



Evaluation of a virtual reality training tool for firefighters responding to transportation incidents with dangerous goods

Maxine Berthiaume^{1,2} · Max Kinateder^{1,3} · Bruno Emond¹ ·
Natalia Cooper¹ · Ishika Obeegadoo¹ · Jean-François Lapointe¹

Received: 30 May 2023 / Accepted: 10 November 2023
© Crown 2024

Abstract

Access to dangerous goods training for firefighters in remote areas is limited for financial and logistical reasons. Virtual reality (VR) is a promising solution for this challenge as it is cost-effective, safe, and allows to simulate realistic scenarios that would be dangerous or difficult to implement in the real world. However, rigorous evaluations of VR training tools for first responders are still scarce. In this exploratory user study, a simple VR training tool involving two dangerous goods scenarios was developed. In each scenario, trainees learned how to safely approach a jack-knifed truck with a trailer and how to collect and communicate information about the transported materials. The tool was tested with a group of 24 professional firefighter trainees ($n = 22$) and instructors ($n = 2$), who each completed the two training scenarios. The main goal of the study was to assess the usability of the VR tool in the given scenarios. Participants provided feedback on cybersickness, perceived workload, and usability. They also filled out a knowledge test before and after the VR training and gave feedback at the end of the study. The VR tool recorded task completion duration and participants' navigation and use of tools events. Overall, the tool provided good usability, acceptance, and satisfaction. However, a wide range in individuals' responses was observed. In addition, no post-training improvement in participants' knowledge was found, likely due to the already high level of knowledge pre-training. Future directions for improving the VR tool, general implications for other VR training tools, and suggestions for future research are discussed.

Keywords Dangerous goods · First responders · Training · Usability · Virtual reality · Virtual reality evaluation

Extended author information available on the last page of the article

Published online: 16 January 2024

1 Introduction

Dangerous goods are common on Canadian roads, and so are accidents. Transport Canada estimates that, in terms of tonnage, approximately 70% of dangerous goods are transported by road, compared to 24% by rail, 6% by marine, and 1% by air (Transport Canada, 2020c). The relatively high incidence of road accidents is a cause for concern. An average of 138 accidents per year involving the road transportation of dangerous goods occurred between 2010 and 2019 (Transport Canada, 2020b). One challenge for first responders is that many of such accidents occur in remote locations. In 2018, for example, most of the accidents occurred in rural areas (76%) compared to other areas (Government of Canada, 2020). Furthermore, in 58% of the accidents in 2018, the dangerous goods in question were flammable liquids, which are also the most common form of dangerous goods on Canadian roads (Government of Canada, 2020).

These statistics highlight the continued need for adequate training for first responders to manage these incidents, particularly for firefighters. An American study found that 63% of first responders who sustained injuries from incidents involving the release of hazardous materials were firefighters (Melnikova et al., 2018). Another study showed that first responders were more likely to be injured when responding to these types of incidents in rural areas (Berkowitz et al., 2004).

1.1 Gaps in dangerous goods training

In general, beyond the guidelines prescribed by governing bodies, individual fire departments could vary in their level of training. For example, dangerous goods training might not be offered in small/isolated communities, and firefighters would need to travel to receive on-site training if the course provider is unable to deliver training in remote areas. Increasing access to dangerous goods training for firefighters in remote areas could help reduce costs related to travel and logistics. Another gap in training is present between career and volunteer firefighters. Volunteer firefighters often have another career to support themselves and their families, which forces them to divide their priorities between firefighting and their employment. A recent census conducted by the Canadian Association of Fire Chiefs (CAFC) determined that 29% of firefighters across Canada were career firefighters and 71% were volunteer firefighters (Canadian Association of Fire Chiefs, 2022). Firefighters are often either career professionals or volunteers based on the size of the municipality in which they reside, such that, urban areas tend to have more career firefighters (Haynes & Stein, 2018; Ramsden et al., 2018; Taylor et al., 2022). The CAFC reported that composite fire departments (i.e., departments that have both career and volunteer firefighters) offer services that may require additional training, funding, and equipment, including responding to incidents involving hazardous materials (Canadian Association of Fire Chiefs, 2021).

1.2 Training levels

Authorities across North America developed the Emergency Response Guidebook (ERG) for first responders arriving at a transportation incident involving dangerous goods. The ERG provides information to first responders with varying levels of training on identifying hazards and how to protect themselves and bystanders during the initial response phase to a transportation incident involving dangerous goods (Transport Canada, 2020a). A survey of Canadian provinces and jurisdictions indicates that all firefighters across the country are trained to respond to incidents involving hazardous materials following standards for firefighter professional qualifications (e.g., NFPA 1001, Transport Canada, 2016b). Based on NFPA 472 and 1072, firefighters in Canadian provinces are trained at three levels: Awareness, Operations and Technician (National Fire Protection Association, 2017, 2018). At the **Awareness level**, firefighters are required to confirm and identify the presence of hazardous materials at an incident site, isolate and secure the area, and communicate with the appropriate authorities for further assistance. Firefighters at the **Operations level** will further respond to the incident, including, but not limited to, victim rescue and recovery, product monitoring, decontamination and product control. At the **Technician level**, firefighters will further coordinate, survey, and assist in the emergency response activities (National Fire Protection Association, 2018).

1.3 Training methods

A range of training methods exist to prepare first responders for incidents involving hazardous materials, including live, virtual, or in classroom training, where information is provided by an instructor and hands-on behavioural training. However, the delivery of adequate dangerous goods training to firefighters, particularly in remote areas, can be challenging for financial (e.g., cost of training equipment and material) and logistical reasons (e.g., availability of trainers and training material).

The method of training can have effects on its efficacy. One study evaluating in classroom training for incidents involving hazardous materials found that participants felt the amount of time spent on the more technical aspects related to hazardous materials was insufficient and recommended additional days for practising with hands-on activities and an increased availability of “refresher” training (Cohen, 1998). However, hands-on live fire training can be difficult to do repeatedly, due to the financial costs associated with using an acquired structure for fire testing, as well as the physical risks of illness, injury, or death of firefighters (National Fire Protection Association, 2012). In contrast, online training is an efficient method of learning, but it can lack realism and user engagement/interaction. An important disadvantage of online training is that firefighters might not be encouraged to consider applications of the material outside of the modules (Vasquez, 2018).

1.4 Training in virtual reality

Virtual reality (VR) is a promising tool for disaster and emergency response training thanks to its potential to bridge the gaps with other commonly established training formats (Hsu et al., 2013; Liu et al., 2023). One could argue that VR training offers a middle ground between real-life behavioural training and other forms of training, such as lectures. Specifically, the immersive, participatory, and collaborative nature of VR training offers a unique realistic quality that is not generally present in classroom-based or web-based training (Hsu et al., 2013). There is increasing acceptance of VR training given considerable cost advantages over large-scale real-life exercises and other modalities (Cook, 2018). For high-risk incidents in varied and complex environments, VR can potentially enable safe, immersive, and cost-effective ways of training (Hsu et al., 2013; MacLeod, 2020; Smith & Steel, 2000; Summers, 2012) to prepare firefighters, physically as well as mentally, for real-life incidents (Engelbrecht et al., 2019). One paper (Engelbrecht et al., 2019) identified strengths, weaknesses, opportunities, and threats of VR-based training for firefighters (see Table 1).

Many studies examining the use of VR to train firefighters have focused on search and rescue operations training (Backlund et al., 2007; Bliss et al., 1997; Tate et al., 1997; Xu et al., 2014), simulating fire extinguishing and other firefighting skills (Bellemans et al., 2020; Braun et al., 2022; Corelli et al., 2020; Hadj Sassi et al., 2022; Narciso et al., 2020; Ooi et al., 2019; Reim et al., 2022; Vichitvejpaisal et al., 2016; Wijkmark et al., 2021a), aerial firefighting training (Clifford et al., 2018a, b), and commander training (Cohen-Hatton & Honey, 2015; Hammar Wijkmark et al., 2019; Wijkmark et al., 2021b). Fewer studies have used VR in the context of dangerous goods training. In one study, participants completed training on unloading dangerous chemical goods in VR or on a desktop. The results suggest that overall user experience was higher in VR than when using the desktop (Chover et al., 2022). Other organizations have developed VR training tools for first responders involving

Table 1 Reproduction of Table 1 of Engelbrecht et al., 2019 listing strengths, weaknesses, opportunities, and threats of VR-based training for firefighters

<p>Strengths</p> <ul style="list-style-type: none"> - Cost effectiveness - Complex and varied training scenarios - High ecological validity - Increased safety for high-risk training - Trainee engagement - Data recording 	<p>Weaknesses</p> <ul style="list-style-type: none"> - Lack of specialization and testing of systems - Immaturity of technology - Technology barrier - Lack of multi-user fidelity
<p>Opportunities</p> <ul style="list-style-type: none"> - System engineering progress - Transfer of findings from other domains - Increase in physical fidelity - Increase resilience against adverse effects 	<p>Threats</p> <ul style="list-style-type: none"> - Uncertain skill transfer - Worsening of overall net-effects of training - Adverse effects of habituation - Adverse effects of engagement stimulation

dangerous goods (Gainer et al., 2020; Haskins et al., 2020; Lofca et al., 2022); however, no corresponding empirical data seem to have been published yet.

1.5 Virtual reality training tool

The VR training tool developed for this study consists of a simulation of an incident scene involving dangerous goods. The tool provides training for first responders at the Awareness level. The training goals target the recognition, identification, and communication of hazards and are based on the safety precautions defined in the ERG. The current training scenario focuses on safely gathering and communicating information and securing the hazard area (see Materials and methods section for additional details). The main aim of the tool was to show how VR can be used for training, especially when access to real-world training for specific scenarios can be difficult to achieve due to resources, safety and risks, or funding.

The present paper reports on two studies: first, a pilot study to optimize the virtual training scenario (described in detail in Emond et al., 2022) followed by a main user study with firefighters at a local fire service to test the VR training tool and collect their feedback using a variety of objective and subjective measures. The work had the following objectives:

- 1) Assess the usability of the VR training tool with a convenience sample (pilot study) and firefighters (main study);
- 2) Assess whether the knowledge test captured learning effects; and,
- 3) Assess whether any adjustments should be made to the VR training tool and/or research protocol based on participants' feedback.

2 Materials and methods

The pilot study and the main user study were approved by the National Research Council Canada's Research Ethics Board (protocol #2021-157). All experimental procedures involving human participants were in accordance with the institutional research committee ethical guidelines. All participants signed their informed consent form prior to participating.

2.1 Participants

The pilot study ($N=9$) aimed to assess the overall usability of the VR training tool. Participants explored the virtual environment and took note of any usability and technical issues that were then addressed in the next iteration in the development of the virtual environment. Some of the improvements included adding audio-recorded instructions about the task, adding cones to secure the scene, and improving

interaction with objects. Once all changes were implemented, the main study was conducted. For additional details about the pilot study, see Emond et al. (2022). For the main study, twenty-four firefighters participated in the study (22 recruits in training and 2 instructors). Participants were referred by a training instructor at the Ottawa Fire Service. The research team wanted to obtain first responders' feedback as they are the intended audience for the VR tool. Participants had to be at least 18 years old, have normal or corrected to normal vision, be vaccinated against COVID-19, and report no symptoms of COVID-19. Potential participants were excluded if they had any health conditions that could be exacerbated by virtual reality (VR), such as a history of cardiovascular or vestibular conditions, epilepsy, severe motion sickness while travelling, eye disease, or recurrent migraines. These criteria did not lead to any exclusions (i.e., all recruited individuals participated in the study). However, the data of six participants were excluded from some of the statistical analyses for the following reasons: (1) VR scenario ended due to severe cybersickness ($n=1$), and (2) headset data loss due to technical issues ($n=5$). See Section 3.1 for sociodemographic results.

