than SP group as suggested by previous studies comparing immersive and non-immersive VR setups [29], [42], [43]. The study aims at possibly extending results as described for engagement.
3) Usability should be higher with the two VR setups than the printed materials because there are well known issues in the comprehension of safety cards [33], [44]. Moreover, usability might be higher in the SP group than VH group because of widespread familiarity with smartphones and possible unfamiliarity of trainees with VR headsets and tracked controllers.

Regarding satisfaction, this is an exploratory study, because the previous studies that compared satisfaction between two VR setups for training did not find a significant difference in this measure [31] or found a significant difference only in one subscale [27]. Regarding knowledge gain and retention, this is an exploratory study because a study that compared VR with safety cards found VR to be superior for knowledge gain [41], while another one did not find a difference in knowledge gain but in knowledge retention [39]. Moreover, as described in Section II, VR setups offering more immersive visualization and more natural interaction features had positive effects on knowledge gain only in some of the studies, and none of them compared a VR headset with a smartphone for procedural training. Finally, this is an exploratory study with respect to confidence, on which results of previous studies varied with the tasks [14], [22], [25].

A. Materials

The VE with the virtual instructor was implemented in Unity 5.5. The VH group used the VE with an HTC Vive VR headset and its two tracked hand controllers. The VR headset was connected to a PC equipped with an Intel i7 processor, 16 GB RAM, and a NVidia GTX 980 graphic card. Audio was delivered through Sennheiser HD 215 closed earphones. The SP group tried the VE with a ZTE Axon 7 smartphone equipped with a 5.5-inch touchscreen. Audio was delivered through the built-in speakers of the smartphone. The PM group used six A4-sized, color-printed safety cards, one for each of the six overwing exits considered in the VE. The safety cards contained a pictorial for each step of the procedures described in Section III. Each pictorial showed the virtual instructor performing the step and red arrows to indicate the action to perform. To minimize graphic presentation differences with respect to VR conditions, pictorials were created by using the same computer-generated graphics of the VE.

B. Participants

The study involved a sample of 72 participants (40 male, 32 female), who were volunteers recruited through personal contact. Their age ranged from 18 to 28 (M = 21.88, SD = 1.88). All participants owned a smartphone and were able to use mobile applications. We asked them to rate their frequency of use of mobile applications with 3D graphics (e.g., games) on a 7-point scale (1 = never, 2 = less than once a month, 3 = about once a month, 4 = several times a month, 5 = several times a week, 6 = every day for less than an hour, 7 = every day for more than an hour). Answers ranged from 1 to 6 (median = 2; 33 participants never used mobile applications with 3D graphics, 15 used them less than once a month, 5 used them about once a month, 9 used them several times a month, 8 used them several times a week, and 2 used them every day for less than an hour). No participant owned a VR headset, and number of times of previous use of VR headsets ranged from 0 to 4 (M = 0.49, SD = 0.90). We assessed participants' differences in frequency of air travel by asking them to count their number of flights in the last two years, as in [33]. Each flight had to be counted individually (e.g., a round trip from airport A to airport C via a connection through airport B results in four flights). Answers ranged from 0 to 15 (M=3.60, SD=3.72). Finally, we used the 32-items Flight Anxiety Situations questionnaire (FAS) developed by [45] to assess participants' anxiety in flight-related situations, which might affect their emotions also during aviation safety training. Each FAS item is rated on a 5-point scale, ranging from 1 (no anxiety) to 5 (overwhelming anxiety). The FAS score is obtained by summing all item scores and can thus range from 32 to 160. In our sample, the FAS score ranged from 32 to 87 (M = 52.60, SD = 14.29).

Participants were assigned to the three groups in such a way that: (i) each group had 24 participants (13M, 11F in the VH and SP groups; 14M, 10F in the PM group); (ii) the three groups were similar in terms of age (VH: M = 21.79, SD = 1.41; SP: M = 21.79, SD = 2.19; PM: M = 22.04, SD = 2.01), frequency of use of mobile applications with 3D graphics (VH, SP: median = 2.00; PM: median = 1.50), number of times participants used VR headsets (VH: M = 0.46, SD = 0.72; SP: M = 0.71, SD = 1.20; PM: M = 0.29, SD = 0.69), number of flights (VH: M = 3.67, SD = 4.45; SP: M = 3.50, SD = 3.60; PM: M = 3.63, SD = 3.16), and FAS score (VH: M = 53.42, SD = 14.60; SP: M = 54.13, SD = 14.35; PM: M = 50.25, SD = 14.23). Each of these variables was submitted to a one-way ANOVA that confirmed the lack of statistically significant differences between groups.

C. Measures

1) Engagement

To measure engagement with the VE or the safety cards, we used the Game Engagement Questionnaire (GEQ) proposed in [46], which was employed by McMahan et al. in their study about the effects of VR setups [42]. The GEQ includes 19 statements and participants are asked to rate each of them on a 3-point scale (1 = No, 2 = Sort of, 3 = Yes). Seven of the GEQ statements are specifically about games or playing, so we adapted them to make them suitable for rating the VE and the safety cards, by changing "game" into "tool" and "playing" into "using". For example, "I really get into the game" becomes "I really get into using the tool", and "I feel like I just can't stop playing" becomes "I feel like I just can't stop using it". The ratings of the 19 statements were summed up to form a scale that could range from 19 to 57.

2) Satisfaction

To measure participants' satisfaction with the VE or the safety cards, we used the seven statements about satisfaction of the well-known USE Questionnaire [47], which rates level of agreement on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The ratings to the seven statements were summed up to form a scale that can range from 7 to 49.

3) Usability

To measure usability of the VE and the safety cards, we employed the well-known System Usability Scale (SUS) [48]. SUS asks participants to rate their level of agreement with ten statements on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). Results are reported as a score that can range from 0 to 100.

4) Presence

To measure presence with the two VR setups, we administered the Igroup Presence Questionnaire (IPQ)[1] [49]. The IPQ asks participants to rate 14 items on a 7-point scale, ranging from 0 to 6. The questionnaire includes a general item related to the sense of "being there", and three subscales for the following independent dimensions: spatial presence (5 items), involvement (4 items) and experienced realism (4 items).

