Design and assessment of a virtual reality learning environment for firefighters

Steven G. Wheeler[1], Simon Hoermann[1,2]*, Stephan Lukosch[1] and Robert W. Lindeman[1]

[1]Human Interface Technology Lab NZ, University of Canterbury, Christchurch, New Zealand, [2]School of Product Design, University of Canterbury, Christchurch, New Zealand

The use of virtual reality (VR) in firefighter training is promising because it provides cost-effective, safe environments that arouse similar behavioral responses to real-life scenarios. However, the pedagogical potential of VR and its impact on learning outcomes compared to traditional methods is currently an under-explored area. This research investigates how well VR can support learning compared to traditional methods in the context of training firefighters in combating vegetation fires. A VR learning environment was developed, informed by a "design for learning" framework providing a pedagogical underpinning. A between-subjects experiment was conducted with 40 participants to measure the knowledge transfer of the VR learning environment against the official textbook. In addition, VR's theorized learning benefits of intrinsic motivation, situational interest, and self-efficacy were compared with textbook-based learning. Lastly, the design quality of the learning environment was assessed based on its learning and user experience. We employed a primarily quantitative approach to data collection and analysis, using a combination of knowledge test results and questionnaires, with supporting qualitative data from semi-structured interviews and observation notes to answer our hypotheses. The results found a significant difference between the knowledge transfer of both conditions, with textbook-based learning more effectively transferring factual and conceptual knowledge than VR. No significant difference was found in reported self-efficacy between the two conditions but was found in reported levels of intrinsic motivation and situational interest, which were higher in the VR condition. The design was found to have facilitated a good user and learning experience, assessed via questionnaire responses. During interviews, VR participants reported high levels of satisfaction with the experience, praising the hands-on learning approach and interactivity, while reporting frustration with the lack of knowledge reinforcement and initial difficulties with the controls. A key finding was that presence was found to be negatively associated with knowledge transfer, which we theorize to be caused by the novelty of the realistic VR environment distracting participants from the more familiar lesson content. This research contributes to the body of work related to knowledge transfer within VR in this domain while highlighting key pedagogical and design considerations that can be used to inform future design implementations.

KEYWORDS

training, firefighting, virtual reality (VR), human-computer interaction (HCI), pedagogy, instructional design, human factors

1 Introduction

Around the world, wildfires are occurring more frequently with greater severity, a trend set to continue in the future (Huang et al., 2015; Yoon et al., 2015). As such, adequate training in combating and controlling vegetation fires is of the utmost importance, now more than ever. However, due to the inherent dangers of firefighting, practical training can be hazardous to both instructors and trainees alike (Baduel et al., 2015; Kirk and Logan, 2015) and expensive due to the cost of resources and equipment (Engelbrecht et al., 2019). As such, the current approach to teaching how to work safely at vegetation fires relies primarily on textbooks supplemented with video and limited live demonstrations.

Virtual reality (VR) is a promising technology that has been successfully used in training in many fields, such as in the teaching of correct safety procedures in hazardous situations (Ha et al., 2016; Oliva et al., 2019; Ooi et al., 2019), stress inoculation training with soldiers (Wiederhold and Wiederhold, 2004; Stetz et al., 2007), and as a teaching tool in manufacturing (Mujber et al., 2004). Moreover, a high sense of presence in virtual environments has been shown to arouse behavior and physiological responses on par with real-life (Wiederhold et al., 2001; Meehan et al., 2003). VR has the potential to transport the user to an almost-real, safe virtual training environment that can be modified and observed by the educator, opening up the possibility of the technology being an economical, effective, and ecologically valid alternative to real-life training exercises.

While promising, the body of work investigating the feasibility of using VR in firefighter training remains small (Engelbrecht et al., 2019; Wheeler et al., 2021). Furthermore, of the studies investigating the use of VR technology to train firefighters, most focus solely on a system's pragmatic benefits (e.g., cost-effectiveness compared to real-life practical training) and functional capabilities (e.g., its ability to replicate real-life scenarios) (Wheeler et al., 2021). While important, we consider that functional considerations alone do not give a cohesive picture of how effective or applicable a system is within a domain. Instead, holistic investigations considering human factors and a system's impact on psychological, behavioral, and learning outcomes are necessary to assess the effectiveness of VR as a training tool for firefighters (Wheeler et al., 2021).

A systematic review on the assessment of learning outcomes and human factors in the context of VR firefighting training found that the majority of studies chose an urban setting and focused on simulating search and rescue tasks (Bliss et al., 1997; Tate et al., 1997; Backlund et al., 2007; Jeon et al., 2019). Two studies differed in scope: Cohen-Hatton and Honey (2015), which evaluated the training of cognitive skills of commanders, and Clifford et al. (2020), which focused on aerial firefighting and simulating radio communication disruptions in a rural setting. In terms of domains of knowledge acquisition (Krathwohl, 2002), apart from Cohen-Hatton and Honey (2015), none of the reviewed studies focused on factual, conceptual, or metacognitive knowledge, instead focusing on the acquisition of procedural knowledge and the accurate replication of real-world tasks. Rural firefighting settings and factual, conceptual, or metacognitive knowledge acquisition remain largely underexplored.

Furthermore, where current research considers knowledge transfer in firefighter training, none outlines a pedagogical framework or specific learning theory that underpins the system design. This corroborates the findings of Mikropoulos and Natsis (2011), who reviewed 53 papers spanning ten years (1999–2009) concerning virtual learning environments (VRLEs) and concluded that the research reviewed commonly failed to discuss or consider their pedagogical approach when creating the environment. This is supported by Radianti et al. (2020), who reviewed 38 articles focusing on immersive virtual reality for higher education with a search window from 2016–18 and found that 68% of papers did not explicitly mention any learning theory. These findings are significant as using a particular technology alone will not automatically result in deeper learning (Koehler and Mishra, 2005; Fowler, 2015). Rather, learning is an intricate process, and the employed technology is just one component of the overall learning activity (Beetham, 2007). Instead, technological affordances should be considered in relation to the learner and the desired learning outcomes (Gaver, 1991). Understanding the unique learning affordances of VR and their cognitive and affective benefits on the learner can lead to more suitable and effective VR learning activities (Makransky and Petersen, 2021). As such, a key challenge in the creation of effective VRLEs is identifying the unique characteristics of VR and how to exploit them pedagogically (Salzman et al., 1999; Dalgarno and Lee, 2010; Mikropoulos and Natsis, 2011).

This research adds to the growing body of work investigating the feasibility of using VR for firefighting training. A VRLE was created using a holistic 'design for learning' approach, which examines the interrelationship between human factors, pedagogy, and the learning affordances of VR. The VRLE's content focuses on the official *Fire and Emergency New Zealand* (FENZ) training module "Working Safely at Vegetation Fires," a fundamental training requirement for all firefighters in New Zealand. The VRLE's ability and suitability to transfer factual and conceptual knowledge are compared to current textbook-based routines. Additionally, we assess the efficacy of the VRLE's design in terms of its ability to facilitate VR's learning affordances, support psychological and needs satisfaction, and deliver a good user experience. Unlike previous research, which primarily investigated urban firefighting and the acquisition of procedural knowledge in VR, we assess VR's potential to transfer factual and conceptual knowledge in a rural firefighting context. In addition, this study provides a foundation to inform the design of effective VRLEs underpinned by pedagogical considerations and human factors. Through these contributions, we aim to better assess the suitability of integrating VR technology into current firefighting training routines and contribute to the broader domain of research that explores the interaction between VR technologies and learning.

2 Materials and methods

2.1 Participants

Participants were recruited through advertisements posted on social media and University of Canterbury e-mail channels or were contacted directly if they registered their interest in participating in future studies via an official "sign up

sheet" in the HIT Lab NZ. Before taking part, participants were screened for eligibility based on the inclusion criteria (no previous firefighting experience, can speak English at an Upper-Intermediate level (a "B2" as per the *Common European Framework of Reference*) or above, and must have standard or corrected to-standard vision and hearing.) An inducement of a $20 (NZD) gift card was given as a reward for participating.

2.2 Experiment design

Our hypotheses are as follows:

- H1: There will be a difference in knowledge gained by the VR and control groups as measured by the pre/post-intervention knowledge test.
- H2: The level of situational interest reported by the VR group will be higher than the control group.
- H3: The level of intrinsic motivation reported by the VR group will be higher than the control group.
- H4: There will be a difference between the mean differences of the pre/post-intervention levels of self-efficacy reported by the VR and control groups.
- H5: The design of the VRLE facilitates learning.

  - H5.1: The VRLE leverages the unique features of VR and their learning affordances, as measured by participants reporting high levels of presence, agency, intrinsic motivation and situational interest, and low levels of extraneous cognitive load interaction/environment and embodied learning.
  - H5.2: The VRLE delivers a positive user experience, as measured by high reported levels of its pragmatic and hedonic qualities.
  - H5.3: The VRLE sufficiently supports psychological need satisfaction, as measured by high reported levels of autonomy and competency.
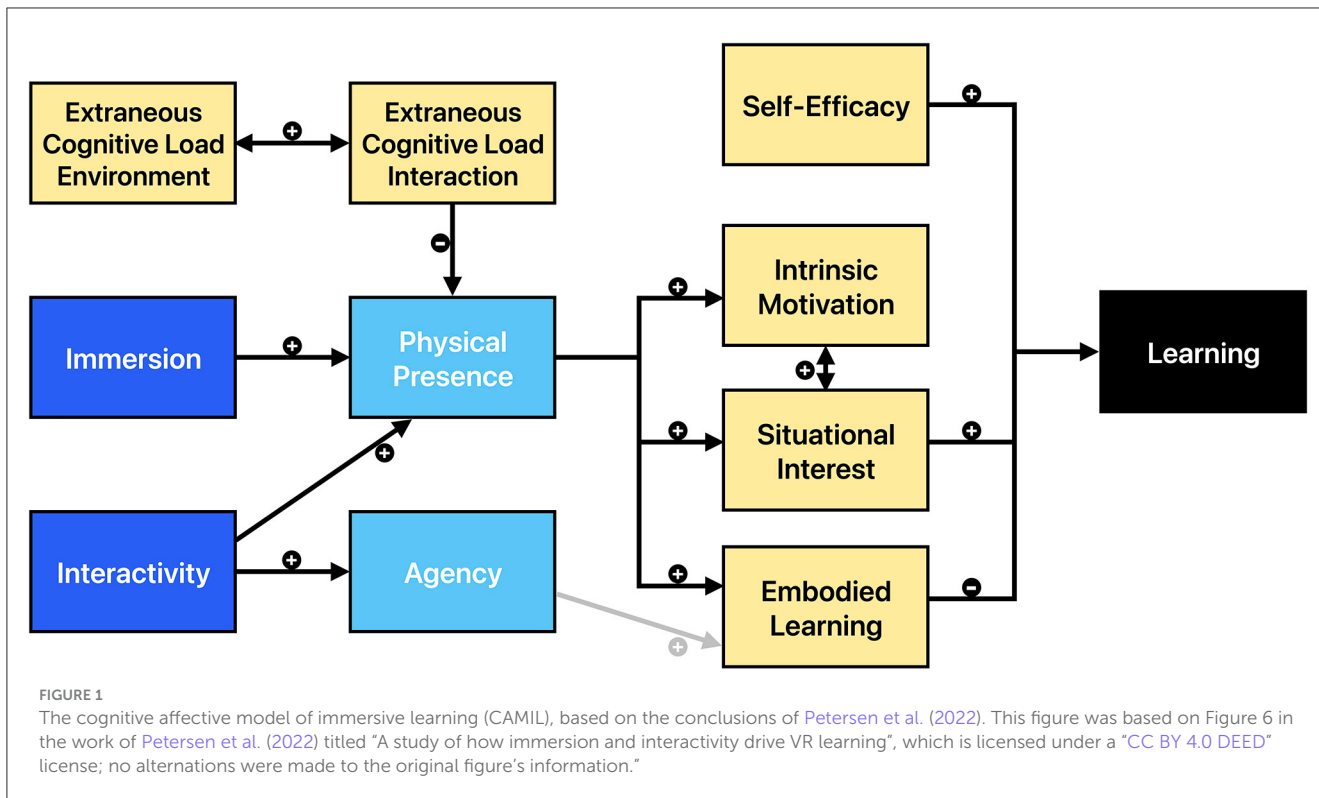
H1 explores the level of knowledge transfer of the VRLE compared to the traditional textbook method. However, beyond knowledge transfer, we wanted to provide a more comprehensive comparison by considering the potential additional learning benefits of the VRLE over textbook learning. To do this, we applied the Cognitive Affective Model of Immersive Learning (CAMIL) (Makransky and Petersen, 2021), which outlines VR's unique features, affordances, and their effect on cognitive and affective factors in relation to learning. The CAMIL was chosen over other models (Ai-Lim Lee et al., 2010; Dalgarno and Lee, 2010) due to pertaining specifically to immersive virtual reality (in place of desktop virtual reality) and representing an up-to-date synthesis of current educational research (Meyer et al., 2019; Makransky and Petersen, 2021). We follow the revised model as proposed by Petersen et al. (2022) in their validation study, which identifies the general affordances of immersive VR's features ("immersion" and "interactivity") as

"presence," "agency," and "representational fidelity"[1] which have the following cognitive and affective factors: Extraneous Cognitive Load Environment (the degree to which the virtual environment contributes to unnecessary cognitive load); Extraneous Cognitive Load Interaction (the degree to which the user interface and interactions contribute to unnecessary cognitive load); Self-Efficacy (the extent to which the users report being confident in the concept covered in the learning environment); Intrinsic Motivation (the degree to which users are motivated to engage with the virtual environment for its own sake); Situational Interest (the degree to which the virtual environment is perceived as interesting by the user); and Embodied Learning (the degree to which the users' actions are mapped to a virtual body in the learning context).

