The Effect of Virtual Environment Training on Participant Competence and Learning in Offshore Emergency Egress Scenarios

by

© Jennifer Smith

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

May 2015

St. John's Newfoundland and Labrador

Abstract

Within the offshore and maritime industries, simulation training can enhance conventional safety training by providing orientation training before workers have been deployed offshore and by improving the workers' level of preparedness for emergency situations. As a precursor to simulation transfer studies, this research aimed to determine the level of competence in basic offshore safety gained through a virtual environment training program, investigate the training time required to achieve competence, and develop a strategy to assess human performance in simulated offshore emergency situations.

The experiment demonstrated that offshore egress learning objectives can be taught using the virtual environment training program with some limitations. Basic offshore safety competence was not demonstrated for all learning objectives. Time to competence for some learning objectives was achieved within the study but for other learning objectives the time to competence is still unknown. Due to individual differences in spatial and procedural learning, some individuals required more exposure to the virtual setting to ensure knowledge retention.

Modifications to the training and technology design are recommended in order to prepare for future transfer studies and offshore applications.

Key terms: offshore emergency egress, safety training, virtual environments, spatial learning, procedural learning, performance assessment

Acknowledgements

Many thanks to my supervisors, Dr. Brian Veitch and Dr. Scott MacKinnon. It has been a privilege to work with them throughout this project. They have been a constant source of direction and encouragement. I'm very grateful to them for providing me with this research opportunity and for supporting my decision to change career paths to better suit my research interests. I have learned a great deal from their guidance and without them this project would not have been realized.

I would like to acknowledge with thanks the financial support provided by the Atlantic Canada Opportunities Agency (ACOA), Research and Development Corporation of Newfoundland and Labrador (RDC), and Virtual Marine Technology (VMT).

Thanks to Mashrura Musharraf, my co-investigator, for her contribution and tremendous support during the project. Mashrura was an excellent collaborator, her positive spirit and dedication contributed to an excellent work environment and her input was invaluable.

I would like to acknowledge with thanks the Virtual Environments for Knowledge Mobilization Project (VEKM) Team, especially the Software Team for all their hard work and technical support in developing the AVERT technology. Special thanks to Andrew House, Patrick Linehan, David Bradbury-Squires, and Andrew Caines.

Thanks to the participants who volunteered for this study. Without their interest and time this work would not have been possible.

Thanks to all my family and friends for their support over the years, without their encouragement I would not have pursued this opportunity. Special thanks to my parents, Bill and Joan, and brothers, Jonathan and Timothy, for their support and patience. Special thanks to my friend Krista Hale, for working many late evenings with me. Finally, I owe so much to Geoffrey Hoffe, for his endless encouragement, unconditional support and his motivational words when needed most.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AVERT	All-hands Virtual Emergency Response Trainer
BST	Basic Survival Training
CAPP	Canadian Association of Petroleum Producers
DOE	Design of Experiments
ECG	Electricardiography
EER	Escape, Evacuation, Rescue
GSR	Galvanic Skin Response
HR	Heart Rate
HRA	Human Reliability Assessment
HMD	Head Mounted Display
IMO	International Maritime Organization
ITQ	Immersive Tendency Questionnaire
LMS	Learning Management System
MAC	Manual Alarm Call Point
ODSE	Open-source Distributed Simulation Engine
OIM	Offshore Installation Manager
PA	Public Address
PQ	Witmer & Singer Presence Questionnaire
PSF	Performance Shaping Factor
ST	(Peripheral) Skin Temperature
RR	Respiration Rate
SME	Subject Matter Expert
SSQ	Simulator Sickness Questionnaire
STCW	Standards of Training, Certification and Watchkeeping
TSR	Temporary Safe Refuge
VE	Virtual Environment
VGEQ	Video Game Experience Questionnaire
VMT	Virtual Marine Technology

List of Appendices

Chapter 1 : Introduction

1.1: Relevance of Work

The offshore oil and gas industry is regarded as a high-risk and safety-critical work environment. Remote marine locations are often the site of essential oil and gas operations such as exploration, production, and transportation. Installations situated hundreds of kilometers off the coast present challenges to these critical operations (Sneddon, Mearns and Flin, 2013), and can result in a variety of hazards to the personnel working onboard. The hazards range from reportable incidents and near misses to major accidents such as loss of stability or position, collisions with marine traffic, well blowouts, process gas leaks, systems fires, and explosions (Norsok, 2001; Skogdalen, Khorsandi, and Vinnem, 2012 citing IADC, 2009). These hazardous conditions can be triggered by the culmination of factors including extreme weather, equipment failure, organizational influences and human error (Reason, 1990). No matter the root cause, offshore accidents are dangerous situations requiring comprehensive emergency preparedness and response procedures. Case studies of offshore accidents, like that of the *Deepwater Horizon Macondo Blowout* in 2010 strongly indicate that the escape, evacuation, and rescue (EER) operations are imperative in safeguarding personnel from the dangers of offshore emergency situations (Skogdalen et al., 2012). In emergency situations, as part of the EER safety protocol the emergency response teams are responsible for controlling the damage caused by the hazardous conditions, accounting for all personnel onboard, and coordinating the evacuation or abandonment of the platform. Flin, Slaven, and Stewart (1996) highlight that these offshore emergency response situations are "characterized by time-pressure, uncertainty and danger". As a result, offshore accidents present many challenges for the response teams in managing the emergency.

Offshore operations are only made practical because of the industry's high regard for safety and its focus on emergency preparedness and response management to mitigate the consequences of accidents. To properly prevent, mitigate, and respond to the risks, the Canadian offshore oil and gas industry is heavily regulated at the international level (e.g. International Maritime Organization's Convention on Standards of Training, Certification and Watchkeeping), the national level (e.g. Transport Canada), and industry level (e.g. Canadian Association of Petroleum Producers (CAPP)). These standards and regulations dictate emergency preparedness and response procedures and minimum competency requirements for personnel working onboard.

1.1.1: Conventional Emergency Preparedness and Response Training:

Offshore management relies on the escape, evacuation, and rescue (EER) safety protocol to respond to the dynamic and unpredictable conditions of emergency situations. EER protocol requires many operations personnel to switch to emergency response team duties in order to respond and diffuse emergency situations. For this reason, personnel onboard offshore petroleum installations are expected to have proficiency in both the technical and safety protocols and procedures of working offshore.

Due to the limited resources on board, everyone plays a role in managing the situation. Experienced personnel are trained to assume specific emergency duties as part of the emergency response team and are responsible for managing the hazardous situation offshore while planning and carrying out the safe abandonment of the installation. For general personnel (which includes new employees, short-term contractors, and visitors) the main duties in an emergency are to follow general emergency protocol as outlined on the installation's station bill. These include recognizing alarms, knowing how to escape by following egress routes to an area of temporary safe refuge, and registering at the designated muster points.

The main duties of general personnel are taught during installation-specific orientation training called offshore safety induction. Once new personnel have completed the mandatory basic survival training certifications (Basic Survival Training, Helicopter Underwater Escape Training, Offshore Survival Introduction, Hydrogen Sulphide, etc.) they are required to participate in orientation training (CAPP, 2013). Offshore safety induction is important in preparing personnel for the urgency of emergency egress and severity of emergency situations, particularly in relation to time constraints, situation uncertainty and the overall complexity of offshore emergencies. Typical approaches to offshore platform orientation training include: 1) offshore orientation videos and supervised orientation periods on initial shifts at an offshore installation, and 2) regularly planned muster drills to practice emergency egress and response procedures.

Many offshore oil and gas companies in Canada establish internal policies requiring new employees to complete a mandatory installation safety induction that is conducted upon their first arrival offshore. New-to-vessel personnel can be at greater risk on the first few weeks on the job offshore due to their lack of knowledge of the environment and the installation-specific safety protocols. Installation safety induction covers familiarization with offshore hazards, important safety equipment and procedures for all personnel to follow in the event of emergency (CAPP, 2013). This on the job orientation program is known as “escorted orientation” training and involves a buddy system for pairing experienced personnel with new employees for an assigned duration while onboard. During the escorted orientation period the new employees are closely supervised to help establish spatial and procedural knowledge of the work environment and ensure they are familiar with the emergency protocols.

While stationed onboard, regular emergency exercises and drills are used to establish routines in new personnel and to practice essential emergency response protocols. Emergency drills are performed frequently to assess whether emergency preparedness can be practically demonstrated by personnel. According to CAPP standard practice, offshore personnel are required to perform weekly muster and fire drills, monthly man-overboard and first aid drills and quarterly abandonment drills (CAPP, 2013). Although emergency drills are often performed in optimal conditions, fatal accidents have occurred in abandonment drills involving lifeboat launching in benign conditions (Veitch, Billard and Patterson, 2009). For this reason, safety concerns limit the extent in which drills can replicate real emergency situations and whether they should be performed onboard at all.

Installation specific escorted orientation training and emergency drill activities are considered conventional training methods for preparing personnel for emergency situations. Veitch, Billard and Patterson (2008) explain that conventional training methods are constrained by ethical, logistical, and financial concerns in providing training that prepares personnel for the difficult and safety-critical emergency response situations. Therefore conventional training programs may be enhanced with simulation-augmented training. The following are some limitations of conventional training programs where simulation training can be helpful:

- *Training time commitments and reduced accessibility to experienced resources* – Escorted orientation training is a considerable time and resource commitment and as a result is a costly form of training for the industry. This form of orientation training requires experienced personnel to provide several weeks of closely supervised buddy system training. In addition, general personnel, such as short-term contractors, are often exposed to multiple platforms which can result in a loss of orientation on each vessel.
- *Cost and frequency of recurrent training* – Installation safety induction is mandatory for all new employees upon arrival and after a six month absence (CAPP, 2013). Although refresher training is essential, it requires significant experienced resources to provide recurring training which can be costly.
- *Drills can be unsafe and do not represent real emergencies* – Performing emergency exercises and drills onboard in optimal conditions may not be the

safest means and may not sufficiently prepare personnel for the real situation due to lack of correspondence to real emergency situations.

1.1.2: Simulation and Virtual Environment Training:

For many industries (aviation, medical, industrial, offshore and marine shipping) simulation training has been proven effective in providing a safe and effective means to practice safety-critical operations in preparation for real-world applications. According to Manovani and Castelnuovo (2003), virtual training can provide learning opportunities by offering both numerous first-person experiences and an environment to experience failure safely. Bradbury-Squires (2013) further suggests that simulation training complements conventional training by allowing trainees to gain artificial experience in managing stress-inducing situations.

Simulation has been used in offshore safety training to enhance conventional training including lifeboat launching (Veitch et al., 2008); lifeboat navigation in ice conditions (MacDonald et al., 2011); and platform hazard and risk assessment (Lang et al., 2007). Simulation could also be used to enhance escorted orientation training by improving the level of preparedness for emergency situations and orientation training before personnel go offshore. Veitch et al. (2008) described a forward look at offshore simulation to include larger scale, simulated muster drills, and Escape, Evacuation and Rescue (EER) exercises. The All-hands Virtual Emergency Response Trainer (AVERT) software was developed by Memorial University to realize the exploitation of virtual environments in EER training and to target all-hands emergency preparedness and response exercises in a simulated environment.

AVERT is a first-person virtual environment (VE) intended to train personnel within a naturalistic representation of their offshore work settings (House et al., 2014). The AVERT software was designed for training workers in general safety practices, onboard familiarization, and emergency preparedness. The AVERT software is coupled with a learning management system (LMS) to deliver guided training programs and administer dynamic test scenarios (House et al., 2014). When a trainee completes a scenario (practically demonstrating his/her competence), performance metrics are recorded to a report file in AVERT. These combinations of functionality allow for the creation of scenarios to target specific learning objectives, as well as the in-depth analysis of performance beyond simple success or failure.

Adding AVERT training to conventional training may reduce the time required for the escorted orientation training by providing spatial knowledge of the vessel in advance of going offshore. Allowing personnel to familiarize themselves with their offshore work environment before being deployed offshore could assist in reducing loss of orientation for personnel required to travel to multiple installations and also support in administering refresher training. To assist with emergency exercises and practice drills, virtual environments can also provide a means to safely practice egress procedures through a series of offshore emergency scenarios. In particular, AVERT can provide credible emergency scenarios such as platform blackouts and fire or explosion situations for personnel to safely practice tasks that could not otherwise be performed safely in drills and provide artificial experience in emergency exercises.

In addition to providing credible situations and safe training opportunities that could otherwise not be performed, Salas et al. (2009) suggested that simulation training can “provide systematic and structured learning experiences”. This enables virtual environments to act as virtual laboratory settings for researchers to measure practically demonstrated competencies. Virtual environments thus are not only to practice navigating in emergency situations but also laboratories for instructors to measure the trainee’s performance. The virtual environment setting allows researchers to look at a situation or incident from the “inside out” when assessing accidents (Dekker, 2005). Virtual environments like AVERT can be used to enhance conventional training and as a tool to assess of performance. For the latter, it is important to demonstrate the efficacy of AVERT to provide confidence in its utility. In order to assure basic offshore egress competence in both site familiarization and emergency offshore egress can be achieved through AVERT simulation training, the virtual environment technology requires validation testing.

1.1.3: Competence in Basic Offshore Safety using AVERT

Introducing a new virtual environment technology for offshore training requires the technology undergo validation testing to demonstrate its utility as a means to assure trainee competence. Past research using AVERT focused on user presence and spatial learning in a virtual environment of an offshore oil installation (Bradbury-Squires, 2013).

The present research builds on Bradbury-Squires' (2013) study and goes beyond user presence and spatial learning to consider teaching and assessing more learning objectives related to emergency egress.

This research studied the effect of the AVERT virtual environment training system on participant competency (task performance) and learning in emergency response during offshore evacuation scenarios. The primary research questions of the study aimed to answer:

- Can people learn basic safety tasks using the AVERT platform? Which learning objectives are best taught using the AVERT platform?
- What competencies can this technology address? Can basic safety competence in spatial awareness, alarm recognition and hazard identification be demonstrated using virtual environment technology?
- How much training time is required to reach a competency in basic safety using AVERT? How does this training time translate into the ramp-up time required to deploy this training prior to going offshore?
- What attributes of the virtual environment and learning management system are impactful on the training experience and essential to activate learning in basic offshore safety training?

In addition, the research aimed to establish a baseline of task performance for novice individuals in virtual egress training scenarios. The secondary research questions of the study aimed to investigate:

- How novices learn and behave in virtual emergency response scenarios?

- How effective is the training program at equipping participants to demonstrate competence in basic elements of offshore egress in AVERT?

Overall, these data should inform future virtual environment studies on spatial and procedural knowledge transfer to real offshore environments. This research also included collecting data on the effects of varying performance shaping factors, training, visibility, and complexity, in the virtual scenarios upon physiological responses and cognitive human performance of individuals during simulated offshore emergency situations. These data were not included in the formal reporting of this work but will be included in Appendix A and B.

1.2 Statement of Claim

The objectives of this work are to:

1. investigate the training time required to reach a competency in basic safety in offshore emergency egress using AVERT.
2. develop a strategy to assess performance in simulated offshore emergency situations using a virtual environment.
3. establish a baseline in task performance for novice individuals in virtual egress training and simulated offshore emergency situations.
4. measure learning through a virtual environment and determine how attributes of virtual environments contribute to learning.

1.3 Hypotheses

The following null hypotheses were tested:

1. H1: Participants will reach basic competency in offshore emergency egress using AVERT within the allocated time of the study (after completing the three sessions).
2. H2: Participants who spent more time training (repeat exposure to training group) will perform better from a task performance perspective in the emergency response testing scenarios than participants who spent less time training (the single exposure to training group).
3. H3: As participants from both groups advance through the training and testing sessions, their task performance in all learning objectives will improve.

Chapter 2 : Literature Review

There are many factors that influence the training effectiveness of virtual environment technology, from the technical aspects to the training goals, and the performance assessment requirements. This literature review will discuss the following main factors affecting the overall skill acquisition and knowledge transfer using virtual environments:

1. Virtual environment fidelity, trainee engagement and a sense of presence
2. Learning processes and concepts driving training
3. Comprehensive human performance assessments

The first section will provide a brief explanation of how virtual environment fidelity, trainee engagement and achieving presence are important for facilitating learning in a virtual environment. The second section will describe the main learning processes and training concepts suitable for virtual environment training as they relate to the offshore training application. The third section will review the methods used to measure human performance through virtual environment technology.

2.1: Virtual Environment Fidelity, Trainee Engagement and a Sense of Presence

Effective virtual environment and simulation training tools are those that are able to support knowledge acquisition within the specific training context and ultimately provide skill transfer to the real-world training application. Researchers suggest that repeated experiences in simulator training that provides similar operational conditions as the real-world work settings can help prepare people for the associated stressful working conditions (Tichon and Wallis, 2010). Thus, virtual environment training technologies are designed to recreate real-world conditions to assist in training correct procedures. In contrast, some researchers have found that the process of using virtual environments to replicate the real-world have exposed deficiencies in virtual environments and decreased the overall training quality (Witmer et al., 2002 citing Witmer, Bailey and Knerr, 1995; Witmer, et al., 1996). These deficiencies include: virtual environments cannot support navigational tasks as some trainees became lost or disoriented (Sebok et al., 2004; Witmer et al., 2002); practice is not equal to training in both real and virtual training settings (Klein, 1997); virtual environments induced simulator sickness in some trainees (Kennedy et al., 1993; Stanney et al., 1999; Witmer et al., 2002); and for short-term exposure, virtual environments do not provide any advantage over conventional maps for spatial knowledge transfer (Darken and Peterson, 2001; Waller et al., 1998). It is important to understand and address the factors that influence virtual environment training in order to maximize the utility of the technology for both skill acquisition and knowledge transfer.

In order for virtual environments to be considered an effective training tool, there are several features that must be considered. From a pedagogical point of view, virtual environments provide a unique set of technical and ‘learner experience’ specific characteristics (Dalgarno and Lee, 2010). The technical characteristics of a virtual environment include the visual representation, situational context, and interface fidelity. The ‘learner experience’ characteristics include the level of engagement or interaction from the trainee, and the trainee’s motivation to learn from the environment. The virtual environment fidelity and the situational context provided by credible scenarios in the virtual environment influence the overall level of immersion achieved in a virtual setting. Dalgarno and Lee (2010) best explain immersion as being related to the simulation technology as the measurable properties of the system or virtual environment that lead to a sense of presence (where presence is more related to the state of mind of the participant). Presence is when a trainee in a virtual setting “behaves like they are in a real life situation even though cognitively they know they are not” (Lee et al., 2010). The virtual environment fidelity and trainee engagement are both required in order to evoke a sense of presence from the trainee. Presence has been theorized as a mediator to effective learning and training transfer from a virtual environment (MacKinnon et al., 2010; Mantovani and Castelnuovo, 2003; Scoresby and Shelton, 2010; Slater, 2009; Witmer et al., 2002). Each of the following features will be reviewed in the literature to understand the influence they have on what makes good simulation training: 1) virtual environment fidelity, 2) trainee motivation and engagement, and 3) achieving a sense of presence.

2.1.1: Fidelity of Virtual Environment Technology

Virtual environment technology consists of the visual representation of a life-like or real-world environment and the interface to interact with the environment virtually. Simulation fidelity refers to how closely a simulation represents the real-world environment from both visual representation and environment interaction perspectives. Bradbury-Squires (2013), citing Gallagher et al. (2005), explains that simulation training involves using a virtual setting that resembles the real work environment to assist in the development of cognitive and psychomotor skills. The idea is that virtual environments that are similar in visual, auditory, and haptic interaction support situated or experiential learning and provide the trainee with a safe environment to learn-by-doing (Huang et al., 2010).

The fidelity of simulation technology is really a technical measure of how sound a virtual environment is at reproducing the real-world (from visual representation, contextual and interactive perspectives). Simulation fidelity includes how well the virtual environment represents the real-world environment and how well the interface mimics interacting with the real-world. In the case where the interface does not mimic real-world interaction, the simulation fidelity includes the ease of operation of the interface.

> Dalgarno and Lee (2010) describe the elements that contribute to simulator fidelity as:
>
> Representational Fidelity: realistic display of the environment, smooth view changes and realistic object motion through the virtual environment, consistency of object behaviour, user representation, spatial audio representation, and kinaesthetic and tactile force feedback.
>
> Learner Interaction: embodies actions, embodies verbal and non-verbal communications, control of environment attributes and behaviours, and construction/scripting of objects and behaviours. (p. 15)

Although simulation fidelity seems to be an obvious requirement for providing training, without careful development and consideration for these elements the utility of the simulation can be lost. These visual and interactive elements are an important part of simulation training because if they are not properly represented it would be distracting for the trainee and affect the trainee's sense of engagement and immersion.

There are different levels of fidelity in simulation technology and they usually range based on the type of technology used. Virtual environment and simulation technology have a variety of forms depending on the level of immersion required from the simulation. Simulation technology range from desk-top based virtual environments to head mounted display (HMD) environments and full-mission surround visual systems. Desktop based virtual environments have been described as the least immersive simulation technology, while surround visual or full-mission simulators are more representative of the real-world application in visual representation and interactive interface. As a result, HMD's are considered more immersive. The level of immersion required of a virtual environment is dependent on the intended training application and corresponding learning objectives. For example, Waller et al., (1998) have shown that neither environment fidelity nor interface fidelity had much effect on the acquisition of spatial route knowledge.

The Tichon and Wallis (2010) study used two different simulator types (an enclosed replica cab system and a visual and force feedback control system) to determine the impact of fidelity and immersion on knowledge transfer and decision-making under stressful situations. This study found that high fidelity is not essential in order to provide transferable skills related to stress training and proposed that smaller simulations may reduce distraction and aid in trainee concentration (Tichon and Wallis, 2010). This suggests that full-mission immersive simulators are not always necessary for spatial acquisition and that virtual environment desktop simulations are equally capable at providing spatial knowledge essential for real-world transfer. Mantovani and Castelnuovo (2003) suggest that virtual environment technology should provide a learning environment that resembles the real-world enough to be relevant and engaging so that the trainee develops artificial experience. They also suggest a safe environment should be provided so that mistakes can be made and participants can reflect on the experience and be ready to seek guidance (Mantovani and Castelnuovo, 2003).

2.1.2: Trainee Engagement

High fidelity (visual, contextual, and interface) simulations alone will not be successful at developing skill transfer. Similarly, virtual environments that are designed to provide learning opportunities and means to assess competence will also miss the mark at developing skill transfer if the simulation itself or the training goals are not taken seriously. Elements that detract from the simulation training utility by distracting the trainees from the learning objectives or causing them to not take the simulation seriously include: a lack of realism visually and contextually (Dalgarno and Lee, 2010; Huang et al., 2010; Tichon, 2007, Waller et al., 1998); cumbersome or counter-intuitive functionality (Huang et al., 2010); latency or the delay between user actions to the displayed response, (Dalgarno and Lee, 2010; Meehan et al., 2003); simulations that induce simulator sickness (Waller et al., 1998; Witmer et al., 2002); poorly designed practice scenarios (without consideration of learning objectives) (Klein, 1997; Tichon, 2007); and ill-timed or irrelevant training feedback (De Freitas, 2009; Grantcharov and Reznick, 2008; Michel et al., 2009).

Overall, these technical and training distractors negatively influence the user experience (or learner experience as outlined by Dalgarno and Lee, 2010). If any of the elements that detract from the simulation occur, the trainee can become disinterested and the learning lessons can be lost. As a result, the trainee's behaviour and task performances being measured in the virtual environment are also inaccurate. Even if these distractors are addressed, there is still a chance for learning lessons to be lost due to limitations in the trainee's attention span. If the trainee is not focused on the task, it is unlikely that learning will occur effectively. There is a certain level of engagement required of the trainee for learning to take place in a virtual setting.

Trainee motivation and engagement is essential for the success of any virtual environment training program. There has been significant research on how to improve trainee motivation (Huang et al., 2010; Lee et al., 2010). Huang et al., (2010) describe three main characteristics that motivate learners in virtual environment settings: imagination, immersion and features of interaction. Visual and contextual realism can help with learner motivation and engagement by supporting the learner's imagination and immersion. From a visual and contextual realism perspective, research has addressed how real is real enough to provide training value (Dalgarno and Lee, 2010; Waller et al., 1998). Mantovani and Castelnuovo (2003) suggest that reality and plausibility judgment of the virtual environment are more important than reconstructing a real-world environment with a high level of visual fidelity. Reality and plausibility judgement involves replicating the situational context in the virtual environment so that the trainee can relate to the situation or believe that it is realistic (Mantovani and Castelnuovo, 2003). Lee et al. (2010) suggest that there is a correlation between academic achievement and students that are intrinsically motivated (without reinforcement) by the environments and activities. If virtual environments can sufficiently represent the real-world training application in both visual and contextual aspects, this would help in promoting student motivation and as a result learning outcomes.