2.2 Equipment

The VR equipment used in this study was the Meta Quest 2, which consists of a head-mounted display and its associated handheld controllers, that offers a fully immersive experience in VR. The training scenarios were developed using the Unity game engine (<https://unity.com/>).

This study used four Meta Quest 2 headsets with the scenarios installed and four pairs of handheld controllers (see Fig. 1 for an illustration of the VR equipment). In addition, four laptops running the Meta Developer Hub application (Meta Quest, 2022) were used to cast the headset view so that the experimenters could help participants if they experienced any technical issues while completing the scenario tasks. The headsets were connected to the laptops via USB-C cables. Participants completed the study questionnaires on Apple iPads. In order to reduce cybersickness symptoms, all participants completed the tasks in a seated position (sitting in a chair).

2.3 Virtual reality scenarios

The training system consists of a VR simulation of a scene involving dangerous goods. The dangerous goods transport truck involved in the incident corresponds to one that can be commonly found on Canadian roads. It comprises a cabin as well as a trailer (for flammable liquids TC406). The trailer has diamond shaped placards on the front, back, and on each side of the trailer, showing the ERG ID for the dangerous goods used in the scenario (see Fig. 2 for an illustration).



Fig. 1 Illustration of the VR equipment used in the study and of the virtual environment. The red icon in the upper right corner represents a wind rose



Fig. 2 Illustration of the transport truck in the virtual environment

Trainees were able to view the environment in a first-person perspective and navigate the scene using handheld controllers. In the background, a formative tutoring system provided trainees with feedback as they navigate the virtual environment. For

example, if a trainee moved too close to the incident in the scenario, they received a warning message shown in their headset and hear a pre-recorded audio message. In addition, the system logged events in the simulation (e.g., participant entering or exiting zones that are upwind from the incident).

Participants were presented with two scenarios during the study: Scenario 1 – included a placard with the ERG ID number for gasoline (1203, common), and Scenario 2 – included a placard with the ERG ID number for a gasoline and ethanol mixture (3475, rare). It was assumed that Scenario 1 would be easier, given the common occurrence of gasoline. For more details on the scenarios, see Appendix 1. The order of the scenarios was randomly assigned and counterbalanced (Scenario 1 followed by Scenario 2 for even Participant ID numbers and vice versa for odd Participant ID numbers). The main tasks in both scenarios were to (1) secure the scene using cones to ensure that no individuals entered the incident zone, (2) find an appropriate location to assess the scene and identify the hazardous material and risks, and (3) complete a report of the incident.

In both scenarios, participants had access to inventory items to identify hazards, collect, and communicate information. These inventory items included: (1) binoculars to identify the placard on the transport truck, (2) a copy of the ERG, and (3) a mobile device to report the incident (multiple-choice dialogue system). See Fig. 3 for an illustration of the inventory items. In the upper right corner of the headset screen, participants could also see a wind rose, which indicated the wind direction relative to the participant's heading (see Fig. 1). Prior to being immersed in the scenarios, participants practised navigating and interacting with the inventory objects and other static objects in the virtual environment (e.g., selecting a police vehicle to exit the scenario).

2.4 Measures

Participants completed several online questionnaires throughout the study via Qualtrics (<https://www.qualtrics.com>). Participants also completed a paper-and-pencil knowledge test before and after the training in VR. During the VR training, various

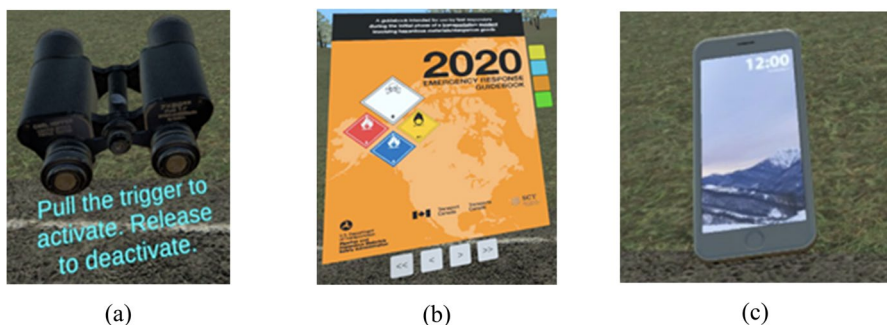


Fig. 3 Illustration of the inventory items available to participants in both VR scenarios: **a** binoculars to identify the placard on the transport truck, **b** the Emergency Response Guidebook (ERG), and **c** a mobile device to report the incident

data were recorded (Participant ID, timestamp, scenario, events, inventory items used).

2.4.1 Sociodemographic questionnaire

The sociodemographic questionnaire included questions about participants' age, gender, level of experience with VR, health conditions, and level of experience with firefighting and hazardous materials training. It was administered prior to the VR training.

2.4.2 Knowledge test

The knowledge test assessed participants' general knowledge about the procedures to follow at the Awareness Level when responding to a hazardous materials incident. The test comprised 15 multiple-choice items (see Appendix 2) and was prepared with the guidance of a senior firefighting officer, who also confirmed the correct answers. Each correct answer was worth one point. To calculate a total score, all the correct answers were summed up. The knowledge test was administered pre- and post-VR training and was designed to capture a possible learning effect.

2.4.3 Task completion duration

Timestamps for each participant were recorded on the VR headsets during both VR scenarios. Task completion duration was computed for each participant by subtracting the scenario start time from the scenario end time.

2.4.4 Error rate

Events (e.g., ENTER-Upwind Zone, EXIT-Upwind Zone) were recorded on the VR headsets for each participant during both VR scenarios. Prior to data analysis, the research team classified the events as positive, neutral, or negative. Events were considered positive when the action taken by the participant was in line with the training objectives (e.g., securing the scene, entering an upwind zone), neutral when the action taken was not part of the training objectives (e.g., scenario beginning and scenario end), and negative when the action taken was hazardous (not securing the scene, entering a downwind or low ground zone, getting too close to the jackknifed trailer). Negative events were considered errors. Error rate was computed by counting the number of events classified as negative.

2.4.5 Simulator Sickness Questionnaire (SSQ)

The Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993) contains 16 items assessing unwanted side effects induced by VR (e.g., nausea, dizziness) and is rated on a 4-point scale (from "None" to "Severe"). The items are divided into three subscales: (1) Nausea, (2) Oculomotor, and (3) Disorientation. The total score was calculated using Kennedy et al.'s (1993) formula. A higher score represents more

severe cybersickness. The SSQ was administered pre-VR training and after each VR scenario.

2.4.6 Presence questionnaires

The Presence Questionnaire (PQ) (Witmer & Singer, 1998) and the Slater-Usoh-Steed Questionnaire (SUS) (Slater et al., 1995; Usoh et al., 2000) measure user's presence in the virtual environment, or the feeling of "being there" in the virtual environment. Both questionnaires were administered after the second VR scenario. The results from these questionnaires are not presented in this paper as a more detailed analysis of the data is described in another paper.

2.4.7 NASA Task Load Index (NASA-TLX)

The NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988) contains six subscales (i.e., Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration) and is rated on a 21-point scale (from "Very Low" to "Very High"). The NASA-TLX was administered after the second VR scenario. The mean of each subscale was calculated. Higher ratings on the subscales indicate higher perceived demand, except for the performance subscale where a higher rating indicates lower self-perceived success on the task.

2.4.8 System Usability Scale (SUS)

The System Usability Scale (SUS) is a reliable and valid assessment tool that takes into account the context of the system use (Bangor et al., 2008; Brooke, 1996; McLellan et al., 2012). The SUS is a recognized tool that can quickly and easily collect a user's subjective rating of a product's usability (Bangor et al., 2008). The SUS consists of 10 statements; half of the statements are worded positively and the other half negatively. The SUS is rated on a 5-point Likert scale (from "Strongly disagree" to "Strongly agree"). The total score was calculated using Brooke's formula (Brooke, 1996). A higher score represents higher usability. The research team slightly modified the questionnaire by replacing "system" with "VR system". The SUS was administered after the second VR scenario.

2.4.9 Post-Study System Usability Questionnaire (PSSUQ)

The Post-Study System Usability Questionnaire (PSSUQ) was used to validate the usability of the VR training tool further (Lewis, 1995). The PSSUQ is acceptable to use with smaller sample sizes and its capability to detect differences is a consequence of its high sensitivity (Tullis & Stetson, 2004). The PSSUQ is rated on a 5-point scale (e.g., "Strongly disagree" to "Strongly agree") and has three subscales: (1) System Usability, (2) Interface Quality, and (3) Information Quality. The mean of each subscale was calculated. Participants completed the PSSUQ after the second VR scenario. For brevity, the research team reduced the number of items to 12, but they ensured that each subscale contained the items that were identified by previous

research as carrying the most weight when calculating a factor's score (Lewis, 1995; Sauro & Lewis, 2005). The research team also replaced “system” with “VR system”.

2.4.10 Open-ended feedback survey

Feedback from participants was collected after the second VR scenario and included three questions about (1) participants' overall satisfaction with the VR training, (2) whether they thought any improvements could be made to the VR scenarios, and (3) any other comments on their experience. Participants' feedback will be used to improve the next version of the scenarios.

2.5 Procedure

Data collection took place during working hours on one day in June 2022 at an Ottawa Fire Service location. Four experimenters concurrently conducted the study. Note that the research team always followed the COVID-19 protocols approved by the National Research Council Canada and the Ottawa Fire Service.

Prior to the study, the research team gave a presentation to participants with information about the study and procedure (consent process, voluntary participation, VR equipment used, cybersickness, brief overview of scenarios). Participants had the opportunity to ask questions during and at the end of the presentation.

Following the presentation, participants could sign up for the study by indicating their preferred time to participate on a sign-up sheet. When participants arrived at their assigned time-slot, they read and signed the informed consent form. Afterwards, they completed the knowledge test followed by the sociodemographic questionnaire and the SSQ. After they finished completing the questionnaires, the experimenters explained and demonstrated how to use the VR equipment to participants.

Participants were then immersed in VR for a short practice trial (approximately 5 min) to familiarize themselves with the VR equipment and navigating in the virtual environment. Once participants confirmed being comfortable using the VR equipment and ready to begin the first scenario, the experimenters asked them to exit the practice scenario and start Trial 1 (either Scenario 1 or Scenario 2, depending on the randomly assigned order).

Participants were asked to follow the brief pre-recorded audio instructions (approximately 5 s), which instructed participants to assess the situation and complete a report of the incident. Once their report was complete, participants were to return to the police vehicle to exit the scenario. The experimenters observed the participants via screen casting while they completed the scenario and helped them if they encountered any technical issues. However, they refrained from telling participants how to complete the task as they should already be familiar with the correct procedures to follow. Once participants completed the report via the mobile device, they returned to the starting location (police vehicle) to exit the scenario. If their report contained any mistakes, a pre-recorded audio message informed them that their report contained mistakes or was incomplete, and participants had the opportunity to correct the mistakes before trying to exit the scenario again. Once all answers were correct, a pre-recorded audio message

congratulated them on successfully completing the task and they were automatically returned to the main menu to select the next scenario. The immersion lasted approximately 5–10 min.

After completing Trial 1, the experimenters asked participants if they were experiencing any cybersickness and had them fill out the SSQ. Participants then began Trial 2 (Scenario 1 or Scenario 2, depending on the randomly assigned order). Their task was identical to the previous scenario. The immersion lasted approximately 5–10 min.

After Trial 2, the experimenters again asked participants if they were experiencing any cybersickness and had them fill out the SUS, the NASA-TLX, the SSQ, and the presence questionnaires (PQ and SUS-presence). Participants also completed the PSSUQ and provided feedback on their experience and the VR training tool in a short online survey. Participants then completed the knowledge test for a second time. Finally, the experimenters thanked participants for their participation and asked if they had any questions.