In their study, Petersen et al. (2022) found that both features of VR had a positive impact on physical presence (see Figure 1) but no direct path to learning. However, the authors identified an indirect path via situational interest (covaried with intrinsic motivation), which positively predicted learning and was, in turn, influenced by the reported levels of physical presence, which was found to increase with higher levels of immersion and interactivity in the virtual environment. The situational interest path corroborates previous research (Johnson-Glenberg et al., 2021) showing that student engagement predicted learner's knowledge levels in the post-test. The authors also identified an indirect path via Embodied Learning, which negatively predicted learning. The authors theorized that this negative relation was caused by the lack of congruency of the embodied actions in the lesson's content and recommended further research into this area. Furthermore, the authors found that self-efficacy predicted learning, but VR's features did not influence it. Based on these findings, we analyzed the relative reported levels of situational interest (H2) and intrinsic motivation (H3) between the two conditions. With H4 we aimed to analyse whether there was a difference between reported levels of self-efficacy before and after interacting with either condition. Unlike H2 and H3, the CAMIL shows that the features of VR do not influence self-efficacy. Therefore, we lacked a basis to pose a directional hypothesis. H5 assesses to what degree the VRLE's design facilitated learning. While learning design efficacy can arguably be assessed via its outcomes (H1), many underlying factors can potentially affect the learning experience. As such, we measure the factors outlined in the CAMIL in H5.1 to establish whether the VRLE applied the model appropriately in its design, while H5.2 and H5.3 concern user experience, engagement (with the technology), usability, and user satisfaction, which we consider to be proven measures to establish design quality (Laugwitz et al., 2008; Peters et al., 2018).

We used a between-subjects experiment design, dividing the participants into the control (textbook) or VR group. Due to our hypotheses dealing with many subjective factors related to user and learning experiences, We employed a primarily quantitative approach to data collection and analysis with supporting qualitative data to give us a more nuanced understanding and better answer

---

1 In their validation study, Petersen et al. (2022) chose to keep 'representational fidelity' constant to limit the number of necessary conditions and, therefore, was not validated and is absent from the model.

FIGURE 1
The cognitive affective model of immersive learning (CAMIL), based on the conclusions of Petersen et al. (2022). This figure was based on Figure 6 in the work of Petersen et al. (2022) titled "A study of how immersion and interactivity drive VR learning", which is licensed under a "CC BY 4.0 DEED" license; no alternations were made to the original figure's information."

our hypotheses. The quantitative component involved collecting and analyzing the knowledge test results in combination with questionnaire data. The qualitative component used observation notes and transcripts of post-intervention unstructured interviews on the user and learning experience of the VRLE. Thematic analysis (Nowell et al., 2017) was conducted on the qualitative data to discover relevant crossovers with the collected quantitative data to yield greater insight.

2.3 Measures

We used a knowledge test containing seven questions based on the official assessments used by FENZ for the "Working Safely at Vegetation Fires" module. The knowledge test results were calculated using the official marking rubric. The rubric includes example answers and the pass criteria required for each question. Examples of the pass criteria include "the explanation must include at least two valid points about the effect of each fuel size on fire behavior" or "[answer] must correctly name each stage and give a similar description." For any potentially ambiguous pass criteria, we noted our interpretation of what constitutes a pass to provide additional clarity. We report and compare the knowledge test results in two different ways. First, we report the "official" mark (maximum possible score: 7) the participant would have received if taking the test under real circumstances. Second, we report the 'granular' mark (maximum possible score: 31), which considers any partially correct answers until the threshold of the "official" mark.

The affordances of VR's features, presence (P) and agency (A), and their associated cognitive and affective factors, extraneous cognitive load interaction (ECL_I) and environment (ECL_E),

intrinsic motivation (I), situational interest (SI), self-efficacy (S) and embodied learning (EL), were measured using a five-point Likert scale ("Strongly Disagree," "Disagree," "Neither Agree or Disagree," "Agree," and "Strongly Agree") for each factor, with a minimum of three Likert-Type Items per scale (P: 5, A: 3, ECL_I: 4, ECL_E: 4, I: 5, SI: 6, EL: 3). The items used for each factor were taken from the questionnaire used in the validation study of the CAMIL by Petersen et al. (2022). A composite score for each participant was calculated per scale using the averages of the scores of all items which was used in combination with standard deviation to measure central tendency. Class intervals were calculated for interpreting the composite score, with the length being 0.8 with intervals of 1–1.80, 1.81–2.61, 2.62–3.42, 3.43–4.23, and 4.24–5.04. These class intervals are interpreted as "Very Low," "Low," "Moderate," "High," and "Very High." As with the knowledge test results, self-efficacy was measured by using the difference of the means of the pre/post-intervention results.

User experience was measured by using the *User Experience Questionnaire (Short)* (UEQ-S) (Schrepp et al., 2017), a 7-point Likert Scale scaled from -3 (fully agree with the negative term) to +3 (fully agree with the positive term) consisting of eight items for the short version. The UEQ-S measures the pragmatic (related to being able to achieve tasks or goals) and hedonic (related to the enjoyment or pleasure derived from using the product) qualities of a product (four items each).

The *Technology-based Experience of Need Satisfaction Interface* (TENS-Interface) questionnaire was used to measure engagement (with the technology), usability, and user satisfaction (Peters et al., 2018). The TENS-Interface questionnaire focuses on "the experience of interacting with a technology via its interface during use." The fundamental elements of the questionnaire are based on *Self-Determination Theory* (Ryan and Deci, 2000), which

identifies a set of basic psychological needs essential for designs to promote self-motivation and well-being. These needs are autonomy ("feeling agency, acting in accordance with one's goals and values"), competency ("feeling able and effective"), and relatedness ("feeling connected to others, a sense of belonging"). The questionnaire has individual subscales for each need and is specifically directed toward the technology interface. The questionnaire uses a 5-point Likert Scale, with which we measured each item from 1 (Strongly Disagree) to 5 (Strongly Agree). Each subscale has five items, is randomized, and all are weighted equally in scoring.

At the end of each interview, the participants were asked to rate their overall satisfaction with their respective interface from 1 (Extremely Dissatisfied) to 5 (Extremely Satisfied). A "Customer Satisfaction" (CSAT) score was calculated from all participants' responses with the following formula per group: (Number of scores 4/Number of responses) * 100. CSAT scores have been used effectively in business and are consistently associated with the financial performance of a firm (Mittal and Frennea, 2010). While this measure will not be employed to answer any hypothesis directly, we consider it beneficial due to being an effective indicator of overall participant satisfaction with each interface while eliciting additional comments or lines of discussion during interviews.

2.4 Procedure

Participants were tested individually in a private dedicated study room located in the HIT Lab NZ. The participants were randomly assigned to either the VR or control (textbook) group. Participants were briefed on the purposes of the study, handed an information sheet explaining the experiment protocol, and then allowed to ask questions. Once satisfied, the participant was asked to sign an informed consent form. Participants then answered a demographic questionnaire, four items on self-efficacy, and the pre-intervention knowledge test. VR participants answered two additional questions on simulation sickness and VR experience.

Textbook participants were given their study material in color, photocopies of the self-assessment test, and paper for taking notes. The participants were informed they had thirty minutes to study and would be told when they had fifteen and five minutes left. For VR participants, safety protocols were explained, and the equipment was demonstrated. All VR participants completed a 5–10 min tutorial explaining the virtual environment and controls. Participants had to perform the requested actions in the VRLE to complete the tutorial. Once finished, the participant had thirty minutes to explore the environment and engage with all the content. The time remaining and number of tasks completed were displayed on the participant's wrist within the virtual environment. Observation notes were taken, and the researcher did not intervene unless for the participant's safety or wellbeing.

Post-intervention, all participants were again presented with the same self-efficacy and knowledge test items, followed by two additional sets of questions measuring the participant's level of intrinsic motivation and situational interest. VR participants answered additional questions regarding the learning and user experience of the VRLE. Minus the demographic questionnaire, the order of all questionnaire items and knowledge test questions were randomized. The study finished with a short three-question semi-structured interview where the participant could make additional comments. The participant was asked to list three things they liked and did not like about their respective interface and asked the aforementioned (see "Measures") customer satisfaction question to end the interview.

2.5 Equipment

The experiment used an *HP Reverb G2: Omnicept Edition* VR headset with foveated rendering enabled. The experiment used a laptop, the *Alienware M15*, with an i9-10980HK @ 2.4GHz CPU, 32GB of RAM, and an NVIDIA GeForce RTX 3080 graphics card running Windows 10.

2.6 Textbook

The textbook used was a redacted version of the official "Working Safely at Vegetation Fires" textbook (Fire Emergency New Zealand, 2020) that is issued to trainees in preparation for a training weekend where they receive in-person instruction. The textbook used in the experiment took material from the first two chapters of the official material and only contained content relevant to the knowledge test. The textbook contains a combination of textual explanations, diagrams, content summaries, and self-assessment activities.
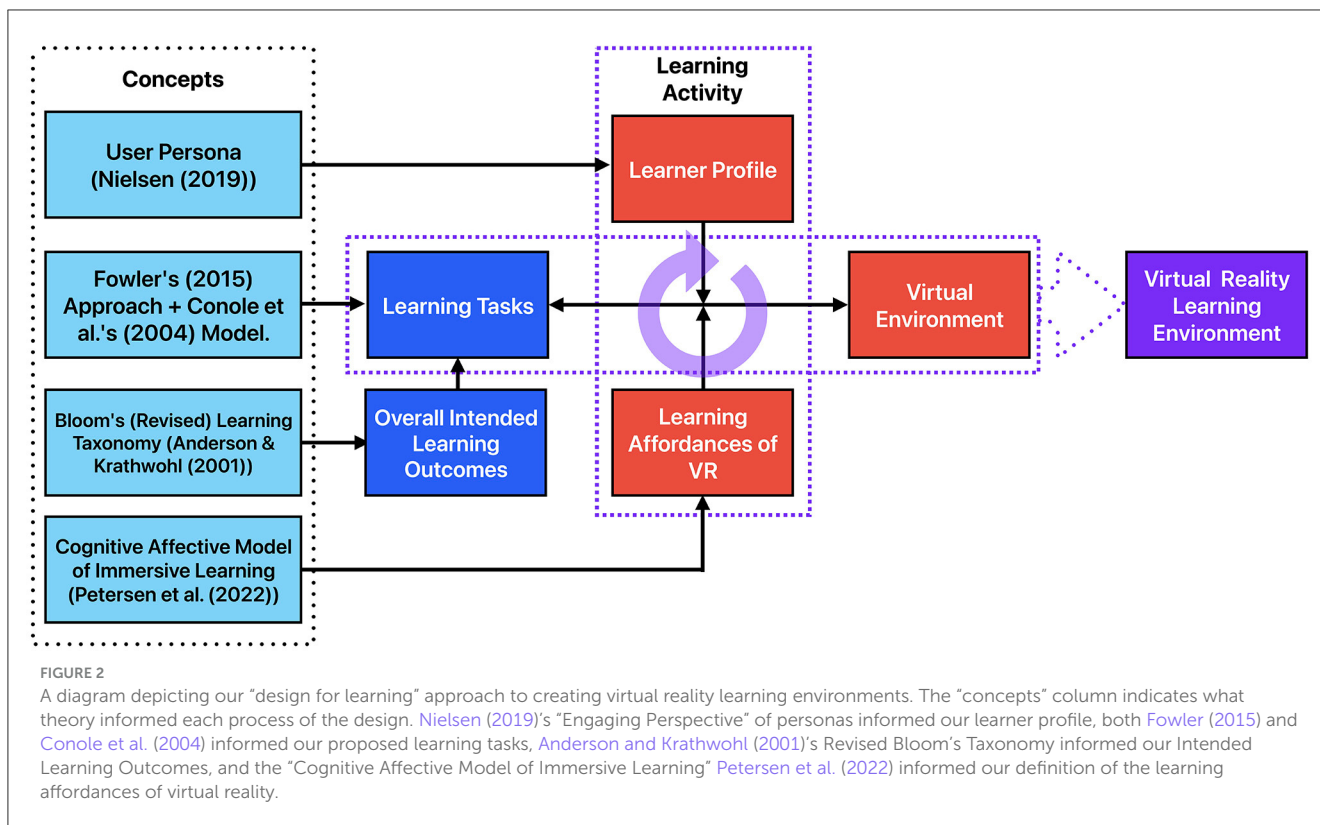
2.7 Virtual reality learning environment

For the purposes of the experiment, we developed a VR learning environment that aims to achieve the same learning outcomes as the textbook, whose material forms the basis of the concepts covered in the VRLE.