2.1.3: Creating a Sense of Presence

The concept of presence has been frequently used to inform simulation training transfer. Researchers have studied the correlation between perceived user presence and task performance as well as training transfer (MacKinnon et al., 2012; Scoresby and Shelton, 2010; Slater, 2009; Mantovani and Castelnuovo, 2003; Insko, 2003; Witmer et al., 2002). Presence is primarily reported by subjective measures through presence questionnaires that elicit the user's perspective on how present they were in the virtual environment (Slater, 1999; Witmer & Singer, 1998; Witmer & Singer, 2005). Researchers have also used two other forms of measuring presence to support subjective measures: behavioural and physiological (Insko, 2003). Behavioural measurement techniques involve observing the trainee performance in the virtual environment and aim to assess how people react and manage situations presented in the virtual environment. Physiological measures involve recording the trainee's responses (heart rate, skin temperature, galvanic skin response and respiration rate) to the stimuli in the virtual environment and compare them to the trainee's self-reported presence questionnaires (Meehan, 2002). All three means of measuring presence are focused on gaining an understanding of how immersed the participant was.

There has been discussion in the literature over the concept of presence and whether there is a correlation between subjectively reported presence and training transfer (David Bradbury-Squires, 2013). In general, presence has been found to be more related to user acceptance or engagement than learning or training transfer. Therefore, presence at most is indirectly related to learning and training transfer through the influence of trainee engagement and acceptance of the training medium. For this reason, the relationship of presence, learning and training transfer falls beyond the scope of the current research.

2.2: Learning Processes and Concepts Driving Training

This section is focused on the elements necessary to support learning through a virtual environment. The first element to support learning is to understand the main learning processes necessary for effective virtual environment training. The second element is to understand the training concepts as they relate to the offshore training. The final element is to align the training goals with the transfer of knowledge and skill to the real-world applications.

Considerable effort was spent defining the learning objectives relevant to the AVERT offshore training application within the capabilities of the virtual environment technology. It was determined that since the AVERT offshore training application is targeting onboard familiarization and emergency egress preparedness training, both spatial and procedural knowledge will be required to provide relevant knowledge transfer. This section will discuss in detail: 1) how learning styles (active and passive) should be employed to shape the development of virtual environment training programs and 2) how to develop spatial and procedural knowledge through virtual environment training.

2.2.1: Learning Styles - Active and Passive

There are two prevalent forms of teaching styles represented in the literature that are relevant to virtual environment course design: active and passive teaching styles. Active learning, as defined by Michel et al. (2009), encompasses several instructional models to support a trainee self-driven learning process in which the trainee is responsible for his/her own learning. Active learning instructional models include: experiential, problem-based, participatory, and cooperative. Passive learning is defined as an instructor-driven learning process that is in line with the traditional lecture-style model of instruction (Wingfield, 2005).

It is important to highlight the use of the term "active learning" in the literature as there is some confusion in relation to the definition of active learning. Most of the literature regarding active and passive learning in virtual environments (particularly related to spatial learning) use the broad term "active" learning when referring to experiential learning or learning by experiencing or exploring. Experiential learning is a process in which students learn from relevant experiences and are able to create knowledge by reflecting on the experiences (Michel et al., 2009 and Wingfield, 2005 citing Kolb 1984; Zapalska et al., 2012). The relevant experiences are often provided from carefully designed student-centered course instruction. Virtual environment technology is designed to provide artificial experiential opportunities through relevant practice scenarios. In most cases, research describing active learning virtual environments are actually referring to experiential learning, and not the whole suite of instructional models that accompany active learning (problem-based, participatory, and cooperative). In this literature review, the term “active learning” is used when referring to the experiential learning instructional model.

Passive Learning Applications:

Passive learning strategies are instructor-driven (Wingfield, 2005), which allows the instructor to lead the learning process and dictate the direction of the lesson. Following traditional lecture style approaches to learning, they are usually conducted by set course curriculum and lectures, video instruction (Michel et al., 2009), and more recently page-turner e-learning venues (Clark and Mayer, 2008). Passive learning strategies are helpful in establishing foundational information in the form of factual and conceptual knowledge.

Active Learning Applications:

Active learning strategies are trainee-centered (Wingfield, 2005), which allows the trainee to have a more hands on approach to his/her own knowledge acquisition.

According to Wingfield (2005) and Michel et al. (2009), both citing Kolb (1984), the definition of active-experiential learning involves six elements in making an experience a learning opportunity: 1) learning is a process, not an outcome; 2) derives from experience; 3) requires the individual to resolve dialectically opposed demands; 4) holistic and integrative; 5) requires interplay between the person and the environment; and 6) results in knowledge creation. (p.119 & p.399)

Active learning strategies can be especially useful in spatial knowledge acquisition as this process involves interacting with the environment and reflecting on the experience in order to develop the necessary route and survey knowledge for environment navigation and wayfinding. In Bradbury-Squires' (2013) review of active and passive spatial learning, there was some evidence that active exploration resulted in "participants constructing a more accurate internal representation of the environment" and "more efficient strategies for wayfinding". Thus, active learning may be better suited for the development of survey representation of spatial knowledge (Bradbury-Squires, 2013). Active learning styles may be more suitable for procedural knowledge development because part of learning a procedure and storing it to memory requires the trainee to perform the task and gain experience in the task.

Active versus Passive Learning:

There has been a lot of research into comparing the utility of active learning versus passive learning styles in virtual environments (Bradbury-Squires, 2013; Carassa et al., 2002; Chrastil and Warren, 2011; Clark and Mayer, 2008; Gaunet et al., 2001;

Melanson et al., 2002; Peruch and Wilson, 2004; and Michel et al. (2009) citing Stewart, Wingfield and Black, 2005).

The results of most studies were unable to quantitatively show that active learning can surpass the learning outcomes of passive learning strategies (or prove that active learning is better at knowledge transfer than passive learning) (Michel et al., 2009; Wingfield, 2005). Michel et al. (2009) compared 15 active and passive learning studies (8 qualitative and 7 quantitative). The qualitative research found active learning was superior to passive learning in regards to learning outcomes, but that quantitative research found no difference between active and passive learning in regards to learning outcomes. In the qualitative studies, the active learning has shown improved learning compared to passive learning in terms of qualitative results including student satisfaction or attitudinal responses to active teaching styles (Michel et al., 2009). Bradbury-Squires (2013) also reviewed six research studies that compared active and passive learning strategies for spatial knowledge acquisition and found that active and passive strategies usually resulted in equal task performance, in some cases active strategies provided superior performance, but there were no instances where passive strategies were better than active strategies. Clark and Mayers (2008) presented evidence that passive strategies are equally as valuable as active learning strategies.

Overall, the research on active and passive learning styles is inconclusive on which learning strategy is better in terms of learning outcomes and task performance. Of the studies that found evidence to suggest that the active strategy benefited over passive strategies, this advantage was sometimes due to the level of engagement of the participant (Michel, 2009). Rotgans and Schmit (2011) support this finding and suggest that active learning strategies encourage increased student satisfaction, which may also stimulate learning. If this in fact the case, developers should encourage the development of student satisfaction through employing active learning strategies in virtual environments. Clark and Mayers (2008) suggest that the learning utility of each strategy (active and passive) is dependent on the application or learning concept being taught. To develop an effective virtual environment training program, it is important to use the right learning strategy for the intended learning application. In some cases it is best to employ both active and passive strategies.

Combination of Active and Passive Learning Styles

Combining active and passive learning styles for virtual environment training programs can be useful in providing basic lessons using passive outlets and allowing for experiential learning and reflecting in active exercises. Clark and Mayer (2008) suggest a combination of examples and practice exercises is more useful in knowledge acquisition than just providing all practice scenarios. A combination of passive presentation of information and active application of information is beneficial for knowledge transfer. Clark and Mayer (2008) suggest that “when the learner is viewing an example, working memory is free to build a new mental model of the skill and then the learner can try out the new mental model in practice problem”.

2.2.2: Spatial Knowledge

Offshore Orientation Requires Spatial Knowledge

The AVERT virtual environment's utility to enhance conventional onboard familiarization training involves facilitating the development of spatial knowledge. Spatial knowledge acquisition of a new environment through virtual environments is one of the advantages of the technology (Farrell et al., 2003; Sebok et al., 2004). This is particularly important for helping personnel in developing a comprehensive spatial understanding of his/her virtual offshore work environment before going offshore.

The literature in this field suggests that virtual environments can be effective in providing spatial knowledge (Farrell et al., 2003 citing Waller et al., 1998; Lathrop and Kaiser, 2005; Witmer et al., 2002) with some limitations. In fact, the research community has shifted focus from whether or not virtual environments can provide spatial knowledge, to investigating how to optimize the spatial acquisition, and how to improve the transfer to real-world applications. Researchers are interested in answering questions looking at how to speed up development of spatial knowledge using virtual environments, and what variables mediate the transfer of spatial learning in virtual environment to real-world applications (Farrell et al., 2003 citing Waller et al. 1998; Ruddle et al., 1998).

What is Spatial Knowledge?

Spatial knowledge is necessary for wayfinding in any environment, virtual or real-world. Darken and Peterson (2001) defines spatial knowledge as "the development and use of a cognitive map" of an environment through a combination of wayfinding (cognitive) and motion (locomotion) in a virtual environment. Motion in this case refers to passive transportation within the virtual environment. Wayfinding relates to the strategic components of movement and the development of a cognitive or mental map to represent the environment (Darken and Peterson, 2001). The wayfinding component of navigation depends on the establishment of a spatial mental model of the environment. According to Sebok et al. (2004) one of the major contributing factors of a virtual environment's utility is in its ability to support navigation and enable the trainee to "determine their present location, plan their route and traverse the route". Therefore spatial comprehension involves "perceiving, understanding, remembering and recalling spatial information" including landmark, route and survey knowledge for future navigational use (Darken and Peterson, 2001).

There is some conflict in the literature over whether or not the cognitive map theory is a true representation of spatial learning. Weisberg et al. (2014) highlight the critiques of the cognitive map (Foo et al., 2005; Shettleworth, 2009) and suggests that the cognitive map exists but that there are noticeable individual differences as to how the cognitive map is constructed (Weisberg et al., 2014 citing Ishikawa & Montello, 2006 and Schinazi et al., 2013).

Acquisition and Representation of Spatial Knowledge

The development of spatial knowledge is often characterized by how the knowledge is collected, through different perspectives and over-time. A well-known model for explaining how spatial knowledge is gained from different spatial perspectives is the Land-Route-Survey model (Seigel and White, 1975). The Land-Route-Survey model has “three stages of development of an individual’s cognitive representation of large-scale navigable spaces” (Waller et al., 1998 citing Seigel and White, 1975). Waller et al. (1998) explain that landmark identification is the primary stage of the mental map and involves identifying important locations in the environment (distinct and disconnected from one another). The second stage, known as route representation, involves connecting the landmarks with routes. Route knowledge (or representation) occurs when a trainee has had more exposure to the environment and allows them to “follow a known path from one location to another” (Chrastil, 2012). The third stage is survey representation and involves the development of a map-like depiction of the environment (Waller et al. 1998). Some researchers suggest that survey knowledge development refers to the process of path integration both within-route understanding and between-route understanding of the environment (Weisberg, 2014). According to Waller et al. (1998) the survey representation completes the spatial model by providing the trainee with “an understanding of the relationship between various landmarks independently of the routes that connect them.” Chrastil (2012) explains that acquiring survey knowledge allows the trainee “to take novel shortcuts and detours between locations, traversing paths that have never been taken before.” The transition from route to survey knowledge is hard to achieve for some people and requires more exposure to the environment to complete the spatial model of the environment (Waller et al. 1998; Weisberg et al., 2014).

Spatial knowledge can be acquired directly from the real environment or through other secondary sources such as video, photos, verbal directions and virtual environments (Darken and Peterson, 2001). Farrell et al. (2003) explain that spatial relationships can be represented as egocentric (local) or allocentric (global). The egocentric spatial representation is localized to the participant and where they move through the environment. Bradbury-Squires (2013) explains that egocentric spatial representation is a reflection of the participant's experience navigating the environment. This spatial framework is involved in route-perspective learning (Brunye et al., 2012; Farrell et al., 2003). The allocentric representation (also known as exocentric in some literature) is more to do with the global location of landmarks from an external perspective, such a top-down maps and this external framework is involved in survey-perspective learning (Brunye et al., 2012; Farrell et al., 2003). The allocentric representation is important in wayfinding of large-scale complex environments as it provides an effective and encompassing overview (Bradbury-Squires, 2013 citing Carassa et al., 2002).

Egocentric spatial knowledge gathered from the environment directly from a first person perspective view is known as route knowledge (Brunye et al., citing Levelt, 1982, Linde and Labov, 1975). Trainees can gain route perspective learning "from salient cues in the environment such as landmarks, routes, nodes, and districts" (Wicken and Holland, 2000; Witmer et al., 2002; and Lathrop and Kaiser, 2005; citing Lynch, 1960). Allocentric spatial knowledge gathered from maps or aerial views of the environment is referred to as survey knowledge. Trainees can gain survey knowledge by consulting a map of the environment or through longer exposure to the environment to develop a map-like mental configuration of the environment.

Bradbury-Squires (2013), citing Wilson et al. (1997), and Seigel and White (1975), explains that the spatial learning process is the progression from egocentric to allocentric frame of reference (also known as route to survey map representation). This explanation may lead one to conclude that spatial acquisition requires an individual to establish land and route knowledge first and then gain survey knowledge. Bradbury-Squires (2013, citing Wilson et al., 1997) suggests that the route to survey knowledge linear progression does not always support spatial knowledge acquisition as survey knowledge sometimes does not develop in this sequence. Route and survey representation involved in spatial learning is not necessarily confined to a linear progression. Brunye et al. (2012) explain that spatial knowledge is acquired by switching between multiple spatial perspectives, most commonly route and survey perspectives (Brunye et al., 2012 citing Levelt, 1982; Linde and Labov, 1975).

As trainees are exposed to an environment over time, route and survey knowledge of that environment is developed. This is the case for real and virtual environment training mediums. However, the representation of spatial knowledge in a person's mental model of an environment depends on how knowledge is collected (Darken and Peterson, 2001). Lathrop (2005) suggests that spatial representation develops "as a function of the amount and type of exposure to the environment". Although researchers have not identified a specific pathway or sequence of learning required to be more effective at developing spatial knowledge, they have identified what characteristics impact the development of spatial knowledge. The different forms of spatial knowledge (land, route, survey) develop depending on how the trainee navigates or explores the new environment, the effort of the trainee to learn the environment and the structure of the environment itself (Chrastil, 2012). In addition, researchers (Darken and Peterson, 2001; Waller et al., 1998; Witmer et al., 2002) have found that participants obtained better survey knowledge using traditional maps than a virtual environment for short exposure time applications. Witmer et al. (2002) explain that the relative effectiveness of maps compared to virtual environments in enabling participants to develop spatial knowledge depends on several factors, including but not limited to the type of spatial tasks, the environment complexity, as well as the exposure time and quality of the virtual environment and maps.

If spatial knowledge representation is impacted by how the knowledge is collected, then different training mediums (real-world, virtual and maps) could result in different spatial knowledge acquisition patterns. According to Waller et al. (1998) it is important to determine whether the spatial representation development process is the same in virtual environments as in real-world situations because the effectiveness of virtual environment training is dependent on the trainee's ability to apply this training to a real-world environment. These researchers (Waller et al., 1998) hypothesised that the training medium (real or virtual) that provided a better mental model would result in better performance in a real-world maze assessment. Waller et al. (1998) investigated two mediators (environment fidelity and interface fidelity) on developing a mental model by comparing six training environments (no training, real-world, map, virtual environment desktop, virtual environment immersive and virtual environment long immersive) and assessing trainee performance in a real-world maze. This research found that the environment and interface fidelity had little effect on development of route knowledge, suggesting that lower fidelity virtual environments could also be used for developing a spatial representation. This study also measured the effectiveness of virtual environment training based on short term and long term virtual environment exposure and found that virtual environments were not as effective as maps for short term exposure, but may have long term exposure advantages (Waller et al., 1998). Similar results were found by Darken & Peterson (2001) and by Witmer et al. (2002). Witmer et al. (2002) found similar results when assessing the influence of cognitive and perceptual variables on the speed and quality of spatial acquisition provided by virtual environments. Darken and Peterson (2001) found similar results when investigating the exposure time in a virtual environment required to gain spatial knowledge. There seems to be a benefit in virtual environment technology in terms of effectiveness for long term exposure.

Virtual Environment Exposure Time Required to Develop Spatial Knowledge?

Bradbury-Squires (2013) citing Peruch et al. (1995) suggests that increased frequency and duration of exploration experiences with an environment improves spatial knowledge development of the environment. Waller et al. (1998) found that for short term exposures, virtual environment training may be no more effective than training in a map (and immersive may be no more effective than desktop training). If enough time is spent in virtual environment training, then virtual environment training may surpass map training for developing spatial knowledge and be indistinguishable from training in the real world (Darken and Peterson, 2001; Waller et al., 1998). The duration participants were exposed to map and virtual environment training in Waller's study was between 1-2 minutes for short term exposure and 5 minutes for a long term exposure comparison. Darken et al. (2001) compared Waller's study to three other studies (Darken et al., 2001; Koh et al., 2000; Waller et al. 1998; Witmer et al., 2002) with regards to training transfer and the total exposure time to virtual environments. Each study targeted different levels of participant experience, route or survey type tasks and different variations of navigational aids. The studies also ranged in virtual environment exposure time from 2 minutes to 60 minutes. Darken et al. (2001) found that participants rarely had enough time to acquire meaningful survey knowledge. In all studies it was determined that with "sufficient" or long term exposure to the virtual environment it was possible to eventually surpass map survey knowledge acquisition (Darken and Peterson, 2001).

Bradbury-Squires' (2013) study used three thirty-minute sessions for a total of 40 to 90 minutes of exposure (approximately 40 minutes for the structured route replication group and approximately 90 minutes for the unstructured active exploration group), which is considerably longer than the other studies reviewed. Participants in that study were evaluated based on two performance measures for spatial knowledge acquisition: 1) their demonstration of wayfinding and 2) their time to complete the scenario in the virtual environment. Bradbury-Squires' study did not use spatial performance measures specifically for survey knowledge acquisition, so it is unclear if more or less exposure time is necessary to complete the survey knowledge representation required for the large-scale environment.

The gap that sometimes occurs in acquiring survey knowledge through virtual environments may not only be attributed to exposure time or the order of route or survey knowledge acquisition, but also in part due to individual learning styles. Weisberg et al. (2014) explain that individuals acquire survey knowledge in particular “at very different rates and require different levels of information.” Brunye et al. (2012) citing Thorndyke and Hayes-Roth (1982), suggest that different learning perspectives (individuals who prefer navigation or route learning versus individuals who prefer map or survey learning) may also impact the development of route and survey knowledge. Bradbury-Squires (citing Waller et al. 1998) explains that individual differences in visualization skills, orientations skills and interface proficiency correlate with spatial learning in a virtual environment. Everyone is different in their ability to acquire spatial learning and the time it takes to acquire it (Weisberg et al, 2014). Waller et al. (1998) citing Tversky (1981) suggests that some people are unable to complete the survey knowledge of real-world environments even after they have been exposed to the environment for several years. Weisberg et al. (2014) found similar results, suggesting that some participants can quickly gather all stages of spatial knowledge while others do not proceed beyond route knowledge and cannot demonstrate a survey knowledge understanding.

Designing Virtual Environments to Maximize Spatial Knowledge Learning?

Although researchers (Waller et al., 1998) have demonstrated that relatively low fidelity virtual environment training allows people to develop useful spatial representation of a large-scale environment, there are some suggestions to consider that can improve spatial knowledge acquisition using a virtual environment. This section will describe some of the more prominent suggestions in the literature that can help enhance spatial learning through virtual environments. These are organized based on representational fidelity and learner interaction for virtual environments:

<u>Representation Fidelity</u>: From a technical perspective, low fidelity simulators are limited to restricted fields of view and simplistic control interfaces. Waller et al. (1998) investigated how restricted field of view and non-intuitive or distracting interfaces impacted the development of survey representation. They found that neither environment fidelity nor interface fidelity had much effect on the acquisition of route knowledge (Waller et al., 1998). This further highlights the issue of developing survey knowledge through virtual environments. As survey knowledge acquisition is impacted by individual differences, there could be many ways to address this deficiency in virtual environments: reduced complexity of the environment, increased exposure time for some trainees, and the addition of navigational aids in the virtual environment.

According to Witmer et al. (2002), the success of some researchers in demonstrating virtual environment developed spatial knowledge was due to the research focusing on simplified building layouts (to successfully demonstrate transfer of survey knowledge from virtual environments to the real-world) (Witmer et al., 2002 citing Wilson et al., 1997). This finding reiterates Darken and Petersons' (2001) conclusion that virtual environments have the potential to help trainees develop survey knowledge if given enough exposure time. It is useful to increase trainee exposure to virtual environments and perhaps even move towards an individualized training approach to spatial learning. To further assist with spatial knowledge acquisition, developers could implement navigational aids within the virtual environment, such as a mini-map representation, directional cues or arrows, personalized landmarks or checkpoints and trails or footprints displayed while navigating the virtual environment (Aldrich, 2009; Darken and Peterson, 2001). The overuse of navigational aids to enhance performance in the virtual environment could hinder knowledge transfer to real-world settings if trainees become dependent on the navigational aids (e.g. map displays in the virtual environment can become a crutch for people who find developing survey knowledge challenging).

<u>Learner Interaction</u>: From a user perspective, motivating the trainee to actively learn the spatial organization of the environment is necessary. Waller et al. (1998 citing Linberg and Garling, 1983) suggest that developing survey representation requires conscious effort. Although survey knowledge acquisition is in part dependent on individual learning styles, the process requires the trainee to actively participate in its development. Virtual environments can maximize spatial knowledge acquisition by improving the learner interaction with the virtual environment. This can be accomplished by supporting an active learning approach to teaching spatial knowledge and by employing a step-by-step

training progression to establish and build spatial knowledge (to help manage the trainee's attentional resources and cognitive demands).

2.2.3: Procedural Knowledge

Offshore Emergency Egress Training Requires Procedural Knowledge

The AVERT virtual environment's utility to improve conventional offshore emergency preparedness, and response exercises or muster drills involves facilitating the development of procedural knowledge in emergency egress. Along with spatial knowledge acquisition, the ability to safely acquire procedural knowledge through virtual environments is another advantage to virtual environment technology. This is particularly important for helping personnel understand and practice complex emergency preparedness exercises and drills that cannot otherwise be safely performed in the real-world.

To this end, well-designed virtual environment training identifies the cognitive and psychomotor learning objectives of the training procedure in order to ensure the training application is properly focused on learning these concepts related to the operational tasks within the virtual setting. The procedural tasks related to offshore safety training for this study are largely cognitive tasks. Due to the limitations of a desktop based virtual environment interface in providing psychomotor learning, most psychomotor related procedural tasks involved in the emergency preparedness and response training were removed from the scope of the experiment. These include the physical operation a manual alarm call point (MAC) and donning an immersion suit. The focus of the procedural knowledge literature review will be on the cognitive aspects of training operational procedures. In the context of offshore training these include recognizing the alarm type and selecting the safest path for the emergency situation.

What is Procedural Knowledge?

Dahlstrom et al. (2009) explain (in the context of the aviation industry) that much of the training in operational industries focuses on the development of content-specific technical skills, or developing an inventory of procedural steps to increase safety. These technical skills or procedural knowledge are usually taught using a learning-by-doing approach. Procedural knowledge is knowing the criteria involved in recognising when to apply the appropriate procedure and the specific step-by-step procedure for the given situation (Bloom's Taxonomy knowledge dimensions as defined by Anderson et al., 2001). In this case, the criteria could include cues, triggers or patterns to recognise the situational conditions, while the procedures could include tasks, techniques or methods that may requiring both cognitive and psychomotor skills. Procedural knowledge in the context of emergency egress training is any predefined responsibilities or tasks that must be performed in a recognized circumstance (e.g. what to do in the event of an emergency – raise the alarm and follow egress route to designated muster station). According to industry standards, a trainee's understanding of procedural knowledge is usually assessed through practical demonstration of the sequential tasks or procedures in a real or simulated environment (US Coast Guard Research and Development Center, 2000).

Acquisition of Procedural Knowledge

Procedural skill acquisition is a participant's ability to gain competency in one or more tasks through real or virtual experiences. Virtual environment and simulation training can help facilitate procedural knowledge acquisition by providing a means for controlled learning experiences (Salas et al., 2009). Much of procedural knowledge development is based on experiential learning (both learning by viewing and by doing, Clark and Mayer, 2008; Dahlstrom et al., 2009). Huang et al., (2010 citing Dewey theory) suggests that students should learn by doing and improve their skills through practice on a real task. Virtual environments can support the learning-by-doing process by providing credible situations and safe training opportunities that could otherwise not be performed in a real life setting. There are many factors that influence procedural skill acquisition in virtual environment training, including the learning style used by the virtual environment training program, the clarity of training goals, the representational and interface fidelity of the virtual environment technology to support cognitive and psychomotor skill development, the complexity of the tasks being trained, the practice scenario design, the duration and frequency of practice scenarios, as well as the quality of feedback received (Dalgarno and Lee, 2010; Grantcharov and Reznick, 2008).