3 Results

R (version 4.2.2, R Core Team, 2022) and RStudio (version 2022.12.0 Build 353, RStudio Team, 2022) were used to perform statistical analyses and to generate plots. Nonparametric tests were conducted due to the small sample size.

3.1 Sociodemographic questionnaire

For statistical analyses, participants from the main study were separated into two groups defined as “Novice” and “Expert”. The group separation was based on self-reported level of experience with the procedures at the Awareness level. Participants had the option to indicate whether they were a novice, an advanced beginner, an intermediate or an expert. Participants also had the option not to answer the question. The participants who responded that they were a novice or an advanced beginner were assigned to one group, called “Novice” ($n = 12$); participants who indicated to be an intermediate or an expert were assigned to another group, called “Expert” ($n = 12$). The separation across age groups, gender, Awareness level, VR experience, and Firefighter experience is presented in Table 2.

The data from the sociodemographic questionnaire showed that there were no significant differences between groups in terms of age, gender, and other demographic variables, based on the results from Fisher’s exact test (see Table 2). This test was performed instead of chi-square tests of independence as the smallest expected frequencies for all variables were below 5.

Table 2 Sociodemographic data of Novice and Expert groups

Variable	Response options	Novice	Expert	<i>p</i> value
Age Group	18–24 years old	1	0	0.291
	25–34 years old	6	4	
	35–44 years old	4	8	
	45–54 years old	1	0	
Gender	Man	9	10	0.590
	Woman	3	1	
	Non-binary	0	1	
VR Experience	Never used virtual reality	10	11	1
	1–5 h per week	2	1	
	6–10 h per week	0	0	
	More than 10 h per week	0	0	
Awareness Level Knowledge	Never heard of Awareness level training	0	0	0.371
	Low (i.e., previously encountered the term "Awareness level" but unsure what it means)	0	0	
	Medium (i.e., previously read materials about Awareness level, familiar with the procedures to follow)	4	2	
	High (i.e., previously received Awareness level training/took an Awareness level course)	7	10	
	Prefer not to answer	1	0	
	Awareness level	1	2	
	Operations Level	11	8	
Hazmat Experience	Technician level	0	2	0.785
	Commander level	0	0	
	Yes, for 1–3 years	1	1	
	Yes, 4–9 years	1	1	
Military Experience	Yes, 10 years or more	0	0	1
	No	10	10	

3.2 User experience and usability

3.2.1 Task completion duration

The overall duration to complete the task was recorded for each participant. Only participants with data for both trials were included in this analysis (total $n=17$). Two participants spent a relatively longer time securing the scene during Scenario 2 (this was the first trial for both participants, i.e., they completed Scenario 2 followed by Scenario 1; see outliers in Fig. 4B). However, despite the relatively longer completion duration, their data were included in the analyses. The final analysis included data from 17 participants.

On average, the overall duration to complete both scenarios was slightly faster in the Novice group, $M=8.09$ (8 min 09 s), $SD=3.57$ (3 min 57 s), when compared to the Expert group, $M=9.25$ (9 min 25 s), $SD=2.95$ (2 min 95 s). However, a Mann–Whitney U Test showed that this difference was not significant, $W=106$, $p=0.199$. See Fig. 4 for task completion duration across groups.

Visual inspection of the data suggests that participants were more efficient in completing the second trial regardless of the scenario (1 or 2). The order in which Scenario 1 or 2 were presented had no effect on the overall time, as generally

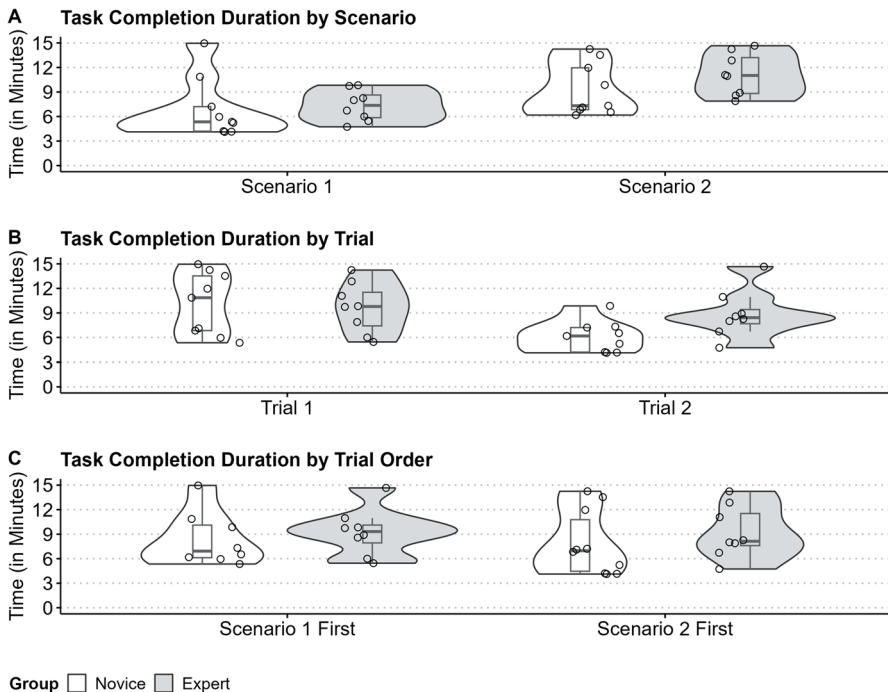


Fig. 4 Task completion duration (in minutes) by scenario, by trial, and by trial order across groups. The line in the centre of the boxplots represents the median. Note that Scenario 1 first and Scenario 2 first is a subsample

participants took longer to complete Scenario 2 than Scenario 1, regardless of their order of presentation (see Fig. 4).

3.2.2 Error rate

Events during the VR training (in both scenarios) were also recorded for each participant (e.g., securing the scene, entering or exiting a particular zone, using inventory items). Prior to statistical analysis, the research team classified the events as positive, neutral, or negative. Analyses were performed on negative events (errors). Only participants with data were included in this analysis (i.e., some participants did not make any errors during either scenario; others made errors during one or both scenarios). The final analysis included data from 16 participants.

The Expert group's error rate during both VR scenarios was higher than the Novice group. However, a Mann–Whitney U Test showed that this difference was not significant, $W = 58.5$, $p = 0.292$. Visual inspection of the data suggests that the error rate in both groups was similar during both trials (generally between 1 and 3 errors). See Fig. 5 for error rate across groups.

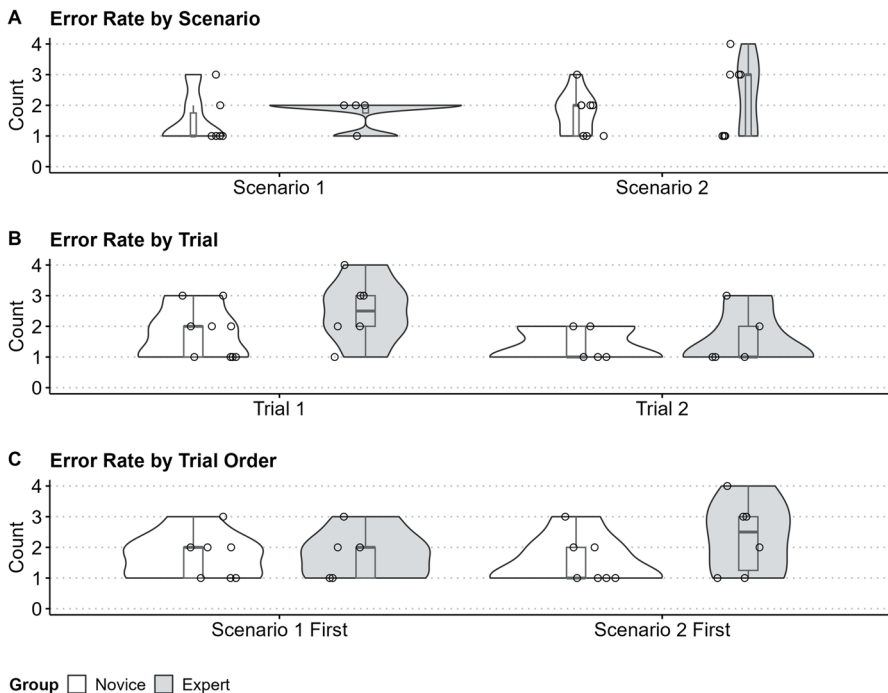


Fig. 5 Error rate by scenario, by trial, and by trial order across groups. The line in the centre of the box-plots represents the median. Note that Scenario 1 first and Scenario 2 first is a subsample

3.2.3 Cybersickness

The SSQ data were computed by using calculation formulas outlined by Kennedy et al. (1993). The results of the overall SSQ showed that both groups had similar baseline levels of comfort. However, the Expert group reported higher levels of overall cybersickness after both VR scenarios compared to the Novice group. A repeated-measures ANOVA of overall cybersickness scores in the Expert group showed that the increase was not significantly higher. For the Novice group, the overall cybersickness levels decreased the longer the participants spent in the virtual environment. The follow-up analysis revealed that the decrease in cybersickness was significant. See Table 3 and Fig. 6 for overall SSQ and subscale scores (i.e., Nausea, Oculomotor, and Disorientation) over time for each group.

When exploring individual sub-scales for each group (i.e., Nausea, Oculomotor, and Disorientation), a similar statistical approach was adopted to explore which dimension contributed to overall cybersickness levels the most. A repeated-measures ANOVA showed that Oculomotor-related discomfort decreased significantly over time in the Novice group. However, nausea and disorientation did not significantly change over time. In the Expert group, the results indicated that Oculomotor and Disorientation scores increased the longer the participants spent in the virtual environment, specifically between pre- and post-training ratings. However, this difference was not significant for either Oculomotor or Disorientation scores. Nausea also did not significantly change over time. When examining scores between the Novice and Expert groups for each sub-scale, none of the comparisons showed significant differences between groups.

3.2.4 Perceived workload

The NASA-TLX data were analyzed by first computing a mean score for each workload subscale and then an overall workload score (i.e., the mean of all six subscales). See Fig. 7 for overall NASA-TLX and sub-scale scores (i.e., Mental Demand, Physical Demand, Temporal Demand, Effort, Frustration and Performance) across groups.