5) Knowledge Gain and Retention

To measure knowledge about the overwing exit procedures, we showed participants a picture of each exit and we asked them to verbally describe in the correct sequence all the steps of the procedure to open the exit, possibly indicating the relevant parts of the picture (e.g., "Pull down the lever under this cover"). We recorded audio and video (focused on participants' hands over the pictures). The experimenter reviewed the videos after the experiment and assigned one of the following codes to each step that belonged to the considered procedure: i) "Correct", if participants described that step correctly and completely, ii) "Incomplete", if participants described that step correctly, but incompletely, e.g., participants said to pull a lever, but they did not specify in which direction, iii) "Partially wrong", if participants described that step, but it was partially wrong, e.g., participants correctly said to pull a lever, but in a wrong direction, or iv) "Omitted", if participants did not mention that step. By summing up all the steps with the same code in all procedures, we obtained four measures, respectively called *correct steps*, *incomplete steps*, *partially wrong steps*, and *omitted steps*. Each of these measures can thus range from 0 to the total number of steps in the six procedures, which is 33 (5 procedures have 6 steps, and one procedure has 3 steps). In addition, the experimenter counted possible

[1] available at http://www.igroup.org/pq/ipq

misplaced steps: each correct, incomplete or partially wrong step described by participants, was considered misplaced if it was described before (respectively after) at least one step that should have preceded (respectively followed) the considered step. For example, if participants described all the six expected steps of a procedure in the correct order, except for step 3 and 4 that were swapped, then the number of misplaced steps is 2 for that procedure. Finally, we defined *completely wrong steps* as the total number of described steps that should not have appeared in the procedure at all, e.g., if participants said to hold the bottom handle of the overwing exit and pull the hatch inside the cabin in the three-steps procedure, then completely wrong steps is 2 for that procedure. To assess knowledge gain as well as retention, we asked participants to take the knowledge test twice: immediately after using the VE or the safety cards (post-test), and two weeks later (retention-test).

6) Confidence

To measure participants' confidence in performing the procedures, we took inspiration from the methods used in [14], [22], [25]. More precisely, after each step mentioned by participants, we asked them to rate their level of confidence in that specific step on a 5-point scale (1 = not sure, 5 = certain). After each procedure, we also asked participants to rate their overall level of confidence in that procedure on the same scale. We calculated a *step confidence* measure as the average of the confidence levels for all the steps that participants remembered correctly, and *procedure confidence* as the average of the overall levels of confidence for all procedures. Confidence was calculated at post-test as well as retention-test.

D. Procedure

Participants were told that we were testing a tool that illustrates how to open the overwing exits of different aircraft types. Written consent for participation and for recording answers to the knowledge test and a final interview was obtained from participants. Participants were also informed that they were going to be contacted again two weeks later for additional questions (without specifying what those further questions were going to concern), and that they could refrain from continuing their participation at any time, without providing a reason to the experimenter. Then, participants filled a demographic questionnaire (gender, age, frequency of use of mobile applications with 3D graphics, ownership of a VR headset, number of times they used a VR headset, number of flights in the last two years), as well as the FAS questionnaire.

Participants in the VH group were helped to wear the VR headset and adjust it until they could see well and felt comfortable. Participants in the VH and SP group were informed about the controls to interact with the VE by means of a tutorial. The tutorial used a VE that contained only a red sphere, a blue cylinder, and a green cube, placed respectively on the right, on the left, and in front of the participants' initial position in the VE. After explaining participants about how to look around in the VE by using the hardware available to them, the experimenter invited them to look at the red sphere, grasp the blue cylinder, and move the green cube. If participants understood the controls, they were invited to press a button to exit the tutorial. Otherwise, they could press another button to restart the tutorial until they felt familiar with the controls.

Then, participants used the VE or the safety cards to learn the six exit opening procedures. All participants studied the procedures in the same order. The experimenter told them that they could spend how much time they liked on each procedure, but once they started a new procedure they could not go back to previous ones. More precisely, in the VH and SP group, participants could repeat the demonstration phase as many times as they wanted, then they started the practice phase and could repeat it as many times as they wanted or go the next exit. To keep conditions similar, participants in the PM group received the safety cards one at a time, and they had to give back the currently examined card to the experimenter in order to receive the next one. While forcing the same amount of time of tool use in all conditions would have helped in comparing knowledge gain in a laboratory setting, we took the decision to allow participants to spend how much time they liked for the ecological validity of the study. In particular, forcing a predefined amount of time with safety cards would have made it more difficult to generalize results to the real-world setting, where people are left free to examine the safety card, and many persons actually do not pay due attention to it [32].

After the experimental condition, all participants filled the engagement, satisfaction, and usability questionnaires. Participants using the VR setups also filled the presence questionnaire. Then, all participants took the knowledge test including their level of confidence as described in Section IV.C.5 and IV.C.6. In the knowledge test, the overwing exits were shown to participants in the order they had studied them. Finally, the experimenter briefly interviewed participants, asking them about the VE or the safety cards: what they liked or disliked, what they found difficult, and what they would possibly change. Two weeks later, knowledge retention and confidence were re-assessed by having participants take the knowledge test again.

V. Results

Table II reports mean and standard deviation of all measures for each group. For measures assessed only at post-test (engagement, satisfaction, usability, and presence), we analyzed the results using a between-subjects one-way ANOVA. For measures assessed at post-test and retention-test (knowledge and confidence), we analyzed the results using a 3 x 2 mixed design ANOVA, in which *group* served as the between-subjects variable, and *measurement instant* served as the within-subjects variable. The significance level was set at 0.05. Effect sizes are reported as partial eta squared (η_p^2). In case of a statistically significant effect of group, we proceeded with pairwise comparisons using Bonferroni. Following [50], in case of a statistically significant interaction between group and measurement instant, we analyzed each simple effect using Bonferroni correction, considering the effects of measurement instant separately for each group and the effects of group separately at each measurement instant.

A. Engagement

For engagement (Fig. 4a), the analysis revealed a statistically significant difference, $F(2, 69) = 7.10$, $p < 0.005$, $\eta_p^2 = 0.17$. Pairwise comparisons revealed two statistically significant differences: between the VH and PM groups, $p < 0.005$, and