2.7.1 Design framework

To create a VRLE that supports learning, the VRLE was designed with the following premises. Firstly, merely using a technology is not enough to afford learning; its content must be designed with the learning outcomes in mind (Koehler and Mishra, 2005). Secondly, the instructional method used in the virtual environment will be specifically effective if it facilitates the unique affordances of the medium (Makransky and Petersen, 2021). Thirdly, consideration of the end-user and their needs must be a primary focus in the design of the application (Agre, 1995; Cooper and Bowers, 1995).

To implement these three principles, the VRLE was designed with a framework based on the recommendations of Beetham (2007), Biggs and Tang (2009), and Fowler (2015) using a 'designing for learning' approach which centers the learner and aims to provide ideal conditions for learning to occur. In this approach, a distinction is made between a learning activity and a learning task. Beetham (2007) defines "learning activities" as "a specific interaction of learner(s) with other(s) using specific tools and

FIGURE 2
A diagram depicting our "design for learning" approach to creating virtual reality learning environments. The "concepts" column indicates what theory informed each process of the design. Nielsen (2019)'s "Engaging Perspective" of personas informed our learner profile, both Fowler (2015) and Conole et al. (2004) informed our proposed learning tasks, Anderson and Krathwohl (2001)'s Revised Bloom's Taxonomy informed our Intended Learning Outcomes, and the "Cognitive Affective Model of Immersive Learning" Petersen et al. (2022) informed our definition of the learning affordances of virtual reality.

resources, orientated toward specific outcomes." Thus, a learning activity can be unpacked into individual components: a learning environment, the learner, "others" (deemed out of scope for this project), and learning tasks orientated toward achieving learning outcomes.

Advocates of this approach stress that learning and task design work indirectly, and the learning activity itself is what takes place during a lesson. That is, although educators may orchestrate certain tasks in advance intended to guide the learner in engaging in certain types of activity, the learners themselves are ultimately autonomous (Conole and Jones, 2010). As such, learning cannot be "designed"; rather, it can only be "designed for" (Beetham, 2007; Laurillard, 2013). Therefore, designing for learning involves an iterative process of guiding and helping the learner engage in certain activities by synergizing the pedagogical requirements of the tasks with the available affordances of the technology and the virtual environment while considering their characteristics and needs. For example, the posited learning task may need to be changed if unsuitable for the technology, or the virtual environment may need to be adjusted to furnish what the task requires. If done adequately, the iterative process should promote well-balanced and effective learning activities (Goodyear and Carvalho, 2019).

The approach outlined in this paper is generic and does not suggest how to derive the necessary information for each section of the framework. Instead, our preferred methods are listed in the "concepts" column in Figure 2 with arrows indicating what each concept informs, which can be substituted if required. The subsequent section will explain the concepts chosen to inform our design.

2.7.2 Concepts

Learning outcomes are usually framed in terms of a noun phrase and a verb phrase (Krathwohl, 2002), with the cognitive process dimension of Bloom's revised taxonomy of learning often used to provide a basis (Beetham, 2007; Biggs and Tang, 2009). We followed the recommendations of Biggs and Tang (2009), who suggest phrasing "intended learning outcomes" (ILOs) as demonstrable actions that the student can perform after being taught. For example, "the student will be able to list the names of the different stages of a fire" or "the student will be able to describe what occurs at each stage of a fire."

To posit suitable tasks to achieve our overall ILOs, we followed a similar approach as Fowler (2015) by listing the learning stages and associated learning outcomes followed by listing general learning tasks [based on the work of Conole et al. (2004)] that could implement the learning outcome (see Table 1). However, Conole et al. (2004) refer to general learning tasks as "mini-learning activities," which we refrain from to maintain the distinction between learning tasks and learning activities as previously defined. As with Fowler (2015), we used the learning stages outlined by Mayes and Fowler (1999), which consist of: "the user's initial contact with other people's concepts" (conceptualization), "the process of building and combining concepts through their use in the performance of meaningful tasks" (construction), and "testing of understanding, often of abstract concepts" (dialogue).

The list of general learning tasks can be applied to Conole et al. (2004)'s model to assess the suitability of the proposed learning task regarding the available tools or resources, the learner, or the desired pedagogical approach. The model consists of three pairings of opposing learning characteristics: Information, Non-reflective, and

TABLE 1 The posited learning tasks arrived at via the same approach as Fowler (2015) and their associated pedagogical alignment based on the toolkit provided by Conole et al. (2004).

Learning stages	Learning outcomes	Generic learning tasks	Specific learning tasks	Conole's model		
				Indv.−social	Non-refl.−Refl.	Expr.−Info.
Conceptualization	To be exposed to new concepts, theories or facts.	Presentation of material.	Live demonstration of concepts using simulations.	-x--------	----x-----	-----x-----
Construction	To build/test/apply theories/concepts.	Apply knowledge. Predict outcomes. Plan experiments.	Adjustable simulations where the user can set their own parameters and observe the results.	-x--------	-------x---	-x--------
	To solve or analyze problems.	Investigate a problem. Apply principles. Theories in new situations.	Interactive simulation games where the user must use previous knowledge to achieve a goal.	-x--------	---------x-	-x--------
Dialogue	To reflect critically.	Self-assessment of level of competence, critique own performance, recognize own limitations.	True/False quizzes based on the knowledge the user has previously learned.	-x--------	--------x--	---------x-

Individual, which are contrasted with Experience, Reflective, and Social. Conole et al. (2004) maps this model onto common learning theories—such as experiential learning or constructivism—which serve as a point of comparison to judge the suitability of the task.

To facilitate a learner-centered approach to inform our VRLE's design, we created a persona of our potential end user (see Figure 3) following the "Engaging Perspective" outlined by Nielsen (2019). Personas are commonplace in many industries (Nielsen, 2003) and are defined as "fictitious, specific, concrete representations of target users" (Pruitt and Adlin, 2010). By putting a face to the user, personas help create empathy and understanding of the user's values, worries, and goals (Grudin and Pruitt, 2002), which results in a consistent conception of the user to the designers (Mikkelson and Lee, 2000).

Design of the virtual environment and its tasks took into account many principles outlined by Clark and Mayer (2016), who provide evidence-based guidelines for designing effective e-learning environments. Each principle was considered based on its theorized applicability to the VRLE's design. The specific details of implementing each principle will be explained in subsequent sections. The considered principles were the following:

- Multimedia principle–"Use words and graphics rather than words alone."
- Redundancy principle–"Explain visuals with words in audio OR text but not both."
- Contiguity principle–"Align words to corresponding graphics."
- Coherence principle–"Adding extra material can hurt learning."
- Modality principle—"Present words as speech, rather than on-screen text."

- Personalization principle–"Use conversational style, polite wording, human voice, and virtual coaches."
- Segmenting principle–"Managing complexity by breaking a lesson into parts."

2.7.3 Learner profile

Our persona (see Figure 3) was based on interviews and findings of previous qualitative research on firefighter identity involving members of the United Kingdom (Baigent, 2001; Hall et al., 2007; Thurnell-Read and Parker, 2008), United States (Kirschman, 2004; Rumsey and Le Dantec, 2019), Swedish (Olofsson, 2013; Holmgren, 2014; Harrison and Olofsson, 2016), Australian (Cowlishaw et al., 2008; Perrott, 2019), and Canadian (Sommerfeld et al., 2017) fire services. The persona notes that introducing any new technology could be met with caution, as there is an emphasis on time-proven, traditional methods over newer alternatives. Furthermore, firefighting training should emphasize hands-on activities with live demonstrations that suitability prepare the recruit for the realities of the profession. Our persona highlights that firefighters take pride in being willing and able to serve the public despite the profession's dangers. As such, the VRLE must be seen as a serious training application instead of a recreational activity or a video game (which 3D virtual environments are commonly associated with).

2.7.4 Intended learning outcomes

The intended learning outcomes listed below express the most optimistic outcomes that could reasonably be expected after interacting with the VRLE. This is based on the expected time the student will spend interacting with the VRLE and their prior knowledge of the subject matter.

Jason Smith

45 years old

Firefighter of 20 years

New Zealand

"The profession of fire fighting is about teamwork, it is about trusting each other. You do not work as a smoke diver without trusting your friend,"

About

Jason joined the fire service due to feeling a strong sense of duty and pride in serving his community and helping the public, despite the dangers involved. While cautious not to glamourise risking one's life, Jason feels there is a dividend to saving lives and helping the public and values the social prestige that comes from the public image of firefighters as heroes and life-savers.

Due to the potential danger and sense of duty of the profession, Jason values being reliable, physically fit and dislikes partaking in any activity that could be seen as him not working. Jason is proud that he has never stepped foot in the 'Rest and Relaxation' room at his fire station and is always ready to answer the call of duty. Jason enjoys working with his fellow firefighters and feels a strong sense of camaraderie with them. For Jason, being dependable and a team player is of the utmost importance.

Jason values tradition and practices that have proven effective over time and is quite cautious of new technologies, initiatives, and external directives. He believes that decisions affecting the fire service should be made by individuals who appreciate the unique challenges and demands of the profession.

Jason feels that current training routines may not best reflect the realities of the profession and leave current recruits underprepared. As such, Jason would prefer training that encompasses the fire services values and needs that focuses on the realities of firefighting via live demonstrations and hands-on activities that involve seasoned instructors.

FIGURE 3

A user persona of the potential end-user of the VRLE based on multiple sources of qualitative research in firefighting identity in the United Kingdom, United States, Canada, Sweden, and Australia. Quote taken the work of Olofsson (2013).
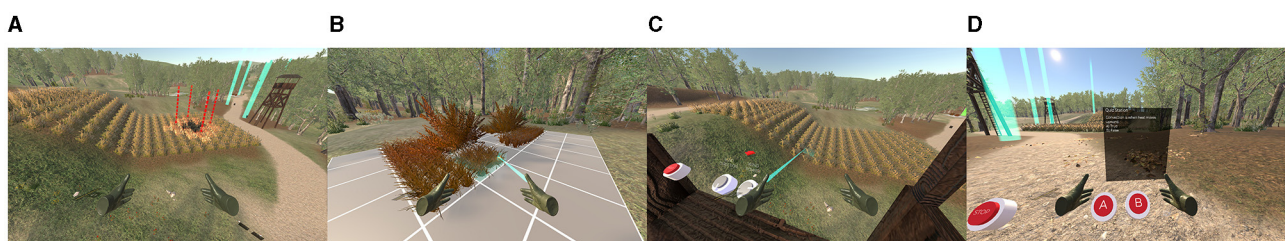

FIGURE 4

The created Virtual Reality Learning Environment. **(A)** A conceptualization lesson: a visualization of how wind and topography affect fire spread. **(B)** The first construction-stage lesson: an interactive simulation where the learner must configure the grid in such a way as to ignite the pre-placed bushes. **(C)** The second construction-stage lesson: the learner is able to change the wind power and direction and choose a point of origin for the fire. **(D)** A dialogue-stage lesson: One of two quizzes in the environment designed to give feedback and encourage reflection.

1. To characterize the different stages and components of a fire, considering their interrelationships and impact on fire behavior.
2. To predict fire behavior and the associated risks based on weather and topography.