Table 3 Results of repeated-measures ANOVA of SSQ in Novice and Expert groups

Group	Variable	<i>df</i> (between)	<i>df</i> (within)	<i>F</i> ratio	<i>P</i> value
Novice	Overall	2	22	4.44	0.024*
	Nausea	2	22	0.76	0.478
	Oculomotor	1.15	12.6	8.93	0.009*
	Disorientation	2	22	0.38	0.689
Expert	Overall	1.08	11.9	1.02	0.339
	Nausea	1.14	12.5	0.05	0.863
	Oculomotor	1.07	11.7	1.26	0.288
	Disorientation	1.11	12.2	1.90	0.194

* $p < 0.05$

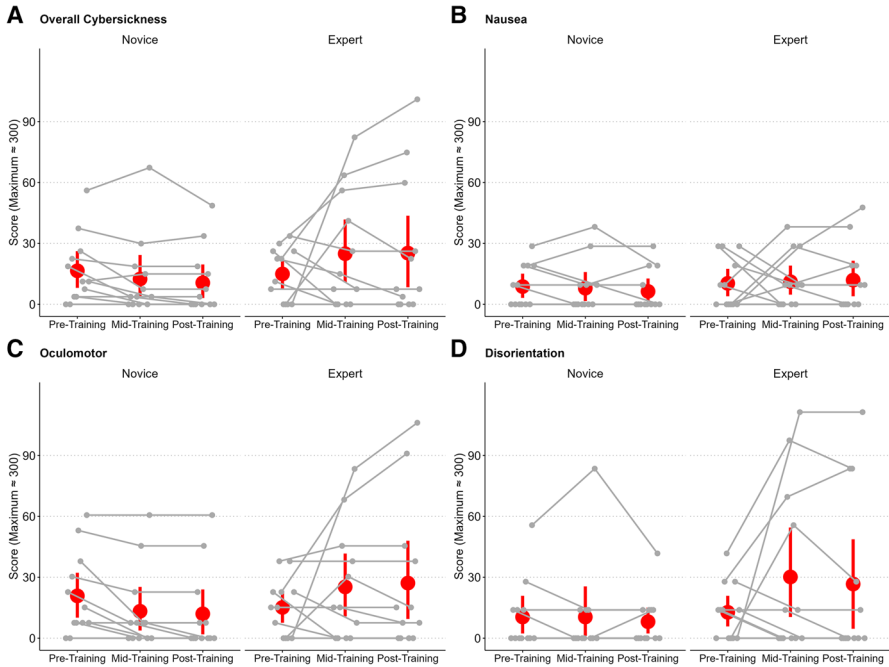


Fig. 6 Overall SSQ (A) and subscale scores (B, C, and D) prior to the VR training (Pre-Training) and following both the first trial (Mid-Training) and the second trial (Post-Training) for individual participants in the Novice and the Expert groups. Red point ranges indicate mean and bootstrapped standard error

A Mann–Whitney U Test performed on overall workload scores showed that both groups experienced similar levels of workload across scenarios with no significant differences between groups, $W=33$, $p=0.076$. A Mann–Whitney U Test was conducted on all subscale scores as well. The results showed that the difference in levels of physical and temporal demand between groups was not significant: $W=62$, $p=0.539$ and $W=70$, $p=0.823$, respectively. The differences in perceived performance, effort, and frustration between both groups were also not significant: $W=62.5$, $p=0.602$, $W=70$, $p=0.822$, and $W=40$, $p=0.178$, respectively. The difference between groups on the mental demand subscale was also not significant, $W=38$, $p=0.054$.

3.2.5 Usability of the VR tool

Descriptive results of the SUS showed that both groups rated the usability of the VR tool as acceptable with mean scores above 68, which is the baseline score for average usability of a system (Brooke, 1996, 2013). The usability ratings from the Expert group, $M=77.92$, $SD=11.42$, were slightly higher than from the Novice

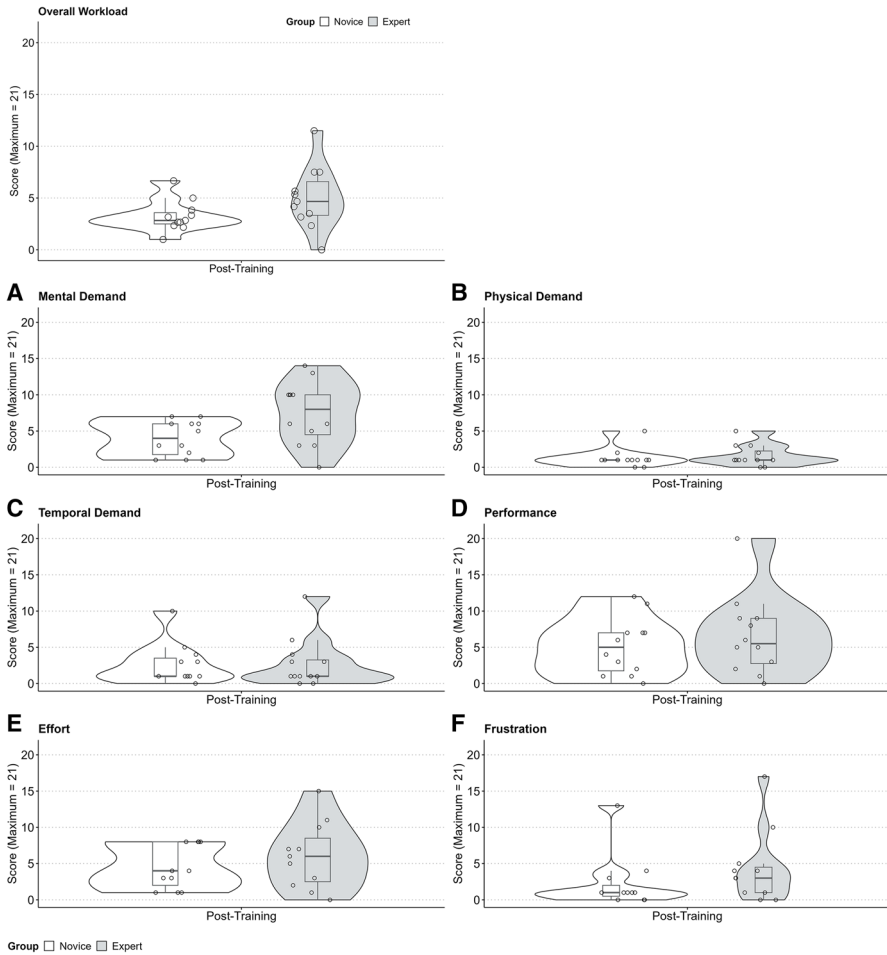


Fig. 7 Overall NASA-TLX and subscale scores (A to F) across groups. The line in the centre of the box-plots represents the median

group, $M = 70.42$, $SD = 14.18$, but the overall difference between the groups was not significant, $W = 51$, $p = 0.235$. See Fig. 8 for overall SUS scores across groups.

At the end of the SUS, participants could list positive and negative aspects of the VR training tool. Content analysis of the responses showed that most of the negative comments were referring to the technical aspect of the simulation – several participants reported that they experienced issues with navigating in the virtual

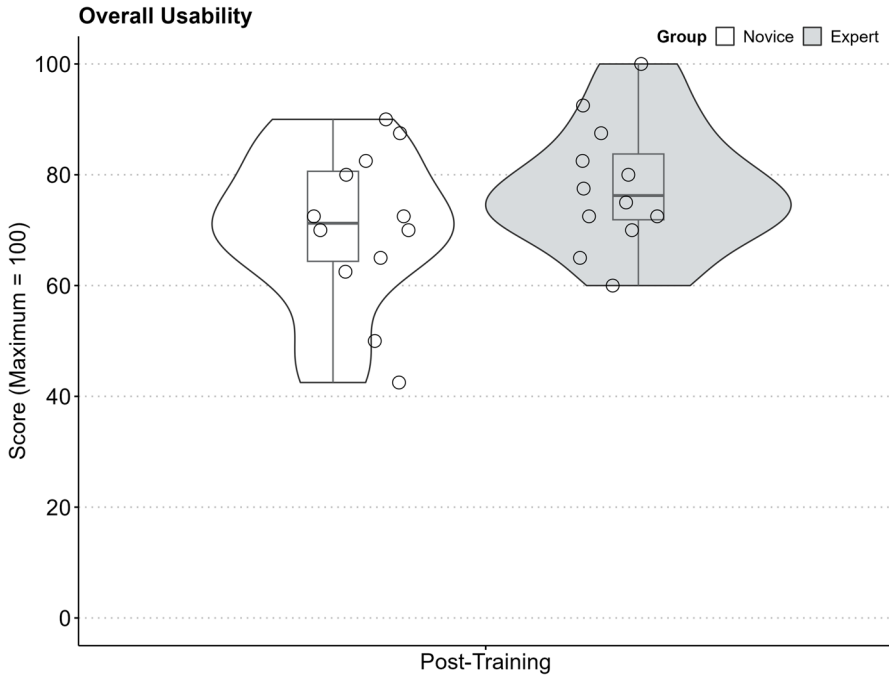


Fig. 8 Overall usability scores across groups. The line in the centre of the boxplots represents the median

environment (e.g., *difficult to use the controls*), specifically the use of binoculars (e.g., *binoculars very sensitive, difficult to focus, hard to see*); and other inventory items (e.g., *difficult to read the ERG as it appeared blurry*); and overall graphics and realism (e.g., *lower visual quality graphics*). For task-related aspects, participants mostly wanted more realistic environment (e.g., *lack of everyday distractions-noises, cars, radio communications, more reporting options available*). In terms of negative aspects from participants' perspective, some reported eye strain, dizziness and feelings of restricted freedom.

The positive aspects of the simulation training from a technological perspective were mostly related to the realism and interactivity of the simulation, (e.g., *interactive, portable, immersive, real scenario, good for replicability*). For task-related aspects, participants liked that they can get great overall practise by using an interactive learning tool (e.g., *good learning opportunity, great scenario practice, good familiarization with hazmat scenarios*), that can be conducted in a safe environment (e.g., *great training while in a safe training environment, educational and hazard free, inexpensive training costs*). From a personal perspective, many participants noted that the training tool was practical, fun, unique, and easy to use in a safe and cost-effective environment.

The results of the Mann–Whitney U Test of PSSUQ scores showed that the Novice group provided a higher rating for overall system usability and information quality than the Expert group. However, the difference was not significant, $W=76$, $p=0.549$ and $W=74$, $p=0.635$, respectively. The Expert group rated the overall interface quality of the system as more satisfactory than the Novice

group, but this difference was also not significant, $W=47$, $p=0.247$. See Fig. 9 for PSSUQ scores across groups.

3.3 Knowledge test

Most participants scored relatively high, both pre- and post-training, with the majority of participants giving correct answers to more than 12 out of 15 questions. The results of a Mann–Whitney U Test showed that there were no significant differences between Novice and Expert groups' scores on the knowledge test before VR training, $W=73$, $p=0.976$. Similarly, there were no significant differences between the Novice and Expert groups' scores from the knowledge test after the training, $W=73$, $p=0.976$. Most (13) participants achieved the same score pre- and post-training, while four participants improved, and seven participants had lower scores post-training. The results of a Wilcoxon Signed-Rank Test showed that the Novice group's knowledge test scores were not significantly different from pre- to post-training, $V=12$, $p=0.270$. The Expert group's scores were also not significantly different from pre- to post-training, $V=15$, $p=0.374$. See Fig. 10 for knowledge test results across groups.

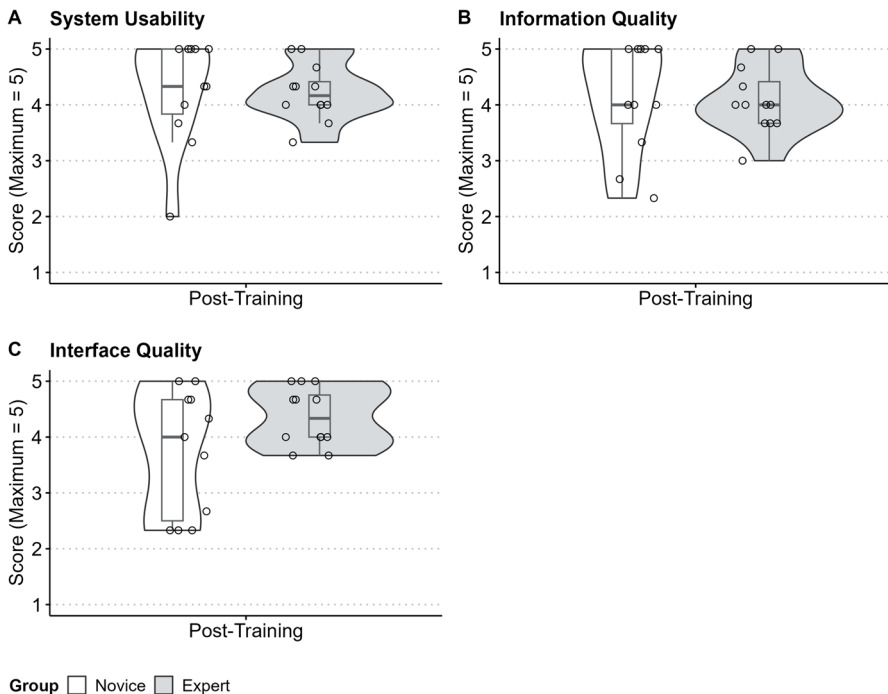


Fig. 9 PSSUQ subscale scores across groups. The line in the centre of the boxplots represents the median