TABLE II
MEAN AND STANDARD DEVIATION OF ALL MEASURES FOR EACH GROUP

Measure	PM		SP		VH	
	M	SD	**M**	SD	**M**	SD
Engagement	**32.00**	6.63	**32.04**	4.31	**36.96**	4.46
Satisfaction	**27.92**	9.90	**30.54**	8.27	**38.33**	5.68
Usability	**70.42**	17.30	**72.71**	11.44	**81.46**	10.11
Presence						
IPQ total score			**2.54**	0.91	**3.76**	1.08
Being there			**2.71**	1.73	**4.92**	1.21
Spatial			**2.73**	1.37	**4.26**	1.03
Involvement			**2.38**	1.14	**3.78**	1.28
Experienced realism			**2.44**	0.80	**2.84**	1.20
Knowledge gain						
Correct steps	**17.43**	4.78	**27.79**	2.92	**25.88**	4.05
Incomplete steps	**1.17**	2.08	**0.71**	1.76	**0.58**	1.18
Partially wrong steps	**3.57**	2.50	**2.67**	1.52	**1.88**	1.15
Omitted steps	**10.83**	4.15	**1.83**	2.22	**4.67**	4.02
Misplaced steps	**4.78**	2.68	**1.00**	1.87	**0.83**	1.55
Completely wrong steps	**3.30**	2.53	**1.33**	2.06	**1.42**	1.38
Knowledge retention						
Correct steps	**18.52**	5.67	**27.75**	3.39	**24.58**	5.24
Incomplete steps	**1.17**	2.08	**1.08**	2.15	**0.54**	1.64
Partially wrong steps	**2.30**	1.33	**2.42**	1.74	**1.88**	1.26
Omitted steps	**11.00**	5.58	**1.75**	2.15	**6.00**	4.29
Misplaced steps	**5.30**	3.70	**1.83**	3.12	**3.25**	3.17
Completely wrong steps	**2.74**	1.94	**1.13**	1.33	**1.67**	1.55
Confidence						
Step - Post-test	**4.02**	0.47	**4.58**	0.35	**4.52**	0.44
Procedure - Post-test	**3.38**	0.60	**4.03**	0.58	**3.88**	0.60
Step - Retention-test	**4.05**	0.64	**4.36**	0.48	**4.44**	0.58
Procedure - Retention-test	**3.51**	0.69	**3.73**	0.64	**3.87**	0.78

between the VH and SP groups, $p < 0.01$. The difference between the SP group and PM group did not reach significance.

B. Satisfaction

For satisfaction (Fig. 4b), the analysis revealed a statistically significant difference, $F(2, 69) = 10.64$, $p < 0.001$, $\eta_p^2 = 0.24$. Pairwise comparisons revealed two statistically significant differences: between the VH and PM groups, $p < 0.001$, and between the VH and SP groups, $p < 0.005$. The difference between the SP group and PM group did not reach significance.

C. Usability

For usability (Fig. 4c), the analysis revealed a statistically significant difference, $F(2, 69) = 4.59$, $p < 0.05$, $\eta_p^2 = 0.12$. Pairwise comparisons revealed that the difference between the VH group and PM group was statistically significant, $p < 0.05$. The differences between the VH group and SP group and between the SP group and PM group did not reach significance.

D. Presence

For presence (Fig. 5), the analysis revealed statistically significant differences between the two VR setups in IPQ total score, $F(1, 46) = 22.44$, $p < 0.001$, $\eta_p^2 = 0.33$, in the general item about the sense of "being there", $F(1, 46) = 26.19$, $p < 0.001$, $\eta_p^2 = 0.36$, in spatial presence, $F(1, 46) = 19.24$, $p < 0.001$, $\eta_p^2 = 0.30$, and in involvement, $F(1, 46) = 16.19$, $p < 0.001$, $\eta_p^2 = 0.26$. In the experienced realism subscale, the difference between the two groups did not reach significance.

E. Knowledge Gain and Retention

One participant had to be excluded from analysis of knowledge gain and retention because her retention-test video was missing due to a technical issue. The analyses were thus conducted on the remaining 71 participants.

Considering *correct steps* (Fig. 6a), the analysis revealed a main effect of group, $F(2, 68) = 36.63$, $p < 0.001$, $\eta_p^2 = 0.52$, no main effect of measurement instant, and no interaction. Pairwise comparisons revealed that the differences in correct steps between VH and PM as well as between SP and PM were statistically significant ($p < 0.001$ for both), while the difference between VH and SP was not statistically significant.

No significant main effect or interaction was found for *incomplete steps* (Fig. 6b).

Considering *partially wrong steps* (Fig. 7a), the analysis revealed a main effect of group, $F(2, 68) = 4.78$, $p < 0.05$, $\eta_p^2 = 0.12$, no main effect of measurement instant, and no interaction. Pairwise comparisons revealed that the difference in partially wrong steps between VH and PM was statistically significant ($p = 0.01$), while the differences between SP and PM as well as between VH and SP were not statistically significant.

Considering *omitted steps* (Fig. 7b), the analysis revealed a main effect of group, $F(2, 68) = 37.01$, $p < 0.001$, $\eta_p^2 = 0.52$, no main effect of measurement instant, and no interaction. Pairwise comparisons revealed that the differences in omitted steps between VH and PM as well as between SP and PM were statistically significant ($p < 0.001$ for both), and the difference between VH and SP was statistically significant ($p < 0.005$).

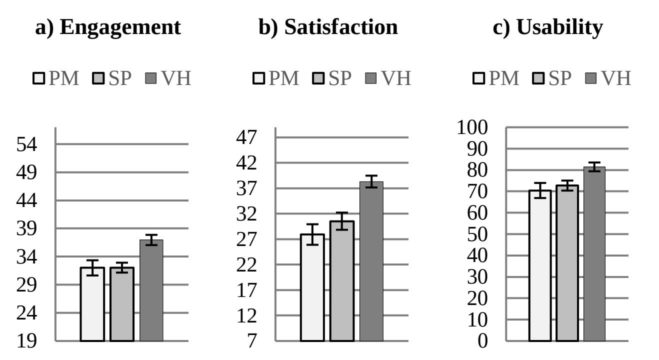

Fig. 4. Means of a) engagement, b) satisfaction, and c) usability. Capped vertical bars indicate ± SE.

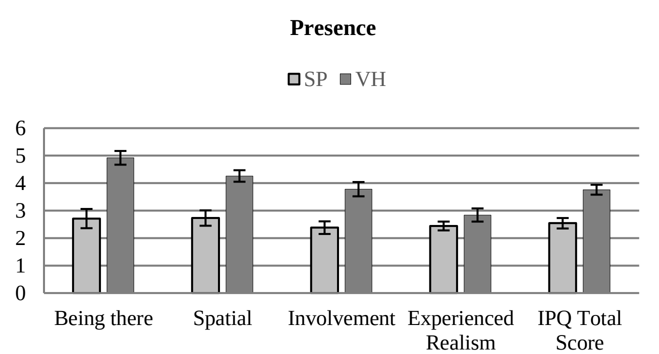

Fig. 5. Means of presence. Capped vertical bars indicate ± SE.