A trainee can develop procedural knowledge in a virtual environment if the simulation training is designed using a step-by-step approach to training in order to promote procedural knowledge development (Grantcharov and Reznick, 2008). This active learning teaching approach would help manage the amount of content the trainee is required to process through the virtual environment training. These exploratory and contextual learning opportunities through active learning strategies in virtual environments can help solidify the procedural knowledge and allow the trainee to retain the learning objectives. Grantcharov and Reznick (2008) describe the process of developing procedural training using contextual and experiential learning strategies in a virtual or simulated setting (in the context of clinical training) to include the following development steps:

1) provide opportunity for the trainee to practice basic psychomotor skills until trainee proficiency criteria has been achieved
2) allow the trainee to acquire knowledge specific to the procedure,
3) demonstrate the procedure to the trainee,
4) break the procedure into key steps,
5) allow the trainee time for comprehension/reflection,
6) allow the trainee to perform single components of the procedure,
7) allow the trainee to perform the entire procedure,
8) provide comprehensive assessment and feedback throughout the learning process

The development steps outlined by Grantcharov and Reznick, (2008) suggest developing the technology to support cognitive and psychomotor skills relevant to the training objectives. Once the technology is designed for these learning aspects, the training itself should provide the trainee the opportunity to be in control of his/her learning process. First the trainee is encouraged to understand the individual steps of the procedure independently and in manageable amounts before they are required to demonstrate the full procedure (Grantcharov and Reznick, 2008). This training strategy also suggests providing immediate constructive feedback so that the trainee is able to understand the incorrect aspects of their performance in the virtual environment in order to correct their understanding.

This strategy is best applied as a whole to the training curriculum design. However, it is important to look more closely at virtual environment scenario design. As Klein (1997) suggests that practice can translate into skill development, but this is dependent on both the design and quality of scenario. Situational context, learning objectives and careful design of practice scenarios are necessary to help the trainee gain skills in a virtual environment. There are two approaches most prominently documented in the literature for designing scenarios in virtual environments: 1) goal based scenario design and 2) event-based scenario design.

Goal-based Scenario Design:

Goal-based scenarios are centered on the trainee achieving a main goal or mission over the course of the scenario. De Freitas and Neumann (2009) citing Squires (2006) describe the goal-based approach to developing simulation training in seven key components:

1) The learning goals – should be intrinsically motivating.
2) The mission – that can be accomplished by using specific skills and knowledge.
3) The cover story – creates the need for the mission to be accomplished.
4) The role – the player as protagonist.
5) The scenario operations – the level design.

6) Resources – tools and resources available.
7) Feedback – both negative and positive feedback is inherent and automatic.

Essentially these recommendations involve establishing goals for the trainee that are enveloped by a mission for the scenario, which requires the trainee to use the procedural and spatial knowledge gained from the training. De Freitas and Neumann (2009) also suggest developing a cover story that provides a need for the scenario to be completed. Included in the cover story is the role of the trainee and the duties s/he is responsible for in the scenario. Once these elements are in place, then the scenario itself requires design, which in the case of offshore safety includes the context of the emergency situation or practice drill. The final step is proving feedback to the trainee. Feedback (both positive and negative) is essential for instilling the lessons of the exercises. Without proper intervention and timely feedback, the trainee can become confused about the rules and adopt poor or complacent behaviours. De Freitas and Neumann (2009) suggest that without proper feedback intervention, the trainees do not learn that their deviations from the standard performance are incorrect.

Student debriefing or feedback comes in many automated forms, some examples include: player comparison statistics, after-action reviews (AAR), and in-game tips or guiding avatar. In the book *The Complete Guide to Simulations and Serious Games*, Aldrich (2009) explains that simulation features like that of the "player comparison panel" are useful forms of feedback that allow the trainee to compare their performance to his/her peers and adjust his/her strategy in the scenario. The performance compared in player comparison panels is usually the time to complete a scenario or the number of correctly complete actions in a scenario. This form of feedback can increase user engagement however, Aldrich cautions that overemphasis of performance scores can negatively impact the trainee's focus on learning the training content (Aldrich, 2009). After-action reviews are another form of feedback that summarizes the trainee's performance at the end of a simulation. This is a form of scenario debriefing that allows the trainee to review their performance based on multiple parameters specific to the procedural tasks being assessed in the scenario. Klein (1997) argues that AAR should include cognitive feedback related to decision making in addition to summarizing performance of specific actions because if the trainee does not understand why the mistakes were made then the trainee could propagate poor habits in subsequent training exercises. As recommended by Grantcharov and Reznick, (2008), immediate feedback in the form of in-game tips and guiding supervisors are designed to instill in-game learning and reinforce the consequences of errors and incorrect behaviours in real-time in the exercise. All forms of automated feedback have benefits and disadvantages however a commonality identified by researchers is the importance of allow the trainee time to process the feedback (also known as forced reflection) (Aldrich, 2009; Klein, 1997).

Event-based Scenario Design:

According to Fowlkes et al. (1998) the event-based approach to training provides opportunities to observe specific actions and reactions from events that have been introduced in the training program. These scenarios are carefully designed to provide hands-on practice in likely emergency situations. This is similar to introducing surprising events to the trainee, which is a characteristic of emergency situations faced in real-world situations. Dahlstrom (2009) explains that many industries have the potential for unforeseeable situations to arise that will require trainees to think critically and draw upon training and experiences to respond to incidents that may fall outside the realm of conventional training. To some extent, this is where virtual environment scenario design can help bridge the gap of conventional training. Scenario based training in particular can help enumerate the multitude of variations that can occur in emergency situations and provide a wide variety of artificial experience through practice scenarios. In addition to this, Van der Spek et al. (2013) suggest that experiencing surprising events in a virtual setting can foster deeper learning. Thus event-based scenarios that employ surprising events in their design may reinforce the development of procedural knowledge.

A common misconception with virtual environment training is the assumption that practice through virtual scenarios results in training. Klein (1997) explains (in the context of naturalistic decision making) that practice does not necessarily equate to training and "that by simply providing the trainee with an opportunity to practice does not necessarily translate into better or more meaningful training." To this end, it is also important to have a firm understand of the learning processes being targeted by the virtual environment training in order to provide meaningful training through practices scenarios. Mantovani and Castelnuovo (2003) further stress that for any application, the type of learning concepts and the learning process itself need to be the focus of the virtual training. This is necessary throughout the development of all virtual environment applications. The learning process and learning objectives should drive the development of the virtual environment technology in order to ensure the appropriate learning outcomes. Researchers also suggest that properly defining and pairing the learning objectives with the virtual environment capabilities is a critical step to ensuring that skill acquisition and knowledge retention occur (Kobes et al., 2010; Mantovani and Castelnuovo, 2003; Witmer, Sadowski & Finkelstein, 2002; Tichon and Wallis, 2010 citing Cannon-Bowers and Salas, 1998 and Wallis et al., 2007). It is important that the learning objectives match the real-world training goals and are realistically within the scope of the technology's capabilities.

2.3: Measuring Human Performance using a Virtual Environment

Comprehensive and standardized assessment of performance using virtual environments is one of the many benefits of employing virtual environment technology in training and assessing operational procedures. Performance measures in virtual environments are directly related to the concepts being taught and the desired knowledge transfer. Real-world mariner assessment methods are explored followed by the virtual environment specific assessment methods. The performance measures most relevant to the AVERT offshore safety application in this study are also highlighted.

Measuring Spatial Knowledge:

There are several different techniques described in the literature to assess the development of spatial knowledge. Spatial performance measures in real and virtual environments include: time to find a location (Bradbury-Squires, 2013; Darken et al., 2001; Farrell et al., 2003; Ruddle and Lessels, 2006; Waller et al., 1998; Witmer et al., 2002), route followed and deviations from route (Darken et al., 2001; Farrel et al., 2003; Ruddle and Lessels, 2006; Witmer et al., 1996), bearing/range estimation (Darken and Peterson, 2001, citing Witmer et al., 2002 and Koh et al., 2000), pointing tasks (Weisberg et al., 2014), time idle and time surveying an area, as well as errors such as misidentified landmarks and bumps into walls. Other means of measuring a participant's spatial knowledge is through drawing paper maps (Darken and Peterson, 2001) or model building (Weisberg et al., 2014), and blindfold tests and mazes in a real environment (Peruch and Wilson, 2004; Waller et al. 1998).

All spatial performance measures are aimed at testing the participants overall understanding of the environment and whether both route and survey knowledge was achieved. Route knowledge is more commonly tested using time to find known location, and route deviations. These performance measures test whether or not the participant was able to repeat actions they already performed in the training. Survey knowledge is tested by looking more closely at global comprehension and includes bearing/range estimations, pointing tasks, drawing paper maps of the environment post exposure, and performance in real-world applications (e.g. blindfolded or maze mock-ups). These performance measures are designed to have the participant navigate a novel area or route in the environment in order to challenge the participant's global understand of the environment.

All techniques are useful as both route and survey knowledge is required to successfully navigate an environment. The most practical techniques that can be demonstrated in a virtual environment were adopted for the AVERT study. These included time to complete, route selection, and off-route or deviations from route.

Measuring Procedural Knowledge:

Measuring procedural knowledge or performance can be categorized into three forms of assessment: oral/written questions (e.g. multiple choice quizzes/theory tests), practical demonstration exercises in a simulated job setting (e.g. virtual environment scenarios) and practical demonstration exercises in an actual shipboard job setting (e.g. real-world drills or scenarios) (US Coast Guard Research and Development Center, 2000). These forms of assessment address both the knowledge or understanding dimension and the application dimension of learning. The US Coast Guard Research and Development Center, further explain that knowledge is assessed using questions and the actual shipboard job setting (or virtual setting) is used to "assess mariner's demonstration of the method or ability to apply knowledge in the operational setting." (US Coast Guard Research and Development Center, 2000). According to industry standards the real world (or conventional) measures of procedural knowledge are broken down into the level of competence or mastery of the procedure required: knowledge, understanding, application and integration (DNV, 2005). The levels indicate whether the procedure requires basic understanding or demonstration of the skill to achieve minimum competence.

There are many different techniques described in the literature to assess the development of procedural knowledge in both real-world and virtual environment settings. Most of the procedural knowledge assessment techniques are domain specific and include observing the trainees perform a sequence of actions related to an operation (US Coast Guard Research and Development Center, 2000); comparing trainee performance against a standard performance measure; or individual trainee performance or team performance (Salas et al., 2009). The information collected to measure performance can include reaction time, time to complete, number correct or number of errors and composite or correctness scores (Gawron, 2008).

Virtual environments can measure trainee performance on procedural tasks by pre-defining the learning outcomes and proficiency criteria for each task before recording the trainee's performance. This is very similar to real-world assessment of task performance. Real-world task performance measurements are developed by defining the learning objective, performance outcomes and proficiency or assessment criteria (US Coast Guard Research and Development Center, 2000). By breaking down the task into these elements it is easier to identify the correct and incorrect behaviours associated with the task. Once this information has been determined, the scenario design can be influenced by the elements being measured. For example, if mustering at your muster station depends on route knowledge and procedural knowledge, then the scenario can be designed to assess both and collect information related to both. For route knowledge the following parameters would be recorded: route chosen, distance traveled, time to complete. For procedural knowledge the following parameters would be recorded: correct alarm recognized, correct location reached, correct movement of T-card.

Once the procedural parameters have been defined and the scenario is developed to exploit those factors then performance can be quantitatively measured and compared.

Same as in a real-world application through test scenarios, exercises, or drills. The benefit of virtual environment technology over real-world is that it allows for multiple measures (time stamped/synced) and perhaps a more in-depth analysis on performance and behaviour than in a real environment. Virtual environments can provide more consistent measure of performance in comparison to real-world assessment measures because as real-world applications have more room for human error on the part of the instructor (attentional resources likely taken to the max capacity).

The most practical procedural performance measures that can be demonstrated in a virtual environment were adopted for AVERT study. These included registering at the temporary safe refuge area (reaching the correct location and moving appropriate T-card), alarm recognition, hazard avoidance, and general safe work practices (not running on the platform and closing all fire and watertight doors).

Chapter 3 : Methodology

3.1. Experimental Overview

The experimental design for this research is a part of a larger experiment developed to accommodate two research purposes: to assess trainee competence in emergency egress using virtual environment technology and to provide supplemental data for human reliability assessment (HRA) in emergency egress situations (Musharraf, 2014). This section will first describe the Design of Experiments (DOE) methodology used and then focus on the competency assessment subset of the design, which is the research focus of this thesis. The human reliability portion of the experiment will not be reported in this thesis work. However this collaboration is important to point out as it influenced the overall design and execution of the experiment.

3.1.1. Experimental Design

3.1.1.1. Design of Experiment Methodology

Following Douglas Montgomery's design of experiments approach (Montgomery, 2012) a full factorial experiment with three replications was used to investigate the impact of three performance shaping factors on the participants' task performance and their likelihood to successfully respond to an emergency situation in AVERT. Figure 3-1 depicts the overall experimental design including the performance shaping factors (PSF) as inputs and the measured responses as outputs. To investigate the impact that performance shaping factors have on the human reliability or success likelihood of the trainee in the emergency scenario. Performance shaping factors were varied for each test scenario and the participants' task performance was recorded as responses for each test scenario (Musharraf, 2014).

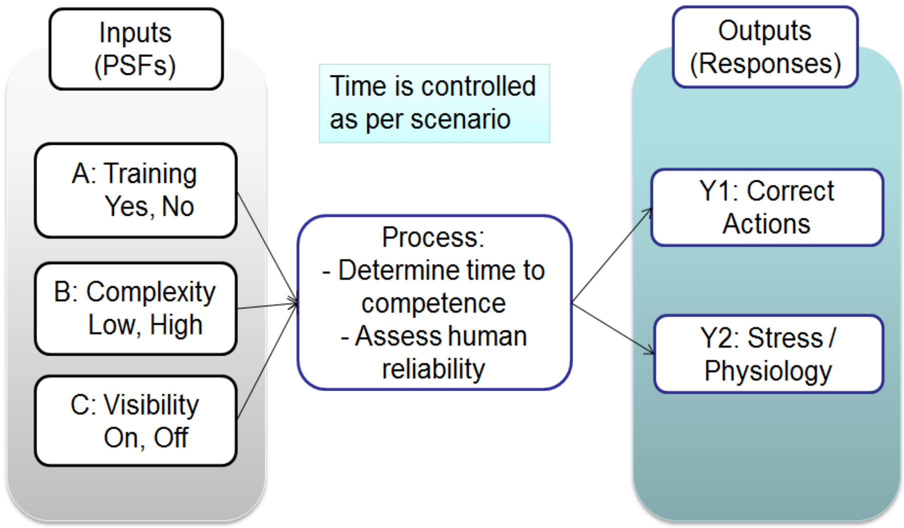

Figure 3-1: Schematic of overall experimental design for the study.

3.1.1.2. Performance Shaping Factors

To accommodate the human reliability analysis in virtual emergency situations, three performance shaping factors were identified: training, complexity and visibility. Each factor was varied at two levels: high and low. The factor levels and ranges chosen are described in Table 3-1. The training effect is the primary interest for this research while the effect of visibility and complexity are secondary interests and part of the human reliability assessment experimental design. Training (Factor A) was varied based on two levels of exposure: low or single exposure and high or repeated exposure. Overall training increased through participant exposure to the test scenarios and learning-by-doing as the participants progressed through the sessions. Following recommendations by Grantcharov and Reznick (2008), participants were gradually introduced to procedures to allow the trainee to practice each procedural component individually and then once more as a combined procedure.

Table 3-1: Factors and levels chosen for both experiments.

Factors	Name	Low (-)	High (+)
A	Training	Single exposure	Repeated exposure
B	Complexity	Cabin Routes - Alarm - Severity	Worksite Routes - Alarm - Severity
C	Visibility	Normal	Blackout

Complexity (Factor B) was varied based on simple spatial structures such as hallways in cabin egress situations to more complex spatial structures such as dense machinery spaces in the engine room/worksite egress situations. Scenario complexity also increased across sessions with regards to the emergency situation severity this ranged from muster drills to full platform evacuation situations. Visibility (Factor C) was varied based on two levels: normal visibility (clear visibility and normal lighting) and blackout conditions. Blackouts simulated full loss of lighting and night time condition situations.

3.1.1.3. Full Factorial Experimental Design

To accommodate both research purposes (time to competence and human reliability analysis) a full factorial 2^3 design was chosen. This design resulted in 8 tests (four identical tests for each training group) for each session and was repeated for a total of three sessions. This design enables a statistical analysis of the main effects and interaction effects for both competency assessment and human reliability assessment. Table 3-2 depicts the test variations for this experiment.

Table 3-2: The 2^3 Full Factorial Design.

Test	Factors			Treatment Condition
	A	B	C	
1	-	-	-	(1)
2	-	+	-	b
3	-	-	+	c
4	-	+	+	bc
5	+	-	-	a
6	+	-	+	ac
7	+	+	-	ab
8	+	+	+	abc

The high and low levels of each factor are represented by +/- in the table. The high and low levels of each factor change between repetitions. The level of training (Factor A) increases across sessions due to the participants' learning through testing. The high and low levels of complexity (Factor B) also change and increase incrementally across sessions to support learning objectives being tested. The visibility levels (Factor C) remained at the same light levels (normal or blackout conditions) across the repetitions.

3.1.1.4. Responses

Two responses were chosen to determine whether the three factors in this experiment have an effect on participant task performance, user experience and the overall participant success/failure likelihood (for human reliability analysis). The responses selected were: Y1: Correct actions (task performance), and Y2: Stress experienced by the participant. Correct actions relate to the subtasks defined in the training curriculum section and include for example time to muster, moving the t-card and closing doors (see Section 3.1.2). The level of stress experienced by the participant involved recording their physiological responses to the virtual stimulus during the in-simulation scenarios.

The correct action responses (involving the training and learning effects) are of primary interest for this research. This design allowed for further investigation into within-group performance differences due to the effects of visibility, complexity and their interaction. The design also allowed for a comparison between task performances (Y1), and the physiological responses (Y2) measured during the scenarios. These data were not included in the formal reporting of this work but will be included in Appendix B

3.1.1.5. Competency Assessment Subset of the Experimental Design

To assess the efficacy of AVERT training in emergency egress an independent group (between participants) and repeated measures experimental design was used. The assessment focused on two aspects: 1) the level of training required to achieve competency in emergency egress and 2) the amount of time required to learn and sufficiently demonstrate competency in emergency egress using AVERT.

Independent groups

To assess the level of training required to achieve competence (Factor A from the full factorial design) two independent groups were designed, each receiving different amounts of exposure to AVERT training: 1. repeated exposure to basic offshore safety training (high level training) and 2. a single exposure to basic offshore safety training (low level training).

The training and testing procedures for each group is described in detail in section 3.4. After completing the AVERT training both groups were required to complete three test sessions:

- Session 1 – Basic Safety Induction
- Session 2 – Advanced Alarm Recognition
- Session 3 – Advanced Hazard Awareness

The single exposure group (group 2) received the Basic Safety Training Tutorials during the first session and did not receive any other form of training thereafter. Figure 3-2 depicts the difference in training exposure each group received. The repeated exposure group (group 1) received the Basic Safety Training tutorials during the first session as well as practice scenarios in AVERT and additional training tutorials and practice scenarios at the beginning of each subsequent session. Each group received feedback on their performance after each test scenario.

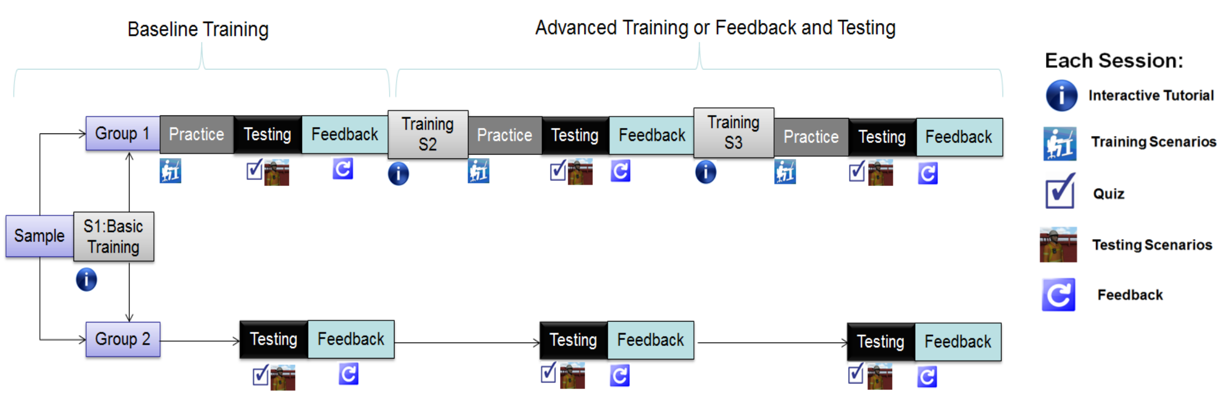

Figure 3-2: Breakdown of training exposure for each group.

Repeated Measures

A repeated measures design was used to measure learning (time to competency). The same participants in each group were tested repeatedly on their competency over the course of three separate sessions. Their task performance was measured across the sessions to measure learning (level of improvement or retention). Learning was observed if the task performance improved for each learning objective across the test sessions. To accommodate the step-by-step learning process, the three sessions were designed to present increasing numbers of learning objectives as the participants progressed through each session. The three sessions were entitled: Session 1 Spatial Awareness of Environment, Session 2 Alarm Recognition, and Session 3 Hazard Avoidance. Table 3-3 lists the learning objectives tested during each session.

Table 3-3: AVERT learning objectives tested for each session.

Session 1 – Environment	**Session 2 – Alarms**	**Session 3 – Hazards**
Spatial Awareness Routes and Mapping Register at TSR Safe Practices	Spatial Awareness Alarms Recognition Routes and Mapping Register at TSR Safe Practices	Spatial Awareness Alarms Recognition Routes and Mapping Hazard Avoidance Register at TSR Safe Practices

For the purposes of this experiment, each participant performed the test scenarios and sessions in the same order. This was chosen as it supported the learning process from a pedagogical standpoint. Participants were tested on the training content in various situations and the scenario complexity increased across the sessions. The step-by-step pedagogical approach used to test the participants' competency is depicted in Figure 3. Training Curriculum Development section (Section 3.1.2) provides more information regarding the learning objectives and training content.

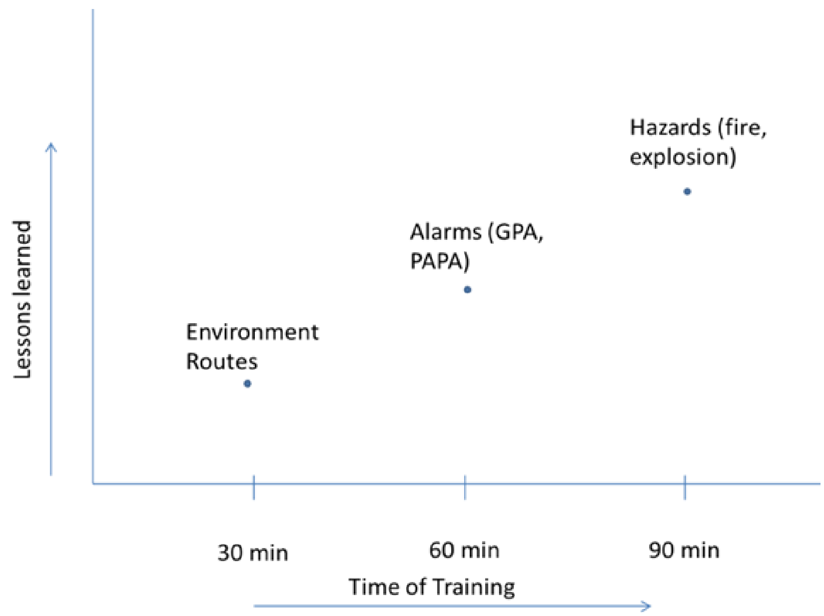

Figure 3-3: Pedagogical step-by-step testing approach.

It is important to note that the scenarios were not randomized to accommodate the incremental learning. This was done from a summative learning perspective to model the experiment training on how the AVERT Egress trainer would be deployed in a real training program. By doing so, the learning factor is now confounding the other factors.

3.1.2. Training Curriculum Development

Industry representatives and standards were referenced in selecting demonstrable competencies to be tested in AVERT. Mantovani and Castelnuovo (2003) explain that virtual environment training often accompanies other conventional training forms and it is important to determine what part of the training is most suitable for virtual environments and what is best kept in traditional methods from a cost/benefit perspective. For the AVERT training curriculum it was important to identify which learning objectives were best suited for AVERT and what learning objectives are not practically demonstrated in a virtual environment.

The AVERT Egress Trainer was designed to target eight major learning objectives: Spatial Awareness, Cognitive Awareness, Alarm Recognition, Routes and Mapping, First Actions, Situation Assessment, Standard Operating Procedures for Muster Station and Lifeboat Station and General Safe Practices. These learning objectives were defined by industry representatives in a workshop held in May 2012. Based on the existing functionality of the AVERT prototype, six of the learning objectives were targeted for this research study. The learning objectives are represented in Table 3-4.