To assess whether our proposed ILOs encompass the desired range of cognitive processes and knowledge categories, we overlaid the cognitive process dimension of Bloom's Revised Taxonomy (Krathwohl, 2002) onto the knowledge dimension and placed our ILOs within this matrix. We concluded that both ILOs fit best

within the "analyze" cognitive domain and have a combination of both factual and conceptual knowledge domains. We believe these outcomes to be viable and aim for a sufficiently deep level of learning, which will allow us to explore offering a more hands-on approach to the learner, a need identified in the previous subsection.

2.7.5 Learning tasks

Due to the identified technological and learning affordances of VR, which are more conducive to learner-focused, self-directed activities oriented around self-discovery in meaningful and authentic contexts, we identified that "constructivism" would be the most appropriate learning theory (Conole et al., 2004; Radianti et al., 2020). Furthermore, constructivism seems appropriate according to our created persona of the potential end-user (see Figure 3) who would favor more hands-on, active approaches. A task has been posited for each learning stage to achieve each overall ILO (Table 1): a task to expose the user to the concepts (conceptualization stage), two tasks where the user applies and tests these concepts (construction stage), and then a task where they can reflect critically on their knowledge and approach in the previous tasks (dialogue stage). Each task primarily aligns with our learning theory by leveraging experience and reflection as its primary focus. Implementing the specific learning tasks resulted in nine "lessons" (areas in the VRLE where the learner undertakes one of the posited tasks applied to a specific context): five aimed at the conceptualization stage, two at the construction stage, and two at the dialogue stage.

The five lessons aimed at the conceptualization stage (see Figure 4A) cover key concepts via simulated visualizations, with explanations from the official textbook material. The first construction-stage lesson is (see Figure 4B) an interactive simulation where the user must use previously learned concepts regarding fuel behavior to achieve a predefined goal. The second construction-stage lesson (see Figure 4C) occurs atop a watchtower, where the learner may start a fire anywhere in the paddock below and can alter parameters such as wind direction and power. In both construction-stage lessons, the learner is encouraged to reflect during and after the simulation on whether their predictions or strategy were correct and why. The two dialogue-stage lessons (see Figure 4D) are interactive quizzes where the learner answers true or false questions. On selecting the correct answer, the virtual instructor elaborates on why the answer was correct. These lessons allow users to self-assess their ability and receive feedback on their answers, encouraging reflection and self-critique of their knowledge.

All explanations in the lessons were spoken, with minimal text, and delivered by an AI-generated voice of a New Zealand instructor. Spoken, as opposed to textual, explanations were chosen based on the recommendations of Clark and Mayer (2016) with the "modality principle", which recommends speech over text to enhance learning. Likewise, as the "redundancy principle" recommends not using text and audio simultaneously, the audio explanations are not subtitled. The choice of the instructor's voice and accent and the phrasing of the dialogue were influenced by the "personalization principle," which states that environments should use a conversational but polite style with a human, relatable

tone. Where applicable, *signaling* was used to reinforce audio explanations. Signaling uses cues to direct the learner's attention to key areas to foster connections between the explanations and visualizations (Fiorella and Mayer, 2021).

Based on our persona, we assessed that the potential learner would likely be inexperienced with VR and unfamiliar with its controls. We also concluded that they potentially may value tradition while being cautious toward newer technologies and approaches. Therefore, we implemented more traditional, familiar tasks requiring less demanding environmental interaction. Consequently, the conceptualization and dialogue tasks are less interactive but still align with constructivism by focusing on experience and reflection as the main learning mechanism. Both construction tasks require more active interaction with the environment but emphasize reflection and critical thinking over mechanical skills or hand-eye coordination. Therefore, we believe these tasks strike a balance between leveraging our chosen learning theory, meeting the requirements of the potential learner, and taking advantage of the learning affordances of virtual reality.

2.7.6 Virtual environment

Representational fidelity has been theorized to influence levels of presence within virtual reality environments, which has an indirect relationship with learning (Makransky and Petersen, 2021). Representational fidelity includes variables such as the realistic and smooth display of the environment and object behavior consistent with real-life (Dalgarno and Lee, 2010). As such, we wanted to strike a balance between an ecologically valid setting, adequate application performance, and sufficiently realistic fire behavior.

We chose an authentic and realistic setting, an outdoor forest environment, using photo-scanned models from real locations. However, a point of concern was that implementing an ecologically valid and highly detailed setting could add extraneous cognitive load via visual clutter that could distract the user from focusing on the learning activity. This concern was echoed by Makransky and Petersen (2021), and adding external details not supporting the instructional goal goes against the "coherence principle" outlined by Clark and Mayer (2016). However, the affordance of presence is one of the defining characteristics of immersive VR (Psotka, 1995; Mikropoulos, 2006; Johnson-Glenberg, 2019) and, according to the CAMIL, has a positive influence on learning. Therefore, with respect to the virtual environment, we believed aiming for higher levels of representational fidelity and presence afforded by the aesthetics and behavior of the environment would outweigh the potential risk of additional cognitive load. Care was still taken not to introduce extraneous information that could add to the cognitive load with the visual design of the lessons' content. Furthermore, providing an authentic and realistic context is a core principle of constructivism (Dalgarno, 2002) and would further align the VRLE with it.

Regarding the realistic behavior of the environment, the fire in the VRLE is not based on a mathematical fire model; therefore, its behavior in the environment will not perfectly replicate real life. However, the fire has been designed to behave consistently with the concepts covered by the lessons. While not mathematically

accurate, we consider this level of accuracy to be serviceable in achieving our ILOs. To avoid inducing cybersickness, which has been theorized to be related to lower refresh-rate displays and low framerates (LaViola, 2000), we aimed to provide sufficiently realistic visuals and fire behavior while maintaining the maximum framerate the refresh rate of the VR headset would allow; in our case, this was 90hz.

The tasks within the VRLE are arranged so that the learner focuses on tasks associated with the first ILO and progresses through each stage of learning before moving on to the content related to the second ILO. The content is arranged in such a way as to build upon previously explained concepts as much as possible to help the learner construct a cohesive map of knowledge. The user must travel a short distance to each of these lessons, a design decision made to elicit a sense of presence in the environment, increase the user's familiarity with the VR controls and give the user a chance to reflect and internalize the previous lesson before moving on, further aligning with constructivist principles (Radianti et al., 2020). Furthermore, the approach of dividing topics into discrete lessons where the user can progress at their own pace follows the "segmenting principle" which can better aid learners in processing key information (Clark and Mayer, 2016; Fiorella and Mayer, 2021).

3 Results

3.1 Quality of measures

Cronbach's alpha indicated the scale quality for all measures except the knowledge test results. We defined a threshold of 0.6 being the lowest acceptable alpha, with 0.7 and above being preferable, based on the recommendations of Nunnally (1994). Presence ($\alpha= 0.669$), agency ($\alpha= 0.607$), extraneous cognitive load interaction ($\alpha= 0.681$), extraneous cognitive load environment ($\alpha= 0.673$) and embodied learning ($\alpha= 0.639$) all were found to be acceptable. The internal consistency of situational interest (SI) and intrinsic motivation (I) was assessed for control (SI = $\alpha= 0.897$, I = $\alpha= 0.876$) and VR groups (SI = $\alpha= 0.683$, I = $\alpha= 0.641$) and were judged to be acceptable. The self-efficacy subscale was assessed and observed to be acceptable for pre-/post-intervention in control ($\alpha= 0.932$, $\alpha= 0.887$) and VR groups ($\alpha= 0.883$, $\alpha= 0.806$). The UEQ-S's pragmatic ($\alpha= 0.75$) and hedonic ($\alpha= 0.84$) scales were acceptable. The TENS-Interface of competency was found to be acceptable ($\alpha= 0.72$), but autonomy fell below the required reliability threshold ($\alpha= 0.49$).

3.2 Sample

Forty participants took part in the experiment (20 in each condition). The majority of participants were students from the University of Canterbury. 21 participants were aged between 18 and 24 years old (control: 10, VR: 11), 10 between 25–34 (control: 5, VR: 5), 7 between 35–44 (control: 4, VR: 3), and a single participant between 45–54 in the VR group and one between

55–64 in the control group. 27 participants were native English speakers (control: 13, VR: 14), 6 were advanced English speakers (control: 1, VR: 5), and the remaining 7 participants reported an "Upper-Intermediate" proficiency (control: 5, VR: 1) minus one textbook participant who reported "Intermediate" proficiency but was assessed by the researcher to have a sufficiently high enough level of English to participate. Of the 20 participants in the VR group, only three identified that they use VR more often than once a month, with the majority either having no experience ($n = 4$) or less than once a month ($n = 10$). Only 6 participants in the VR group responded that they had previously experienced simulation sickness when using virtual reality.

3.3 Demographic data

As all demographic variables were ordinal data, we used a Kruskal-Wallis H test to establish a difference in the distribution of knowledge test results across the various demographic categories. VR experience was not found to have any difference in the distribution of knowledge test results ($p = 0.168$). However, a statistically significant difference in knowledge test results was found between age categories among VR participants ($p=.043$); no similar statistically significant difference was found in the textbook group ($p = 0.919$). Dunn's pairwise tests were carried out for the different age ranges adjusted using the Bonferroni correction. However, no pairing reached statistical significance, with the lowest p-value being 0.025 before being adjusted to 0.151 between the 25–34 and 18–24 groups. A statistically significant difference in knowledge test results was found between participants' English proficiency in the control group ($p = 0.028$); no significant difference was found in the VR group ($p = 0.150$). Dunn's pairwise tests were conducted for the different English proficiency levels adjusted using the Bonferroni correction. The pairwise tests showed Upper-Intermediate English speakers ranked lower ($z= -7.877$) than Native English speakers with a borderline statistically significant difference after adjustment ($p = 0.058$). No statistically significant difference in distribution was found between simulation sickness groups and knowledge test results ($p = 0.691$).

3.4 H1: there will be a difference in knowledge gained by the VR and control groups as measured by the pre/post-intervention knowledge test

Hypothesis 1 was supported. In the "official" knowledge test results, the control group obtained a mean difference between pre-/post-knowledge tests of 4 (S.D. 1.62) and VR 3.05 (S.D. 1.63). The means met assumptions of equal variances ($p = 0.248$). A two-tailed independent-sample Student's **t**-test failed to reach statistical significance, indicating there is insufficient evidence to reject the null hypothesis ($t_{(38)} = 1.84$, $p= 0.073$).

However, we performed a two-tailed independent-sample Student's t-test to compare the mean difference between the

pre-/post-intervention knowledge tests' "granular" results, which found a significant difference ($t_{(38)}$ = 2.12, p = 0.04). In the "granular" results, the control group obtained a mean difference of 19.8 (S.D. 5.28), higher than the VR group, which obtained a mean difference of 16.1 (S.D. 5.72). The means met assumptions of equal variances (p = 0.827). One-tailed paired-sample t-tests were conducted between the means of the pre and post-intervention "granular" knowledge test results for both groups. The t-test found a significant difference in pre/post-intervention results for both control (p = <0.001) and VR (p = <0.001) groups. Therefore, while there was no significant difference in the "official" knowledge test results, H1 is substantiated based on the significant difference between control and VR groups found in the "granular" knowledge test results. Furthermore, it was found that participants of both groups performed better in the post-intervention knowledge test, with participants of the control group showing the most improvement.

To investigate further, we divided the "granular" knowledge test results into whether the question concerned "factual" or "conceptual" knowledge, based on the knowledge domain definitions provided by Krathwohl (2002). The marking criteria remained the same as the "granular" test results (explained in Section 2.3), except factual knowledge was no longer bound by the official marking criteria's threshold. In practice, this meant that "question 2," which asked participants to name the parts of a fire, had eight possible points, whereas this is limited to four points in the official marking criteria. The factual questions had a total possible score of 11, and the conceptual questions had a score of 28. A two-tailed independent-sample Student's t-test was conducted on the mean differences of the pre/post-knowledge test results for factual and conceptual knowledge between control and VR groups. There was a significant difference between the mean differences of both factual ($t_{(38)}$ = 2.527, p = 0.016) and conceptual ($t_{(38)}$ = 2.243, p = 0.031) knowledge results between both groups. One-tailed paired-sample t-tests were conducted on the means of the pre and post-intervention knowledge test results. The t-test found a significant difference in pre/post-intervention results for both factual (control: p = <0.001, VR: p = <0.001) and conceptual knowledge questions (control: p = <0.001, VR: p = <0.001). For factual knowledge, the mean difference of the control group (M 8.65, S.D. 2.25) was higher than that of the VR group (M 6.4, S.D. 3.28). For conceptual knowledge, the mean difference of the control group (M 14.3, S.D. 4.62) was higher than that of the VR group (M 11, S.D. 4.68).