Considering *misplaced steps* (Fig. 8a), the analysis revealed a main effect of group, $F(2, 68) = 17.77$, $p < 0.001$, $\eta_p^2 = 0.34$, a main effect of measurement instant, $F(1, 68) = 10.13$, $p < 0.005$, $\eta p2 = 0.13$, and no interaction. Pairwise comparisons revealed that the differences in misplaced steps between VH and PM as well as between SP and PM were statistically significant ($p < 0.001$ for both), while the difference between VH and SP was not statistically significant. The main effect of measurement instant revealed that there was a statistically significant increase in misplaced steps between post-test ($M = 2.17$, $SD = 2.74$) and retention-test ($M = 3.44$, $SD = 3.58$).

Considering *completely wrong steps* (Fig. 8b), the analysis revealed a main effect of group, $F(2, 68) = 8.06$, $p = 0.001$, $\eta_p^2 = 0.19$, no main effect of measurement instant, and no interaction. Pairwise comparisons revealed that the difference in completely wrong steps between VH and PM was statistically significant ($p < 0.01$), the difference between SP and PM was statistically significant ($p = 0.001$), and the difference between VH and SP was not statistically significant.

F. Confidence

One participant had to be excluded from this analysis, as explained in the previous section. Considering *step confidence* (Fig. 9a), the analysis revealed a main effect of group, $F(2, 68) = 7.24$, $p = 0.001$, $\eta_p^2 = 0.18$, no main effect of measurement instant, and no interaction. Pairwise comparisons revealed statistically significant differences between VH and PM ($p < 0.005$) and between SP and PM ($p < 0.01$), while the difference between VH and SP was not statistically significant.

Considering *procedure confidence* (Fig. 9b), the analysis revealed a main effect of group, $F(2, 68) = 4.23$, $p < 0.05$, $\eta_p^2 = 0.11$, no main effect of measurement instant, and an interaction between group and measurement instant, $F(2, 68) = 3.15$, $p < 0.05$, $\eta_p^2 = 0.09$. Pairwise comparisons revealed that the differences between VH and PM as well as between SP and PM were statistically significant ($p < 0.05$ for both), while the difference between VH and SP was not. The analysis of simple main effects revealed that in the SP group, the difference between post-test and retention-test was significant ($p < 0.05$), while in the other two groups it was not. There were statistically significant differences between VH and PM ($p < 0.05$) as well as between SP and PM ($p < 0.005$) at post-test, and no statistically significant difference between the three groups at retention-test.

VI. Discussion

Results confirmed that a VR setup can be more engaging than printed materials for procedural safety training, but the difference in engagement was statistically significant only between the VR headset and the safety cards, not between the smartphone and the safety cards. This extends the results of [39] where an immersive training game on a VR headset was compared with a safety card and found to be more engaging. In the present study, we found that an immersive VE on a VR headset is more engaging than safety cards even when it is not a game, but a series of lessons with a virtual instructor. A training game for smartphones was found to be more engaging than a safety card in [40], while the lessons with the virtual instructor in the present study were not significantly more engaging than safety cards when they were experienced on a smartphone. These results might suggest that a training VE experienced on a VR headset could be more engaging than safety cards, regardless of the inclusion of gaming content, while a non-game VE on a smartphone might not be enough to get a significant advantage in engagement over safety cards. Additional studies are needed to identify which features of VR training content (e.g., gamification aspects) can make it more

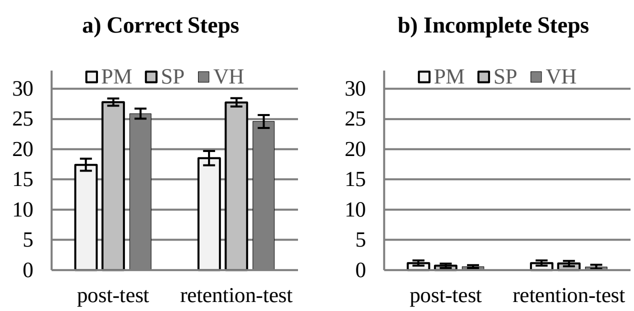

Fig. 6. Means of a) correct steps, b) incomplete steps at post-test and retention-test. Capped vertical bars indicate ± SE.

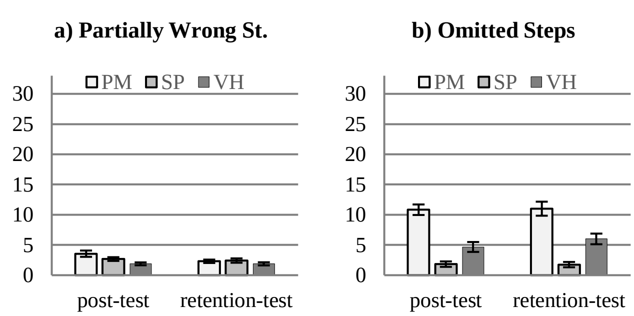

Fig. 7. Means of a) partially wrong steps, b) omitted steps at post-test and retention-test. Capped vertical bars indicate ± SE.

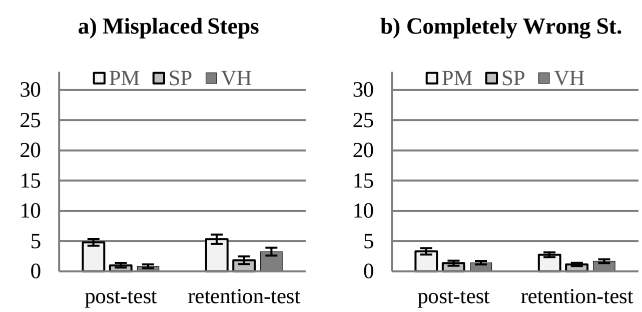

Fig. 8. Means of a) misplaced steps, b) completely wrong steps at post-test and retention-test. Capped vertical bars indicate ± SE.