Subject matter expert guidance and industry standards were used in the development of subtasks, training curriculum content and assessment criteria for each of the six competencies. The subtasks for each of the learning objectives are represented in Table 3-5.

Table 3-4: Learning objectives for the AVERT Egress Trainer.

Learning Objectives	**General Description**
1. Establish Spatial Awareness of the Environment	Know location of primary and secondary muster stations from two specific areas (cabin in the accommodation block and one primary work area).
2. Alarm Recognition	Objective 1 - Recognize and respond to different alarms (General Platform and Prepare to Abandon Platform Alarms) Objective 2 - Understand role and urgency of alarms (cognitive awareness of emergency situation).
3. Routes and Mapping	Know primary and secondary egress routes to muster station and lifeboat station from two specific areas (cabin in the accommodation block and one primary work area).
4. Continually Assess Situation and Avoid Hazards on Route	Understand the need to initially and continually assess situation.
5. Register at Temporary Safe Refuge (TSR) – Perform Muster and Lifeboat Station Procedures	Know muster station protocol and individual responsibility. Follow orders from muster checker/lifeboat coxswain (who must be clearly visible through appearance and behaviours) in preparation for the clear call or abandonment of the vessel.
6. General Safe Practices	Know how to keep a safe workplace. General practices (i.e. use of fire doors, areas of access, personal protective equipment to mitigate hazards).

Table 3-5: Learning objectives and subtasks for the AVERT Egress Trainer.

Learning Objectives	**Subtasks**
1. Establish Spatial Awareness of Environment	Understand the Terminology and Mapping of Main Vessel Zones
	Identify Location of Primary Work Area
	Identify Location of Vessel Living Areas including: Cabin and Mess Hall
	Identify Primary Muster Station
	Identify Lifeboat Station (Secondary Muster Station)
2. Alarms Recognition: Understand role of alarms and urgency of situation	Identify General Platform Alarm (GPA)
	Identify Prepare to Abandon Platform Alarm (PAPA)
3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Accommodation Cabin to Primary Muster Station
	Accommodation Cabin to Lifeboat Station
	Main Work Area to Primary Muster Station
	Main Work Area to Lifeboat Station
4. Assess Emergency Situation and Avoid Hazards on Route	Assess surroundings, recognize hazards on the platform and know when and how to raise the alarm
	Follow direction of Public Announcement (PA), plan egress routes and re-route if necessary (due to hazards blocking route or updates from PA)
5. Register at Temporary Safe Refuge (TSR)	Muster Station and Lifeboat Station Procedures: Perform T-Card Procedures
	Remain at muster station and obey instructions from muster checker (anticipate further PA announcements or clear call)
	Remain at lifeboat station and follow directions for Lifeboat Coxswain (anticipate boarding of lifeboat)
6. Safe Practices	Do not run on the platform
	Recognize and Use of Fire Doors and Watertight Doors (specific to egress routes)

A training curriculum was created to teach the six learning objectives and educate new personnel in basic offshore safety practices to accompany the AVERT simulations. DiMattia, Khan, and Amyotte (2005) describe the emergency muster situation as having four sequential steps: Awareness, Evaluation, Egress and Recovery. Each phase has specific muster actions for the general personnel to follow. The AVERT training curriculum focused on the Awareness, Evaluation and Egress stages of an emergency situation in the virtual environment. Table 3-6 depicts the muster actions for each phase.

Table 3-6: Muster phases and corresponding muster actions (DiMattia et al., 2005).

Awareness Phase	
1	Detect alarm
2	Identify alarm
3	Act accordingly
Evaluation Phase	
4	Ascertain if danger is imminent
5	Muster if in imminent danger
6	Return process equipment to safe state
7	Make workplace as safe as possible in limited time
Egress Phase	
8	Listen and follow PA announcements
9	Evaluate potential egress paths and choose route
10	Move along egress route
11	Assess quality of egress route while moving to TSR
12	Choose alternate route if egress path is not tenable
13	Collect personal survival suit if in accommodations at time of muster
14	Assist others if needed or as directed
Recovery Phase	
15	Register at TSR
16	Provide pertinent feedback attained while en route to TSR
17	Don personal or TSR survival suit if instructed to abandon
18	Follow OIM’s instructions

Three modules were developed to teach awareness, situation evaluation and safe egress:

- Platform Orientation module,
- Keeping a Safe Workplace module and
- Responding to Emergencies module.

The interactive training modules allowed participants to learn the specific steps required to assess the situation and to safely follow egress routes in emergency muster situations. A sample of the Offshore Safety training tutorial is depicted in Figure 3-4.

Figure 3-4: Screen capture of Basic Offshore Safety training tutorial.

The AVERT training curriculum consisted of:

- narrated tutorial modules, platform maps,
- a video walkthrough of platform, and
- site specific cabin and worksite route videos.

The interactive tutorials enabled users to click through the content at their own pace. This training material was created using Storyline © 2012 (version 1) software by Articulate Global Inc. The walkthrough and route videos were created using FRAPS ™ 2013 (version 3.5.99) by Beepa for video capture, Audacity ® (version 2.0.0) for audio editing and finally Premier Pro CC 2013 by Adobe Systems Inc. for video editing.

3.1.3. Testing Scenarios Development

Subject matter experts in offshore training were consulted in the development of the performance measures and test scenarios to assess trainee competency. The experts provided credible real-world situations for the research team to model in AVERT. Hazard types and likely locations for the hazards to occur on the platform were based on the circumstances provided by the subject matter experts. Detailed public address announcements were recorded to describe important information about the emergency situation to the participants for each scenario. The scenarios were tested and refined prior to starting the experiment to see how the hazard placement impacted trainee behaviour and route options.

The method used to establish the assessment methods and performance criteria for the AVERT training curriculum was established using the US Coast Guard Research and Development Center's Method for developing mariner assessments – Chapter 3 - Manual for assessment of developers. The performance measures outlined for each learning objectives are presented in Table 3-7.

Table 3-7: Performance measures identified for each learning objective in AVERT.

Learning Objectives	**Performance Measure**
1. Establish Spatial Awareness of Environment	Correct location
	Total time to muster at correct location Total distance travelled to correct location
2. Alarms Recognition: Understand role of alarms and urgency of situation	Correct location (GPA = Mess Hall, PAPA = Lifeboat)
3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Route selected (primary, secondary, or other)
	Total backtracking time Total backtracking distance
4. Assess Emergency Situation and Avoid Hazards on Route	Successful interaction with manual alarm call point (MAC)
	Route selection (primary, secondary, tertiary or other) Re-route in event of alarm change, PA update or encounter hazards blocking path
	Exposure time to hazard (smoke) Exposure time to hazard (fire) Incurred injury – Probability of 1st degree burns
5. Register at Temporary Safe Refuge (TSR)	Correct location and Move t-card correctly
6. Safe Practices	Speed of trainee (% running)
	Number of fire/watertight doors left open (closed)

Significant collaboration with the AVERT software development team was required to ensure the performance measures listed in Table 3-7 were suitably implemented in AVERT. Some of the performance measures were recorded and assessed automatically in-simulation (e.g. correct location, time and distance to muster, and exposure to hazards). Other performance measures required the research team to track the participants' performance using an observation log and by reviewing replay files (e.g.

route selection and re-routing). This is discussed further in section 3.5.1 - Performance Measures in AVERT.

The design of experiments (DOE) design matrix (see Table 3-2) was used to design twelve test scenarios. Subject matter experts assisted in the development of credible emergency scenarios to allow trainees to demonstrate their understanding of the six core learning objectives. The test scenarios covered a range of activities, from basic muster drills that required the trainees to go to their muster station, to a full emergency evacuation that required trainees to avoid hazards that blocked their paths and then to muster at their lifeboat stations. As participants progressed through each session, the complexity of the test scenarios increased. A detailed description of the test scenarios for each session and the required action from the participant can be found in Appendix C.

To accommodate the extent of targeted learning objectives, three separate sessions were used. In session 1, participants were required to listen to their supervising avatar and show that they knew how to get to their muster station or lifeboat station from their cabin and worksite. In session 2, participants advanced to responding to various alarm situations ranging from muster drills to real emergency situations. In session 3, participants were required to use vigilance in assessing their situation and to respond to the emergency evacuation signals. These scenarios involved blackouts, alarm changes and hazards such as galley fires, and engine room explosions. Throughout each scenario, participants were tested on their ability to identify the emergency situation based on the alarm type and PA announcements, and to respond according to the situation by selecting the safest route available to one of two designated locations at the Temporary Safe Refuge area.

3.2: Sample Size and Description of Participants

An iterative process was used to determine the sample size for this experiment. The formula used to determine the sample size is shown in Equation 3.1.

$$n = \frac{\left(t_{\alpha/2}\right)^2 \sigma^2}{B^2} \tag{3.1}$$

where:

n = sample size

$t_{\alpha/2}$ *= t-score for a 95% confidence interval*

σ *= estimated standard deviation of the population*

B = acceptable margin of error

To determine the *t-score* ($t_{\alpha/2}$), a confidence interval of 95% and 30 degrees of freedom was used. At least 30 degrees of freedom was required in order to approximate a standard normal distribution of the population. The task performance results from the previous AVERT study (Bradbury-Squires, 2013) were used to inform the standard deviation estimation of total time and distance travelled to muster. An estimated standard deviation (σ) for time to muster of +/-1.5 minutes (30% of the mean time to muster) was used for the sample size calculation. Using an iterative process, a margin of error of +/-10% was deemed acceptable. Based on the sample size calculations, a corresponding target sample size of 40 participants was used as the recruitment strategy. Another expectation was that a minimum of 15 participants for each training treatment condition would be required complete the study.

A total of 40 participants were recruited for the study. Four participants withdrew from the study. Two participants withdrew due to reported simulator sickness symptoms and two other participants withdrew due to scheduling challenges. Thirty-six volunteers in total participated in this study: 27 males and 9 females. A statistical test was performed to determine if the two groups were balanced before stopping data collection at 36 participants (i.e. to ensure there was no statistical difference between the groups based on video game experience and age). Seventeen participants were assigned to the repeat exposure to training treatment condition. Nineteen participants were assigned to the single exposure to training treatment group. No additional participants were recruited. As a result, the margin of error increased to +/-11%.

Participants were primarily recruited from Memorial University's campus by email (Appendix D), posters (Appendix E), and by word of mouth following protocol approved by Memorial University's Interdisciplinary Committee on Ethics in Human Research (ICEHR). As a result, the majority of participants were undergraduate and graduate students. Their ages ranged from 19-39 years (26.5 ± 4.4 years).

All volunteers who participated were naïve subjects with no significant prior offshore experience and no exposure to the AVERT simulator prior to the study. To measure time to competence and learning, participants with equal abilities and no experience offshore were required. The population of offshore personnel in Newfoundland possess a broad spectrum of experience and spatial knowledge of similar offshore environments. The recruitment strategy focused on naïve participants instead of using experienced offshore workers because it is difficult to recruit offshore personnel with an equal level of experience.

Upon arrival at the laboratory, participants were given a standardized explanation of the research experiment, the purpose of the research and what their participation would involve. They were informed of the possible benefits and risks to their participation and of their right to withdraw from the study at any point. Once all the information was presented and the participant had the chance to ask questions, they were then asked to provide their informed consent (Appendix F).

Part of the experimental set-up involved participants completing three questionnaires before commencing the experiment (Appendix G, H & I):

- Video Game Experience Questionnaire (VGEQ),
- Offshore Experience Questionnaire (OEQ) and
- Immersive Tendencies Questionnaire (ITQ).

Participants reported their gaming experience and marine experience prior to participating in the study and this background information was used to assign groups. The two groups were balanced based on age, gender and reported video game experience. Five questions regarding video game experience were used to balance the participants' abilities at the beginning of the experiment. The following information from the VGEQ (Appendix G) was used: whether or not participants played video games (VGEQ1), number of years playing video games (VGEQ2), number of hours per week (VGEQ3), familiarity with the controller/interface (VGEQ6) and experience level using first-person vantage point games (VGEQ7).

3.3: AVERT Simulator and Learning Management System

The All-hands Virtual Emergency Response Trainer (AVERT) is a first-person vantage point simulation prototype intended to train workers in emergency egress in offshore environments and was developed by Memorial University's Virtual Environments for Knowledge Mobilization project. The AVERT simulation environment coupled with a learning management system (LMS) was used to deliver guided training content and administer the test scenarios.

3.3.1. AVERT Environment

The AVERT Egress Trainer provides a three dimensional visual representation of what one would experience when located on an offshore platform. From the perspective of the trainee, AVERT offers an immersive virtual environment to provide them with hands-on workplace familiarization and a safe setting to practice emergency egress procedures. The AVERT Egress Trainer prototype was developed using the Virtual Environments for Knowledge Mobilization project's custom designed open source distributed simulation engine (ODSE). The graphical representation was developed using the following open source libraries: OGRE3D© by Torus Knot Software, Sky X© and Hydrax© by Verguin Gonzalez, Caelum© by Caelum Team, Boost© (version 1.0) by Boost Software and Sixsense C++ libraries. Audio capabilities were developed using FMOD Audio © sounds system. The in-game physics was made available using PhysX©

by NVIDIA. The 3-D model for the virtual environment was developed using SoftImage© software by Autodesk.

For the purposes of this study, the virtual environment was modeled after a generic Floating Production Storage and Offloading vessel (FPSO). The participants had access to the following areas on the virtual platform:

- Accommodation Block,
- Deck at the aft of the vessel,
- Engine Room, and
- Steering Gear Room.

3.3.2. Learning Management System and Automated Brief/Debrief System

A learning management system (LMS) was used to interface with the AVERT simulation. The LMS is an open source web-based e-learning platform called Moodle™. The Moodle™ LMS linked to tutorial content and AVERT scenarios and allowed participants to login to the session course material. Through the LMS, participants had access to guided power point presentations, platform walkthrough videos, and quizzes. Depending on the group assignment (single or repeated training exposure), participants could select practice scenarios to perform. AVERT recorded in-game task performance to a report file that the LMS used to provide feedback and track performance. Participants could also deploy testing scenarios in AVERT through the LMS. The automated brief/debrief system provided the participants with after action review feedback on their performance. Using the LMS coupled with the automated brief/debrief system, participants were free to select the degree of feedback on their performance after completing each scenario.

3.3.3. Equipment Set-up

The AVERT workstation consisted of a desktop 19-inch flat screen monitor, a standard off-the-shelf dual joystick video game controller (Xbox 360© controller) and headphones. The instructor station was situated directly behind the AVERT workstation. This allowed researchers to observe what the participant was doing while monitoring the participant's physiological responses to the test scenarios.

Participants interacted with the AVERT offshore simulation using the video game controller (Xbox 360© controller). The left joystick controlled the user's field of view (FOV) while the right joystick controlled the user's movement through the virtual space. To perform actions within the environment (for example open doors or interact with the muster boards) the user could press the (A) button. To use the flashlight, the user could press the (B) button.

Prior studies found that controller proficiency has some impact on task performance. Effort was made to reduce the familiarization time required with the controller and the impact that controller proficiency had on the overall success of the user interaction. For this reason, controller sensitivity was fixed at a moderate movement and participants were given the option to adjust the camera view (y-axis) to their preference (from normal to inverted). Movement options included stationary, walking (4.8km/h) and running (12 km/h).

3.4: Procedure (Offshore Training and Simulation Testing)

Over the course of the experiment, participants were asked to attend three sessions on three separate days. The total time commitment ranged from 8 to 10 hours. This section will discuss the treatment each group received during the experiment.

3.4.1. Session 1: Environment Awareness (4-5 hours)

The first session involved the initial set-up, participant consent, and group assignment. After completing the briefing, the remainder of the session consisted of two phases: training phase and testing phase.

Session 1: Training Phase:

The training content for session 1 focused on establishing environment knowledge (spatial awareness) and the basics to properly prepare for responding to egress situations. During this session, participants were instructed to imagine it was their first day of training for a job on an offshore platform. They were informed that in the event of an emergency offshore they would be required to successfully evacuate in a limited time period. With this in mind, their goal was to learn the material presented through tutorials to the best of their ability (i.e. learn escape routes from their accommodation and worksite to the primary and alternate muster points).

The bulk of session 1 involved familiarizing the participants up with the basic offshore safety practices. The AVERT Basic Safety Briefing training content for this session covered the Platform Orientation, Keeping a Safe Workplace and Responding to Emergencies modules as described in section 3.1.2. To help participants process the information, the training was divided into two parts. Part 1 included the Platform Orientation with a video tour of the platform and the Keeping a Safe workplace module. Part 2 involved the Responding to Emergencies Module. Part 1 and part 2 were separated by a 30-minute platform exploration period. This allowed participants to use the information in part 1 (spatial learning) during their exploration and review the content in part 2 (emergency protocols) with a better spatial understanding of the virtual platform.

After completing part 1 of the tutorial, participants were given an intermission followed by a five to fifteen minute controller familiarization period. The controller familiarization provided participants with an opportunity to learn the controls associated with AVERT. Participants with prior video game experience received 5 minutes of familiarization. Participants with little to no game experience were allowed 15 minutes of familiarization. The environment used during the controller familiarization was not the environment the participants were required to know in the study.

Once participants were comfortable with the controls they were given thirty minutes to freely explore their offshore environment. To guide their active exploration they were given a list of important locations to find. They were instructed to find their cabin, worksite (in the engine room), as well as primary and secondary muster points.

Part two of the AVERT Basic Safety Briefing tutorial focused on the Responding to Emergencies Module. This introduced the concept of emergency alarm signals and the specific steps required to assess the situation and to safely follow egress routes in emergency situations. Participants watched five route videos highlighting important egress routes: two available routes from their cabin and three available routes from their worksite in the engine room.

After completing the tutorial training, the participants were assigned to one of two groups: 1. a single exposure to basic offshore safety training, or 2. repeated exposure to basic offshore safety training. The groups were balanced based on participants' reported video game experience. The group assignment determined the participant's level of exposure to AVERT and the schedule of training and testing for each session. The single exposure group subsequently proceeded directly into the testing phase (they did not receive any other form of training), while the repeated exposure group received four practice scenarios prior to the testing phase. Figure 5 shows the training and questionnaire distribution timeline for both groups in session 1.

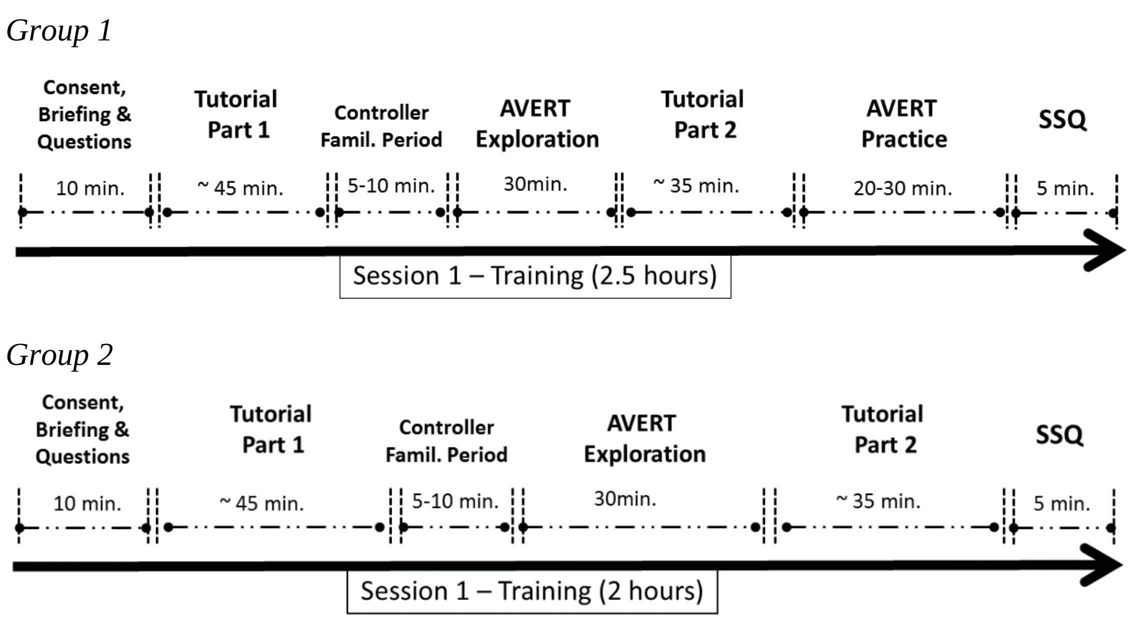

Figure 3-5: Session 1 training and questionnaire distribution timeline for each group.

Session 1: Testing Phase

The testing phase began after the second intermission. Participants had 15 minutes to complete a twenty question multiple choice quiz. Upon the completion of the quiz they received a score and were prompted to review the correct answers.

Before starting the test scenarios in AVERT, the participants were asked to complete a simulator sickness questionnaire (Appendix J) and were equipped with the physiological measuring system (see Section 3.5 for detailed set up). The physiological responses measured included:

- Heart Rate (HR),
- Skin Temperature (ST),
- Galvanic Skin Response (GSR), and
- Respiration Rate (RR).

For comparison, a resting baseline of the participant's physiology was collect before the start of each test scenario. During the baseline the participants were asked to relax as much as possible and to avoid talking or moving.

Participants were instructed to take the simulation seriously and to react to the situations in the simulation as they would in a real life setting. Test scenarios were completed in a specific order using a step-by-step pedagogical approach. Figure 6 depicts the timeline for the test scenarios. Participants received automated feedback on their performance after completing each scenario.

Group 1 & Group 2:

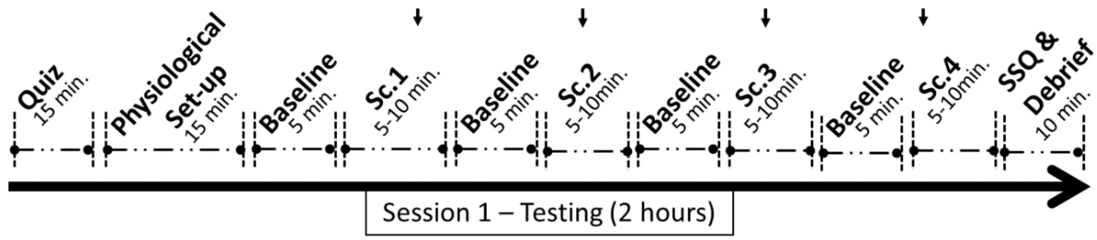

Figure 3-6: Session 1 testing and questionnaire distribution timeline for each group.

Once participants completed all test scenarios, they were asked to fill out a second simulator sickness questionnaire to provide a measure of the simulator sickness symptoms attributed to their navigation in the virtual environment. Participants also completed a post session debrief questionnaire (Appendix K) asking them to reflect on their performance and to describe what went well and what was most challenging about the exercises. They were also asked to rate the utility of the training material and AVERT training aids (feedback, checkboxes and videos).

Participants were asked to return for session 2 and session 3 on separate days. The sessions were spaced by a minimum of 24 and a maximum of 48 hours. In two cases individuals exceeded the 48-hour maximum time between their second and third session due to scheduling challenges (snow storms).

3.4.2. Sessions 2 & 3: Alarm Recognition and Hazard Avoidance (1-2 hours each)

The second session involved assessing the participants' understanding of alarms and how to response to more complex emergency situations. The third session involved assessing the participants' ability to respond to emergency situations with dynamic hazards and unexpected changes to the severity of the emergency situation.

Sessions 2 & 3: Training Phase (for group 1 only):

Participants in group 1 returned for refresher training while participants in group 2 not receive any other training and went straight into the testing phase. Group 1 training involved a thirty minute refresher tutorial on alarms, review of the route videos and four practice scenarios in AVERT before the testing phase. In total, group 1 received an additional 4 hours of tutorial review and 12 scenarios of practice than group 2 over the course of the three sessions. Figure 7 depicts the timeline of additional training group 1 received in session 2 and session 3.

Group 1

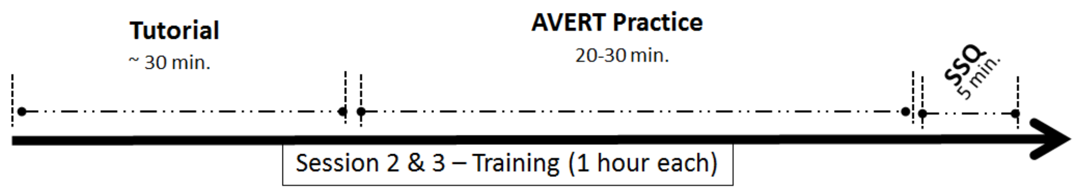

Figure 3-7: Sessions 2 and 3 training timeline for group 1.

Sessions 2 & 3: Testing Phase

After the intermission the participants of both groups followed the same protocol from session 1: completing the twenty question multiple-choice quiz, and four test scenarios. In the final session, participants were also required to complete two of subjective assessment questionnaires (Appendix L). Figure 8 depicts the timeline for the test scenarios.