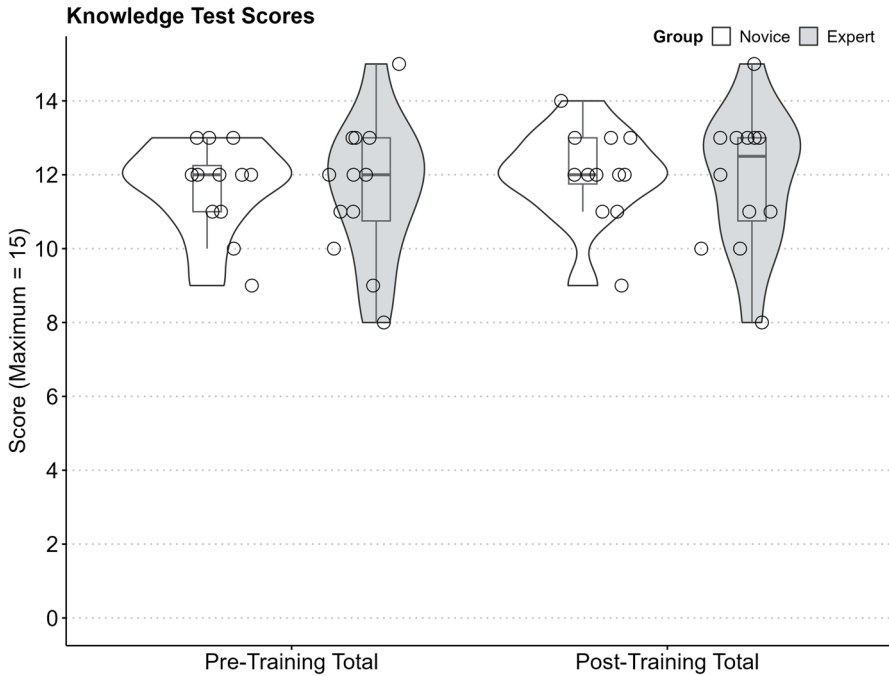


Fig. 10 Knowledge test results at pre- and post-training across groups. The line in the centre of the box-plots represents the median

3.4 Participant feedback

To investigate the perceived value of the VR training tool, participants provided their feedback regarding overall satisfaction with the tool, the potential effectiveness of the tool in contributing to positive training transfer into real-world environments, the perceived value of the tool, and whether they would recommend the tool for training. The results showed that most participants rated the VR tool very positively, with the Expert group ranking the tool slightly higher on overall satisfaction, $M=4.09$, $SD=0.87$ vs. $M=4.00$, $SD=0.89$, but slightly lower on suitability for training transfer compared to the Novice group, $M=3.64$, $SD=1.29$ vs. $M=3.82$, $SD=1.33$. The Expert group also provided higher ratings for willingness to recommend the tool, $M=4.23$, $SD=0.92$ vs. $M=3.91$, $SD=1.04$, as well as for perceived value of the tool compared to the Novice group, $M=4.45$, $SD=0.80$ vs. $M=4.18$, $SD=0.98$. See Fig. 11 for the results of the post-training feedback survey across groups.

The feedback survey also included two open-ended questions where participants could share improvements that they thought could be made to the training tool and any other comments about their experience. Content analysis was conducted on the collected responses and showed that the answers could be categorized as either technology-related improvements or task-related improvements. For the technology-related improvements, most participants mentioned a need to improve on overall realism of the scenario in order to enable distance judgments (e.g., “Possibly

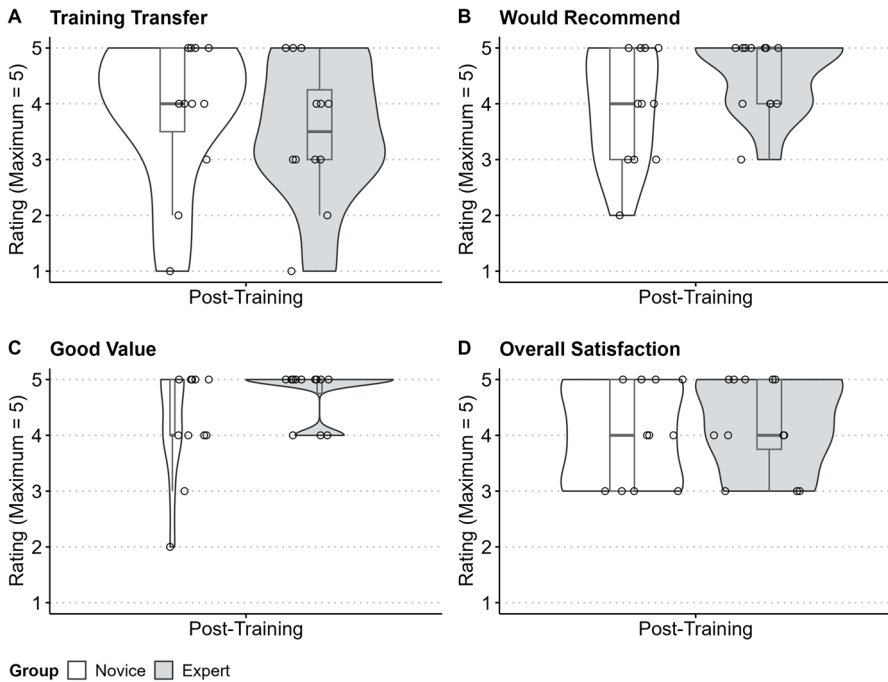


Fig. 11 Post-training feedback survey subscale ratings across groups. The line in the centre of the box-plots represents the median

indicate the initial evacuation distance where the cones are placed.”, and “*Improve clarity and consistency with the landscape to help judge distances*”); additional sounds and foliage movement to simulate wind to further improve the overall realism of the scenario (“*Make it more realistic, more like an actual scene..*”); improving frame rate and resolution to further improve interaction with the inventory items (“*Make the binoculars easier to use as I found it difficult to find specific items quickly*”, and “*Improve resolution/frame rate, functionality of binoculars and focus. Graphics to include more immersive environments (realistic)*”); and providing additional feedback to guide a successful completion of the report (“*Add feedback of what was missing in you report*”).

For the task-related improvements, participants reported wanting a more detailed scenario with additional features, such as a driver (“*There are aspects that are not in the VR scenario that happen in real life...*”), more detailed instructions before the task (“*... more direction of task to be done*”, and “*...review what the red circle and wind direction thing was...*”), as well as additional complexity levels that would enable further progress in training scenarios (e.g., extended reporting features, “*More difficulty levels to continue education*” and “*More extensive questioning in the report, such as PPE suggested, distance for perimeter, what type of agent to apply if leak or spill were to occur...*”).

In the overall comments, most participants expressed enthusiasm about using the tool as they reported that the training tool fun to use (e.g., “*Was very fun to use*”, and “*Great experience*”, and “*This is the future of training!!!!!!*”). Many

participants also reported that the tool was effective for training (e.g., “*Very effective tool for training. I would love to see this initiative continue and expand to other training scenarios.*”) and was valuable for many training scenarios (“*I believe that this type of training is highly valuable for many, many reasons.*”).

Some participants also reported feeling cybersickness symptoms, such as dizziness, nausea, eye strain, and blurred vision. Participants’ feedback will help to make adjustments to the training tool to reduce cybersickness and increase the complexity of the scenarios. Table 4 shows frequencies of responses across categories from qualitative feedback.

4 Discussion

The current study had several objectives: (1) assess the usability of a VR training tool involving dangerous goods with firefighters, (2) assess whether the knowledge test captured learning effects, and (3) assess whether any adjustments should be made to the VR training tool and/or research protocol based on participants’ feedback. Overall, most participants found the VR training tool useful and practical. However, important limitations of the specific to the current study that likely translate to other VR training tools were identified.

4.1 User experience and usability

To meet the first objective, several measures were used to assess the usability of the VR training tool. The results of each measure are discussed below.

Table 4 Content analysis of qualitative responses, showing frequencies of responses for identified categories

	Frequency
Technology-related improvements	
• More realistic (to judge distance, sounds, wind in the trees)	9
• Improve interaction (frame rate, resolution) with tools	8
• Provide additional feedback	2
Task-related improvements	
• Include additional information/instruction before the task	5
• More difficulty levels to continue education (extended report)	3
• Include more detailed scenario (driver)	2
General comments	
• Great for training, highly valuable	8
• Really fun	3
• Additional task would be useful, more complexity	3

4.1.1 Task completion duration

No significant differences were observed in task completion duration. While it visually appeared that the Expert group required more time to complete the scenarios than the Novice group and that completing Scenario 2 took longer, none of these comparisons reached statistical significance. This is consistent with the study conducted by Vankipuram et al. (2010) that evaluated a VR training tool for medical procedures. In this study, expert surgeons took longer to complete a VR training than novices and residents. In a post-experiment interview, experts mentioned that they took more time during the trials to ensure that they completed the tasks accurately whereas novices and residents reported focusing more on the time elapsed.

Visual inspection of the data also suggested that participants completed the second trial faster than the first trial regardless of the scenario. This result is also consistent with Vankipuram et al. (2010) who found that participants tended to complete the second, third, and fourth trials faster than the first. The reduced task completion duration indicates that participants improved in completing the task at hand. This could be accounted for by both familiarity (participants becoming more familiar with the VR tool) and a potential learning effect (e.g., participants having learned that the scene needed to be secured before assessing the scene).

4.1.2 Error rate

The results showed that the Expert group made more errors during both scenarios than the Novice group. This is an unexpected result as previous studies have shown that experts tend to make fewer errors than novices during VR training (Haluck et al., 2001; Vankipuram et al., 2010). Some participants reported during the VR scenarios that they wanted or tried to approach the transport truck to make sure that the driver was not injured as this was part of firefighter training. However, the VR tool displayed a proximity error message and an audio recording when participants were too close to the incident. It would be interesting in a future version of the scenarios to refine the tutoring system and to include a driver in or near the transport truck (already developed but not included in this experiment) or other elements, such as a visible leak or spill, additional inventory items (e.g., 4-gas monitor), different environmental conditions (e.g., smoke, flames), etc. Further development of the tool is planned to improve its interface quality and responsiveness.

4.1.3 Cybersickness

In VR-specific applications, it is important to assess user cybersickness to minimize user discomfort. Overall, self-reported negative side effects of the VR tool were low to moderate. The Expert group experienced higher overall cybersickness during both VR scenarios than the Novice group. While the Novice group's overall cybersickness decreased over time, the Expert group's overall cybersickness increased, especially oculomotor and disorientation symptoms.

After the pilot study, measures were taken to reduce cybersickness, such as participants performing the task seated, preventing participants from moving when

using the binoculars. Despite these measures, some participants reported relatively high levels of cybersickness. The data on cybersickness point to an important general limitation of VR training tools; some participants were noticeably more susceptible to experiencing nausea and other symptoms (see Fig. 6) than others; one even to the degree that they had to terminate their participation. This indicates that VR training may not be suitable for *all* potential trainees. While future work is needed to ensure that negative side effects are minimal, any training application should be tested for unintended side effects prior to deployment.

The cybersickness experienced by participants could potentially be due to conflicting visual and vestibular information during the immersions in VR (Reason & Brand, 1975). Specifically, participants were physically seated during the immersions and used the VR controllers to navigate in the scenarios. This discrepancy between seeing movement in the VE but not physically moving could have contributed to participants' cybersickness.