3.5 H2: the level of situational interest reported by the VR group will be higher than the control group

Hypothesis 2 was supported. The means did not meet assumptions of equal variances (p = 0.02). As such, we conducted a comparison of means with the Mann-Whitney U test, which found a statistically significant difference between the means (U = 80, p = 0.001), with a higher reported level of situational interest in the VR group (M 4.48, S.D. 0.39) compared to the control group (M 3.54, S.D. 0.93).

3.6 H3: the level of intrinsic motivation reported by the VR group will be higher than the control group

Hypothesis 3 was supported. The means met assumptions of equal variances (p = 0.293). A one-tailed independent samples Student's t-test ($t_{(38)}$ = 2.509, p = 0.0085) in the direction of the VR group revealed a statistically significant higher reported level of intrinsic motivation in the VR group (M 4.27, S.D. 0.57) compared to the control group (M 3.69, S.D. 0.85).

3.7 H4: there will be a difference between the mean differences of the pre/post-intervention levels of self-efficacy reported by the VR and control groups

Hypothesis 4 was not supported. The mean difference between reported levels of self-efficacy pre- and post-intervention for the VR group was 1.41 (S.D. 1.14) and 1.34 (S.D. 1.16) for the control group. The means met assumptions of equal variances (p = 0.965). A two-tailed independent samples Student's t-test failed to reach statistical significance to reject the null hypothesis ($t_{(38)}$ = 0.2051, p = 0.8385). One-tailed paired-sample t-tests were conducted on the means of the pre and post-intervention self-efficacy results. The t-test found a significant difference in pre/post-intervention results (control: p = <0.001, VR: p = <0.001.) Therefore, while H4 was not supported as no difference between the mean differences could be found, participants from both conditions reported higher levels of self-efficacy post-intervention.

3.8 H5: the design of the VRLE facilitates learning

3.8.1 H5.1: the VRLE leverages the unique features of VR and their learning affordances

Hypothesis 5.1 was supported. The Presence subscale obtained a composite score of 4.08 (S.D. 0.62), Agency a score of 4.13 (S.D. 0.7), and Embodied Learning a score of 4 (S.D. 0.89), placing the subscales within the "High" interval (3.43–4.23). Extraneous Cognitive Load Interaction received a score of 1.83 (S.D. 0.81) and Extraneous Cognitive Load Environment a score of 1.55 (S.D. 0.72), placing both scores within or close to the "Very Low" interval (1.00–1.80). All subscales were in their hypothesized intervals with the exception of Embodied Learning (which was expected to be "Low" to "Very Low"). To explore this, linear regression analysis was conducted between Embodied Learning and the knowledge test results. However, while the reported coefficient suggests a negative relationship (t = −1.395), the p-value (0.180) suggests it is not statistically significant. Combined with the previously covered composite scores for intrinsic motivation (M 4.27, S.D. 0.58, "High" interval) and situational interest (M 4.48, S.D. 0.85, "Very High" interval), these findings align with the expected ranges outlined previously, thereby supporting the hypothesis.

3.8.2 H5.2: the VRLE delivers a positive user experience

Hypothesis 5.2 was supported. Responses to the UEQ-S resulted in an overall mean score of 2.006 (S.D. 0.69), with both pragmatic and hedonic qualities receiving responses of 1.925 (S.D. 0.73) and 2.088 (S.D. 0.82) respectively. These findings fall within the 10% best results ("excellent") range based on a benchmark data set involving 21,175 participants across 468 studies, thereby indicating a positive user experience.

3.8.3 H5.3: the VRLE sufficiently supports psychological need satisfaction

Hypothesis 5.3 was partially supported. Participants reported high levels in both competency (M 4.27, S.D. 0.57) and autonomy (M 4.01, S.D. 0.58) subscales, which place competency within the "Very High" class interval and autonomy within "High," thereby indicating strong support for psychological need satisfaction as reported by participants. However, the internal consistency of the autonomy subscale fell below acceptable thresholds and cannot be used to assess this facet of the TENS-Interface questionnaire reliably.

3.8.4 Customer satisfaction evaluation

Participants' responses to the customer satisfaction question outlined in the "Measures" section resulted in the VRLE receiving a 95% satisfaction rate and the textbook 65%.

3.9 Qualitative data

Based on the approach of Nowell et al. (2017), thematic analysis was conducted on the interview transcripts and observation notes. The data were codified and organized broadly based on the participant's group and whether the code was positive or negative in sentiment. From these codes, common themes were identified and refined, resulting in the following: visualization, reinforcement and dialogue, immersion and representational fidelity, interactivity and active learning, intrinsic motivation and situational interest, and technical issues and real-world factors. The following section will explain these themes in turn.

3.9.1 Visualization

A common theme between both groups was that participants appreciated the use of visual aids. Textbook participants enjoyed the diagrams and images, which helped them visualize the information in the text. Participants described the images and diagrams as "very helpful" and "clear," with one participant noting that "if you just tell me that with text. I wouldn't understand." Another participant notes that they liked the diagrams as they are a "visual learner" and another that they appreciated the images or diagrams as opposed to a "wall of text." However, while positively praised, certain diagrams were criticized for being unclear or contradictory. Participants stated that they thought some diagrams were "not very clear," that they were "confused" or that they "misunderstood them," with some stating the diagrams

or images were "ambiguous" or "inconsistent." Participants also stated they would like to supplement the information with videos and realistic images to enhance their understanding and avoid confusion.

Similarly, VR participants liked the use of "live" demonstrations via simulations, with participants stating "it felt much more salient than just picturing and reading the words and having to kind of imagine the scenario," "I liked being able to see what it was telling me actually happen." Participants also stated they enjoyed being able to change the frame of reference by moving around the environment to view the simulation from different perspectives, with participants stating they liked "to see things more clearly and everything is very close set up," "I liked to be able to move around and look at the different parts of it" and "I could go right into the fire. I quite liked that."

3.9.2 Reinforcement and dialogue

VR's quiz lessons (dialogue-stage lessons) were praised (n = 5) for "making sure you're actually retaining the information," with one participant stating that "[the quizzes] allowed me to go over what I learned." However, VR participants also expressed frustration (n = 6) at being unable to remember core pieces of information adequately, especially the specific names of concepts, and wanted further reinforcement beyond the quiz lessons available. Participants commented that some facts stated in lessons were "never mentioned again," others expressed frustration that they understood the concept but not the specific name; e.g., "I knew what they did, but I couldn't recall the names" and "I know what object that is but I can't remember the exact term for it... it makes it harder to perhaps pass a quiz or test."

A common suggestion (n = 7) was to augment the VR visualizations (conceptualization-stage lessons) with accompanying text and summaries to better relate the visuals with the underlying theory or to better process the information being relayed by the voice-over. One participant commented that "it's difficult for you to remember the specific names of the things because you haven't seen them." This feeling is shared by another participant who stated that "I need to read things as well as hearing them, to actually remember them." Furthermore, one participant suggested that "I would probably do better if I was to study the [textbook] and then go and do [the VRLE] afterward because then I could put the two together."

Conversely, textbook participants (n = 10) highly praised the summaries for helping reinforce and focus on the necessary concepts. Participants commented that the summaries were "helpful," "positive," "highlight the important information," and gave "a quick overview of everything." Furthermore, textbook participants enjoyed its "browsability," which allowed them to quickly return and revise topics they may not have fully understood or internalized; e.g., "I think it's easy to flick through and go back and forth" and "I could go back and read over things... if I wanted to review my notes."

VR participants noted that the VRLE does not afford such browsability, with no opportunity to return to previous lessons (due to the experiment's time constraints) or easily skip through a lesson's content, with the quiz lessons being the only way to revise and reinforce their knowledge; e.g., "I need to go back... and

go over the lesson." As such, a suggestion from VR participants was to have the ability to skip back and forth through the lesson's content, as is possible with video playback, instead of having to start each lesson from the beginning. One participant mentioned they would have liked to "go back if you didn't hear something properly;" another was frustrated that they "can't go back to re-hear what [the voice] was saying without stopping and starting the lesson."

Likewise, the textbook self-assessment quizzes were a common source of praise (n = 7), allowing the participants to focus, condense the previously covered content, and reflect critically; e.g., "it helped me. review and find what I hadn't learned. So it narrows down information a bit" and "the quiz asked something which I couldn't get from the summary, [I] had to go back to the details." However, the participants felt frustrated because the textbook had no feedback, which hindered their ability to engage in the dialogue-stage of learning. That is, the self-assessments in the textbook had no answers, and the participants had to determine whether they were correct by revising the relevant information. While potentially positive as it encourages reflection and critical thinking, the lack of confirmation of whether their answers were correct was frustrating; e.g., "I can't ask questions and have them answered in real-time," and "There were no answers. I had to infer things from the text... inductive reasoning." This was not a criticism of the VRLE's quizzes, which gave the correct answer after each question attempt.

3.9.3 Immersion and representational fidelity

During interviews, over half of VR participants discussed immersion, presence, and representational fidelity (n = 14). Participants stated that they were "immersed in the world" and felt "physically being there in the actual environment," as if they "co-existed with nature, fire and all the elements there." Participants described the graphics of the virtual environment as "real" or "realistic," and appreciated the forest setting, referring to it as "beautiful" and saying that they were "blown away." Others stated that the aesthetics "matched the topic" well and that "all of the trees were part of the context," which helped contextualize the learning and made participants appreciate "the importance of the environment to the firefighters." Participants did not explicitly discuss fire behavior, minus one participant saying they would have preferred less grid-like fire spread in the second construction lesson (Figure 4C) and another that they had difficulty trusting the realism of the fire and would have liked to see video footage of a real fire as a point of comparison.

As previously theorized in Section 2.7.6, the high ecological validity and sense of presence were reported and observed to have a negative effect. Some participants (n = 4) commented that they found it difficult to focus on the subject matter content due to wanting to investigate and explore the rich environment. One participant stated that the "only negative is because I was so busy being amazed and blown away [by the virtual environment]," and another stated that they "wanted to look here and there. So, nature is kind of this distraction in a way." These comments are consistent with observations, with one participant remarking that they could not pay attention to the lesson due to the scene's beauty, and two

others were observed looking around the environment instead of toward the lesson's content.

3.9.4 Interactivity and active learning

The majority of VR participants (n = 13) reported they appreciated the active, hands-on learning that the VRLE afforded and the interactive elements of the environment (n = 10). The most commonly praised areas were the construction-stage lessons, where participants enjoyed the freedom to explore and construct their own knowledge. Participants commented that they enjoyed the experimental nature of these lessons, such as "being able to control the wind speed and where the fire started and building the fire" and "being able to interact with a few different possibilities and seeing how they affect the end result." Others commented that "I could test out stuff that the lesson didn't show me" and "I can really see how [the fire] begins and ends, and it's very educational." The hands-on interactivity of these lessons was praised, with participants commenting that they enjoyed "being able to place the things" and that they liked VR as they "actually started the fires" instead of only observing. Another participant commented that interacting with the environment "made the practical side of things more applicable." It was observed that most participants reflected aloud and made conscious choices and strategies during construction-stage lessons based on knowledge previously covered in the VRLE.

The interaction technique used in the environment was an initial source of frustration for participants (n = 11). Switching between two modes of locomotion was a commonly cited and observed frustration, with participants commenting that it was "hard," "weird," "strange" or "confusing." Others cited the control scheme (i.e., the actions assigned to each button) as unintuitive, with four participants mentioning they often confused the actions of the "grab" and "trigger" buttons. Confusion regarding the control scheme was a common observation, especially during high-interactivity tasks such as in the first construction-stage lesson (Figure 4B). However, after an initial period of confusion, all participants were notably more comfortable with the controls and were more confident in their movements and interactions by the end of the intervention period. During interviews, participants stated that they "very quickly got the hang of [the controls]," and "I just had to spend some time figuring out how to do the interactive portions. But yeah, that's not a big problem."

3.9.5 Intrinsic motivation and situational interest

Most textbook participants (n = 14) reported that the textbook was "familiar," "ordinary," or "average." However, many (n = 6) stated that they found the content interesting in itself (situational interest), which motivated them to study. Participants stated that "most of my excitement was because of the content," the subject was "so interesting" and that "it was kind of cool to read about it." However, some (n = 6) textbook participants were observed lacking focus in the final ten minutes of their study time, with some stopping studying completely. The majority of VR participants cited that VR was novel (n = 9) (e.g., "state of the art," "it's my first time") and "engaging," "fun" or "exciting" (n = 7). VR participants cited the interactivity and environment as motivating factors (n = 4) (e.g., "it felt like you were playing a game, but you were

learning," "it was fun because it was interactive and you had an environment so it wasn't just like you're in a grey box") with others citing novelty ($n = 3$) (e.g., "because it's new it just holds your attention a lot better than something that's a bit more boring"). Discussion regarding the intrigue of the subject itself was minimal (i.e., situational interest), with only one VR participant explicitly mentioning they enjoyed the topic. While no VR participant asked to stop the intervention before the time limit was reached, there was only a minority of occasions where a significant surplus of time remained after completing all tasks.