Group 1 & Group 2:

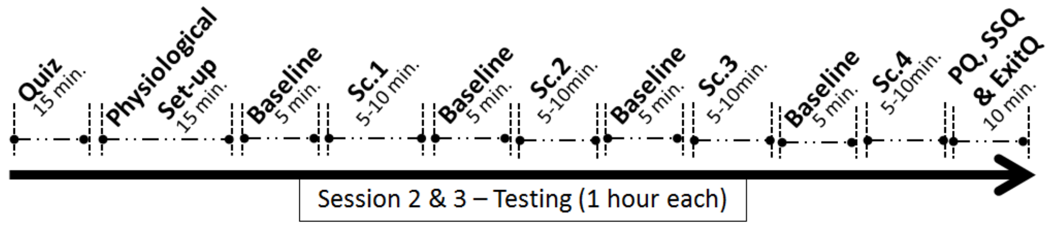

Figure 3-8: Sessions 2 and 3 testing timeline for each group.

3.5: Data Collection Protocol

Three forms of data were collect during the experiment: 1) AVERT task performance, 2) physiological responses, and 3) subjective reports in questionnaire form.

3.5.1: Performance Measurements in AVERT

When a trainee plays through a scenario, AVERT records performance metrics to a report file. This report file allows for quantitative analysis of performance. The AVERT report file recorded the following information for all test scenarios:

- Correct location, (GPA dictates Mess Hall, PAPA dictates Lifeboat)
- Total time to muster at correct location
- Total distance travelled to correct location
- Total backtracking time
- Total backtracking distance
- Successful interaction with manual alarm call point (MAC)
- Exposure time to hazard (smoke/fire)
- Incurred injury – Probability of 1st degree burns
- Move t-card correctly
- Speed of trainee (percent running)
- Number of fire/watertight doors left open (closed)

To accompany the report file, an observation log was recorded describing the actions of each participant during the test scenarios for qualitative analysis. The observation log was used as a means to confirm the accuracy of the AVERT report file data and provided the following qualitative performance metrics:

- Route selection (primary, secondary, tertiary or other)
- Re-route in event of alarm change, PA update or encounter hazards blocking path

The AVERT report file data combined with the observation logs were used to develop an aggregated task performance score for each test scenario. This aggregated score refers to the participants' overall Competency Score in achieving the core learning objectives (see results section for details).

3.5.2: Physiological Measurements

During the testing scenarios, the participant’s physiology was measured as an indicator of their immersion and stress due to the virtual emergency situations. Physiological data collection took place during the test phase of each session. To control diurnal variation effects upon the physiological measurements, each participant was tested at the same time of day. The physiological responses were recorded using the NeXus-10 Mark II measuring system accompanied with BioTrace+ software (version V2013A) by Mind Media B.V. Netherlands. Electricardiography (ECG) was measured by placing electrodes on the participant’s chest and abdomen. The ECG raw data was transformed into heart rate (HR) using the BioTrace+ software. Respiration rate (RR) was measuring by placing a sensor strap over the participant's ribcage. Peripheral skin temperature (ST) was measured by taping a thermocouple between the participant's thumb and the index finger on their left hand. Finally, galvanic skin response (GSR) was measured by placing two electrodes on the ring finger of each hand.

3.5.3: Subjective Assessments

Participants were asked to complete a number of questionnaires: Simulator Sickness Questionnaire, Immersive Tendencies Questionnaire, Presence Questionnaire, and Simulation Utility Questionnaire.

The simulator sickness questionnaire was distributed before and after the participants' exposure to the AVERT simulation for all test sessions. A total of six before and after simulator sickness reports were collected for each participant. This data provided information on whether or not navigation through AVERT causes participants to experience symptoms of simulator sickness.

Chapter 4 : Results

4.1: Task Performance

4.1.1: Task Performance with Training Type as Grouping Variable

4.1.1.1. Time spent training for both groups

Thirty-six participants (27 males and 9 females) completed the study. Seventeen participants received repeated exposure to training through three tutorial and practice sessions (group 1) and nineteen participants received a single exposure to training through one tutorial session (group 2). As part of the repeated exposure training, group 1 received more time in aggregate to review the tutorials and practice with AVERT than group 2. Figure 4-1 depicts the cumulative time spent reviewing the tutorial content across the three sessions (S1, S2, S3). Both groups have almost identical exposure time to tutorial content in the first session (S1). As group 2 is not exposed to tutorials in subsequent sessions S2 and S3, the gap in cumulative exposure time to tutorial content increases step-wise as group 1 is exposed to tutorials (T2 and T3) in sessions S2 and S3.

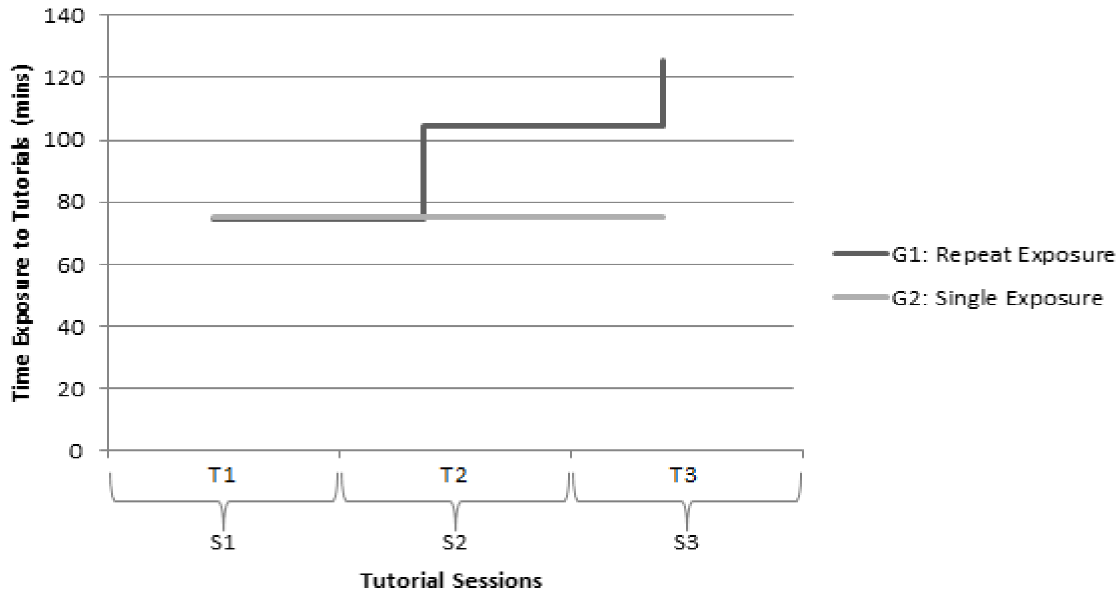

Figure 4-1: Total time exposure to learn and review tutorial material.

Specifically, group 1 and group 2 participants used a mean total of 74.58 minutes and 75.40 minutes respectively during session 1 to learn the tutorial material. Group 1 participants used a mean of 30.12 additional minutes to review tutorial material in session 2 and a mean of 20.63 minutes more in session 3, for a mean total exposure time to tutorial material of 125.33 minutes. Part of the tutorial in session 1 involved watching a platform orientation video and videos of five egress routes: two from the accommodations and three from worksites. Both groups reviewed 100% of the route videos available during session 1. Group 1 also had the opportunity to watch the egress videos again during the tutorial review at the beginning of session 2 and session 3. On average, group 1 participants also reviewed 79% of the route videos during session 2 and 64% of the route videos in session 3.

Figure 4-2 shows the time participants spent actively engaging in AVERT through practice (P1, P2, P3) and test scenarios (E1, E2, E3). Each of the three sessions is broken out into components. For example, session 1 is comprised of an active exploration exercise (denoted as A in the figure), practice exercises (denoted as P1) and testing scenarios (denoted as E1). The tutorial training is not illustrated in Figure 2 as the tutorials did not include active engagement with AVERT by the participants.

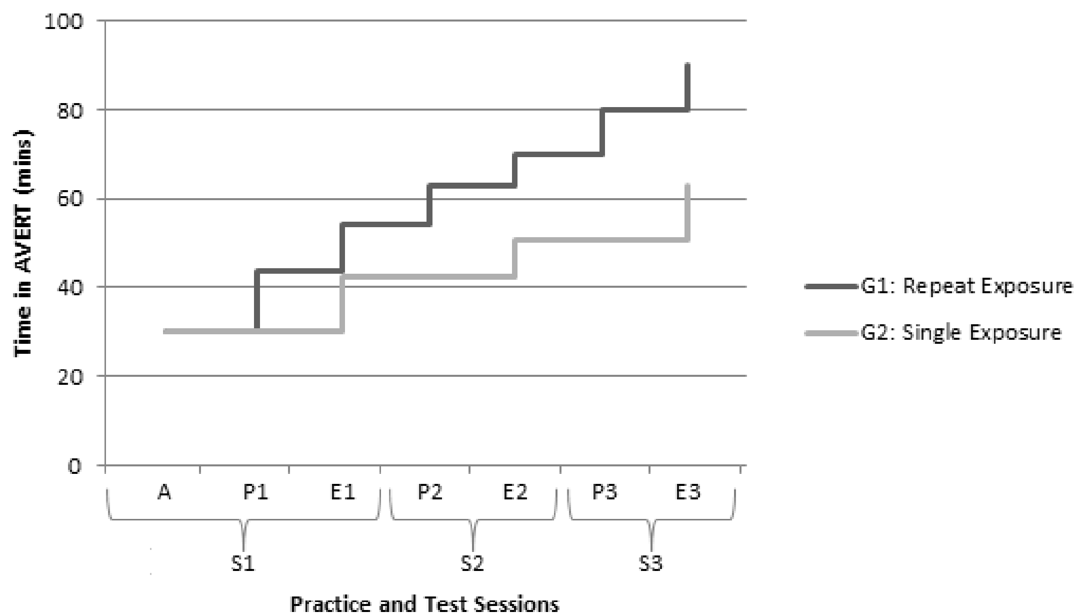

Figure 4-2: Time spent in AVERT across the three sessions.

As illustrated in Figure 4-2, both groups had the same initial active engagement with AVERT through a thirty minute active exploration exercise in session 1. The active exploration scenario was designed to provide the participants thirty minutes to actively explore and familiarize themselves with the accommodations and worksites on the platform. The scenario also highlighted a list of important places they should know and locate during their exploration.

Table 4-1 shows that both group 1 and group 2 participants received the same active exploration time (A) in session 1. Group 1 participants were subsequently given 40 minutes to practice in four AVERT scenarios, prior to the tests at the end of the session. In session 1, group 1 participants used a mean of 13.80 minutes to practice in the exercise scenarios (P1). In sessions 2 and 3, they used a mean of 8.71 and 9.99 minutes, respectively to practice (P2 and P3). In total, group 1 participants spent a mean of 32.49 minutes over the three sessions in AVERT practice scenarios. Group 2 participants were not given the option to practice, so their time in the AVERT environment after the initial exploration exercise was limited to testing in each of the three sessions.

Table 4-1: Time (in minutes) spent in active exploration, practice and test exercises across the three sessions.

Group	S1			S2		S3		∑(Pi)	∑(Ei)	Total
	A	P1	E1	P2	E2	P3	E3			
G1	30.0	13.8	10.6	8.7	6.9	9.9	10.2	32.5	27.7	90.2
G2	30.0	0.0	12.4	0.0	8.4	0.0	11.9	0.0	32.8	62.8

At the end of each session, participants in group 1 and group 2 were given four test scenarios with up to 40 minutes to demonstrate their competency in AVERT. Each test scenario had a 10 minute time limit. As indicated in Table 4-1, group 1 participants spent a mean of 10.58 minutes to complete the session 1 test scenarios (E1), a mean of 6.89 to complete the session 2 test scenarios (E2) and a mean of 10.19 minutes to complete the session 3 test scenarios (E3). Group 2 participants spent a mean of 12.43 minutes to complete session 1 test scenarios (E1), a mean of 8.41 to complete the session 2 test scenarios (E2) and a mean of 11.94 minutes to complete the session 3 test scenarios (E3). In total, group 1 and group 2 participants spent a mean of 27.66 minutes and 32.77 minutes, respectively over the three sessions in AVERT to complete the twelve test scenarios.

As indicated in Figure 4-2, group 1 received a mean total exposure time to AVERT (includes A, P and E) of 90.15 minutes, while group 2 received a mean total exposure time to AVERT (includes A and E) of 62.77 minutes. This difference in cumulative exposure time to AVERT can be attributed to the additional practice time group 1 received and the fact that group 1 completed the test scenarios in less time than group 2. Group 1 received on average an additional 32.49 minutes to practice (P1, P2, and P3) than group 2. Group 2 spent a cumulative 5.11 minutes more time than group 1 to complete the test scenarios (E1, E2, and E3). On average, group 1 completed the test scenarios 1.70 minutes more quickly than group 2 across all three sessions.

4.1.1.2. Task performance comparison between groups

The AVERT learning objectives were outlined by industry representatives in a workshop. Subject matter expert guidance helped design assessment criteria for each of the six competencies. Group task performances for the six AVERT learning objectives were measured repeatedly across the three sessions. Each session targeted specific learning objectives and incrementally assessed the corresponding task performance measures related to Spatial Awareness of Environment (S1), Alarm Recognition (S2) and Hazard Avoidance (S3). Table 4-2 shows which learning objectives were targeted and the associated task performance measures that were tested during each session.

The results are reported based on the AVERT Learning Objectives and their corresponding Performance Measures as outlined in Table 4-2. The Spatial Awareness of the Environment learning objective has three performance measures: Correct Muster Location, Time to Muster and Distance Travelled during Mustering. The Alarm Recognition learning objective has one performance measure: Correct Muster Location. The Routes and Mapping learning objective has three performance measures: Route Selection, Backtracking Time and Backtracking Distance. The Assessing Situation and Hazard Avoidance learning objective has three performance measures: Raise the alarm, Re-route in the event of hazard or alarm change, Avoid Exposure to Hazards (includes time spent in smoke, time spent in contact with fire and injury due to fire exposure). The Register at the Temporary Safe Refuge (TSR) learning objective has only one performance measure: Correct Muster Location. Finally, the Safe Practices learning objective has two performance measures: Percent Running, and Closing Fire and Watertight doors.

Table 4-2: AVERT learning objectives and task performance measures tested for each session.

Session 1 - Environment	**Session 2 - Alarms**	**Session 3 - Hazards**
1. Spatial Awareness • Correct location • Total time to muster • Total distance travelled 3. Routes and Mapping • Route selected • Backtracking time • Backtracking distance 5. Register at TSR • Reach correct location and move T-card correctly 6. Safe Practices • Percent time running • Close fire and watertight doors	1. Spatial Awareness • Correct location • Total time to muster • Total distance travelled 2. Alarms Recognition • Identify alarm and go to correct location 3. Routes and Mapping • Route selected • Backtracking time • Backtracking distance 5. Register at TSR • Reach correct location and move T-card correctly 6. Safe Practices • Percent time running • Close fire and watertight doors	1. Spatial Awareness • Correct location • Total time to muster • Total distance travelled 2. Alarms Recognition • Identify alarm and go to correct location 3. Routes and Mapping • Route selected • Backtracking time • Backtracking distance 4. Hazard Avoidance • Successful interaction with MAC (Raise Alarm) • Re-route to avoid hazards blocking path • Exposure time to smoke • Exposure time to fire • Injury incurred (probability of 1st degree burns or death) 5. Register at TSR • Reach correct location and move T-card correctly 6. Safe Practices • Percent time running • Close fire and watertight doors

The task performance measures were recorded for each test scenario through the AVERT report files and the experiment observation logs. The data collected for the task performance measures included several types of data: nominal, ordinal and continuous data. Each type of data required a different method of statistical analysis.

The nominal data collected includes pass/fail task performance measures such as Correct Muster Location, Correct Alarm Recognition, Registering at TSR (moving T-card), Door Count, and Raising Alarm. The ordinal data collected included the ranked Route Selection task performance measure. Given the size and independent sample of the data, Fisher's Exact Test was used to compare proportions for the two groups instead of Pearson chi-square.

The continuous data included the task performance measure such as Time to Muster, Distance Travelled to Muster, Backtracking Time, Backtracking Distance, Percent Running, Exposure time to Smoke and Fire, and Injury Likelihood. Due to the small dataset, the Shapiro-Wilk Test was used instead of Kolmogorov-Smirnov to test the distribution of the data. Significant deviations from normality were found using the Shapiro-Wilk test (p-value 0.05) for the task performance continuous data. Therefore the null-hypothesis that data is normally distributed was rejected. Non-parametric statistics were used to perform statistical testing of the continuous data to compare means and variance. To determine the effect of training group on task performance, a two independent samples Kolmogorov-Smirnov test was used.

Data analysis was conducted using IBM SPSS Statistics (v22.0) Exact Tests software. Results include all 36 participants unless otherwise stated (no outliers have been removed). All results are expressed as mean ± standard error. An alpha value of p-value < 0.05 was used to signify a statistical significance between groups while a p-value < 0.10 was used to signify a statistical trend difference between groups. All tests are two-tailed unless stated otherwise.

1. Spatial Awareness of Environment and Alarm Recognition

Correct Muster Location *(incl. Alarm Recognition and Register at the TSR subtasks)*

The correct location task is an indicator of whether or not the participant was able to reach his/ her primary or secondary muster point. The test scenarios are based on realistic evacuation situations where an alarm is sounded to indicate which location the participant is being directed to muster (this is the Alarm Recognition objective). If a General Platform Alarm (GPA) is sounded, then all personnel are required to gather at the primary muster point, the Mess Hall. If a Prepare to Abandon Platform Alarm (PAPA) is sounded, then all personnel are required to go to their secondary muster point, the Lifeboat Station. Each scenario clearly indicates where the correct muster point is for the participant. Once the participant reaches the correct location (also known as their muster points within the Temporary Safe Refuge) s/he must register at the muster point by moving her/his T-card from the ‘Steady’ to the ‘Mustered’ state. This is the Register at the TSR objective.

In AVERT, to be successful in reaching the correct location task the participant must perform the following:

- reach the designated muster location within the 10 minute time limit, and
- signify s/he has reached the location by moving her/his T-card.

Conversely, the participant fails this task if s/he:

- does not reach the right muster location,
- interacts with another muster board at the wrong location (any muster station other than her/his own),
- or if s/he is unable to reach the designated muster location within the 10 minute time limit.

Group results for the correct alarm recognition, correct muster location and registering at the TSR performance measures are reported in Table 4-3. For reporting purposes, the Alarm Recognition and Temporary Safe Refuge Registering learning objectives are included in this section as they are dependent on a successful muster in order to be considered correct as well. For example, the alarm type dictates the location at which to muster. Without the participant going to the correct location, AVERT has no measure of whether or not the participant knew the meaning of the alarm type. Likewise, to correctly register at the muster station (i.e. moving his/her t-card); the participant must reach the correct location first. For these reasons, the three separate learning objectives will be discussed together in this section.

The percentage of correct and incorrect musters for each group is listed in Table 4-3. Overall, the mean success for both groups was similar across all three sessions. There were three instances where the group difference was 15% or greater (denoted as scenario codes TE1, TE4, and TH3). The first test scenario (TE1) resulted in the biggest difference of 19% between groups and the overall poorest correct location scores: 82% correct for group 1 and 63% correct for group 2. In comparing the group proportions using Fisher's Exact Test, there was no statistical difference found between group participants in reaching the correct muster location for all scenarios across all three sessions.

Table 4-3: Group proportions for correct muster and correct alarm recognition tasks.

Scenario Code	Group	Muster Count		Performance Percentage		P-value
		Correct	Incorrect	% Correct	% Incorrect	Fisher's Exact Test Exact Sig. (2-sided)
TE1	G1	14	3	82	18	.274
	G2	12	7	63	37	
TE2	G1	14	3	82	18	1.000
	G2	16	3	84	16	
TE3	G1	16	1	94	6	.472
	G2	19	0	100	0	
TE4	G1	17	0	100	0	.231
	G2	16	3	84	16	
TA1	G1	16	1	94	6	.472
	G2	19	0	100	0	
TA2	G1	16	1	94	6	1.000
	G2	18	1	95	5	
TA3	G1	16	1	94	6	1.000
	G2	18	1	95	5	
TA4	G1	16	1	94	6	.472
	G2	19	0	100	0	
TH1	G1	16	1	94	6	1.000
	G2	17	2	89	11	
TH2	G1	16	1	94	6	1.000
	G2	17	2	89	11	
TH3	G1	17	0	100	0	.231
	G2	16	3	84	16	
TH4	G1	16	1	94	6	.605
	G2	16	3	84	16	

As shown in Table 4-3, there were seven scenarios where neither group was 100% successful at reaching the correct muster location. The lowest overall score for this task was 63.16% correct muster for group 2 in the first scenario (TE1). The highest overall score for this task was 100% correct muster. This was achieved by group 1 in two scenarios (TE4 and TH3) and for group 2 in three scenarios (TE3, TA1, and TA4). The main reason group 1 does score 100% for correct location task in the majority of the scenarios is due to one outlier who was consistently unable to reach the correct muster location throughout the test scenarios. If the outlier is removed group 1 would receive 100% correct location in every scenario in session 2 and session 3.

Total Time and Distance to Complete the Scenario:

Two performance measures of the participants' spatial awareness of the environment are their total time to muster at the correct location and their total distance travelled to the correct location. These measures provide insight into the participants' knowledge of the physical geography and how proficient they are at navigating to their designated muster points. Participants were required to complete the test scenarios as quickly and efficiently as possible. The time and distance they took to complete the scenarios were recorded in a AVERT report file.

Participants were required to demonstrate they could find their muster points from two main locations on the platform: their cabin and their worksite in the engine room. The cabin routes are shorter distances and have lower navigation complexity (i.e. more direct routes with fewer route options). The engine room worksite routes are longer and involve more complex navigation (i.e. less clearly designated routes and more route options). Therefore, the data in this section is reported based on starting location: cabin egress and worksite egress scenarios.

The mean total time and distances travelled described in this section include the entire sample of participants even if the participant did not go to the correct location in the scenario. The incorrect muster time to complete and distance travelled were included so that no one parameter interfered with the other in general reporting. The Aggregated Competence Score section describes in detail the interaction of the parameters.

Cabin Egress Scenarios:

Figure 4-3 and 4-4 show the task performance by group for time to muster and distance travelled for the cabin egress test scenarios in AVERT. Two scenarios of every session focused on accommodation egress situations. Each of the three sessions is broken out into two cabin egress scenarios: TE1 and TE3 for session 1, TA1 and TA3 for session 2 and TH1 and TH2 for session 3. For example, session 1 and session 2 are comprised of a normal condition test scenario (denoted as TE1 and TA1 in the Figure) and a blackout condition test scenario (denoted as TE3 and TA3). Session 3 is comprised of two hazard condition evacuation test scenarios from the cabin (denoted a TH1 and TH2). See Appendix C for specific scenario details.

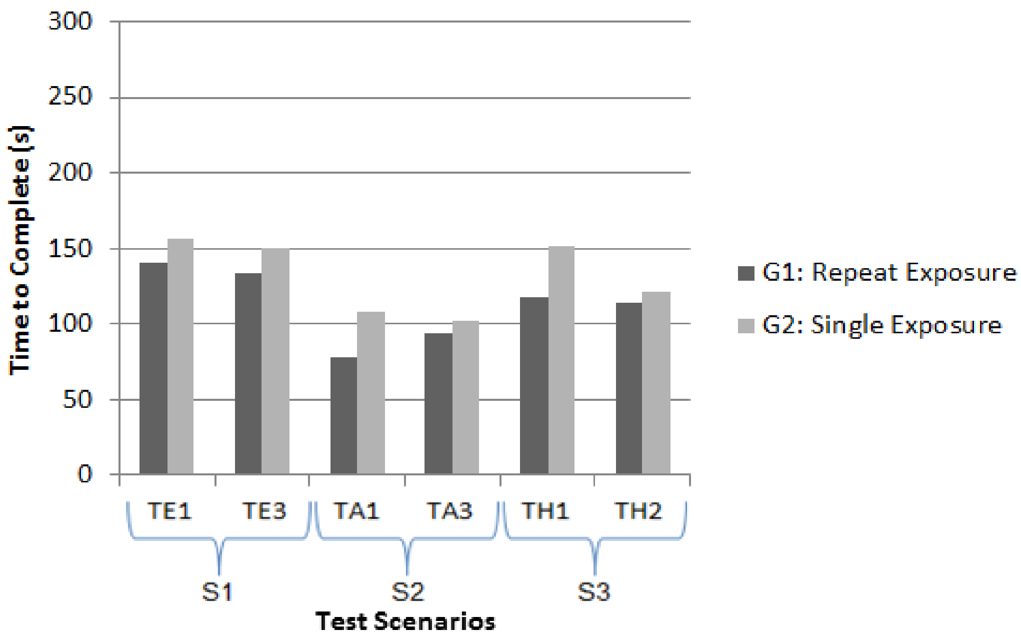

Figure 4-3: Time to complete cabin egress scenarios across the three sessions.