Some studies have shown that older adults tend to experience higher cybersickness than younger adults (Arns & Cerney, 2005; Knight & Arns, 2006; Petri et al., 2020). Although participants in the Expert group were slightly older than those in the Novice group, age did not seem to have a significant effect on cybersickness. Previous research also suggests that more experienced VR users tend to experience lower cybersickness than first time or less experienced users (Marengo et al., 2019; Tian et al., 2022). Some studies have reported similar results for gaming experience, where users with more video gaming experience tended to experience lower cybersickness than those with less experience (Kourtesis et al., 2023; Weech et al., 2020). However, in the current study, prior VR experience did not seem to significantly affect participants' cybersickness levels, whether they were in the Novice or the Expert group.

4.1.4 Perceived workload

The results revealed that there were no significant differences between the groups in perceived overall workload, physical demand, temporal demand, performance, effort, or frustration. However, the difference between the groups in perceived mental demand approached acceptable levels for statistical significance. Specifically, the results suggest that the Expert group experienced higher mental demand than the Novice group. This result is in line with Grabowski (2021) who found that firefighters experienced higher mental demand when completing VR training than cadets. The author suggested that this may be due to being less familiar with this type of technology.

In the open-ended feedback survey, some participants reported that the task in the VR scenarios was unclear and also reported having difficulty placing the virtual cones on the road to secure the scene. This could have led to increased perceived mental demand. Following the study, the research team made improvements to the interaction method for placing the virtual cones. In future phases of development, written instructions about the task to complete could be included at the beginning of the VR scenarios.

4.1.5 Usability

The total scores from the usability measures showed that overall, the VR training tool seems acceptable for training. Specifically, the mean SUS scores for both groups were above 68, which is the baseline score for average usability of a system (Brooke, 1996, 2013). The SUS scores are similar to previous studies evaluating VR applications (Caiza et al., 2021; Capasso et al., 2022; Grabowski & Jach, 2021; Rhiu et al., 2020). However, a few adjustments and updates still need to be made to the tool to address the limitations above. The results from the PSSUQ showed that the Novice group generally provided higher ratings for overall system usability and information quality than the Expert group. Furthermore, the Expert group provided higher ratings for overall interface quality than the Novice group. However, some participants provided a relatively low rating for the VR tool's information quality and interface quality in both groups, which suggests that the content of the scenarios and the VR tool's interface could be improved.

One example of a particularly challenging interaction within the virtual environment was the use of binoculars. Binoculars magnify an image, which makes it possible, for example, to decipher a plaque from a large distance. However, image magnification also amplifies hand movements which makes steady movements more difficult. This effect appeared to be even stronger in VR given several participants had noticeable difficulties in using the binoculars. Future research is needed to increase the usability of items in the virtual inventory.

4.2 Knowledge

To meet the second objective (i.e., assessing whether the knowledge test captured any learning effects), participants completed a short test before and after the training in VR (Appendix 2). Participants' scores were then analyzed to examine whether there were any differences between pre- and post-training.

All participants performed relatively well on the knowledge test. No significant differences were found in the statistical analysis of the knowledge test scores between pre- and post-training, which indicates that the test failed to detect a learning effect from the VR training. The results also showed that Novice and Expert groups' knowledge test scores did not significantly change from pre- to post-training. One reason for these results could be that all participants who took part in the study had already completed Operations level training. Trainees need to complete the Awareness level training before receiving Operations level training. In other words, a ceiling effect might have been observed. In future work, the research team plans to recruit participants who will be at a beginner's level in firefighter training and only obtained Awareness level training. This will then allow us to determine whether any new knowledge has been gained from the VR training scenarios. The small sample size also limits the generalizability of the results of the study. A larger and more diverse sample will be recruited in future studies.

Another possible explanation for the lack of significant differences between the Novice and Expert groups or between pre- and post-training is that the knowledge test might not have been sensitive enough to discern novices and experts; further it is also possible that the test assessed general knowledge instead of knowledge gained from the training tool per se. Note that the test was developed with the guidance of a senior firefighting officer. The knowledge test will be reviewed and improved in future work. The research team will also test the knowledge evaluation instrument with a larger sample of firefighters.

4.3 Participant feedback

To meet the third objective (i.e., assess whether any adjustments should be made to the VR training tool and/or research protocol), participants listed negative and positive aspects of the training tool in the SUS and completed a feedback survey after the training in VR. Participants' feedback helped to identify improvements that could be made to the training tool in future work.

The results indicated that overall, the tool was accepted as being suitable for the training of first responders in terms of providing overall satisfaction with the experience. Furthermore, in aggregate, participants reported that they saw good value in the training tool and perceived it as being acceptable for training transfer capabilities. The Expert group provided slightly higher ratings for the tool's recommendation, which suggests they saw some potential in the VR tool to provide safe and realistic training. It should be noted that there was a fairly wide range of responses in both groups, suggesting that there may be user subgroups at the population level.

In the SUS, many participants reported that the tool's most positive aspects were its cost effectiveness, practicality, and safety. In terms of negative aspects, participants mentioned that they experienced issues with navigation and inventory use. In the post-training feedback survey, participants noted that they found the training tool fun to use and thought that it was effective for training. However, some improvements need to be made to minimize cybersickness and refine the tool's interface. These results are in line with Engelbrecht et al. (2019) who noted cost effectiveness and increased safety for high-risk training as being strengths of VR-based training for firefighters. They also identified technology barriers as being a weakness.

4.4 Limitations

The following limitations should be considered. First, participants were assigned to two groups – Novices and Experts. Participants did not only differ in their prior self-reported expertise, but also their age. Overall, participants in the Expert group were slightly older than those in the Novice group. This implies that effects of the level of expertise and age are difficult to differentiate. This is particularly relevant given that VR training tools are still a relatively novel technology and therefore general comfort (or discomfort) with the technology, regardless of the training content, might differ across age groups. For instance, some research has shown that younger adults tend to benefit more from VR training than older adults (Fang & Huang, 2021).

Second, a wide range of responses was observed across all feedback categories (simulator sickness, workload, usability). This indicates, broadly speaking, that while VR training might be embraced by many, it might not be for everyone. However, any effective training needs to be able to engage and help a wide range of trainees.

Finally, the main study tested 24 participants; unfortunately, this implies that the current study was underpowered. A post-hoc power analysis revealed that for a hypothetical effect size of $d=0.5$, the achieved power was only 0.31. Future research needs to collect data from larger samples.

5 Conclusion and future research directions

Although the VR training tool has some limitations that should be addressed in future phases of development, few VR training tools for road transportation incidents involving dangerous goods currently exist. Most VR tools previously developed for firefighters focus on other areas of training (search and rescue operations, Backlund et al., 2007; simulating fire extinguishing and other firefighting skills, Bellemans et al., 2020). The VR tool presented in this paper provides a realistic dangerous goods scenario and training goals based on the ERG 2020 edition (Transport Canada, 2020a), NFPA 470 guidelines (National Fire Protection Association, 2021), and Transport Canada Competency guidelines (Transport Canada, 2016a). The main contribution of the user study is to help improve the training provided to firefighters to safely respond to a dangerous goods incident. Providing adequate training for these types of incidents will help minimize accidents, injuries, and deaths of both firefighters and the public.

Real-world training for firefighters responding to a dangerous goods incident can be dangerous, logistically challenging, and costly, especially in remote areas. VR technology offers practical, cost-efficient, and safe solution to increase accessibility to training in these areas. In the current study, a VR training tool with two scenarios involving dangerous goods was developed for firefighters at the Awareness level. The results suggest that the tool had good overall usability and acceptability. However, improvements need to be made to the tool before the next phase of testing. In future work, the research team plans to further refine the VR training scenarios and ensure that the knowledge test is directly related to the VR training scenarios. Performance should be further assessed in future studies as the current study was unable to conclusively show that the VR training tool significantly improved participants' knowledge. In addition, the training transfer effectiveness of the VR tool should be assessed (e.g., comparing the VR training tool to traditional classroom training).

Appendix 1

Virtual reality scenarios

Scenario 1

The scenario was intended to be relatively simple and a relatively common occurrence. Participants were faced with the following scenario:

- A transport truck with a cabin and a trailer is jackknifed close to a road.
- The truck is transporting an (initially) unknown (potentially) hazardous material.
- There are no obvious signs of fire (e.g., smoke).
- There are no buildings or people in the immediate vicinity of the incident.
- It is daytime and overall visibility is good.
- The environment contains vegetation off the road, and there is a point of elevation not too far away.
- There is wind blowing from the southwest.
- Daytime temperatures are below 30 °C and above 0 °C.
- There are no dynamic changes in the scenario (i.e., no fire development).

Scenario 2

Scenario 2 is identical to Scenario 1, except that the hazardous material is a gasoline and ethanol mixture (3475) instead of gasoline. In addition, the scenario contains different environmental conditions and risks. For example, the wind is blowing from the southeast, the point of elevation is in a different location, and a small river/brook and houses are visible.

Appendix 2

Knowledge test

Note that the correct answers to the knowledge test are in **bold**. The abbreviation “hazmat” refers to “hazardous material” – participants were already aware of this as it is a commonly used abbreviation.

Before training test

After training test

Participant ID: _____

Please answer the questions to the best of your knowledge.

Q1. The APIE process is a simple 4-step process response model that:

- A. **Can guide responders’ action at hazmat incidents.**
- B. Is not effective in large or complex hazmat incidents.
- C. Is most effective in large or complex incidents.
- D. Is required to be used at all North American hazmat incidents.

Q2. The most common route of entry in the body of hazardous material is:

- A. Injection.
- B. Ingestion.
- C. **Inhalation.**
- D. Absorption.

Q3. Which mechanism of harm is the greatest threat at hazmat incidents?

- A. Toxicity.
- B. Corrosivity.
- C. **Energy release.**

D. Acute exposure.

Q4. Competing priorities, distractions, and information overload may contribute to:

A. A successful rescue.

B. Loss of situation awareness.

C. Reduced loss of life and property.

D. Correct identification of the hazardous material.

Q5. The first priority at any incident is the:

A. Isolation of the hazardous materials.

B. Preservation of potential evidence.

C. Protection and preservation of the environment.

D. Protection and safety of emergency responders.

Q6. What colour is a placard for flammable material?

A. White.

B. Orange.

C. Red.

D. Yellow.

Q7. The Emergency Response Guidebook (ERG) is primarily designed to be used at:

A. At sea.

B. Health care facilities.

C. Fixed-facility locations.

D. Incidents occurring on highway or railroad.

Q8. The three hazard control zones are referred to as:

- A. A, B, and C.
- B. One, two, and three.
- C. Hot, warm, and cold.**
- D. Red, orange, and yellow.

Q9. What are the two main product control strategies?

- A. Salvage and overhaul.
- B. Spill control and leak control.**
- C. Hazard-control zones and fire control.
- D. Sheltering in place and defending in place.

Q10. What information needs to be transmitted to the responding hazmat team when reporting road incident involving dangerous materials? (Please select all that apply)

- A. Persons name and contact.
- B. Location and weather conditions.
- C. Source and cause of spill.
- D. Severity of the incident.
- E. Shape/size of a vehicle.
- F. Placard number and colour.
- G. All of the above.**

Q11. What level of response is required in the incidents where the containers that are involved in an accident have no visible damage?

- A. Level I.**
- B. Level II.
- C. Level III.

Q12. What are the 4 main responsibilities at the Awareness level when reporting possible hazmat incident? Please select all that apply.

- A. Recognise the presence of hazardous material.
- B. Secure the area.
- C. Protect yourself and others.
- D. Call for trained personnel.
- E. All of the above.**

Q13. How can certain hazards be identified during the road incident involving transportation of dangerous material? Please select all that apply.

- A. Colour of the label or placard.**
- B. Shape/size of the vehicle.**
- C. Shipping documents.**
- D. Driver's clothes.
- E. All of the above.

Q14. What are the possible tools that first responder can use when assessing road accident involving transportation of dangerous materials. Please select all that apply.

- A. Binoculars.**
- B. Hammer.
- C. ERG book.**
- D. Magnifying glass.
- E. All of the above.

Q15. What is the first responder's best defense when first arriving on scene of a hazardous material emergency? Please choose all that apply.