3.9.6 Technical issues and real-world factors

Bugs or technical difficulties occurred in half of all VR interventions. The most common occurrences of bugs were found within the high-interactivity construction-stage lessons, with the majority being usability issues (problems using the interface to change parameters or configure the environment) and an issue with the fire not behaving correctly during the second construction-stage lesson (Figure 4C). One participant expressed concern about colliding with the real environment, stating that "[the real space] felt too small. So I didn't want to walk toward [the lessons]." Another mentioned the VR headset was uncomfortable, and two participants also cited that they were physically tired from using the VRLE and would have preferred to be seated during the experience. Regarding simulation sickness, some participants felt slightly dizzy or unsteady while using VR ($n = 4$), citing the locomotion technique as the source, with one participant stating that "moving is a little bit faster than I expect" and another stating "Oh, the movement is a little bit dizzy[ing]." Initial unsteadiness when using the "free movement" locomotion technique (as opposed to teleportation) was also a common observation. Lastly, the experiment time limit of 30 min was cited as a concern by three VR users, who felt they had to rush through the experience; e.g., "I felt a bit rushed by the time limit, and having to think inside of a time frame," "I don't know if you could go back to the lessons or not and redo them, but I guess the time frame [was a negative]."

4 Discussion

Our study aimed to assess the ability and suitability of using VR to transfer factual and conceptual knowledge compared to traditional textbook methods for training firefighters. The efficacy of the VRLE was assessed based on the knowledge test results compared to the textbook and the related affective factors of intrinsic motivation, situational interest and self-efficacy. Lastly, the VRLE's design was validated through quantitative and qualitative methods.

Regarding hypotheses 2 and 3, VR participants reported higher levels of situational interest and intrinsic motivation than their textbook counterparts. The qualitative data corroborate our quantitative findings, where most VR participants described the VRLE as novel or engaging, while the textbook was met with less enthusiasm. This trend was also reflected in the "CSAT" score, where VR participants were much more satisfied with the VR environment than the textbook participants. No statistically significant link was found between VR experience and levels of

intrinsic motivation and situational interest, thereby indicating the higher reported levels of these factors cannot be explained by the user being less accustomed to VR. Due to its higher levels of intrinsic motivation, situational interest, and satisfaction, the VRLE could potentially yield greater knowledge transfer over a longer period, especially under self-study conditions, where self-motivation is key. Regarding hypothesis 4, no significant difference was found in reported levels of self-efficacy between the mean differences of either group. However, both groups did show an improvement post-intervention.

Our findings from the first hypothesis (H1) indicate that the textbook was more effective than the VRLE at transferring knowledge, whether factual or conceptual. While many potential explanations exist for this result, a few salient issues arose during quantitative and qualitative analysis. The first potential explanation involves the pedagogical approach and design of the VRLE's content. While the VRLE was praised for being hands-on and interactive, many VR participants expressed frustration with the lack of text accompanying the audio explanations and knowledge reinforcement in the VRLE. Conversely, the textbook group heavily praised its quizzes and summaries, citing that they helped reinforce and retain knowledge. The lack of accompanying text in the VRLE was noted as particularly difficult in learning factual knowledge, with participants stating they would have preferred to see the names of the concepts rather than just listening to them. As a result, participants reported that they understood the concepts but could not remember the exact names. The absence of text accompanying the audio explanation was a design decision made to follow the "redundancy principle" proposed by Clark and Mayer (2016), which states people learn more deeply in multimedia lessons with either audio explanation or text, but not both. However, Clark and Mayer (2016) note that exceptions can be made when the key concepts are not already known to the user, learners are not native speakers of English, or when only a select few words are shown on screen. As such, at times in the VRLE, certain concepts were labeled sparingly to facilitate the learning of factual information but perhaps could have been used to greater effect. Similarly, random and quick access to information and the ability to rapidly revise key topics ("browsability") were praised by textbook participants, with VR participants noting they felt the VRLE lacked this affordance. While such browsability is a fundamental advantage of a "non-transient" medium such as a textbook, where the user is in complete control of the pace and order of the content's consumption, the VRLE could have been designed to take into account the reinforcement of key concepts more explicitly. For example, the VRLE could have furnished the environment with information stations that refresh key concepts or ask the learner to reflect on a concept. However, this frustration could also have been compounded by the experiment design, which only allowed the participant thirty minutes in the VRLE. With prolonged exposure to the VRLE, it is possible that the need for greater levels of knowledge reinforcement or rapid access to information would be diminished, due to less pressure from a strict time limit.

Another potential explanation for the difference in both group's knowledge test results is based on the concept of cognitive load (CL). Cognitive Load Theory (CLT) (Sweller, 2011; Mayer, 2021) focuses on relating how working memory and cognitive resources are used while learning. CLT distinguishes between three types of

CL: intrinsic CL, extraneous CL, and germane CL. Intrinsic CL refers to the inherent complexity of the subject matter, whereas extraneous CL refers to the additional mental effort required due to the delivery of the content. Germane CL refers to "effective" CL, which is the percentage of working memory dedicated to processing intrinsic CL compared to extraneous CL. Research suggests that VR produces higher extraneous CL than less immersive mediums (Makransky et al., 2019; Meyer et al., 2019), making CL an important factor to consider when creating effective VRLEs. Two types of extraneous CL are incorporated into the CAMIL: extraneous CL caused by the environment (ECL_E) and by the interaction technique (ECL_I).

Participants reported "Very Low" levels of ECL_I when responding to the post-intervention questionnaire (Section 3.4), which measured the factors of the CAMIL. Furthermore, the "pragmatic" subscale of the UEQ-S and the "autonomy" subscale of the TENS-Interface (Section 3.6) both pertain to the quality of the interaction technique, with both factors rated highly by participants. However, considering observations and interview responses, answering to what extent the interaction technique potentially contributed to extraneous CL appears more nuanced than the questionnaire responses suggest. Participants were introduced to the basics of the interaction techniques in the pre-intervention tutorial, where they were asked to perform the actions before continuing. However, it was observed that any interaction with the environment was initially confusing, even if the participant was familiar with the technique due to the tutorial. Interaction techniques involving grabbing, turning, or placing objects proved especially difficult for many learners. Movement controls were noted as being initially confusing, and participants felt unsteady or surprised when using the "free movement" locomotion technique, with four participants reporting mild feelings of simulation sickness (Section 3.9.6). While all participants became accustomed to interacting with the VRLE, as observed and reported in interviews (Section 3.9.6), the initial learning period required was notable. As such, it is likely that participants were often focused on understanding how to interact with the environment (extraneous CL) rather than on reflecting on the content itself, especially in interactive-heavy activities, such as either of the construction-stage lessons (Figures 4B, C).

External real-world factors could also potentially cause extraneous CL, such as the user colliding with their real-world surroundings, worrying about their real-world surroundings (e.g., being concerned about collisions), external sounds, technical glitches, or uncomfortable VR equipment. Such factors can be reflected in the presence levels reported by the users, with external interruptions to the VR experience referred to as a "break-in-presence," which can result in lower levels of immersion and presence (Slater et al., 2003). However, the mean composite score given by the participants for presence was in the "High" interval, suggesting that any break-in-presence did not substantially impact immersion or cause severe extraneous CL. No technical issues affected the application's performance, which ran at the maximum frame rate that the refresh rate of the VR headset allowed (120 hz). Furthermore, most participants were not concerned about the size of the physical play area nor reported any discomfort caused by the headset. Therefore, we believe the data and observations indicate that technical issues and real-world factors did not strongly affect the learning and user experience of the VRLE or significantly contribute to extraneous CL.

During the design phase (Section 2.7.6), we theorized that focusing on high representational fidelity could potentially cause extraneous cognitive load by introducing distracting visual information, a sentiment shared by Makransky and Petersen (2021) and warned against by Clark and Mayer (2016) with the "coherence principle." However, as with "Extraneous Cognitive Load Interaction," the reported levels of "Extraneous Cognitive Load Environment" were in the "Very Low" interval (Section 3.8), indicating that the environment did not cause high levels of extraneous cognitive load. Yet, during interviews, four participants reported that, while they enjoyed the environment, it was visually interesting to a fault and drew their attention away from the lesson's content (Section 3.9.3), and two more participants were observed inspecting the environment as opposed to looking at the lesson's content.

While it is possible that the reported and observed distractions were not significant, there were potentially insufficient measures to explore whether the environment was a distraction. Of the four items measuring ECL_E, two items were relevant to the topic of focus or distraction: "The virtual environment was full of irrelevant content" and "The elements in the virtual environment made the learning very unclear." However, it is possible that the participants may not have interpreted the terms "irrelevant content" or "elements in the virtual environment" as discussing superfluous or distracting environment features but rather whether the lessons in the virtual environment were relevant or clear. Another path to explore the topic of the environment distracting the user is to look at "Situational Interest," which covers relevant questions such as "Did the lesson capture your attention?" and "Were you concentrated on the lesson?" Yet, as covered previously (Section 3.8), the reported levels of Situational Interest were in the "Very High" interval, indicating that the participants felt they could pay attention to the lessons and were not distracted.

However, while exploring the topic of the virtual environment distracting the user or contributing to extraneous cognitive load, a significant negative relation was found between presence and knowledge test results ($f^2 = 0.645$, $\beta = -1.652$, $p = 0.003$) which we believe could be related. This finding contradicts previous research (Winn, 1993; McCormick and Wickens, 1995; Psotka, 1995; Wickens and Baker, 1995; Salzman et al., 1999; Mikropoulos, 2006; Johnson-Glenberg, 2019), including the CAMIL (Petersen et al., 2022), which suggests a positive relationship between presence and learning. Our initial theory for this finding was a potential increase in extraneous CL caused by high levels of presence, which could negatively affect learning. However, reported levels of ECL_E were low and, consistent with Petersen et al. (2022)'s findings, linear regression analysis found no significant relation between presence and ECL_E ($p = 0.176$). Therefore, we were unable to sustain this theory with our current measures.

An alternative theory was that presence negatively predicted knowledge test results indirectly via Embodied Learning, as shown in Petersen et al. (2022)'s validation study of the CAMIL. In our own tests, a statistically significant positive relation between Presence and Embodied Learning was found ($f^2 = 0.984$, $\beta = 1.012$, $p =$

<0.001). However, our tests failed to reach statistical significance when conducting a regression analysis between Embodied Learning and the knowledge test results ($p = 0.180$). If such a relation did exist, as suggested by the CAMIL, this could potentially explain the negative relationship of presence on knowledge test results due to the reported levels of Embodied Learning reported by the participants falling in the "High" interval.

Another explanation was a potential negative relation between presence, novelty, and focus. Presence has been theorized to be stimulated by novelty, where people tend to be more aroused and broadly focused in new and unique situations (Fontaine, 1992; Witmer and Singer, 1998). Conversely, more familiar situations and activities require much less focus and inspire less presence. As such, we theorized that a potential explanation of the negative relation between presence and knowledge test results could be due to the "novelty factor" of VR. That is, while the user is highly present and focused on the novel virtual environment, they are potentially less focused on the more familiar scenarios of theoretical lessons or quizzes. Novelty refers to the unexpected, surprising, or new (Huang, 2003). For VR applications, the novelty factor can be particularly high due to VR technology not being as widespread as other mediums (Miguel-Alonso et al., 2023). However, the potential effects of the novelty factor are not only caused by the uncommon VR equipment but also by any new and unfamiliar environment (Huang, 2003). As such, all new users of the VRLE will experience varying degrees of the novelty factor, which could be especially apparent in participants less experienced with VR due to interacting with a greater number of unfamiliar experiences.

As such, we first investigated this theory by examining the relationship between the participants' experience with VR, presence, and knowledge test results. While no difference was observed in the distribution of the knowledge test results between the groups of VR experience in the "Demographic Data" subsection in Section 3.3, we conducted further Kruskal-Wallis H tests on the "VR experience" categorical data to examine the distribution of Presence and Extraneous Cognitive Load Environment and Interaction. However, no statistically significant variance in distribution was observed with Presence ($p = 0.464$), Extraneous Cognitive Load Environment ($p = 0.069$), or Interaction ($p = 0.276$).