As indicated in Figure 4-3, group 1 completed the cabin egress test scenarios on average 18.63 seconds more quickly than group 2 across the three sessions. In two scenarios (TA1 and TH1), the difference between group 1 and group 2 in time to muster was more than 30 seconds. Group 1 is 30.04 seconds faster at completing TA1 ($p = 0.036$) and 33.62 seconds faster at completing TH1 than group 2 ($p = 0.488$). The same trend is not shown in distance travelled to muster. Figure 4-4 shows that group 1 uses less distance to muster (on average 9.70 meters less than group 2) in four out of the six test scenarios. However, in the first scenario (TE1) group 1 travelled on average 42.36 meters more to muster than group 2 ($p = 0.803$). This difference in distance is entirely due to one outlier. This participant became lost during the scenario and went to alternate muster station at the forward section of the vessel. If the outlier is removed the mean distance for group 1 in TE1 changes from 143.16m to 84.34m.

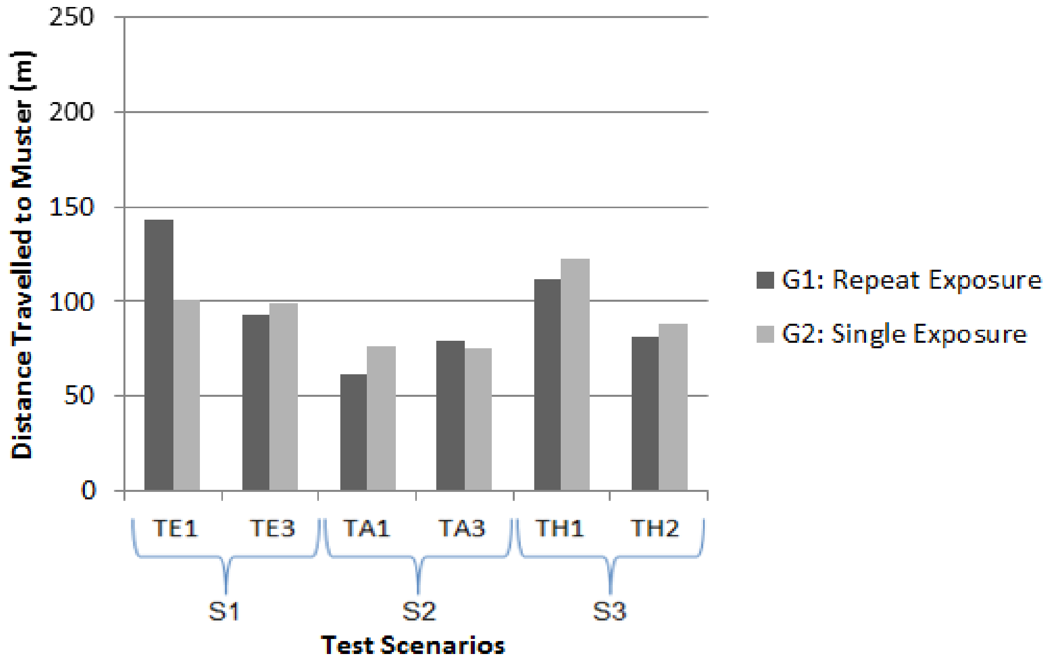

Figure 4-4: Distance travelled in the cabin scenarios across the three sessions.

Table 4-4 shows the task performance by group for time to muster and distance travelled for the cabin egress test scenarios. In the cabin alarm scenario in session 2, (denoted as scenario code TA1) there was a significant difference between groups for time to muster ($p = 0.036$) and distanced travelled during evacuation ($p = 0.005$). Group 2 took on average 30.04 seconds longer and travelled 14.65 meters more to complete the scenario. Otherwise, no other statistical significance was found for the task performance (time and distance) in the cabin egress scenarios.

Table 4-4: Time to muster and distance travelled for the cabin egress scenarios.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TE1	Normal	Time to muster	G1 G2	17 19	140 156	31 27	0.527
		Distance	G1 G2	17 19	143 101	59 15	0.803
TE3	Blackout	Time to muster	G1 G2	17 19	134 151	12 22	0.450
		Distance	G1 G2	17 19	93 99	10 15	0.527
TA1	Alarm	Time to muster	G1 G2	17 19	78 108	7 9	**0.036**
		Distance	G1 G2	17 19	62 77	6 5	**0.005**
TA3	Alarm + Blackout	Time to muster	G1 G2	17 19	94 102	8 9	0.820
		Distance	G1 G2	17 19	80 75	6 6	0.488
TH1	Alarm + Hazard	Time to muster	G1 G2	17 19	118 152	10 21	0.488
		Distance	G1 G2	17 19	111 123	8 13	0.851
TH2	Alarm + Hazard	Time to muster	G1 G2	17 19	114 122	10 15	0.127
		Distance	G1 G2	17 19	81 88	9 8	0.652

Engine Room Egress Scenarios:

Figure 4-5 and Figure 4-6 show the task performance by group for time to muster and distance travelled for the engine room worksite test scenarios in AVERT. Similar to the accommodations scenarios, there are two scenarios in every session that focused on the worksite egress situations. Each of the three sessions is broken out into two engine room egress scenarios: TE2 and TE4 for session 1, TA2 and TA4 for session 2 and TH3 and TH4 for session 3. For example, session 1 and session 2 are comprised of a normal condition test scenario (denoted as TE2 and TA2 in the Figure) and a blackout condition test scenario (denoted as TE4 and TA4). Session 3 is comprised of two hazard condition evacuation test scenarios from the engine room worksite (denoted a TH3 and TH4). See Appendix C for specific scenario details.

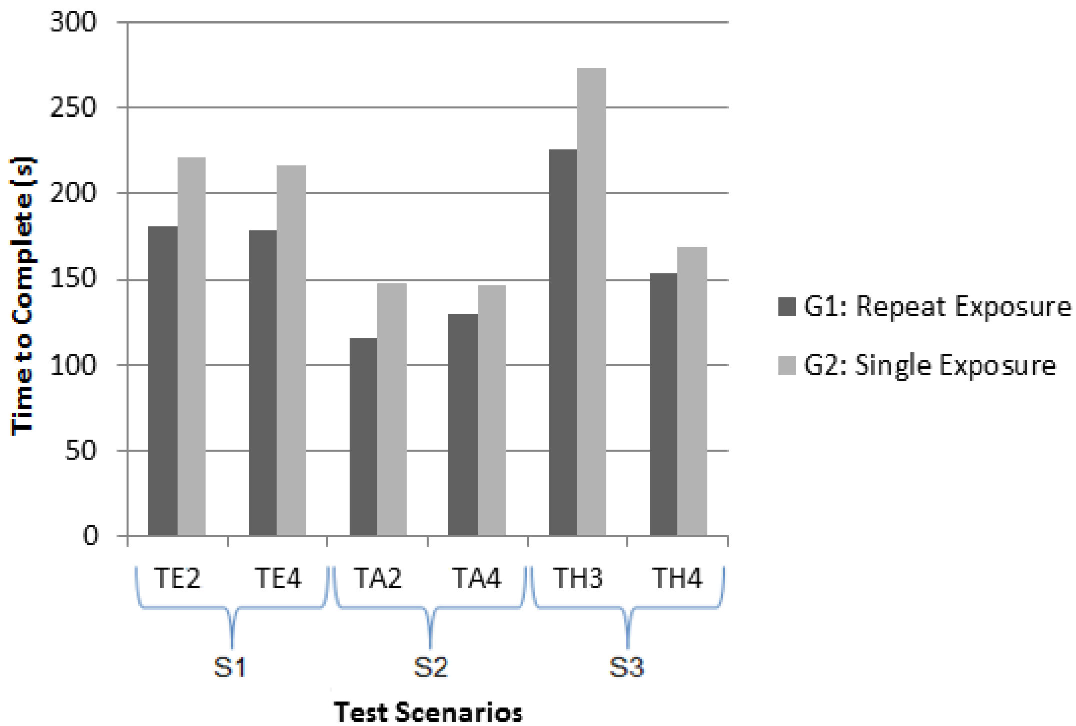

Figure 4-5: Time to complete worksite scenarios across the three sessions.

As indicated in Figure 4-5, group 1 completed the engine room worksite egress test scenarios on average 31.97 seconds faster than group 2 across the three sessions. In two scenarios (TE2 and TH3) the difference between group 1 and group 2 in time to muster is more than 40 seconds. Group 1 is 40.40 seconds faster at completing TE2 ($p = 0.266$) and 47.19 seconds faster at completing TH3 than group 2 ($p = 0.711$).

Figure 4-6 shows the distance travelled to muster for both groups for the engine room worksite scenarios. The same trend for group 1 is shown in distance travelled to muster with the exception of TE4 in which both groups completed the scenario with a mean difference of 3.77 meters. On average, group 1 uses less distance to muster (20.35 meters less than group 2) in five out of the six test scenarios.

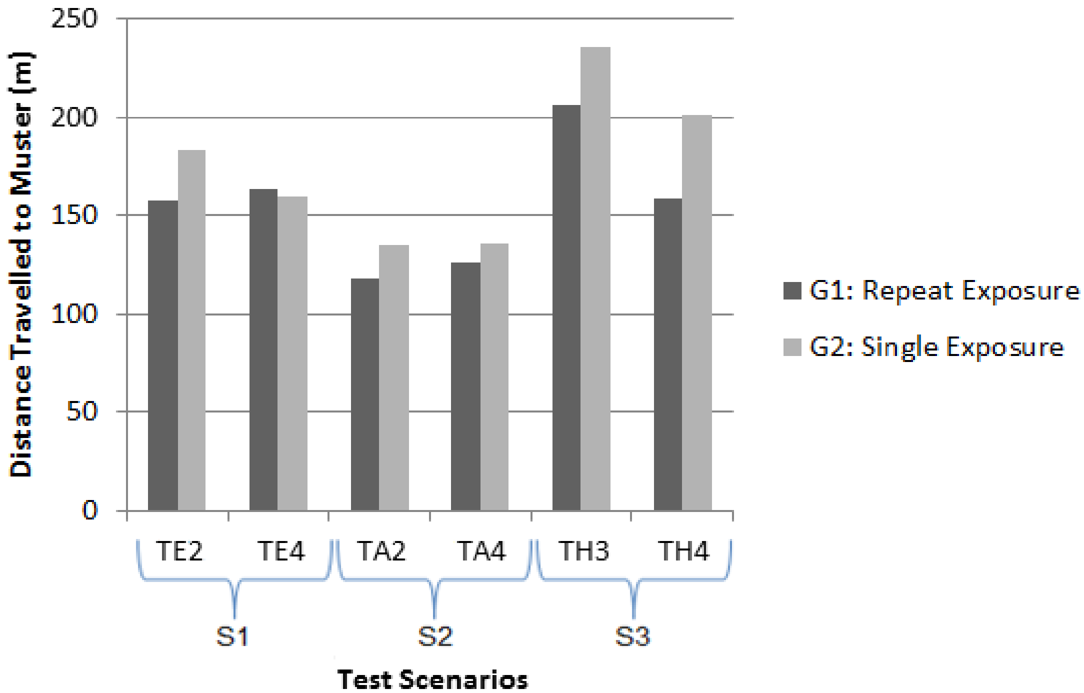

Figure 4-6: Distance travelled in the worksite scenarios across the three sessions.

Table 4-5 shows the task performance by group for time to muster and distance travelled for the engine room worksite egress test scenarios. As indicated in Table 4-5, there was no statistical significance found between groups for the time to muster and distance travelled in the engine room egress scenarios. In TH4, where the average difference in distance travelled between groups is large, this is a result of one participant that became lost in the scenario (creating a statistical outlier). Task performance improvements for both groups across sessions will be discussed in the learning section.

Table 4-5: Time to muster and distance travelled for the engine room egress scenarios.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TE2	Normal	Time to muster	G1 G2	17 19	181 222	29 31	0.266
		Distance	G1 G2	17 19	157 183	18 19	0.416
TE4	Blackout	Time to muster	G1 G2	17 19	179 217	21 21	0.416
		Distance	G1 G2	17 19	164 160	29 12	0.214
TA2	Alarm	Time to muster	G1 G2	17 19	116 148	11 15	0.335
		Distance	G1 G2	17 19	118 135	11 11	0.161
TA4	Alarm + Blackout	Time to muster	G1 G2	17 19	130 147	17 17	0.416
		Distance	G1 G2	17 19	126 136	17 13	0.508
TH3	Alarm + Hazard	Time to muster	G1 G2	17 19	227 274	26 38	0.711
		Distance	G1 G2	17 19	206 236	36 29	0.104
TH4	Alarm + Hazard	Time to muster	G1 G2	17 19	153 169	30 26	0.266
		Distance	G1 G2	17 19	158 202	42 44	0.122

2. Routes and Mapping

Route Selection:

Participants were given five egress routes to learn during the training tutorials: two beginning in the accommodations from their cabin and three starting in the engine room from their worksite. For the accommodations scenarios:

- the primary route was a shorter more direct interior route from the cabin to their muster station or lifeboat station, and
- the secondary route was a less direct exterior route from the cabin to their muster station or lifeboat station. There was also an alternative to the secondary route (known as the tertiary route) that involved taking a longer corridor from the cabin to the exterior route.

For the worksite scenarios:

- the primary route was a direct route through the engine room from the worksite to the accommodation block and muster station or lifeboat station,
- the secondary route was a direct route through the exit ladder from the engine room to the exterior main deck and to the lifeboat station or muster station, and
- the tertiary route was the least direct route through the engine casing to the exterior main deck and to the lifeboat station or muster station.

During the test session, each scenario provided information to help the participant select the most appropriate route to take given the situation. The most appropriate route to take in a scenario changed depending on the circumstance. A successful route selection was one that:

- was any route the participant initially selected and did not deviate from in the first session (to help participants learn the routes early on)
- was the most efficient route (quickest time and shortest distance) to their muster station or lifeboat station,
- the route that did not put the participant in harm's way (i.e. listening to the PA announcement and not following a route that was blocked by a hazard).

Route choices were ranked based on the primary, secondary, tertiary routes available from the cabin and worksite. In sessions 1 and 2, all routes were weighted equally. However, in session 3, route selection and subsequent re-routing are both more important due to the increased scenario complexity of session 3 test scenarios. Session 3 scenarios require participants to avoid hazards and in some cases the hazards could block available routes causing participants to carefully select their route or re-route in the event that they encounter a hazard along their route.

Route selection represents the participant's route choice and whether the participant followed a designated route during the scenario. Route selection was observed and recorded in the experiment observation log by the research team while the participant performed the scenario, or through reviewing the AVERT replay files to confirm any missing values. For reporting purposes, the route selection results are reported in terms of the percentage each route was selected by the groups for each scenario.

Table 4-6 and Table 4-7 summarize the percentage of routes selected by group for the cabin and engine room scenarios, respectively. As indicated in Table 4-6, the route selection for the cabin routes is mainly between the primary or secondary route options. Group 1 selects the primary route from the cabin 69% of the time, the secondary route 27% of the time, the tertiary route 2% of the time, and exhibits lost behaviour 3% of the time across the six scenarios. Group 2 selects the primary route from the cabin 50% of the time, the secondary route 43% of the time, the tertiary route 4% of the time, and exhibits lost behaviour 3% of the time across the six scenarios. Both groups have a percentage of participants that were lost in the first scenario cabin egress scenario (TE1). For the first scenario (TE1) 12% of group 1 and 11 % of group 2 were lost.

In comparing the group proportions using Fisher's Exact Test, there was a trending difference ($p = 0.065$) between group participants in route selection for scenario TA1 in session 2. During the alarm situation in TA1, 88% of group 1 selected the primary route while only 58% of group 2 selected the primary route. Conversely, 12% of group 1 selected the secondary route while 42% of group 2 selected the secondary route. No other statistical difference was found between group participants in route selection for all cabin scenarios across all three sessions.

Table 4-6: Percent route selected by group for the cabin egress scenarios.

Scenario Code	Condition	Group	Percent Route Selected				P-value
			primary	secondary	tertiary	lost	Fisher's Exact Test
TE1	Normal	G1 G2	59 42	24 42	6 5	12 11	.685
TE3	Blackout	G1 G2	65 47	29 37	6 11	0 5	.730
TA1	Alarm	G1 G2	88 58	12 42	0 0	0 0	**.065**
TA3	Alarm and Blackout	G1 G2	59 53	41 42	0 5	0 0	1.000
TH1	Alarm and Hazard	G1 G2	71 53	29 47	0 0	0 0	.322
TH2	Alarm and Hazard	G1 G2	71 47	23 47	0 5	6 0	.165

Session 3 scenarios TH1 and TH2 were designed to be more complex than previous cabin egress scenarios and required participants to select the safest route given the emergency circumstances. The cabin hazard condition scenario TH1 was designed to block the primary muster point due to an adjacent galley fire. Participants who selected the primary egress route ran the risk of exposing themselves to a smoke hazard. The safest route to take given the emergency situation was the secondary exterior route from the cabin to the lifeboat station. As indicated in Table 6 for hazard scenario TH1, 71% of group 1 participants selected the primary route and 24% selected the secondary route from the cabin. For group 2 participants, 53% selected the primary route and 47% selected the secondary route from the cabin.

Similarly, the cabin hazard condition scenario TH2 involved a helideck fire with heavy smoke compromising the exterior egress route from the cabin, which was the secondary egress route. Participants who selected the secondary egress route ran the risk of exposing themselves to a smoke hazard. The safest route to take given the emergency situation was the primary interior route from the cabin to the muster station and lifeboat station. As indicated in Table 6 for hazard scenario TH2, 71% of group 1 participants selected the primary route, 24% selected the secondary route from the cabin and 6% were lost. For group 2 participants, 47% selected the primary route, 47% selected the secondary route and 6% selected the tertiary route from the cabin.

Table 4-7 shows that the route selection for the engine room routes were more distributed across the three route options. Group 1 selected the primary route from the worksite 64% of the time, the secondary route 19% of the time, and the tertiary route 8% of the time across the six scenarios. Group 2 selected the primary route from the worksite 58% of the time, the secondary route 22% of the time, and the tertiary route 4% of the time across the six scenarios. On average, 10% of group 1 participants and 16% of group 2 participants were lost in the engine room scenarios across the three sessions. Participants exhibited behaviours of being lost more frequently in the worksite scenarios. Unlike the cabin scenarios, in all the worksite scenarios there were lost participants in both groups. The highest percentage of lost participants occurred in scenario TH3 where 18% of group 1 and 32% of group 2 were lost. This could be attributed to the TH3 scenario design in which the primary egress route to the engine room was blocked by heavy smoke.

In comparing the group proportions using Fisher's Exact Test, there was no statistical difference found between group participants in route selection for all engine room egress scenarios across all three sessions.

Table 4-7: Percent route selected by group for the engine room egress scenarios.

Scenario Code	Condition	Group	Percent Route Selected				P-value
			primary	secondary	tertiary	lost	Fisher's Exact Test
TE2	Normal	G1 G2	71 53	12 21	6 5	12 21	.772
TE4	Blackout	G1 G2	53 68	24 26	18 0	6 5	.329
TA2	Alarm	G1 G2	71 58	12 32	6 5	12 5	.602
TA4	Alarm and Blackout	G1 G2	76 63	12 16	6 11	6 11	1.000
TH3	Alarm, Hazard and Blackout	G1 G2	29 37	47 32	6 0	18 32	.551
TH4	Alarm, Hazard And Blackout	G1 G2	82 68	6 5	6 5	6 21	.826

Similarly to the cabin egress scenarios of session 3, the engine room egress scenarios TH3 and TH4 were designed to be more complex than previous engine room egress scenarios and required participants to select the safest route given the emergency circumstances.

The engine room hazard condition scenario TH3 was designed to block the primary egress route from the worksite. Participants who selected the primary egress route ran the risk of exposing themselves to heavy smoke and fire hazards. The safest route to take given the emergency situation was the secondary or tertiary routes from the worksite to the lifeboat station. As indicated in Table 4-7 for hazard scenario TH3, 29% of group 1 participants selected the primary route, 47% selected the secondary route, 6% selected the tertiary route from the worksite in the engine room, and 18% of group 1 participants were lost. For group 2 participants, 37% selected the primary route, 32%

selected the secondary route from the worksite in the engine room, and 32% of group 2 participants were lost.

For the engine room hazard condition scenario TH4, there was an engine room fire and explosion that compromised the secondary and tertiary egress route from the engine room. Participants who selected the secondary or tertiary egress routes ran the risk of severe burns or death due to fire and smoke exposure. The safest route to take given the emergency situation was the primary egress route from the worksite to the accommodations to the lifeboat station. As indicated in Table 4-7 for hazard scenario TH4, 82% of group 1 participants selected the primary route, 6% selected the secondary route, and 6% selected the tertiary route from the worksite in the engine room. For group 2 participants, 68% selected the primary route, 5% selected the secondary, and 5% selected the tertiary route from the worksite in the engine room. 6% of group 1 participants and 21% of group 2 participants were lost in the final engine room test scenario, TH4.

Route Deviations (Off-Route):

If the participant deviated from the designated route, an off-route performance measure was also recorded. This measure identified how well the participants followed the designated egress routes they learned in the training tutorials. Only major route deviations were regarded as the participant going off route. Major route deviations included behaviours where the participants exhibited lost behaviour, did not follow any particular route for the situation, clearly made a wrong turn or detour, or arrived at the wrong muster location. Behaviours including briefly going in the wrong direction, or opening the wrong door, that were self-corrected quickly were excluded from the off-route count. Table 4-8 shows the percentage of participants who remained on route and the percentage of participants who deviated from the designated egress routes across the three sessions.

Table 4-8: Group proportions for route deviations in all scenarios.

Scenario Code	**Group**	**Off Route Count**		**Performance Percentage**		**P-value**
		Correctly Follow Route	**Off Route**	**% Correct**	**% Incorrect**	**Fisher's Exact Test Exact Sig. (2-sided)**
TE1	G1	12	5	71	29	1.000
	G2	14	5	74	26	
TE2	G1	13	4	76	24	.302
	G2	11	8	58	42	
TE3	G1	12	5	71	29	.219
	G2	17	2	89	11	
TE4	G1	11	6	65	35	.742
	G2	11	8	58	42	
TA1	G1	15	2	88	12	1.000
	G2	17	2	89	11	
TA2	G1	15	2	88	12	.236
	G2	13	6	68	32	
TA3	G1	13	4	76	24	.391
	G2	17	2	89	11	
TA4	G1	15	2	88	12	.408
	G2	14	5	74	26	
TH1	G1	12	5	71	29	.732
	G2	12	7	63	37	
TH2	G1	15	2	88	12	1.000
	G2	16	3	84	16	
TH3	G1	11	6	65	35	.202
	G2	8	11	42	58	
TH4	G1	15	2	88	12	.236
	G2	13	6	68	32	

Backtracking Time and Distance:

Two complementary performance measures of Route Selection used to assess the participant's understanding of the Routes and Mapping learning objective were backtracking time and backtracking distance. Once a participant selected an egress route, the backtracking performance measure was an assessment of how well the participant could follow the chosen egress route. The participants' backtracking time and distance during the scenario were calculated using the AVERT report file and a post-experiment backtracking algorithm. The backtracking algorithm involved the avatar's geometric positioning data (x, y, z) output by the AVERT report file and a series of horizontal and vertical position point checks. The AVERT report file records the avatar's position (x, y, z) every second of the test scenario. The recorded position points were compared sequentially using the algorithm parameters to determine if backtracking had occurred in the scenario. The parameters used for the algorithm were a horizontal inclusion range of 1.5m, a vertical inclusion range 1.2m and an exclusion window of 4 position points. These parameters were selected and refined through previous studies. As a reference, Bradbury-Squires (2013) used 4.0m, 2.0m, and 4 position points as the parameters for the backtracking algorithm. The final backtracking parameters chosen were considered an acceptable capture of intentional backtracking by a participant.

The horizontal (1.5m) and vertical (1.2m) ranges were used as two individual inclusion window checks to determine if any of the avatar's preceding position points were close enough to the present position point to conclude that backtracking had occurred. If the participant returned to an area already traversed, then the previous position points would cross the horizontal and vertical inclusion range and this action would be recorded as backtracking. The four position point exclusion window exempted the four preceding points from the backtracking analysis and allowed the participants to have a small window to assess the next available route option without it being counted as backtracking. Thus, any movement within the horizontal and vertical range surrounding the avatar was exempt from backtracking calculation for the duration of the exclusion window. The final backtracking time and backtracking distance reported are a cumulative total of backtracking that took place during each test scenario.

Specifically, a participant that concisely followed the designated egress route resulted in little to no backtracking time and backtracking distance. Participants that were uncertain of the egress route and spent time retracing their steps resulted in a larger backtracking time and distance. Larger backtracking time and backtracking distance indicated that the participant was either lost, inefficient at finding her/his way around their environment, or encountered a hazard along her/his selected route and re-routed.

As participants were required to demonstrate they could find their muster points as quickly as possible from two main locations on the platform, the results for backtracking are reported based on the two locations: the cabin and the worksite in the engine room.

Cabin Egress Scenarios:

Figure 4-7 and Figure 4-8 show the backtracking time and backtracking distance by group for the cabin egress test scenarios in AVERT. On average, group 1 in most cases experienced less backtracking time than group 2 in the cabin scenarios. In three scenarios the difference between group backtracking times is greater than 10 seconds (TE1, TA1 and TH1). As indicated in Figure 4-8, there was less between groups difference experienced in backtracking distance with the exception of scenario TH1. In particular, for the cabin hazard scenario (TH1), group 2 spent an additional 27.74 seconds in backtracking time and 10.06 meters of backtracking distance than group 1.