- A. Check your phone for missed call and messages.
- B. Assume worst case scenario.
- C. Prevent anyone entering the initial isolation zone.**
- D. Call friends and tell them what happened.
- E. All of the above.

Acknowledgements The authors wish to thank Eric DeMarbre for his support in developing the virtual reality training tool. The authors would also like to thank H el ene Fournier, Shirley MacLeod, Chunyun Ma, and Cathy Agyemang for their support in the literature review for the study. The authors wish to thank the Ottawa Fire Service for their support in the data collection for the study; in particular, Andrew Ferguson for his support in coordinating the data collection and guidance during the development of the knowledge test. The authors would also like to thank Stephen Downes for supporting a Technical Advisory Group for this project and the members of the Technical Advisory Group for their expertise and feedback throughout the development of the virtual reality training tool. The authors would also like to thank Noureddine B enichou for his guidance and advice throughout this project and for supporting the Technical Advisory Group. This project was supported in part by collaborative research funding from the National Research Council Canada’s Artificial Intelligence for Logistics Program.

Authors’ contribution Maxine Berthiaume: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – Review & Editing. Max Kinaterer: Conceptualization, Methodology, Investigation, Formal analysis, Writing – Review & Editing. Bruno Emond: Conceptualization, Methodology, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. Natalia Cooper: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – Review & Editing. Ishika Obeegadoo: Investigation, Formal analysis, Writing – Review & Editing. Jean-Fran ois Lapointe: Conceptualization, Methodology, Writing – Review & Editing, Funding acquisition.

Funding Open Access funding provided by National Research Council Canada. This work was funded by the National Research Council Canada (Project # A1-016850). Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not reflect the views of the National Research Council Canada’s Artificial Intelligence for Logistics Program.

Data availability Data and R code for the study can be found in the following repository: <https://doi.org/10.4224/40003205>.

Declarations

Ethics declarations The pilot study and the main user study were approved by the National Research Council Canada’s (NRC) Research Ethics Board (protocol #2021–157). All experimental procedures performed involving human participants were in accordance with the institutional research committee ethical guidelines. All participants gave their informed consent form prior to participating.

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arns, L. L., & Cerney, M. M. (2005). The relationship between age and incidence of cybersickness among immersive environment users. *IEEE Proceedings. VR 2005. Virtual Reality, 2005* (pp. 267–268). IEEE. <https://doi.org/10.1109/VR.2005.1492788>

- Backlund, P., Engstrom, H., Hammar, C., Johannesson, M., & Lebram, M. (2007). Sidh – a game based firefighter training simulation. *2007 11th International Conference Information Visualization (IV '07)* (pp. 899–907). IEEE. <https://doi.org/10.1109/IV.2007.100>
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
- Bellemans, M., Lamrnens, D., De Sloover, J., De Vleeschauwer, T., Schoofs, E., Jordens, W., Van Steenhuyse, B., Mangelschots, J., Selli, S., Hamesse, C., Freville, T., & Haeltermann, R. (2020). *Training firefighters in virtual reality*. 2020 International Conference on 3D Immersion (IC3D) (pp. 1–6). IEEE. <https://doi.org/10.1109/IC3D51119.2020.9376336>
- Berkowitz, Z., Horton, D. K., & Kaye, W. E. (2004). Hazardous substances releases causing fatalities and/or people transported to hospitals: Rural/agricultural vs. other areas. *Prehospital and Disaster Medicine*, 19(3), 213–220. <https://doi.org/10.1017/s1049023x00001801>
- Bliss, J., Tidwell, P., & Guest, M. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence*, 6(1), 73–86. <https://doi.org/10.1162/pres.1997.6.1.73>
- Braun, P., Grafelmann, M., Gill, F., Stolz, H., Hinckeldeyn, J., & Lange, A.-K. (2022). Virtual reality for immersive multi-user firefighter training scenarios. *Virtual Reality & Intelligent Hardware*, 4(5), 406–417. <https://doi.org/10.1016/j.vrih.2022.08.006>
- Brooke, J. (1996). SUS: A “quick and dirty” usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Caiza, G., Riofriolo-Morales, M., Gallo, C. V., Alvarez, T. S., Lopez, W. O., & Garcia, M. V. (2021). Virtual reality system for training in the detection and solution of failures in induction motors. In M. Affenzeller, A. G. Bruzzone, E. Jimenez, F. Longo, & A. Petrillo (Eds.), *33rd European Modeling and Simulation Symposium, EMSS 2021* (pp. 199–207). Dime University of Genoa. <https://doi.org/10.46354/i3m.2021.emss.027>
- Canadian Association of Fire Chiefs. (2021). *The great Canadian volunteer firefighter census 2021*. Retrieved February 10, 2023, from https://cdn.ymaws.com/cafc.ca/resource/resmgr/census_2021/EN_Census_Report.pdf
- Canadian Association of Fire Chiefs. (2022). The great Canadian fire census 2022: Report of the Canadian Association of Fire Chiefs answer the call and data committees. Retrieved March 24, 2023 from https://cdn.ymaws.com/cafc.ca/resource/resmgr/grweek2023/English_2022_Census.pdf
- Capasso, I., Bassano, C., Bracco, F., Solari, F., Viola, E., & Chessa, M. (2022). A VR multiplayer application for fire fighting training simulations. In L. T. De Paolis, P. Arpaia, & M. Sacco (Eds.), *Extended Reality. XR Salento 2022* (pp. 130–138). Springer. https://doi.org/10.1007/978-3-031-15546-8_11
- Chover, M., Sotoca, J. M., & Marín-Lora, C. (2022). Virtual reality versus desktop experience in a dangerous goods simulator. *International Journal of Serious Games*, 9(2), 63–78. <https://doi.org/10.17083/ijsg.v9i2.493>
- Clifford, R. M. S., Hoermann, S., Marcadet, N., Oliver, H., Billinghamurst, M., & Lindeman, R. W. (2018a). Evaluating the effects of realistic communication disruptions in VR training for aerial firefighting. *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (pp. 1–8). IEEE. <https://doi.org/10.1109/VS-Games.2018.8493423>
- Clifford, R. M. S., Khan, H., Hoermann, S., Billinghamurst, M., & Lindeman, R. W. (2018b). Development of a multi-sensory virtual reality training simulator for airborne firefighters supervising aerial wild-fire suppression. *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)* (pp. 1–5). IEEE. <https://doi.org/10.1109/VAR4GOOD.2018.8576892>
- Cohen, A. (1998). A multidimensional evaluation of fire fighter training for hazardous materials response: First results from the IAFF program. *American Journal of Industrial Medicine*, 34(4), 331–341. [https://doi.org/10.1002/\(SICI\)1097-0274\(199810\)34:4%3c331::AID-AJIM6%3e3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-0274(199810)34:4%3c331::AID-AJIM6%3e3.0.CO;2-V)
- Cohen-Hatton, S. R., & Honey, R. C. (2015). Goal-oriented training affects decision-making processes in virtual and simulated fire and rescue environments. *Journal of Experimental Psychology: Applied*, 21(4), 395–406. <https://doi.org/10.1037/xap0000061>
- Cook, C. (2018). *Designing a virtual embedded scenario-based military simulation training program using educational and design instructional strategies* (Publication No. 5854) [Doctoral dissertation, University of Central Florida]. Electronic Theses and Dissertations. Retrieved February 9, 2023, from <https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=6854&context=etd>

- Corelli, F., Battegazzorre, E., Strada, F., Bottino, A., & Cimellaro, G. P. (2020). Assessing the usability of different virtual reality systems for firefighter training. *4th International Conference on Human Computer Interaction Theory and Applications* (pp. 146–153). SciTePress. <https://doi.org/10.5220/0008962401460153>
- Engelbrecht, H., Lindeman, R. W., & Hoermann, S. (2019). A SWOT analysis of the field of virtual reality for firefighter training. *Frontiers in Robotics and AI*, 6, 101. <https://www.frontiersin.org/articles/10.3389/frobt.2019.00101>
- Emond, B., Lapointe, J.-F., Kinateder, M., DeMarbre, E., Berthiaume, M., Kondratova, I., Cooper, N., Benichou, N., & Downes, S. (2022). HCI issues, design and development for a first responders VR training system on dangerous goods transportation incidents. In P. Zaphiris & A. Ioannou, *Learning and collaboration technologies: Novel technological environments* (pp. 48–60). Springer, Cham. https://doi.org/10.1007/978-3-031-05675-8_5
- Fang, Y.-M., & Huang, Y.-J. (2021). Comparison of the usability and flow experience of an exercise promotion virtual reality programme for different age groups. *Behaviour & Information Technology*, 40(12), 1250–1264. <https://doi.org/10.1080/0144929X.2021.1938680>
- Gainer, S., Eadara, S., Haskins, J., Huse, W., Zhu, B., Boyd, B., Laird, C., Farantatos, J., & Jerald, J. (2020). A customized input device for simulating the detection of hazardous materials. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 7–12). IEEE. <https://doi.org/10.1109/VRW50115.2020.00288>
- Government of Canada (2020). *Number of dangerous goods commodities involving reportable accidents by class, release location, damage type, containment type, emergency response personnel, phase of transport, accident type, action type and contributing factor*, Transport Canada. Retrieved February 9, 2023, from <https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=3810025601>
- Grabowski, A. (2021). Practical skills training in enclosure fires: An experimental study with cadets and firefighters using CAVE and HMD-based virtual training simulators. *Fire Safety Journal*, 125, 103440. <https://doi.org/10.1016/j.firesaf.2021.103440>
- Grabowski, A., & Jach, K. (2021). The use of virtual reality in the training of professionals: With the example of firefighters. *Computer Animation and Virtual Worlds*, 32(2), e1981. <https://doi.org/10.1002/cav.1981>
- HadjSassi, M. S., Battisti, F., & Carli, M. (2022). Simulation-based virtual reality training for firefighters. *Electronic Imaging*, 34, 194-1–194-5. <https://doi.org/10.2352/EI.2022.34.10.IPAS-194>
- Haluck, R. S., Webster, R. W., Snyder, A. J., Melkonian, M. G., Mohler, B. J., Dise, M. L., & Lefever, A. (2001). A virtual reality surgical trainer for navigation in laparoscopic surgery. In J. D. Westwood, H. Miller Hoffman, G. T. Mogel, D. Stredney, & R. A. Robb (Eds.), *Medicine Meets Virtual Reality 2001* (pp. 171–176). IOS Press.
- Hammar Wijkmark, C., Heldal, I., & Metallinou, M.-M. (2019). Can remote virtual simulation improve practice-based training? Presence and performance in incident commander education. *PRESENCE: Virtual and Augmented Reality*, 28, 127–152. https://doi.org/10.1162/pres_a_00346
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52, 139–183. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Haskins, J., Zhu, B., Gainer, S., Huse, W., Eadara, S., Boyd, B., Laird, C., Farantatos, J., & Jerald, J. (2020). Exploring VR training for first responders. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 57–62). IEEE. <https://doi.org/10.1109/VRW50115.2020.00018>
- Haynes, H. J. G., & Stein, G. P. (2018). *Canadian fire department profile, 2014–2016*. Retrieved February 9, 2023, from <https://www.nfpa.org/-/media/Files/News-and-Research/Fire-statistics-and-reports/Emergency-responders/oscanada.pdf>
- Hsu, E. B., Li, Y., Bayram, J. D., Levinson, D., Yang, S., & Monahan, C. (2013). State of virtual reality based disaster preparedness and response training. *PLoS Currents*, 5. <https://doi.org/10.1371/currents.dis.1ea2b2e71237d5337fa53982a38b2aff>
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- Knight, M. M., & Arns, L. L. (2006). The relationship among age and other factors on incidence of cybersickness in immersive environment users. *Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization* (p. 162). Association for Computing Machinery. <https://doi.org/10.1145/1140491.1140539>