Regardless of VR experience, the unfamiliarity and novelty of the VR environment could distract the user by arousing a higher sense of presence and focus than the traditional theoretical content. The novelty effect can manifest as an increased motivation to use something or increased perceived usability (Koch et al., 2018), with both factors lessening when novelty fades. As such, a plausible way to measure novelty levels among participants of all experiences is by looking at the UEQ-S results, which measure perceived usability via its pragmatic subscale and motivation or excitement to use the VRLE through the hedonic subscale. A linear regression analysis found a statistically significant, positive relation between presence and the hedonic ($f^2 = 0.938, \beta = 0.524$, $p = <0.001$) and pragmatic ($f^2 = 0.253$, $\beta = 0.382$, $p = 0.047$) subscales of the *UEQ-S*. A statistically significant, negative relation was found between the hedonic ($f^2 = 0.230$, $\beta = -0.859$, $p = <0.057$) and pragmatic subscales ($f^2 = 0.451$, $\beta = -1.25$, $p = 0.011$) and knowledge test results. As such, participants who were more positive toward the environment, while more immersed, performed worse on average on the knowledge test.

While inconclusive, we believe these findings show that the novel VR environment caused a decrease in attention to the less novel lesson content, resulting in a decrease in knowledge test results. The emphasis on providing a highly ecologically valid scenario could have compounded the issue by providing a highly interesting environment that was superfluous to the instructional objectives. However, the degree to which this finding can explain the difference in knowledge test results is unclear. These findings corroborate the work of Makransky et al. (2019), who found that immersive VR caused high levels of presence but less learning than its non-immersive counterpart. Similarly, the authors theorized, based on Van Der Heijden (2004)'s perspective, that participants viewed the immersive VRLE as "hedonic," causing them to focus on the enjoyment of the environment instead of the learning material.

While no definitive conclusion can be made of the cause of the negative relation between presence and learning, it does underline the complex relationship between the features of VR and learning outcomes. The design decision to focus on representational fidelity and evoking as high a sense of presence as possible likely did not, as previously thought, enhance learning. Whether the environmental design and its representational fidelity are the source of the negative relation between presence and learning is debatable. However, the assumption that increasing representational fidelity should take priority as presence will always lead to better learning was misguided. This further highlights Fowler (2015)'s statement that technological affordances alone will not necessarily result in higher learning. Deep consideration of how these affordances (such as presence) align with the intended learning outcomes and educational context is crucial.

4.1 Limitations and future research

The primary limitation of this study is that both conditions were only compared over a thirty-minute study period, with the knowledge test being conducted immediately after. As the VRLE is designed to take advantage of VR's unique characteristics by encouraging experimentation and explaining concepts through live demonstrations, the time pressure of the experiment naturally puts the VR participants at a disadvantage over the textbook learners, who have fast, direct, non-transient access to the content. As such, a longitudinal study design would have allowed us to better assess the pedagogical approach of the VRLE in the absence of time pressure, as well as observe the long-term effects of the increase in intrinsic motivation and situational interest provided by the VRLE and the influence of novelty on cognitive load and learning. A second limitation of the study was that the sample consisted primarily of university students, which may not have been a sufficiently representative sample of the target user outlined in the 'learner profile' section. Consequently, there could have been a disconnect between the design and the learning preferences of our sample, which affected the results. We identified that firefighters would favor hands-on, active experimentation over more traditional academic environments. Therefore, a university student who is used to processing large amounts of textual information likely would be used to textbooks and would benefit less from the VRLE. Therefore, this introduces some complexity when applying

our findings to the target population. As such, while we believe our sample sufficiently represents early-career firefighters, future research with direct involvement with firefighters would be beneficial.

A further avenue of research is to examine whether applying the "pre-training principle" (Fiorella and Mayer, 2021) would effectively enhance the knowledge transfer of the VRLE. This principle states that multimedia learning is especially effective when the user already knows the names and factual information of the concepts. In Meyer et al. (2019), the authors found that this principle was especially significant when using VR due to the increased extraneous CL that the medium places on the user. Therefore, leveraging the benefits of both mediums could lead to superior learning outcomes. For example, the learner could take advantage of the browsability and non-transitional nature of the textbook to quickly and easily revise key topics and then use the VRLE to experiment and visualize them. As such, further research into the efficacy of a blended approach in this domain would be beneficial.

Our findings highlighted the difficulty of answering H5 ("The design of the VRLE facilitates learning"). Although all three sub-hypotheses were substantiated, many additional factors could be considered. As a result, we recommend further study into the interplay between environment design, learner attention [with the potential to look into gaze direction (Baron-Cohen et al., 2001; Droulers and Adil, 2015)], the intelligibility and coherence of the content, and learning outcomes. Future research in this area should also have rigorous measures for gauging the level of cognitive load the VRLE places on their user, with specific consideration to its source, such as distinguishing between environmental or task-based cognitive load. Additionally, as also recommended by Makransky and Petersen (2021), future investigation into the relation between Embodied Learning and learning outcomes would be useful due to the potential negative relation between presence and learning mediated by Embodied Learning.

4.2 Conclusion

We proposed and applied a "design for learning" approach to creating a virtual reality learning environment. We tested the implementation's ability to transfer factual and conceptual knowledge against official textbook material. While the VRLE was less effective at transferring knowledge than the textbook, it exhibited higher levels of intrinsic motivation and situational interest in the user, who also reported high satisfaction and positive sentiment toward the VRLE. However, our findings highlight the importance of considering pedagogical and human factors when designing a virtual environment for learning. VR's unique characteristics and affordances can both aid and hinder learning, and holistic design considerations are essential when investigating the feasibility of using a particular technology in a learning context. As such, while these findings can be applied to discuss the feasibility of VR as a medium for learning, our findings also offer a foundation for future research that focuses on how best to facilitate learning

using VR in this domain and what considerations are necessary for successful design execution.

Data availability statement

The datasets presented in this article are not readily available because ethics approval was granted on the basis that no data would be shared with researchers outside of the research team. Requests to access the datasets should be directed to simon.hoermann@canterbury.ac.nz.

Ethics statement

The studies involving humans were approved by Human Research Ethics Committee, University of Canterbury. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

SW: Conceptualization, Data curation, Investigation, Methodology, Software, Writing—original draft, Writing—review & editing, Project administration. SH: Supervision, Writing—review & editing, Funding acquisition, Methodology, Project administration. SL: Supervision, Writing—review & editing, Methodology, Project administration. RL: Supervision, Writing—review & editing, Methodology, Project administration.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomp. 2024.1274828/full#supplementary-material

References

Agre, P. E. (1995). *Conceptions of the User in Computer Systems Design*. Cambridge, United Kingdom: Cambridge University Press 67–106.

Ai-Lim Lee, E., Wong, K. W., and Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Comput. Educ.* 55, 1424–1442. doi: 10.1016/j.compedu.2010.06.006

Anderson, L. W., and Krathwohl, D. R. (2001). A taxonomy for learning, teaching, and assessing: a revision of bloom's taxonomy of educational objectives. *Longman*. 83, 154–159.

Backlund, P., Engstrom, H., Hammar, C., Johannesson, M., and Lebram, M. (2007). "Sidh - a game based firefighter training simulation," in *Proceedings of the 2007 11th International Conference Information Visualization (IV '07), Zurich, Switzerland* (New York, NY, USA: IEEE), 899–907. doi: 10.1109/IV.2007.100

Baduel, C., Paxman, C. J., and Mueller, J. F. (2015). Perfluoroalkyl substances in a firefighting training ground (FTG), distribution and potential future release. *J. Hazar. Mater.* 296, 46–53. doi: 10.1016/j.jhazmat.2015.03.007

Baigent, D. (2001). *Gender relations, masculinities and the fire service: a qualitative study of firefighters' constructions of masculinity during firefighting and in their social relations at work*. PhD thesis, Anglia Ruskin University, Chelmsford, United Kingdom.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., and Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiat. All. Disc.* 42, 241–251. doi: 10.1111/1469-7610.00715

Beetham, H. (2007). "An approach to learning activity design," in *Rethinking Pedagogy for a Digital Age: Designing and Delivering e-Learning*, eds. H. Beetham, and R. Sharpe (London: Routledge), 26–40. doi: 10.4324/9780203961681

Biggs, J. B., and Tang, C. S. (2009). *Teaching for Quality Learning at University: What the Student Does. The Society for Research into Higher Education*. Maidenhead: McGraw-Hill.

Bliss, J. P., Tidwell, P. D., and Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence* 6, 73–86. doi: 10.1162/pres.1997.6.1.73

Clark, R. C., and Mayer, R. E. (2016). *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning*. New York: John Wiley sons. doi: 10.1002/9781119239086

Clifford, R. M. S., Engelbrecht, H., Jung, S., Oliver, H., Billinghurst, M., Lindeman, R. W., et al. (2020). Aerial firefighter radio communication performance in a virtual training system: radio communication disruptions simulated in VR for air attack supervision. *Visual Comput.* 37, 63–76. doi: 10.1007/s00371-020-01816-6

Cohen-Hatton, S. R., and Honey, R. C. (2015). Goal-oriented training affects decision-making processes in virtual and simulated fire and rescue environments. *J. Exper. Psychol.* 21, 395–406. doi: 10.1037/xap0000061

Conole, G., Dyke, M., Oliver, M., and Seale, J. (2004). Mapping pedagogy and tools for effective learning design. *Comput. Educ.* 43, 17–33. doi: 10.1016/j.compedu.2003.12.018

Conole, G., and Jones, C. (2010). "Sharing practice, problems and solutions for institutional change: comparing different forms of representation," in *Technology-Enhanced Learning* (Brill), 277–295. doi: 10.1163/9789460910623_017

Cooper, G., and Bowers, J. (1995). "Representing the user: notes on the disciplinary rhetoric of human-computer," in *The Social and Interactional Dimensions of Human-Computer Interfaces* 48.

Cowlishaw, S., Evans, L., and McLennan, J. (2008). Families of rural volunteer firefighters. *Rural Soc.* 18, 17–25. doi: 10.5172/rsj.351.18.1.17

Dalgarno, B. (2002). The potential of 3D virtual learning environments: a constructivist analysis. *Electr. J. Instr. Sci. Technol.* 5, 3–6.

Dalgarno, B., and Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? Learning affordances of 3-D virtual environments. *Br. J. Educ. Technol.* 41, 10–32. doi: 10.1111/j.1467-8535.2009.01038.x

Droulers, O., and Adil, S. (2015). Perceived gaze direction modulates ad memorization. *J. Neurosci. Psychol. Econ.* 8:15. doi: 10.1037/npe0000029

Engelbrecht, H., Lindeman, R. W., and Hoermann, S. (2019). A SWOT analysis of the field of virtual reality for firefighter training. *Front. Robot. AI* 6:101. doi: 10.3389/frobt.2019.00101

Fiorella, L., and Mayer, R. E. (2021). "Principles for managing essential processing in multimedia learning: segmenting, pre-training, and modality principles," in *The Cambridge Handbook of Multimedia Learning*, eds. R. E. Mayer, and L. Fiorella (Cambridge, United Kingdom: Cambridge University Press) 185–198. doi: 10.1017/9781108894333.019

Fire Emergency New Zealand (2020). *Working Safely at Vegetation Fires*. Wellington, New Zealand: Fire and Emergency New Zealand–Training.

Fontaine, G. (1992). The experience of a sense of presence in intercultural and international encounters. *Presence* 1, 482–490. doi: 10.1162/pres.1992.1.4.482

Fowler, C. (2015). Virtual reality and learning: where is the pedagogy? Learning activities in 3-D virtual worlds. *Br. J. Educ. Technol.* 46, 412–422. doi: 10.1111/bjet.12135

Gaver, W. W. (1991). "Technology affordances," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Reaching through Technology - CHI '91, New Orleans, Louisiana, United States* (New York, NY, USA: ACM), 79–84. doi: 10.1145/108844.108856

Goodyear, P., and Carvalho, L. (2019). The analysis of complex learning environments. *Rethink. Pedag. Digital Age* 3, 49–65. doi: 10.4324/9781351252805-4

Grudin, J., and Pruitt, J. (2002). "Personas, participatory design and product development: an infrastructure for engagement," in *Proceedings of PDC 02, Participatory Design Conference 2002. Malmo, Sweden* [Palo Alto, CA: Computer Professionals for Social Responsibility (CPSR)], 144–152.