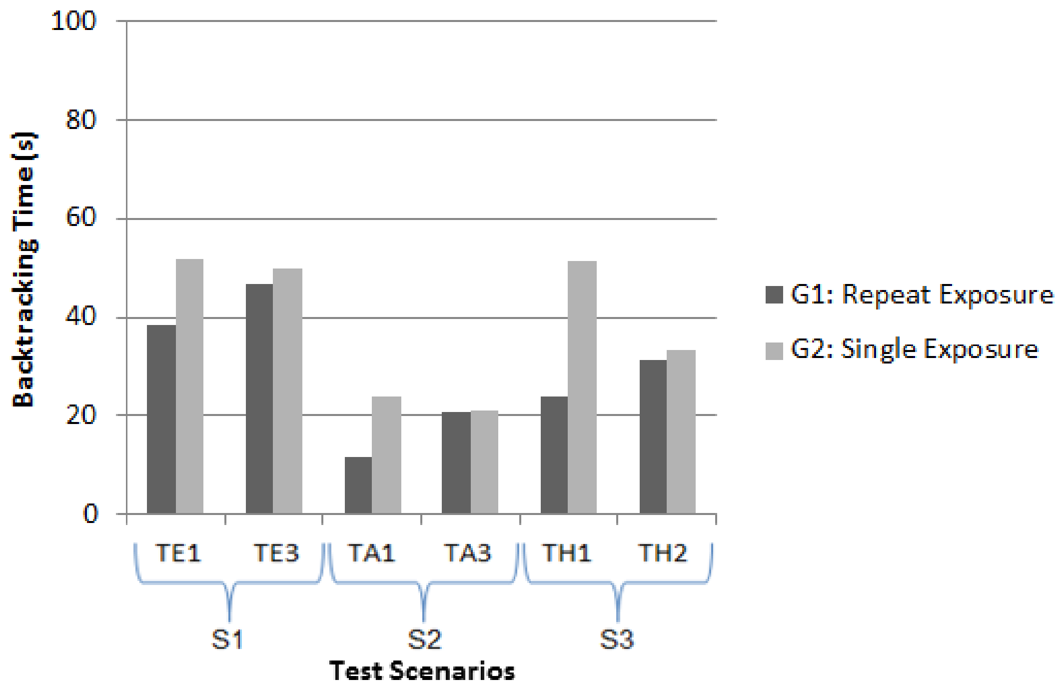

Figure 4-7: Backtracking time for cabin egress scenarios across the three sessions.

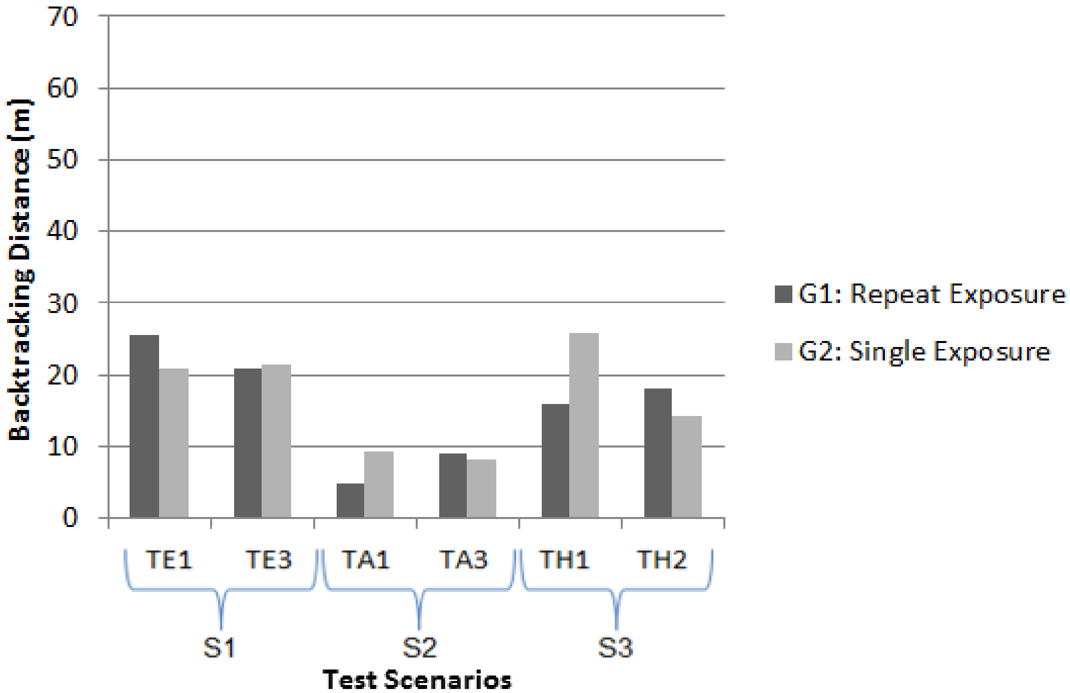

Figure 4-8: Backtracking distance for cabin egress scenarios across the three sessions.

Table 4-9 shows the backtracking time and distance by group for the cabin egress test scenarios. In two scenarios, the between group difference is statistically significant. In the cabin alarm scenario in session 2 (denoted as scenario code TA1) there was a significant difference between groups for both backtracking time ($p = 0.004$) and distance ($p = 0.006$). In the cabin hazard scenario in session 3 (TH1) there was a significant difference between groups for backtracking time ($p = 0.041$). No other statistical significance was found for backtracking time and distance in the cabin egress scenarios.

Table 4-9: Backtracking time and distance by group for cabin egress scenarios.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TE1	Normal	BT time	G1 G2	17 19	38 52	11 12	0.186
		BT distance	G1 G2	17 19	25 21	12 6	0.960
TE3	Blackout	BT time	G1 G2	17 19	47 50	8 14	0.397
		BT distance	G1 G2	17 19	21 21	4 8	0.158
TA1	Alarm	BT time	G1 G2	17 19	12 24	4 4	**0.004**
		BT distance	G1 G2	17 19	5 9	2 2	**0.006**
TA3	Alarm and Blackout	BT time	G1 G2	17 19	21 21	5 5	0.711
		BT distance	G1 G2	17 19	9 8	2 2	0.378
TH1	Alarm and Hazard	BT time	G1 G2	17 19	24 52	4 12	**0.041**
		BT distance	G1 G2	17 19	16 26	3 7	0.786
TH2	Alarm and Hazard	BT time	G1 G2	17 19	31 33	8 9	0.995
		BT distance	G1 G2	17 19	18 14	7 4	0.960

Engine Room Egress Scenarios:

Figure 4-9 and Figure 4-10 show the backtracking time and backtracking distance by group for the engine room worksite test scenarios in AVERT. In most cases, group 1 on average experienced less backtracking time than group 2 in the engine room scenarios. In three scenarios the difference between group backtracking times is greater than 15 seconds (TE2, TE4 and TH3). In particular, for TE2, group 2 spent an additional 17.00 seconds; for TE4, group 2 spent an additional 21.04 seconds; and for TH3, group 2 spends an additional 22.20 seconds in backtracking time than group 1. As indicated in

Figure 4-10, there was less between groups difference experienced in backtracking distance.

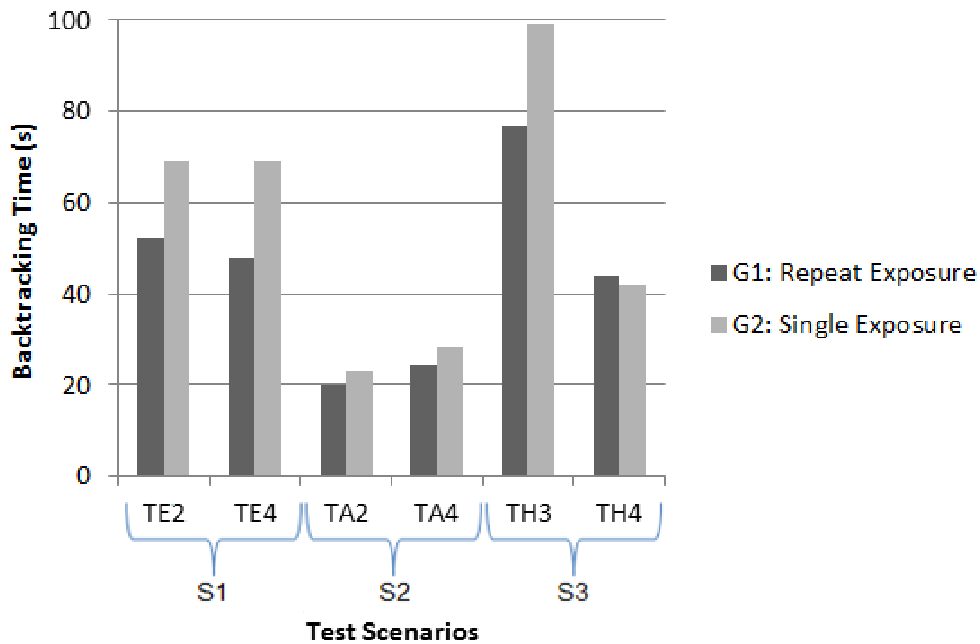

Figure 4-9: Backtracking time for worksite scenarios across the three sessions.

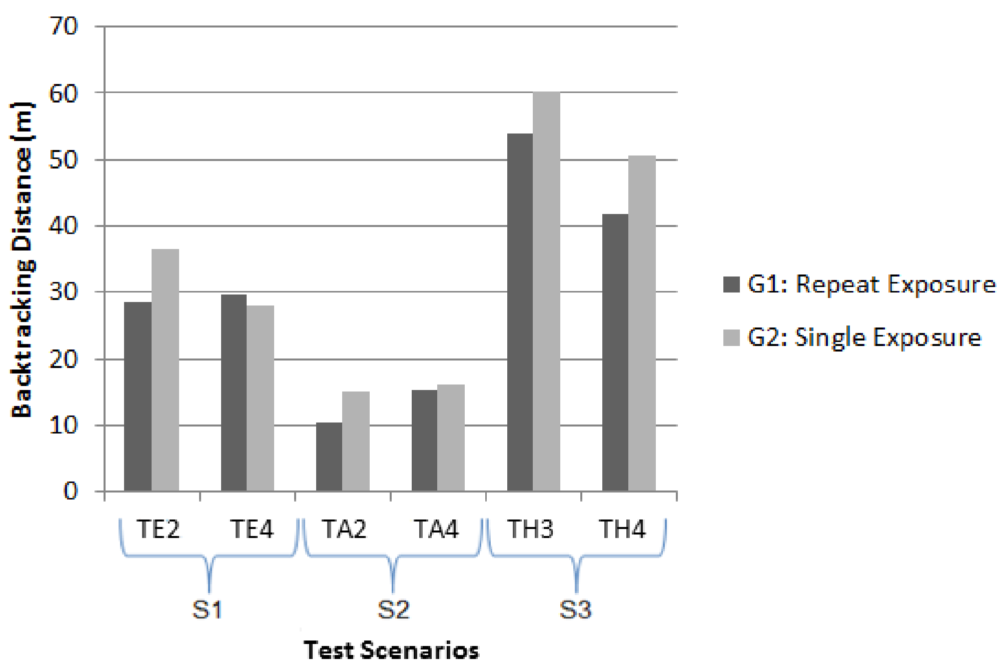

Figure 4-10: Backtracking distance for worksite scenarios across the three sessions.

Table 4-10 shows the backtracking time and distance by group for the engine room test scenarios. There was no statistical significance found for the backtracking task performance in the engine room egress scenarios. However, in two scenarios (TE2 and TA2) there is a trending difference between groups for backtracking distance.

Table 4-10: Backtracking time and distance by group for the engine room scenarios.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TE2	Normal	BT time	G1 G2	17 19	52 69	17 22	0.359
		BT distance	G1 G2	17 19	29 37	10 9	**0.095**
TE4	Blackout	BT time	G1 G2	17 19	48 69	9 11	0.397
		BT distance	G1 G2	17 19	30 28	11 5	0.601
TA2	Alarm	BT time	G1 G2	17 19	20 23	6 4	0.229
		BT distance	G1 G2	17 19	10 15	4 4	**0.086**
TA4	Alarm and Blackout	BT time	G1 G2	17 19	24 28	9 9	0.565
		BT distance	G1 G2	17 19	15 16	8 5	0.508
TH3	Alarm, Hazard and Blackout	BT time	G1 G2	17 19	77 99	19 22	0.953
		BT distance	G1 G2	17 19	54 60	20 14	0.359
TH4	Alarm, Hazard and Blackout	BT time	G1 G2	17 19	44 42	24 17	0.335
		BT distance	G1 G2	17 19	42 51	31 28	0.186

3. Continually Assessing Emergency Situation & Hazard Avoidance

There are three performance measures for the Assessing Situation and Hazard Avoidance learning objective: Raise the alarm, Re-route in event of hazard or alarm change, and Avoid Exposure to Hazards (includes time spent in smoke, time spent in contact with fire and incurred injury due to fire exposure).

<u>Raising the Alarm:</u>

The first part of the Assessing the Situation objective was to be vigilant about the situation and know where and how to raise the alarm in the event of an emergency situation. Scenario TH4 tested the participants' ability to assess a fire situation in the engine room and raise the alarm in a limited time period.

To be successful at raising the alarm the participants would need to do the following:

- see the fire at their worksite,
- go to the correct manual alarm call point (MAC),
- interact with the MAC and enable the alarm within a time limit of 20 seconds.

Other indicators of performance were tracked, including noting participants that went to the MAC location but were unable to engage the alarm in time, and participants that mistook an incorrect location for the MAC station.

Table 4-11 shows task performance measures of raising the alarm by group. Only one participant (from group 1) was able to successfully raise the alarm. Four participants, two in each group, attempted to raise the alarm but were unable to successfully raise the alarm within the time limit for the scenario. Three other participants, all from group 2, tried to raise the alarm but were not at a proper MAC station necessary to raise the alarm.

In comparing the proportions for both groups, no statistical significance was found between groups in raising the alarm in the final test scenario (TH4).

Table 4-11: Comparing group proportions for raising the alarm (correct interaction with MAC).

Variable	Group	Raising Alarm Count		Raising Alarm Percentage		P-value
		Correct	Incorrect	% Correct	% Incorrect	Fisher's Exact Test
TH4 Correctly Raise Alarm	G1	1	16	6	94	.472
	G2	0	19	0	100	
TH4 Attempted to Raise Alarm (at the right location)	G1	2	15	12	88	1.000
	G2	2	17	11	89	
*TH4 effort to Raise Alarm (but at wrong location)	G1	0	17	0	100	.231
	G2	*3	16	*16	84	

Re-routing:

The second part of the Assessing the Situation objectives was to select a route to avoid hazards or re-route in the event of encountering a hazard along the chosen path. Session 3 introduced hazard situations resulting in more complex scenarios than the other sessions. The increased complexity made the subtask of route selection and re-routing more important for session 3. Route selection in the previous section (Routes and Mapping) depicted the percentage of participants that selected primary, secondary and tertiary routes for each of the hazard situations. If the participant selected a route that was compromised by a hazard they were required to re-route. Similarly to the route selection data, the re-routing data was collected through experimenter review of the AVERT replay files. Table 4-12 and Table 4-13 summarize the percentage of re-routing by group for the cabin and engine room scenarios, respectively.

The re-routing analysis looked at what the participants did when encountering a hazard. Three re-routing behaviours were observed in the hazard scenarios:

- Participants whose route choice was a clear unobstructed route did not require re-routing (denoted as N/A for not applicable),
- Participants whose first route choice was a compromised route requiring them to re-route and they did so safely (denoted as successful re-routed),
- Participants whose first route choice was a compromised route requiring them to re-route and they chose not to re-route (denoted as no-re-routing).

Table 4-12: Percent re-routing by group for the cabin egress scenarios.

Scenario Code	**Condition**	**Group**	**Re-Routing**			**P-value**
			No re-routing required (n/a)	**Re-routing Required**		**Fisher's Exact Test (2-sided)**
				Successfully Re-routed (y)	**No re-routing (n)**	
TH1	Alarm, Hazard and Blackout	G1 G2	23 47	65 42	12 11	.277
TH2	Alarm, Hazard And Blackout	G1 G2	71 47	23 21	6 32	.177

Cabin Egress Scenarios:

Table 4-12 summarizes the re-routing for the cabin scenarios TH1 and TH2. For scenario TH1, the secondary route to the lifeboat station was unobstructed. Taking the secondary route resulted in no re-routing required. The primary route was compromised by smoke. Thus taking the primary route caused participants to either recognize or encounter the smoke hazard and re-route to a safer egress path. In scenario TH1, 29% of group 1 and 47% of group 2 selected the secondary route and therefore did not require re-routing. However, 23% of group 1 did not require re-routing. This percentage difference is due to one participant (6%) taking the safer route choice but not mustering at the lifeboat station and instead going through the lifeboat station and proceeding through the smoke filled muster station to muster at the mess hall instead of lifeboat station.

In TH1, 71% of group 1 selected the primary route and 53% of group 2 followed the primary route to the lifeboat station. In group 1, of the 71% that selected the primary route, 65% re-routed and 6% did not re-route and went directly through the smoke. In group 2, of the 53% that selected the primary route, 42% re-routed and 11% did not re-route and went directly through the smoke.

For scenario TH2 the primary route to the lifeboat station was clear of hazards. Taking the primary route to the lifeboat would result in no re-routing required. The secondary and tertiary routes were compromised by smoke in the outside stairwell. Taking the secondary or tertiary route would require the participants to either recognize or encounter the smoke hazard and re-route to a safer egress path. In this scenario, 71% of group 1 selected the primary route and 47% of group 2 selected the primary route and therefore did not require re-routing in TH2.

In TH2, 6% of group 1 were lost in the scenario and 23% of group 1 selected the secondary route where they encountered the smoke hazard and were required to re-route. For group 2, 47% followed the secondary route and were required to re-route. Of the 29% of group 1 that required re-routing (selected the secondary route or exhibited lost behaviour), 23% (that selected the secondary route) successfully re-routed while one (6%) did not re-route and went directly through the smoke. The lost individual attempted to re-route. One of the participants that selected the secondary route did not reroute. Of the 53% of group 2 that followed the secondary route, 21% re-routed and 32% did not re-route and went directly through the smoke.

Engine Room Egress Scenarios:

Table 4-13 summarizes the re-routing for the engine room scenarios TH3 and TH4. For scenario TH3, the secondary and tertiary routes to the lifeboat station were clear of hazards. Taking the secondary or tertiary route to the lifeboat resulted in little to no re-routing required. The primary route was compromised by smoke and fire, requiring the participants to re-route to a safer egress path. For group 1, 29% selected the primary route, 47 % selected the secondary, 6% selected the tertiary and 18% were lost. For group 2, 37% selected the primary route, 32 % selected the secondary, 0% selected the tertiary and 32% were lost. For group 1 the 47 % that selected the secondary route and the 6% that selected the tertiary route did not require re-routing (making up 53% of group 1 that did not require re-routing). Those that selected the primary route (29%) or were lost (18%) were required to re-route in TH3 scenario. All but one participant re-routed when required in group 1. Similarly, for group 2 the 32 % that selected the secondary route did not require re-routing. Those that selected the primary route (37%) or were lost (32%) were required to re-route in TH3 scenario (totaling 69% requiring re-routing). All but one group 2 participant re-routed when required.

Table 4-13: Percent re-routing by group for the engine room egress scenarios.

Scenario Code	**Condition**	**Group**	**Percent Route Selected**			**P-value**
			No re-routing required (n/a)	**Re-routing Required**		**Fisher's Exact Test (2-sided)**
				Re-routed (y)	**No Re-routing (n)**	
TH3	Alarm, Hazard and Blackout	G1 G2	53 32	41 63	6 5	.311
TH4	Alarm, Hazard And Blackout	G1 G2	82 74	6 21	12 5	.493

For scenario TH4, the primary route to the lifeboat station was unobstructed. Taking the primary route resulted in little to no re-routing required (depending on how well they knew the route). The secondary and tertiary routes were blocked by an explosion, as well as fire and smoke. Taking the secondary and tertiary routes, which were compromised by smoke and fire, required participants to either recognize or encounter the hazards and re-route to a safer egress path. For group 1, 82% selected the primary route, 6% selected the secondary, 6% selected the tertiary and 6% were lost. The 82% of group 1 that selected the primary route did not require re-routing. Of the participants in group 1 that selected the secondary route, tertiary route or were lost in the

TH4 scenario, 6% re-routed and 12% did not re-route and were exposed to fire and smoke. For group 2, 68% selected the primary route, 5 % selected the secondary, 5% selected the tertiary and 21% were lost. The 68% of group 2 that selected the primary route and one participant that was lost (6%) did not require re-routing (totaling the 74% of group 2 that did not require re-routing. The participant that was lost in this case successfully navigated out of the engine room using the primary egress route but became lost trying to find the correct muster station within the temporary safe refuge area. Of the participants in group 2 that selected the secondary, tertiary or were lost in the TH4 scenario, 21% re-routed and 5% did not re-route and were exposed to fire and smoke.

<u>Avoid Exposure to the Hazards (Smoke and Fire) and Injury Incurred:</u>

The final part of the Assessing the Situation objectives was to avoid exposure to hazards. AVERT has built-in empirical mathematical models of hazards such as smoke, fire, and explosions (Pula et al., 2005; Assael and Kakosimos, 2010; Woodward and Pitblado, 2010; Babrauskas, 1983; Burgess and Herzberg, 1974). Risk of hazard exposure and severity of injury are estimated in the AVERT based on the avatar's proximity and duration of exposure to the hazards (House et al. 2014). If a participant spent too long in smoke or got too close to the fire, AVERT recorded the time spent exposed to the hazard and calculated the probability of burns and death.

All hazard scenarios were in session 3 (TH1, TH2, TH3, & TH4). The AVERT report file recorded the time spent exposed to smoke and fire in scenarios. This metric is a cumulative total of the time in seconds that the participant spent in the smoke or fire. Results for the exposure to smoke and fire by group are reported in Table 4-14.

Along with the time spent in contact with smoke and fire, AVERT also calculated and recorded the probability of burns and death due to fire exposure. The only scenario where participants were exposed to fire was TH4. Results for the injury incurred by fire exposure by group are reported in Table 4-15.

As indicated in Table 4-14, there was a trending difference (p = 0.051) between the groups in exposure to smoke for scenario TH2. Group 2 was exposed to a mean of 2.49 seconds more smoke in TH2 than group 1. No other statistical significance was found between groups in hazard exposure. Scenario TH3 appears to have a significant difference in mean time spent in smoke, however, this is entirely due to one outlier that received a large exposure to smoke in the scenario.

Table 4-14: Time spent exposed to smoke and fire hazards by group.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TH1	Cabin Hazard	Time Exposed to Smoke	G1 G2	17 19	1.36 0.66	0.53 0.34	0.402
TH2	Cabin Hazard	Time Exposed to Smoke	G1 G2	17 19	0.83 3.32	0.52 0.97	**0.051**
TH3	Engine Room Hazard & Blackout	Time Exposed to Smoke	G1 G2	17 19	0.05 2.21	0.05 2.21	1.000
TH4	Engine Room Hazard & Blackout	Time Exposed to Smoke	G1 G2	17 19	0.32 0.65	0.32 0.43	0.737
		% Fire Exposure	G1 G2	17 19	2.71 2.57	2.05 1.14	0.645

Table 4-15 shows the mean probability of injury (first degree burns) and probability of death due to fire exposure for both groups in the final scenario TH4. Group

2 experienced an increased probability of injury and death. However, no statistical significance was found between groups in probability of injury or death due to fire exposure. This difference in probability of death in TH4 is due to one outlier who received a large cumulative time exposed to the fire hazard.

Table 4-15: Probability of injury and death due to fire exposure in scenario TH4.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TH4	Engine Room Hazard & Blackout	Prob of Injury (Burns)	G1 G2	17 19	9.16 16.40	5.46 7.03	0.615
		Prob of Death	G1 G2	17 19	0.04 4.42	0.03 4.38	0.615

4. Safe Practices

Percentage Running:

Part of the general safe work practices is that participants understand the importance of not running on the platform and the risk associated with running in an emergency situation. To successfully fulfill this requirement, the participants were required to walk during all scenarios. The time spent running was recorded in the AVERT report file and subsequently was normalized with the time spent to complete the scenario, resulting in a percent time spent running in the scenario. To be successful in the scenario participants must spend 0% time running.

Table 4-16 reports the mean percentage of time spent running for both groups during each scenario. In most scenarios, both group participants' percentage running is greater than 10% with the exception of TH2. There is only one instance where either group runs less than 10% of the scenario time. Group 1 in scenario TH2 ran on average 7% of the scenario time.

The overall mean percentage time running across all sessions for group 1 was 14% and for group 2 was 15%. In two scenarios (TH2 and TH4) group 2 runs 5% or more time than group 1. As indicated in Table 16, the difference between groups in percentage time running is trending for those two scenarios (TH2 and TH4).