- Kourtesis, P., Amir, R., Linnell, J., Argelaguet, F., & MacPherson, S. E. (2023). Cybersickness, cognition, & motor skills: The effects of music, gender, and gaming experience. *IEEE Transactions on Visualization and Computer Graphics*, 29(5), 1–11. <https://doi.org/10.1109/TVCG.2023.3247062>
- Lewis, J. R. (1995). IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57–78. <https://doi.org/10.1080/10447319509526110>
- Liu, R., Becerik-Gerber, B., & Lucas, G. M. (2023). Effectiveness of VR-based training on improving occupants' response and preparedness for active shooter incidents. *Safety Science*, 164, 106175. <https://doi.org/10.1016/j.ssci.2023.106175>
- Lofca, K., Haskins, J., Jerald, J., & Kopper, R. (2022). Studying the effect of physical realism on time perception in a HAZMAT VR simulation. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)* (pp. 884–885). IEEE. <https://doi.org/10.1109/VRW5335.2022.00293>
- MacLeod, S. (2020). *AR/VR/XR training for firefighters / first responders* (Project No: SM – 023). NRC-Library and Information Management Services.
- Marengo, J., Lopes, P., & Boulic, R. (2019). On the influence of the supine posture on simulation sickness in virtual reality. *2019 IEEE Conference on Games (CoG)* (pp. 1–8). IEEE. <https://doi.org/10.1109/CIG.2019.8848098>
- McLellan, S., Muddimer, A., & Peres, S. C. (2012). The effect of experience on System Usability Scale ratings. *Journal of User Experience*, 7(2), 56–67.
- Melnikova, N., Wu, J., Yang, A., & Orr, M. (2018). Acute chemical incidents with injured first responders, 2002–2012. *Disaster Medicine and Public Health Preparedness*, 12(2), 211–221. <https://doi.org/10.1017/dmp.2017.50>
- Meta Quest. (2022). *Meta Quest Developer Hub*. Retrieved February 10, 2023, from <https://developer.oculus.com/documentation/unity/ts-odh/>
- Narciso, D., Melo, M., Raposo, J. V., Cunha, J., & Bessa, M. (2020). Virtual reality in training: An experimental study with firefighters. *Multimedia Tools and Applications*, 79, 6227–6245. <https://doi.org/10.1007/s11042-019-08323-4>
- National Fire Protection Association. (2012). *NFPA 1403: Standard on live fire training evolutions*. Retrieved February 9, 2023, from <https://www.nfpa.org/assets/files/aboutthecodes/1403/tia1403-12-1.pdf>
- National Fire Protection Association. (2017). *NFPA 1072: Standard for hazardous materials/weapons of mass destruction emergency response personnel professional qualifications*. Retrieved February 9, 2023, from <https://www.nfpa.org/codes-andstandards/all-codes-and-standards/list-of-codes-and-standards/detail?code=1072>
- National Fire Protection Association. (2018). *NFPA 472: Standard for competence of responders to hazardous materials/weapons of mass destruction incidents*. Retrieved February 9, 2023, from <https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=472>
- National Fire Protection Association. (2021). *NFPA 470: Hazardous materials/weapons of mass destruction (WMD) standard for responders*. Retrieved February 9, 2023, from <https://www.nfpa.org/codes-and-standards/all-codes-and-standards/list-of-codes-and-standards/detail?code=470>
- Ooi, S., Tanimoto, T., & Sano, M. (2019). *Virtual reality fire disaster training system for improving disaster awareness*. *Proceedings of the 2019 8th International Conference on Educational and Information Technology* (pp. 301–307). Association for Computing Machinery. <https://doi.org/10.1145/3318396.3318431>
- Petri, K., Feuerstein, K., Folster, S., Bariszlovich, F., & Witte, K. (2020). Effects of age, gender, familiarity with the content, and exposure time on cybersickness in immersive head-mounted display based virtual reality. *American Journal of Biomedical Sciences*, 12(2), 107–121. <https://doi.org/10.5099/aj200200107>
- R Core Team. (2022). *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing*, Vienna, Austria. Retrieved February 10, 2023, from <https://www.R-project.org/>
- Ramsden, R., Smith, J., Turcotte, K., Garis, L., Kunz, K., Maxim, P., Thomas, L., & Pike, I. (2018). *Determinants of injury and death in Canadian firefighters*. Retrieved February 9, 2023, from <https://cjr.ufv.ca/wp-content/uploads/2018/03/Determinants-of-Injury-.pdf>
- Reason, J. T., & Brand, J. J. (1975). *Motion sickness*. Academic Press.
- Reim, L., Kallus, W., Pszeida, M., Schneeberger, M., Schönauer, C., Peer, A., Feischl, R., Aumayr, G., Wallner, D., Dini, A., Weber, A., Ladstätter, S., Almer, A., & Paletta, L. (2022). Evaluation

- of virtual reality-based first responder training under physiological and cognitive-emotional strain. *Proceedings of 13th AHFE International Conference on Cognitive Computing and Internet of Things*. AHFE International. <https://doi.org/10.54941/ahfe1001840>
- Rhiu, I., Kim, Y. M., Kim, W., & Yun, M. H. (2020). The evaluation of user experience of a human walking and a driving simulation in the virtual reality. *International Journal of Industrial Ergonomics*, 79, 103002. <https://doi.org/10.1016/j.ergon.2020.103002>
- RStudio Team. (2022). RStudio: Integrated Development for R. RStudio, PBC. Retrieved February 10, 2023, from <http://www.rstudio.com/>
- Sauro, J., & Lewis, J. R. (2005). Estimating completion rates from small samples using binomial confidence intervals: Comparisons and recommendations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 49(24), 2100–2103. <https://doi.org/10.1177/154193120504902407>
- Slater, M., Usoh, M., & Steed, A. (1995). Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction*, 2(3), 201–219. <https://doi.org/10.1145/210079.210084>
- Smith, J. D., & Steel, J. (2000). The use of virtual simulation for dismounted infantry training. *Proceedings of the 2nd NATO Modelling and Simulation Conference* (pp. 15-1–15-9). RTO/NATO. Retrieved February 9, 2023, from https://www.academia.edu/18956762/The_Use_of_Virtual_Simulation_for_Dismounted_Infantry_Training
- Summers, J. (2012). Simulation-based military training: An engineering approach to better addressing competing environmental, fiscal, and security concerns. *Journal of the Washington Academy of Sciences*, 98(1), 9–29. Retrieved February 9, 2023, from <https://www.jstor.org/stable/24536565>
- Tate, D. L., Sibert, L., & King, T. (1997). Using virtual environments to train firefighters. *IEEE Computer Graphics and Applications*, 17(6), 23–29. <https://doi.org/10.1109/38.626965>
- Taylor, R., Macy, G., Hwang, J., Golla, V., Cann, C., & Sanyang, E. (2022). Using collaborative partnerships to engage firefighters in rural communities. *International Journal of Environmental Research and Public Health*, 19(4), 2009. <https://doi.org/10.3390/ijerph19042009>
- Tian, N., Lopes, P., & Boulic, R. (2022). A review of cybersickness in head-mounted displays: Raising attention to individual susceptibility. *Virtual Reality*, 26(4), 1409–1441. <https://doi.org/10.1007/s10055-022-00638-2>
- Transport Canada. (2016a). *Competency guidelines for responders to incidents of flammable liquids in transport, high-hazard flammable trains*. Retrieved February 9, 2023, from https://tc.canada.ca/sites/default/files/migrated/tc_competency_guidelines_e.pdf
- Transport Canada. (2016b). *Emergency response task force: Final report and recommendations—A comprehensive strategy for improving transportation of dangerous goods incident response in Canada*. Retrieved February 9, 2023, from <https://tc.canada.ca/sites/default/files/2020-06/ERTF-Report-eng.pdf>
- Transport Canada. (2020a). *2020 Emergency Response Guidebook*. Retrieved February 9, 2023, from <https://tc.canada.ca/sites/default/files/2020-08/PDF%20English.pdf>
- Transport Canada. (2020b). *Transportation in Canada: Statistical addendum*. Retrieved February 9, 2023, from https://tc.canada.ca/sites/default/files/2021-06/transportation_in_canada_statistical_addendum.pdf
- Transport Canada. (2020c). *Transportation of dangerous goods*. Transport Canada. Retrieved February 9, 2023, from <https://tc.canada.ca/en/corporate-services/policies/transportation-dangerous-goods>
- Tullis, T., & Stetson, J. (2004). A comparison of questionnaires for assessing website usability. Usability Professionals Association (UPA) Presentation. Retrieved February 27, 2023, from https://www.researchgate.net/publication/228609327_A_Comparison_of_Questionnaires_for_Assessing_Website_Usability
- Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, 9(5), 497–503. <https://doi.org/10.1162/105474600566989>
- Vankipuram, M., Kahol, K., McLaren, A., & Panchanathan, S. (2010). A virtual reality simulator for orthopedic basic skills: A design and validation study. *Journal of Biomedical Informatics*, 43(5), 661–668. <https://doi.org/10.1016/j.jbi.2010.05.016>
- Vasquez, M. N. (2018). *The experiential development of hands-on skills: A multiple-case study of online training amongst firefighters* (Publication No. 10791017) [Doctoral dissertation, Capella University]. ProQuest Dissertations Publishing. Retrieved February 9, 2023, from <https://www.proquest.com/openview/aace7ae0a6b07ed146ad2ee6b6c463c3/1?pq-origsite=gscholar&cbl=18750>

- Vichitvejpaisal, P., Yamee, N., & Marsertsri, P. (2016). Firefighting simulation on virtual reality platform. *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1–5). IEEE. <https://doi.org/10.1109/JCSSE.2016.7748912>
- Weech, S., Kenny, S., Lenizky, M., & Barnett-Cowan, M. (2020). Narrative and gaming experience interact to affect presence and cybersickness in virtual reality. *International Journal of Human-Computer Studies*, *138*, 102398. <https://doi.org/10.1016/j.ijhcs.2020.102398>
- Wijkmark, C. H., Heldal, I., & Metallinou, M.-M. (2021a). Experiencing immersive VR simulation for firefighter skills training. *Proceedings of the 18th International Conference on Information Systems for Crisis Response and Management* (pp. 913–921). ISCRAM. Retrieved August 8, 2022, from <https://hdl.handle.net/11250/2829058>
- Wijkmark, C. H., Metallinou, M. M., & Heldal, I. (2021b). Remote virtual simulation for incident commanders—Cognitive aspects. *Applied Sciences*, *11*(14), 6434. <https://doi.org/10.3390/app11146434>
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. <https://doi.org/10.1162/105474698565686>
- Xu, Z., Lu, X. Z., Guan, H., Chen, C., & Ren, A. Z. (2014). A virtual reality based fire training simulator with smoke hazard assessment capacity. *Advances in Engineering Software*, *68*, 1–8. <https://doi.org/10.1016/j.advengsoft.2013.10.004>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Maxine Berthiaume^{1,2}  · Max Kinateder^{1,3}  · Bruno Emond¹  ·
Natalia Cooper¹  · Ishika Obegadoo¹ · Jean-François Lapointe¹ 

✉ Maxine Berthiaume
Maxine.Berthiaume@nrc-cnrc.gc.ca

Max Kinateder
Max.Kinateder@nrc-cnrc.gc.ca

Bruno Emond
Bruno.Emond@nrc-cnrc.gc.ca

Natalia Cooper
Natalia.Cooper@nrc-cnrc.gc.ca

Ishika Obegadoo
iobee055@uottawa.ca

Jean-François Lapointe
Jean-Francois.Lapointe@nrc-cnrc.gc.ca

¹ National Research Council Canada, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada

² School of Psychology, University of Ottawa, Ottawa, Canada

³ Department of Cognitive Science, Carleton University, Ottawa, Canada