Ha, G., Lee, H., Lee, S., Cha, J., and Kim, S. (2016). "A VR serious game for fire evacuation drill with synchronized tele-collaboration among users," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology, Munich Germany* (New York, NY, USA: ACM), 301–302. doi: 10.1145/2993369.2996306

Hall, A., Hockey, J., and Robinson, V. (2007). Occupational cultures and the embodiment of masculinity: hairdressing, estate agency and firefighting. *Gender, Work Organ.* 14, 534–551. doi: 10.1111/j.1468-0432.2007.00370.x

Harrison, K., and Olofsson, J. (2016). Becoming a firefighter–on the intra-active relationship between firefighters and their tools. *NORMA* 11, 158–173. doi: 10.1080/18902138.2016.1217693

Holmgren, R. (2014). Reformed firefighter training program in Sweden: conflicting instructor conceptions of professional learning. *Nordic J. Vocat. Educ. Train.* 4, 1–14. doi: 10.3384/njvet.2242-458X.14v4i2a5

Huang, M.-H. (2003). Designing website attributes to induce experiential encounters. *Comput. Hum. Behav.* 19, 425–442. doi: 10.1016/S0747-5632(02)00080-8

Huang, Y., Wu, S., and Kaplan, J. O. (2015). Sensitivity of global wildfire occurrences to various factors in the context of global change. *Atmosphere. Environ.* 121, 86–92. doi: 10.1016/j.atmosenv.2015.06.002

Jeon, S.-G., Han, J., Jo, Y., and Han, K. (2019). "Being more focused and engaged in firefighting training: applying user-centered design to vr system development," in *Proceedings of VRST '19: 25th ACM Symposium on Virtual Reality Software and Technology, Parramatta NSW Australia* (New York, NY, USA: ACM), 1–11. doi: 10.1145/3359996.3364268

Johnson-Glenberg, M. C. (2019). "The necessary nine: design principles for embodied VR and active stem education," in *Learning in a Digital World: Perspective on Interactive Technologies for Formal and Informal Education, Smart Computing and Intelligence*, eds. P. Díaz, A. Ioannou, K. K. Bhagat, and J. M. Spector (Singapore: Springer), 83–112. doi: 10.1007/978-981-13-8265-9_5

Johnson-Glenberg, M. C., Bartolomea, H., and Kalina, E. (2021). Platform is not destiny: embodied learning effects comparing 2D desktop to 3D virtual reality STEM experiences. *J. Comput. Assist. Learn.* 37, 1263–1284. doi: 10.1111/jcal.12567

Kirk, K. M., and Logan, M. B. (2015). Firefighting instructors' exposures to polycyclic aromatic hydrocarbons during live fire training scenarios. *J. Occupat. Environ. Hygiene* 12, 227–234. doi: 10.1080/15459624.2014.9 55184

Kirschman, E. (2004). *I Love a Fire Fighter: What the Family Needs to Know*. New York: Guilford Press.

Koch, M., von Luck, K., Schwarzer, J., and Draheim, S. (2018). "The novelty effect in large display deployments –experiences and lessons-learned for evaluating prototypes," in *Proceedings of 16th European Conference on Computer-Supported Cooperative Work - Exploratory Papers, Troyes, France. European Society for Socially Embedded Technologies (EUSSET)*.

Koehler, M. J., and Mishra, P. (2005). What happens when teachers design educational technology? The development of technological pedagogical content knowledge. *J. Educ. Comput. Res.* 32, 131–152. doi: 10.2190/0EW7-01WB-BKHL-QDYV

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Into Pract.* 41, 212–218. doi: 10.1207/s15430421tip4104_2

Laugwitz, B., Held, T., and Schrepp, M. (2008). "Construction and evaluation of a user experience questionnaire," in *Proceedings of HCI and Usability for Education and Work: 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2008, Graz, Austria, November 20–21, 2008. Proceedings 4* (Berlin, Germany: Springer), 63–76.

Laurillard, D. (2013). *Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology*. London: Routledge. doi: 10.4324/9780203125083

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bull.* 32, 47–56. doi: 10.1145/333329.333344

Makransky, G., and Petersen, G. B. (2021). The cognitive affective model of immersive learning (camil): a theoretical research-based model of learning in immersive virtual reality. *Educ. Psychol. Rev.* 33, 937–958. doi: 10.1007/s10648-020-09586-2

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

Mayer, R. E. (2021). "Cognitive theory of multimedia learning," in *The Cambridge Handbook of Multimedia Learning*, eds. R. E. Mayer, and L. Fiorella (Cambridge, United Kingdom: Cambridge University Press), 31–48. doi: 10.1017/CBO9780511816819.004

Mayes, J., and Fowler, C. (1999). Learning technology and usability: a framework for understanding courseware. *Inter. Comput.* 11, 485–497. doi: 10.1016/S0953-5438(98)00065-4

McCormick, E. P., and Wickens, C. D. (1995). *Virtual reality features of frame of reference and display dimensionality with stereopsis: their effects on scientific visualization*. Master's thesis, University of Illinois at Urbana-Champaign.

Meehan, M., Razzaque, S., Whitton, M. C., and Brooks, F. P. (2003). "Effect of latency on presence in stressful virtual environments," in *Proceedings of IEEE Virtual Reality, 2003. Proceedings, Los Angeles, CA, USA* (New York, NY, USA: IEEE), 141–148.

Meyer, O. A., Omdahl, M. K., and Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: a media and methods experiment. *Comput. Educ.* 140:103603. doi: 10.1016/j.compedu.2019.103603

Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., and Bustillo, A. (2023). Countering the novelty effect: a tutorial for immersive virtual reality learning environments. *Appl. Sci.* 13:593. doi: 10.3390/app13010593

Mikkelson, N., and Lee, W. O. (2000). "Incorporating user archetypes into scenario-based design," in *Proceedings of 9th Annual Usability Professionals' Association (UPA) Conference* (Chicago, Illinois, USA: Usability Professionals' Association).

Mikropoulos, T. A. (2006). Presence: a unique characteristic in educational virtual environments. *Virtual Reality* 10, 197–206. doi: 10.1007/s10055-006-0039-1

Mikropoulos, T. A., and Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Comput. Educ.* 56, 769–780. doi: 10.1016/j.compedu.2010.10.020

Mittal, V., and Frennea, C. (2010). *Customer satisfaction: A strategic review and guidelines for managers*. MSI Fast Forward Series, Marketing Science Institute, Cambridge, MA.

Mujber, T. S., Szecsi, T., and Hashmi, M. S. (2004). Virtual reality applications in manufacturing process simulation. *J. Mater. Proc. Technol.* 155, 1834–1838. doi: 10.1016/j.jmatprotec.2004.04.401

Nielsen, L. (2003). "Constructing the user," in *Human-Computer Interaction: Theory and Practice* 430.

Nielsen, L. (2019). *Personas - User Focused Design*. London: Springer. doi: 10.1007/978-1-4471-7427-1

Nowell, L. S., Norris, J. M., White, D. E., and Moules, N. J. (2017). Thematic analysis: striving to meet the trustworthiness criteria. *Int. J. Qualit. Methods* 16:160940691773384. doi: 10.1177/1609406917733847

Nunnally, J. (1994). *Psychometric Theory*. New York, NY: McGraw-Hill.

Oliva, D., Somerkoski, B., Tarkkanen, K., and Lehto, A. (2019). "Virtual reality as a communication tool for fire safety –Experiences from the VirPa project," in *GamiFIN 2019, Levi, Finland* (Tampere, Finland: Gamification Group), 241–252.

Olofsson, J. (2013). "The profession of firefighting is about teamwork, it is about trusting each other": masculine enactments and generational discrepancies within the swedish fire service. *Cult. Soc. Masculin.* 5, 75–88. doi: 10.3149/CSM.0501.75

Ooi, S., Tanimoto, T., and Sano, M. (2019). "Virtual reality fire disaster training system for improving disaster awareness," in *Proceedings of the 2019 8th International Conference on Educational and Information Technology, Cambridge United Kingdom* (New York, NY, USA: ACM), 301–307. doi: 10.1145/3318396.3318431

Perrott, T. A. (2019). Doing hot and 'dirty' work: masculinities and occupational identity in firefighting. *Gender, Work Organiz.* 26, 1398–1412. doi: 10.1111/gwao.12412

Peters, D., Calvo, R. A., and Ryan, R. M. (2018). Designing for motivation, engagement and wellbeing in digital experience. *Front. Psychol.* 9:797. doi: 10.3389/fpsyg.2018.00797

Petersen, G. B., Petkakis, G., and Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Comput. Educ.* 179:104429. doi: 10.1016/j.compedu.2021.104429

Pruitt, J., and Adlin, T. (2010). *The Persona Lifecycle: Keeping People in Mind throughout Product Design*. New York: Elsevier.

Psotka, J. (1995). Immersive training systems: virtual reality and education and training. *Instruct. Sci.* 23, 405–431. doi: 10.1007/BF00896880

Radianti, J., Majchrzak, T. A., Fromm, J., and Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. *Comput. Educ.* 147:103778. doi: 10.1016/j.compedu.2019.103778

Rumsey, A., and Le Dantec, C. A. (2019). "Clearing the smoke: the changing identities and work in firefighters," in *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA: ACM). doi: 10.1145/3322276.3322292

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55:68. doi: 10.1037//0003-066X.55.1.68

Salzman, M. C., Dede, C., Loftin, R. B., and Chen, J. (1999). A model for understanding how virtual reality aids complex conceptual learning. *Presence* 8, 293–316. doi: 10.1162/105474699566242

Schrepp, M., Hinderks, A., and Thomaschewski, J. (2017). Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *Int. J. Inter. Multim. Artif. Intell.* 4:103. doi: 10.9781/ijimai.2017.09.001

Slater, M., Brogni, A., and Steed, A. (2003). "Physiological responses to breaks in presence: a pilot study," in *Proceedings for Presence 2003: The 6th Annual International Workshop on Presence* (Aalborg, Denmark: International Society for Presence Research).

Sommerfeld, A., Wagner, S. L., Harder, H. G., and Schmidt, G. (2017). Behavioral health and firefighters: an intervention and interviews with canadian firefighters. *J. Loss Trauma* 22, 307–324. doi: 10.1080/15325024.2017.1284515

Stetz, M. C., Long, C. P., Schober Jr, W. V., Cardillo, C. G., and Wildzunas, R. M. (2007). Stress assessment and management while medics take care of the VR wounded. *Ann. Rev. Cyber Ther. Telemed.* 5, 165–171.

Sweller, J. (2011). "Cognitive load theory," in *Psychology of Learning and Motivation*, eds. J. P. Mestre, and B. H. Ross (Cambridge, MA: Academic Press), 37–76. doi: 10.1016/B978-0-12-387691-1.00002-8

Tate, D., Sibert, L., and King, T. (1997). Using virtual environments to train firefighters. *IEEE Comput. Graph. Appl.* 17, 23–29. doi: 10.1109/38.626965

Thurnell-Read, T., and Parker, A. (2008). Men, masculinities and firefighting: occupational identity, shop-floor culture and organisational change. *Emot. Space Soc.* 1, 127–134. doi: 10.1016/j.emospa.2009.03.001

Van Der Heijden, H. (2004). User acceptance of hedonic information systems. *MIS Quart.* 28:695. doi: 10.2307/25148660

Wheeler, S. G., Engelbrecht, H., and Hoermann, S. (2021). Human factors research in immersive virtual reality firefighter training: a systematic review. *Front. Virtual Real.* 2:671664. doi: 10.3389/frvir.2021.671664

Wickens, C. D., and Baker, P. (1995). "Cognitive issues in virtual reality," in *Virtual Environments and Advanced Interface Design*, eds. W. Barfield, and T. A. Furness (New York: Oxford University Press), 514–541. doi: 10.1093/oso/9780195075557.003.0024

Wiederhold, , M. D., and Wiederhold, BK. (2004). Training combat medics using VR. *CyberPsychol. Behav.* 7, 319–319. doi: 10.1037/e705532011-045

Wiederhold, B. K., Jang, D. P., Kaneda, M., Cabral, I., Lurie, Y., May, T., et al. (2001). An investigation into physiological responses in virtual environments: An objective measurement of presence. *Towards Cyberpsychol.* 2, 175–183.

Winn, W. (1993). *A conceptual basis for educational applications of virtual reality*. Technical Publication R-93–9, Human Interface Technology Laboratory of the Washington Technology Center, Seattle: University of Washington.

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence* 7, 225–240. doi: 10.1162/105474698565686

Yoon, J.-H., Kravitz, B., Rasch, P. J., Wang, S.-Y. S., Gillies, R. R., and Hipps, L. (2015). Extreme fire season in california: a glimpse into the future? *Bull. Am. Meteorol. Soc.* 96, S5–S9. doi: 10.1175/BAMS-D-15-00114.1