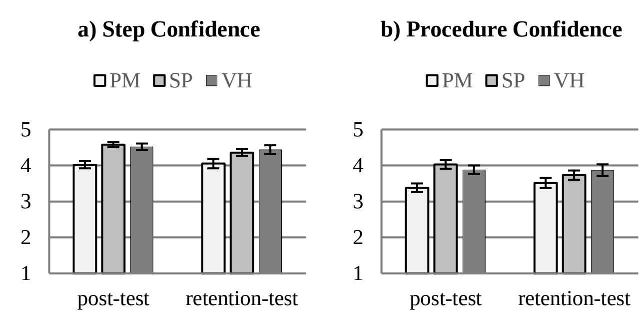

Fig. 9. Means of a) step confidence, b) procedure confidence at post-test and retention-test. Capped vertical bars indicate ± SE.

engaging than printed materials even when VR is delivered by a smartphone. Considering the comparison between VR setups, results confirmed our hypothesis that the immersive setup was significantly more engaging than the non-immersive setup with the same content. This result is consistent with [29] and [42], where respectively a serious game and an entertainment game were found to be more engaging with an immersive rather than non-immersive VR setup. Our study extends previous ones confirming such difference in engagement when the non-immersive VR setup is a smartphone and when the content is not a game, but lessons with a virtual instructor. The present study also investigated trainees' satisfaction and the results were consistent with those for engagement. Participants who tried the VE on the VR headset were more satisfied than those who tried the safety cards and those who tried the VE on the smartphone, while no significant difference in satisfaction was found between the safety cards and the smartphone. Pearson correlation between engagement and satisfaction revealed a medium and statistically significant correlation, $r = 0.32$, $p < 0.01$. The more participants were engaged, the more they were satisfied. The high satisfaction scores received by the VR headset may also be partly due to participants' low familiarity with VR headsets, which could have biased participants in perceiving this VR setup as a futuristic and exciting technology.

Regarding usability, the safety cards score was the lowest as expected, but the difference was statistically significant only in the comparison with VR headset, which obtained the highest usability score. On the contrary, the difference in usability between the smartphone and the safety cards did not reach significance. Although all participants were regular smartphone users, and none of them regularly used (or had ever used) a VR headset, the smartphone was surprisingly no better than the VR headset in terms of usability. This can be possibly due to the interaction features offered by the VR headset: regardless of the fact that participants were not familiar with tracked controllers, the natural gestures to perform the procedure might have been perceived as easier to learn and perform than the more familiar touch and slide gestures on the smartphone touchscreen. Further studies comparing different interaction features in VR setups are needed to support this hypothesis.

Confirming previous studies that compared an immersive and a non-immersive VR setup [29], [42], [43], presence was significantly higher with the immersive one. The present study extends previous results by including a smartphone as the non-immersive VR setup, and considering lessons with a virtual instructor instead of a game. The immersive setup (VR headset) obtained significantly higher scores than the non-immersive one (smartphone) for the overall scale, for the general item related to the sense of "being there", as well as for spatial presence and involvement. No statistically significant difference was found for experienced realism, probably because the two VR setups used the same VE.

Regarding knowledge gain and retention, printed materials were worse than both VR setups because participants who used safety cards correctly remembered a significantly smaller number of steps. Safety cards led to more partially wrong steps than the VR headset and to more omitted, misplaced, and completely wrong steps than both VR setups. Incomplete steps was the only measure for which there was no statistically significant difference. The lack of interaction between group and measurement instant for all knowledge measures indicates that the effect of group overall concerned both post-test and retention-test. In previous studies, a serious game on a VR headset was better than a safety card at (1-week) retention-test and not at post-test in [39], while another study [41] showed that non-immersive VR on a smartphone was better than safety cards in terms of errors made by participants when they performed procedures in the real world immediately after training (retention was not assessed). The present study extends previous results showing that both immersive and non-immersive VR setups result in better knowledge gain as well as retention than safety cards.

Comparing the two VR setups, we found no statistically significant difference for most of the knowledge gain and retention measures, with the interesting exception of omitted steps, which were higher for participants who tried the VR headset. Analyzing this result in more detail, we observed that this higher value in the VH group was due to omissions in a specific step (the 5th) of the six-steps procedures. In the 5th step, the door hatch should be lifted and pulled inside the cabin to open the exit. The step was performed very easily with the VR headset because, after unlocking the exit (4th step), participants had just to move slightly their hands toward themselves to pull the hatch inside, so some participants might have probably lumped the two steps into a single one in their mental model of the procedure. On the contrary, participants using the smartphone had to perform a clearly distinct slide gesture to pull the hatch inside the cabin, and this could have helped to distinguish and remember the step. A previous study of procedural safety training [29] contrasted two types of VR headsets and a desktop monitor, finding a general lack of differences in knowledge gain and retention among them. The present study confirms and extends those findings by contrasting VR headset and smartphone. On the contrary, VR setups offering more immersive visualization features led to better performance in a study where participants had to apply a visual scanning strategy [26], and in a study about a procedure that consisted in moving objects from a location to another [20]. It is worth noting that in another study where the procedure consisted in moving objects between two locations, the difference between two VR setups (laptop monitor and 2-screen projection display) was significant only when the procedure was complex (31 steps) and not when it was simple (10 steps) [19]. Each procedure in the present study was short (3 steps in one case, 6 in the other five cases, 33 is the total number of steps) and some steps were the same in different procedures. Thus the overall knowledge taught in our VE might have been simpler to memorize than the 31-steps procedure for which a significant effect of the VR setup was found in [19]. In addition, unlike the above mentioned studies [19], [20], [26], our VE included a virtual instructor, which could have helped in the memorization of the procedure in both VR setups.

It is important to note that for most knowledge gain and retention measures we found no significant effect of measurement instant. Knowledge gained during training was mostly retained after two weeks. Misplaced steps was the only measure with a significant main effect of measurement time: participants misplaced a significantly greater number of steps at retention-test than at post-test. Previous studies showed significant decay over time in knowledge with safety cards [39], and no significant decay with immersive or non-immersive VR setups [29], [39]. However, lack of significant decay with the safety cards in the present study could be simply due to the fact that knowledge gain with the cards was already lower than the VR setups at post-test, while in [39] knowledge gain with the safety card was not significantly different than VR at post-test.

Results for step confidence showed that participants who trained using the VR setups were significantly more confident in correct steps than participants who trained with the safety cards. No significant difference in step confidence was found between the two VR setups. As for most knowledge measures, the passing of time between post-test and retention-test did not significantly reduce participants' confidence in correct steps. Results for procedure confidence were more articulated: on one hand, the main effect of group and pairwise comparisons showed that both VR setups were better than printed materials overall; on the other hand, the interaction between group and measurement instant and the subsequent analysis of simple main effects revealed that the difference in procedure confidence between safety cards and VR setups is significant at post-test, but not at retention-test, and there is a significant decrease in procedure confidence between post-test and retention-test only for the smartphone.

In summary, both VR setups were better than printed materials for procedural safety training in terms of knowledge gain, retention, and participants' confidence in their replies, while there were only minor differences in these measures between the two VR setups. Regarding the usability measure, only the VR headset was better than the safety cards in terms of usability, and presence was higher with the VR headset than the smartphone. Moreover, VR setup played an important role on engagement and satisfaction, with VR headset scoring better than both printed materials and smartphone. The study was not designed to identify which specific aspects of the VR headset (e.g., better visualization or interaction features) contributed more to the increase in engagement, but some hypotheses can be formulated by considering the framework in [10], which associates the different types of engagement with features of the learning experience in the proposed VE. For example, relevant feedback, dynamic assessment, visual representation of the information, situatedness of the learning experience, gestures and movement are all associated with cognitive engagement [10]. While the virtual instructor provided relevant feedback, dynamically assessed participants' knowledge, and visually demonstrated the taught procedures in the same ways in the two VR setups, situatedness could be higher with the VR headset as confirmed by the higher scores for spatial presence and the general item related to the sense of "being there". Moreover, the mapping of gestures and movements in the VE on the actual gestures and movements in the steps of the procedures was more natural using the tracked controllers, as suggested also by the usability scores. Therefore, these two features of the VR headset could have contributed to increase cognitive engagement with respect to the smartphone, where the touchscreen was used to display the VE (filling a much smaller part of participants' field of view) and was able to detect only movements made with fingertips on the 2D screen surface. Considering behavioral engagement and motivation, the possibility for trainees to fail gracefully and the adaptive help provided by the virtual instructor could encourage an affirmative answer to the first question identified by Eccles et al. [11] ("Can I do this?"), regardless of the VR setup. However, making the natural gestures with the tracked controllers of the VR headset might make participants feel more able to perform the steps of the procedure also in the real world. User interaction based on tracked controllers and the possibility offered by the VR headset to see the steps of the procedure performed by a human-scale virtual instructor in a real-size aircraft might help trainees in answering the question "What do I need to do in order to succeed?". Since these motivation-related questions concern participant's self-efficacy, further studies are needed to assess self-efficacy and its relation with specific aspects of the VE and the virtual instructor to better understand the role of the VR setup on behavioral engagement. As described by Plass et al. [10], the second question is related to participants' intrinsic motivation, a construct that was not assessed in the present study. VR setup might have played a role on affective engagement, because presence was higher with the VR headset than the smartphone, and presence can play a role on emotions [13]. Although the virtual instructor did not use emotion appeals as in [51], its tone of voice and facial expressions could have evoked emotions especially when it provided participants with feedback. Observing the facial expressions of the virtual instructor at human-scale using the VR headset or the same expressions on the scaled-down virtual instructor on the smartphone might have conveyed different intensities of emotions to participants, likely increasing affective engagement in participants using the VR headset. Specific studies of emotions conveyed by virtual instructors in different VR setups are needed to understand these aspects. Overall, the VR setup could have played a role on cognitive, behavioral, or affective engagement, as reflected by the higher engagement score found for the VR headset rather than the smartphone.

The obtained results can support a recommendation of adopting a VR headset rather than a smartphone for procedural safety training, because the former can boost trainees' engagement, satisfaction, and presence. Considering financial implications, recent consumer VR headsets could be an affordable solution for most training centers, but delivering the same training on smartphones could be desirable to reach a wider number of users and for longer times, because trainees could use their own devices. Our study advances knowledge by showing that a smartphone can be similar to a VR headset (and better than printed materials) in terms of knowledge gain and retention of safety procedures, but actions should be taken to make the VE more engaging on the smartphone. For example, lessons learned from the present study guided the design of the instructional part of our "Air Safety World" mobile application

[52], which includes all the lessons about overwing exits considered in the present study. To reach a potentially wide number of users (more than a million at the moment of writing), we made the application publicly available on Google Play and Apple App Store. To boost engagement, we introduced gamification features, including rewards for successfully completing the lessons. Rewards are given as virtual coins that allow users to unlock mini games as well as acquire aircraft and create a virtual airport they can manage in the application.

VII. Conclusion

VR headsets and smartphones are receiving increasing attention as VR setups for training. The present study was the first to compare these two different VR setups in procedural safety training, assessing their effects on trainees' engagement, satisfaction, usability, presence, knowledge gain, 2-weeks retention, and confidence. Moreover, we compared the two VR setups with the traditional printed materials used for the considered training, confirming and extending the results about benefits of VR over printed materials in training. Although the effects of the VR headset and the smartphone were similar on knowledge gain, retention, and confidence, the former can be recommended to increase trainees' engagement, satisfaction, and presence. The results found in the present study could likely be extended to other procedures of comparable complexity in the same or similar safety-related domains. In particular, the use of immersive or non-immersive VR setups could be promising for other domains where safety procedures are commonly taught by means of printed materials that would likely have limitations in comprehension and engagement similar to safety cards [32], [33]. For example, in personal and occupational health and safety, printed materials are used to teach procedures like hand washing and other hygiene precautions, or correctly wearing and using protective equipment.

Future work should compare combinations of the same VR headset with different ways of supporting hand interaction (e.g., using a gamepad, tracked hand controllers, or camera-based hand tracking) to investigate the specific role of interaction on usability and on the different types of engagement identified in [9]. Moreover, new user studies should concern the effects of VR setups on trainees' perception of the virtual instructor to assess if visualizing it in human scale with the VR headset or scaled-down on a small screen plays a role in evoking different types of emotions.

Acknowledgment

We are grateful to David Weed (Civil Aerospace Medical Institute, Federal Aviation Administration, USA) for his feedback and encouragement. We thank Nicola Zangrando and Riccardo Sioni (Human-Computer Interaction Lab, University of Udine, Italy), who carried out 3D modeling and animation activities for the development of the application.

References

[1] S. Haque and S. Srinivasan, "A meta-analysis of the training effectiveness of virtual reality surgical simulators," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 1, pp. 51–58, Jan. 2006, doi: 10.1109/TITB.2005.855529.

[2] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis, "Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis," *Comput. Educ.*, vol. 70, pp. 29–40, Jan. 2014, doi: 10.1016/j.compedu.2013.07.033.

[3] M. C. Howard, "A meta-analysis and systematic literature review of virtual reality rehabilitation programs," *Comput. Human Behav.*, vol. 70, pp. 317–327, May 2017, doi: 10.1016/j.chb.2017.01.013.

[4] L. Jensen and F. Konradsen, "A review of the use of virtual reality head-mounted displays in education and training," *Educ. Inf. Technol.*, vol. 23, pp. 1515–1529, Jul. 2018, doi: 10.1007/s10639-017-9676-0.

[5] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Comput. Educ.*, vol. 147, Apr. 2020, Art. no. 103778, doi: 10.1016/j.compedu.2019.103778.

[6] G. Robertson, S. K. Card, and J. D. Mackinlay, "Three views of virtual reality: Nonimmersive virtual reality," *Comput.*, vol. 26, no. 2, pp. 81–83, Feb. 1993, doi: 10.1109/2.192002.

[7] *Distribution dashboard*. (2020). Google. Accessed: Oct. 22, 2020. [Online]. Available: https://developer.android.com/about/dashboards

[8] C. Wigginton, M. Curran, and C. Brodeur, "Global mobile consumer trends, 2nd edition," Deloitte, 2017. Accessed: Oct. 22, 2020. [Online]. Available: https://www2.deloitte.com/global/en/pages/technology-media-and-telecommunications/articles/gx-global-mobile-consumer-trends.html

[9] S. Domagk, R. N. Schwartz, and J. L. Plass, "Interactivity in multimedia learning: An integrated model," *Comput. Human Behav.*, vol. 26, no. 5, pp. 1024–1033, Sep. 2010, doi: 10.1016/j.chb.2010.03.003.

[10] J. L. Plass, B. D. Homer, and C. K. Kinzer, "Foundations of game-based learning," *Educ. Psychol.*, vol. 50, no. 4, pp. 258–283, 2015, doi: 10.1080/00461520.2015.1122533.

[11] J. S. Eccles, A. Wigfield, and U. Schiefele, "Motivation to succeed," in *Handbook of child psychology*, 5th ed., vol. 3, W. Damon and N. Eisenberg, Eds. Hoboken, NJ, USA: Wiley, 1998, pp. 1017–1095.

[12] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," *Presence Teleoperators Virtual Environ.*, vol. 6, no. 6, pp. 603–616, Dec. 1997, doi: 10.1162/pres.1997.6.6.603.

[13] G. Riva *et al.*, "Affective interactions using virtual reality: the link between presence and emotions," *CyberPsychology Behav.*, vol. 10, no. 1, pp. 45–56, Feb. 2007, doi: 10.1089/cpb.2006.9993.

[14] K. Mania and A. Chalmers, "The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach," *CyberPsychology Behav.*, vol. 4, no. 2, pp. 247–264, Jul. 2001, doi: 10.1089/109493101300117938.

[15] W. Winn, M. Windschitl, R. Fruland, and Y. Lee, "When does immersion in a virtual environment help students construct understanding?," in *Proc. 5th Int. Conf. Learning Sciences*, Oct. 23–26, 2002, pp. 497–503.

[16] H. Aoki, C. M. Oman, D. A. Buckland, and A. Natapoff, "Desktop-VR system for preflight 3D navigation training," *Acta Astronaut.*, vol. 63, no. 7–10, pp. 841–847, Oct./Nov. 2008, doi: 10.1016/j.actaastro.2007.11.001.

[17] K. Johnsen and B. Lok, "An evaluation of immersive displays for virtual human experiences," in *2008 IEEE Virtual Reality Conf.*, Mar. 8–12, 2008, pp. 133–136, doi: 10.1109/VR.2008.4480764.

[18] M. Limniou, D. Roberts, and N. Papadopoulos, "Full immersive virtual environment CAVE™ in chemistry education," *Comput. Educ.*, vol. 51, no. 2, pp. 584–593, Sep. 2008, doi: 10.1016/j.compedu.2007.06.014.

[19] A. Sowndararajan, R. Wang, and D. A. Bowman, "Quantifying the benefits of immersion for procedural training," in *Proc. 2008 Workshop Immersive Projection Technologies / Emerging Display Technologies*, Aug. 9–10, 2008, Art. no. 2, doi: 10.1145/1394669.1394672.

[20] E. D. Ragan, A. Sowndararajan, R. Kopper, and D. A. Bowman, "The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments," *Presence Teleoperators Virtual Environ.*, vol. 19, no. 6, pp. 527–543, Dec. 2010, doi: 10.1162/pres_a_00016.

[21] E. Fassbender, D. Richards, A. Bilgin, W. F. Thompson, and W. Heiden, "VirSchool: The effect of background music and immersive display systems on memory for facts learned in an educational virtual environment," *Comput. Educ.*, vol. 58, no. 1, pp. 490–500, Jan. 2012, doi: 10.1016/j.compedu.2011.09.002.

[22] B. Laha, K. Sensharma, J. D. Schiffbauer, and D. A. Bowman, "Effects of immersion on visual analysis of volume data," *IEEE Trans. Vis.*

Comput. Graph., vol. 18, no. 4, pp. 597–606, Apr. 2012, doi: 10.1109/TVCG.2012.42.
[23] F. Bacim, E. D. Ragan, S. Scerbo, N. F. Polys, M. Setareh, and B. D. Jones, “The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs,” in *Proc. Graphics Interface 2013*, May 29–31, 2013, pp. 25–32.
[24] E. D. Ragan, R. Kopper, P. Schuchardt, and D. A. Bowman, “Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task,” *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 5, pp. 886–896, May 2013, doi: 10.1109/TVCG.2012.163.
[25] B. Laha, D. A. Bowman, and J. J. Socha, “Effects of VR system fidelity on analyzing isosurface visualization of volume datasets,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 4, pp. 513–522, Apr. 2014, doi: 10.1109/TVCG.2014.20.
[26] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan, “Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task,” *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 7, pp. 794–807, Jul. 2015, doi: 10.1109/TVCG.2015.2403312.
[27] R. Shewaga, A. Uribe-Quevedo, B. Kapralos, and F. Alam, “A comparison of seated and room-scale virtual reality in a serious game for epidural preparation,” *IEEE Trans. Emerg. Topics Comput.*, vol. 8, no. 1, pp. 218–232, Jan./Mar. 2020, doi: 10.1109/TETC.2017.2746085.
[28] T. Silva, E. Marinho, G. Cabral, and K. Gama, “Motivational impact of virtual reality on game-based learning: Comparative study of immersive and non-immersive approaches,” in *2017 19th Symp. Virtual and Augmented Reality*, Nov. 1–4, 2017, pp. 155–158, doi: 10.1109/SVR.2017.28.
[29] F. Buttussi and L. Chittaro, “Effects of different types of virtual reality display on presence and learning in a safety training scenario,” *IEEE Trans. Vis. Comput. Graph.*, vol. 24, no. 2, pp. 1063–1076, Feb. 2018, doi: 10.1109/TVCG.2017.2653117.
[30] G. Aeckersberg, A. Gkremoutis, T. Schmitz-Rixen, and E. Kaiser, “The relevance of low-fidelity virtual reality simulators compared with other learning methods in basic endovascular skills training,” *J. Vascular Surgery*, vol. 69, no. 1, pp. 227–235, Jan. 2019, doi: 10.1016/j.jvs.2018.10.047.
[31] G. Makransky, T. S. Terkildsen, and R. E. Mayer, “Adding immersive virtual reality to a science lab simulation causes more presence but less learning,” *Learn. Instruction*, vol. 60, pp. 225–236, Apr. 2019, doi: 10.1016/j.learninstruc.2017.12.007.
[32] D. Cosper and G. McLean, “Availability of passenger safety information for improved survival in aircraft accidents,” Office of Aerospace Medicine, Federal Aviation Administration, Washington, DC, USA, Tech. Rep. DOT/FAA/AM-04/19, 2004.
[33] C. L. Corbett, G. A. McLean, and D. K. Cosper, “Effective presentation media for passenger safety I: Comprehension of briefing card pictorials and pictograms,” Office of Aerospace Medicine, Federal Aviation Administration, Washington, DC, USA, Final Rep. DOT/FAA/AM-08/20, 2008.
[34] M. Zyda, “From visual simulation to virtual reality to games,” *Comput.*, vol. 38, no. 9, pp. 25–32, Sep. 2005, doi: 10.1109/MC.2005.297.
[35] J. Rickel, S. Marsella, J. Gratch, R. Hill, D. Traum, and W. Swartout, “Toward a new generation of virtual humans for interactive experiences,” *IEEE Intell. Syst.*, vol. 17, no. 4, pp. 32–38, Jul./Aug. 2002, doi: 10.1109/MIS.2002.1024750.
[36] A. Gulz, “Benefits of virtual characters in computer based learning environments: Claims and evidence,” *Int. J. Artif. Intell. Educ.*, vol. 14, no. 3–4, pp. 313–334, Dec. 2004.
[37] W. L. Johnson and J. C. Lester, “Face-to-face interaction with pedagogical agents, twenty years later,” *Int. J. Artif. Intell. Educ.*, vol. 26, pp. 25–36, Mar. 2016, doi: 10.1007/s40593-015-0065-9.
[38] D. A. Bowman and R. P. McMahan, “Virtual reality: how much immersion is enough?,” *Comput.*, vol. 40, no. 7, pp. 36–43, Jul. 2007, doi: 10.1109/MC.2007.257.
[39] L. Chittaro and F. Buttussi, “Assessing knowledge retention of an immersive serious game vs. a traditional education method in aviation safety,” *IEEE Trans. Vis. Comput. Graph.*, vol. 21, no. 4, pp. 529–538, Apr. 2015, doi: 10.1109/TVCG.2015.2391853.
[40] L. Chittaro and F. Buttussi, “Exploring the use of arcade game elements for attitude change: Two studies in the aviation safety domain,” *Int. J. Human Comput. Stud.*, vol. 127, pp. 112–123, Jul. 2019, doi: 10.1016/j.ijhcs.2018.07.006.
[41] L. Chittaro, C. L. Corbett, G. A. McLean, and N. Zangrando, “Safety knowledge transfer through mobile virtual reality: A study of aviation life preserver donning,” *Saf. Sci.*, vol. 102, pp. 159–168, Feb. 2018, doi: 10.1016/j.ssci.2017.10.012.
[42] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, “Evaluating display fidelity and interaction fidelity in a virtual reality game,” *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 4, pp. 626–33, Apr. 2012, doi: 10.1109/TVCG.2012.43.
[43] F. Pallavicini, A. Pepe, and M. E. Minissi, “Gaming in virtual reality: What changes in terms of usability, emotional response and sense of presence compared to non-immersive video games?,” *Simul. Gaming*, vol. 50, no. 2, pp. 136–159, Apr. 2019, doi: 10.1177/1046878119831420.
[44] J. K. Caird, B. Wheat, K. R. McIntosh, and R. E. Dewar, “The Comprehensibility of Airline Safety Card Pictorials,” *Proc. Human Factors Ergonomics Society Annu. Meeting*, vol. 41, no. 2, pp. 801–805, Oct. 1997, doi: 10.1177/107118139704100216.
[45] L. J. Van Gerwen, P. Spinhoven, R. Van Dyck, and R. F. W. Diekstra, “Construction and psychometric characteristics of two self-report questionnaires for the assessment of fear of flying,” *Psychol. Assessment*, vol. 11, no. 2, pp. 146–158, Jun. 1999, doi: 10.1037/1040-3590.11.2.146.
[46] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, “The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing,” *J. Exp. Social Psychol.*, vol. 45, pp. 624–634, Jul. 2009, doi: 10.1016/j.jesp.2009.02.016.
[47] A. M. Lund, “Measuring usability with the USE questionnaire,” *Usability Interface*, vol. 8, no. 2, pp. 3–6, Oct. 2001.
[48] J. Brooke, “SUS – A quick and dirty usability scale,” in *Usability evaluation in industry*, P. W. Jordan, B. Thomas, B. A. Weerdmester, and I. L. McClelland, Eds. New York, NY, USA: Taylor & Francis, 1996, pp. 189–194.
[49] T. Schubert, F. Friedmann, and H. Regenbrecht, “The experience of presence: Factor analytic insights,” *Presence Teleoperators Virtual Environ.*, vol. 10, no. 3, pp. 266–281, Jun. 2001, doi: 10.1162/105474601300343603.
[50] B. H. Cohen, *Explaining Psychological Statistics*, 2nd ed. Hoboken, NJ, USA: Wiley, 2001.
[51] F. Buttussi and L. Chittaro, “Humor and fear appeals in animated pedagogical agents: An evaluation in aviation safety education,” *IEEE Trans. Learn. Technol.*, vol. 13, no. 1, pp. 63–76, Jan./Mar. 2020, doi: 10.1109/TLT.2019.2902401.
[52] *Air Safety World*. (2017). HCI Lab, University of Udine. Accessed: Oct. 22, 2020. [Online]. Available: http://hcilab.uniud.it/asw

Fabio Buttussi received his PhD in computer science from the University of Udine. He is an assistant professor in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy. His research interests are in HCI, virtual reality, artificial intelligence, and their applications in health and safety.

Luca Chittaro is full professor of Human Computer Interaction (HCI) in the Department of Mathematics, Computer Science, and Physics of the University of Udine, Italy, where he heads the HCI Lab (http://hcilab.uniud.it). He has authored or co-authored over 200 international academic publications, and he is an ACM Distinguished Speaker. His major research interests are in virtual reality, serious games, persuasive technology, mobile HCI, and their applications in health and safety. He has received research grants from organizations such as the US Federal Aviation Administration (FAA) and the European Union (EU), and companies such as the Benetton Group and the Intesa Sanpaolo bank.