In the cabin hazard scenario (TH2) there was a trending difference ($p = 0.086$) between group 1 and group 2 in percentage time running. Group 2 ran 5% more than group 1 in the cabin hazard scenario TH2. Likewise, in the engine room hazard scenario (TH4), there was a trending difference ($p = 0.076$) between group 1 and group 2 in percentage time running. Group 2 ran 8% more than group 1 in the engine room hazard scenario. No other statistical significance was found between groups in percentage time running.

Table 4-16: Percent running that occurred in each scenario for both groups.

Scenario Code	Condition	Variable	Group	N	Mean	SE	p - value
TE1	Normal	Percent Running	G1 G2	17 19	15 13	3 3	0.870
TE2	Normal	Percent Running	G1 G2	17 19	15 17	2 3	0.596
TE3	Blackout	Percent Running	G1 G2	17 19	11 14	3 3	0.397
TE4	Blackout	Percent Running	G1 G2	17 19	16 15	3 3	0.960
TA1	Alarm	Percent Running	G1 G2	17 19	11 12	2 2	0.768
TA2	Alarm	Percent Running	G1 G2	17 19	17 18	2 2	0.693
TA3	Alarm and Blackout	Percent Running	G1 G2	17 19	14 14	3 2	0.647
TA4	Alarm and Blackout	Percent Running	G1 G2	17 19	15 17	2 2	0.693
TH1	Alarm and Hazard	Percent Running	G1 G2	17 19	14 15	3 2	0.436
TH2	Alarm and Hazard	Percent Running	G1 G2	17 19	7 13	2 2	**0.086**
TH3	Alarm, Hazard and Blackout	Percent Running	G1 G2	17 19	16 17	2 2	0.740
TH4	Alarm, Hazard and Blackout	Percent Running	G1 G2	17 19	13 21	2 3	**0.076**

Closing Fire and Watertight Doors:

The second part of the general safe work practices is for participants to understand the importance of closing fire and watertight doors and the risk associated with leaving them open. To successfully fulfill this requirement, the participant must close all fire and watertight doors while evacuating the platform. If a participant leaves one or more fire or watertight doors open, then this is considered a fail for this task.

Table 4-17 shows the number of participants in both groups that correctly closed all important doors. There were only four instances where 70% or more of the group participants were able to successfully close all the fire and watertight doors. For group 1, 71% of the group participants were able to close all doors in scenarios TH3 and TH4. For group 2, 79% of the group participants were able to close all doors in scenarios TA4 and TH2. The lowest percentage in closing doors was the first scenario in session 1 (TE1), where only 35% of group 1 and 21% of group 2 were successful in closing all fire and watertight doors. The highest percentage in closing doors in any scenario was 79% and this was obtained by group 2 in scenarios TA4 and TH2.

Table 4-17: Comparing group proportions for the closing doors task.

Scenario Code	Group	Door Count		Door Closing Percentage		P-value
		Correct	Incorrect	% Correct	% Incorrect	Fisher's Exact Test
TE1	G1	6	11	35	65	.463
	G2	4	15	21	79	
TE2	G1	10	7	59	41	1.000
	G2	11	8	58	42	
TE3	G1	8	9	47	53	1.000
	G2	8	11	42	58	
TE4	G1	10	7	59	41	.525
	G2	9	10	47	53	
TA1	G1	10	7	59	41	.749
	G2	10	9	53	47	
TA2	G1	11	6	65	35	1.000
	G2	12	7	63	37	
TA3	G1	7	10	41	59	.525
	G2	10	9	53	47	
TA4	G1	9	8	53	47	.158
	G2	15	4	79	21	
TH1	G1	11	6	65	35	.335
	G2	9	10	47	53	
TH2	G1	8	9	47	53	**.082**
	G2	15	4	79	21	
TH3	G1	12	5	71	29	.322
	G2	10	9	53	47	
TH4	G1	12	5	71	29	1.000
	G2	13	6	68	32	

In comparing the group proportions using Fisher's Exact Test, a trending difference of 32% ($p = 0.082$) was found between the two groups regarding door closing for the cabin hazard scenario (TH2). In this scenario, 47% of group 1 was successful in closing all doors; 79% of group 2 was successful in closing all doors. No other statistical significance was found between groups in closing doors for all other scenarios.

4.1.1.3. Quiz performance comparison between groups

The AVERT training program targeted two knowledge dimensions: the understanding and application of spatial and procedural knowledge. The procedural knowledge was assessed based on the participant's understanding (factual or conceptual knowledge) and his/her procedural application. The participant's understanding of procedural knowledge was assessed using multiple-choice questions. The application of procedural knowledge was assessed using the AVERT scenarios. Two measures of procedural knowledge were collected to compare the participant's understanding of the procedures with their application of the procedures using their task performance in AVERT.

An aggregated score was developed for the multiple-choice questions. See Appendix M for a list of the multiple choice quiz questions. The quiz questions were broken down into questions relating to specific learning objectives, such as spatial understanding, alarm recognition, and assessing the situation, for example. Table 4-18 shows the quiz scores by group for the three sessions.

Table 4-18: Quiz results for each group across the three sessions.

Session	Group	N	Mean	Std. Dev	Minimum	Maximum	P-value
1	G1	17	86	13	54	99	**0.099**
	G2	19	81	9	64	93	
2	G1	17	82	8	65	94	**0.086**
	G2	19	75	10	51	92	
3	G1	17	85	11	62	99	0.442
	G2	19	79	12	50	94	

As indicated in Table 4-18, group 1 achieved a higher mean quiz score than group 2 in all three sessions. On average across the three sessions, group 1 scored an overall mean of 84% and group 2 scored an overall mean of 79% on the quizzes. On average, group 1 scored 6% higher on the quizzes than group 2. Overall, there was a trending difference between group scores for quiz 1 (5%, $p = 0.099$) and quiz 2 (7%, $p = 0.086$). There was no statistical difference between scores for quiz 3.

In particular for session 1, group 1 received a mean quiz score of 86% and group 2 received a mean quiz score of 81%. This suggests that the practice scenarios group 1 received solidified their understanding of the procedural tasks and helped them achieve a higher quiz score. For session 2, group 1 received a mean quiz score of 82% and group 2 received a mean quiz score of 75% (trending difference). For session 3, group 1 received a mean quiz score of 85% and group 2 received a mean quiz score of 79%. This indicates that the refresher training received by group 1 helped participants correctly answer questions in the second and third quizzes.

Breakdown of Quiz Scores by Session:

Figure 4-11, Figure 4-12 and Figure 4-13 show the histograms of participants' quiz scores by group across the three sessions. The proportions were broken down into participants that received less than 60%, $60 \leq n < 75$, $75 \leq n < 90$ and $90\% \leq n$.

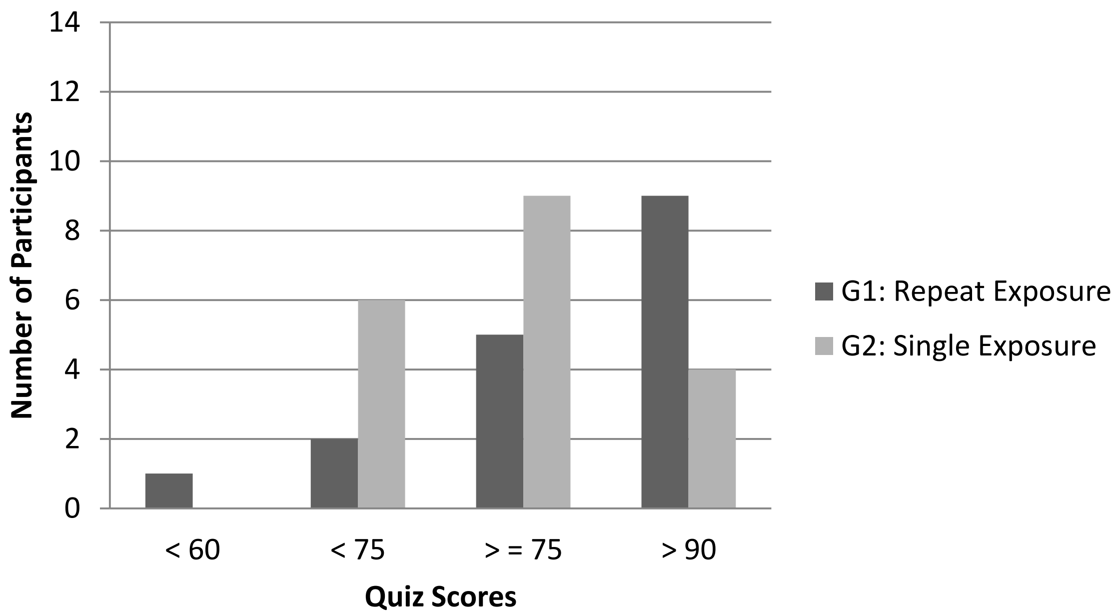

Figure 4-11: Session 1 quiz results for each group.

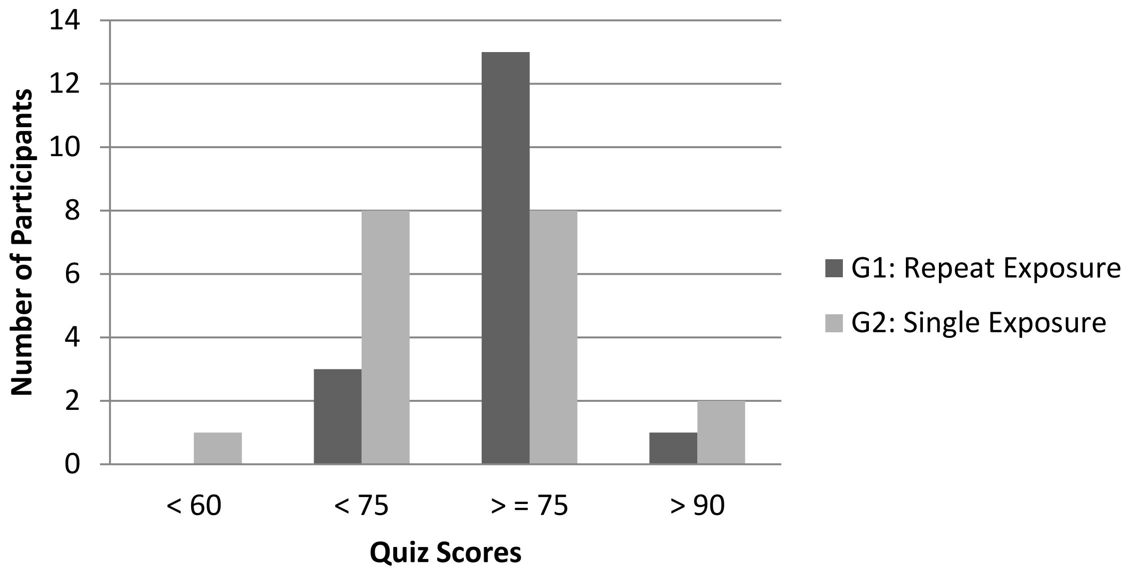

Figure 4-12: Session 2 quiz results for each group.

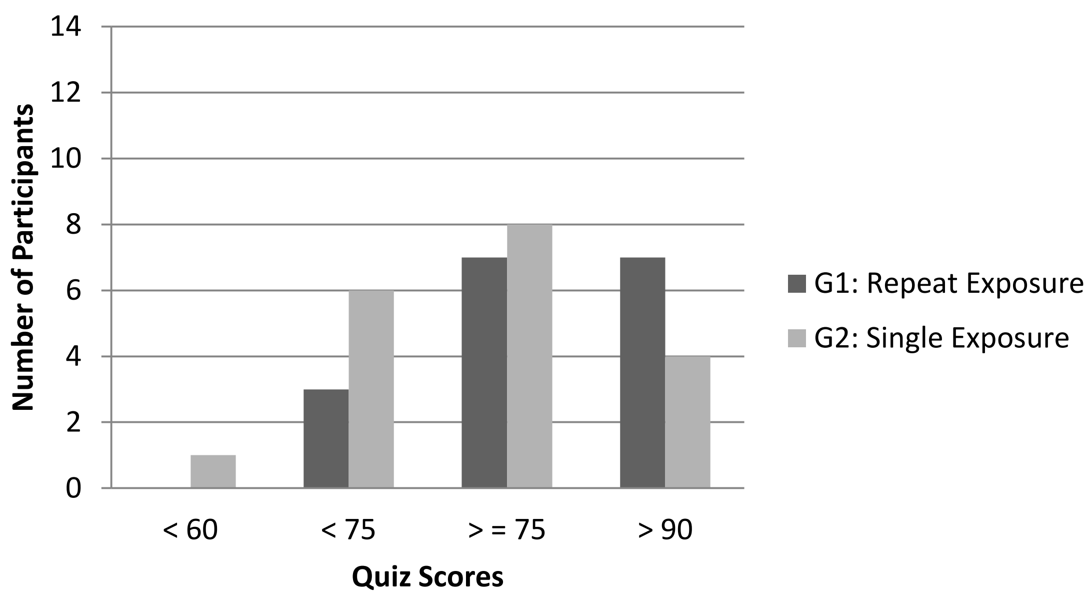

Figure 4-13: Session 3 quiz results for each group.

As indicated in Figure 4-11, in session 1 the mean quiz score for group 1 was 86% and the mean quiz score for group 2 was 81%. Although group 1 had a greater number of participants achieving 90 or above, this group also had the greatest spread of quiz scores.

Looking at specific questions, on average the questions in session 1 were within 4% of the score between the groups. There were three questions where group 1 outscored group 2 with a mean score difference of 15% or greater. When asked “What is the expected chain of events in an emergency?”, group 1 was 20% more successful at answering this question than group 2. For the question “What is the T-card for?”, group 1 was 25% more successful at answering the question than group 2. For the question “Who is in charge at the muster station?”, group 1 was 39% more successful at answering this question than group 2. These questions were related specifically to the following learning outcomes: the cognitive awareness of the emergency situation and muster procedures.

As indicated in Figure 4-12, in session 2, the mean quiz score for group 1 was 82% and the mean quiz score for group 2 was 75%. In session 2, there is a noticeable decrease in quiz performance by both groups.

Looking at specific questions, on average the questions in session 2 were within 6% of the score between the groups. There were three questions where group 1 outscored group 2 with a mean score difference of 15% or greater. When asked “What visual alarm is associated with the General Platform Alarm?”, group 1 was 19% more successful at answering this question than group 2. For the question “What do you do when you see a steady green light?”, group 1 was 20% more successful at answering the question than group 2. For the question “Who is in charge at the lifeboat station?”, group 1 was 26%

more successful at answering this question than group 2. These questions were related specifically to alarm recognition and the cognitive awareness of the emergency situation learning objectives.

As indicated in Figure 4-13, in session 3, the mean quiz score for group 1 was 85% and the mean quiz score for group 2 was 79%. In session 3, the overall quiz performance for both groups is similar in distribution of scores.

Looking at specific questions, there were seven questions where group 1 or group 2 outscored the other with a mean score difference of 15% or greater. When asked "What is the safest exit to take given where the hazard is located [on a map of the engine room worksite]?", group 1 was 16% more successful at answering this question than group 2. For the question "Where is the TSR on the platform?", group 1 was 19% more successful at answering the question than group 2. For the question "What is the safest exit to take given where the hazard is located [on a map of the accommodations]?", group 2 was 20% more successful at answering this question than group 1. For the second question related to the engine room egress "What is the safest exit to take given where the hazard is located [on a map of the engine room worksite]?", group 2 was also 20% more successful at answering this question than group 1. For the question "What would you do in the event of an alarm that wasn't followed by a PA announcement?", group 1 was 21% more successful at answering the question than group 2. For the question "What areas have increased risk on the platform?", group 1 was 37% more successful at answering the question than group 2. For the question "What does EER stand for?", group 1 was 48% more successful at answering the question than group 2.

4.2: Aggregated Competency Score Comparison

Participants were required to be successful in all six AVERT learning objectives in order to achieve competence in basic offshore egress. An aggregated competency score was developed to measure each participant’s overall understanding of offshore egress and his/her ability to demonstrate competence in AVERT on an individual level. The aggregated competency score was developed by combining the AVERT performance measures corresponding to the six learning objectives in each scenario. Combining the AVERT performance measures into a competence score enabled a comparison of overall performance to identify the strengths and weaknesses of a group or an individual.

The aggregated competency score was developed by applying a scoring system (i.e. designate pass-fail criteria or passing benchmark and threshold) to each performance measure and then applying an overall weighting of importance for that particular learning objective for each session. This section outlines the task performance scoring and weighting schemes used followed by the results for the three sessions.

4.2.1 Task Performance Scoring

A score system was assigned to each performance measure for all six learning objectives. The US Coast Guard’s Method for Developing Mariner Assessments was used to guide the development of the scoring system and involved creating a proficiency standard and assessment criteria for each performance measure (US Coast Guard Research and Development Center, 2000). The Mariner Assessment Guidelines suggest that critical actions or tasks require pass-fail proficiency criteria and non-critical tasks follow a graded proficiency criteria (US Coast Guards Research and Development Center, 2000). Subject matter experts were consulted in the development of the proficiency standards and assessment criteria specific to the AVERT performance measures. For the aggregated competency score, a combined pass-fail and a graded proficiency criterion was used.

The subject matter experts deemed all subtask learning objectives as critical actions. The associated task performance measures for these learning objectives were a clear correct and incorrect response. For example, AVERT's corresponding performance metrics included: reached the correct location, and moved T-card at muster station. Therefore, this form of data followed the designated pass-fail proficiency criteria. In addition to the pass-fail critical tasks, AVERT also collected many forms of complementary data that informed whether or not the participants required improvement. The data included performance metrics such as time to complete, route selection, and backtracking distance. Salas et al. (2009) recommended that "performance measures usually work best when expert models of the task are used as standards against which to compare and evaluate performance." Following Salas et al.'s recommendation, the informative data was graded using a proficiency benchmark and passing threshold. The following sections outline the scoring systems used for each of the six learning objectives.

1. Spatial Awareness and Alarm Recognition:

Correct Location (pass-fail)

The correct location metric is a pass-fail task performance measure. The correct location is an indicator of whether or not the participants were able to reach their primary or secondary muster points. In the test scenarios, participants were required to recognize the alarm, egress to the correct muster location and indicate that they mustered by moving the T-card. They were instructed in the tutorials where their primary and secondary muster points were located. They were also taught the meaning of the different alarm types and that the alarm dictated where they should muster in the situation. The participant received a full passing score if they successfully reached the correct location. If the participant did not reach the correct location in time or if they attempted to muster at another location, then they received a failing score. Table 4-19 describes the behaviour observed and the scoring system applied to the correct location task.

Table 4-19: Reaching correct location behaviour and the associated score scheme.

<table>
<tr><th>Behaviour</th><th>Score</th></tr>
<tr><td>Reaching the correct location for situation and successfully moving the T-card on muster board.</td><td>Pass</td></tr>
<tr><td>Reaching correct location but not moving T-card.</td><td rowspan="4">Fail</td></tr>
<tr><td>Unable to reach correct location within time allocated.</td></tr>
<tr><td>Going to the wrong location for the situation and moving T-card at wrong location for situation.</td></tr>
<tr><td>Attempting to muster at the wrong location (looking for T-card and realizing his/her name is not on muster board) and then self-correcting to correct location.</td></tr>
</table>

Time to Muster at Correct Location (time limit and benchmark)

To achieve the requirements of the spatial awareness learning objective, participants were required to reach the correct location for the situation and do so within the time limit. The time to complete the scenario (regardless of whether correct or not) was recorded in AVERT. Participants that completed the scenario incorrectly (e.g. mustered at the wrong location) did not receive a passing score for time to complete the scenario. These times were excluded from the time to complete results. Scoring for time to complete was broken down into two categories: overall time limit and the benchmark and passing threshold. Participants who did not muster within the time limit did not receive a passing score and were classified as requiring improvement. Participants that completed the scenario correctly, but were unable to meet the passing threshold of the benchmark time were also identified as requiring improvement.

Overall Time Limit:

Regulations do not dictate a maximum allowable time for mustering for egress drills or emergency situations. It is up to each offshore installation to set standard protocols on the total time available to muster during emergency exercises and drills. The International Association of Oil and Gas Producers (OGP) in a 2010 report suggested the typical time for mustering was between five to fifteen minutes (International Association of Oil and Gas Producers, 2010). Within this five to fifteen minute period the following events should have occurred: go to muster stations, perform a head count and if necessary order to abandon. For this experiment, a total of ten minutes was allocated as the maximum allowable time for the participants to complete each scenario. If a participant was unable to reach the correct muster location within the allotted ten minutes, they did not pass the scenario.

Participants were instructed that they had a maximum of ten minutes to complete each test scenario and they received feedback after each scenario showing the total time they took to complete the scenario. The time limit of ten minutes was ample time for a participant to reach the correct location in both the accommodation and worksite scenarios. Most scenarios, in practice, ten minutes was considered to be too much time to muster using AVERT. For this reason, a standard benchmark time and passing threshold was developed.

Benchmark Time and Passing Threshold:

To score the participants' time to complete the scenarios, a benchmark time of how long an experienced participant would take to complete the muster scenario was used. A passing threshold or time range was also used to identify the participants who were unsuccessful at meeting the benchmark and required improvement. The benchmark and threshold were developed to rank participant performance and indicate when proficiency was achieved and when improvement was required. The benchmark time in each scenario represented the time required to safely walk to the correct muster location within the temporary safe refuge area. The threshold time in each scenario represented an acceptable time to complete the scenario but where proficiency was not achieved (in either wayfinding or controller operation).

Participants who completed the scenario at or near the benchmark time received passing score. Participants who did not meet the benchmark time but fell within the threshold passed the scenario but were flagged as requiring improvement. Participants whose times fell outside the threshold time received a failing score for the scenario and were flagged as requiring significant improvement. Participants who completed the scenario far below the benchmark time were flagged as likely to have been running in the scenario and required intervention and reinforcement of safety practices learning objective.

In real offshore environments, emergency situations can present complex and hazardous conditions that have the potential to negatively impact muster times by compromising egress routes and rendering areas of the platform inaccessible. Early design iterations of the performance scoring considered increasing the time and distance allowances for the AVERT emergency scenarios modelled in session 3. Additional time and distance allowances were considered to accommodate for the anticipated increase in travel time required due to hazards blocking routes and requiring participants to re-route. However, in real emergency situations, offshore personnel do not get more time to muster or evacuate the platform. Increasing the time allowances for the emergency conditions would not be realistic in a real-life situation. Therefore, the same time benchmark and passing thresholds were used to assess participant performance across all sessions (from basic wayfinding to complex emergency scenarios).

The total time to complete the scenario (time to muster) depended on the starting location of the scenario. Cabin scenarios were on average shorter in duration than engine room scenarios. For this reason, different benchmark times and thresholds were applied for the cabin and engine room scenarios.

Accommodation Scenarios:

The benchmark time range used for the accommodation scenarios was 70 to 110 seconds. This time represents the mean time range to complete the cabin egress route safely at a walking speed. Participants who completed the cabin egress scenarios in 70-110 seconds received a passing score (full points). Any completion time under 70 seconds could be attributed to the participant running to the muster station. As this behaviour is not acceptable, participants with a completion time less than 70 seconds received a failing score (zero points). Participants who completed the egress route in 110-180 seconds have performed satisfactorily but can improve their performance. Participants that completed the scenario between 110-180 seconds received an “improvement required” score and received half points. Participants who took more than 180 seconds (3 minutes) to complete the cabin scenarios were likely to not have a good understanding of the environment or poor proficiency with the controller interface and received a failing score (zero points). Table 4-20 describes the behaviour observed and the scoring system applied to the time to complete the cabin egress task.

Table 4-20: Time to muster behaviour for the cabin scenarios and the associated score scheme.

Time to Muster	Behaviour	Score
< 70s	Unacceptable (likely running the whole scenario)	Zero points
70-110s	Acceptable (meets benchmark time)	Full points
110-180s	Satisfactory (improvement required)	Half points
> 180s	Unacceptable (too slow and improvement required)	Zero points

Worksite Scenarios:

The benchmark time range used for the worksite scenarios was 90 to 130 seconds. This time represents the mean time range to complete the engine room egress route safely at a walking speed by competent individuals. Participants who completed the engine egress scenarios in 90-130 seconds received a passing score (full points). Completion times under 90 seconds was attributed to the participant running to the muster station. As this behaviour is not acceptable, participants with a completion time less than 90 seconds received a failing score (zero points). Participants who completed the egress route in 130-240 seconds have performed satisfactorily but can improve their performance. Participants that completed the scenario between 130-240 seconds received an "improvement required" score and received half points. Participants who took more than 240 seconds (4 minutes) to complete the engine room scenarios are likely to not have a good understanding of the environment or poor proficiency with the controller interface and received a failing score (zero points). Table 4-21 describes the behaviour observed and the scoring system applied to the time to complete the engine room egress task.

Table 4-21: Time to muster behaviour for the worksite scenarios and the associated score scheme.

Time to Muster	Behaviour	Scoring
< 90s	Unacceptable (likely running the whole scenario)	Zero points
90-130s	Acceptable (meets benchmark time)	Full points
130-240s	Satisfactory (improvement required)	Half point
> 240s	Unacceptable (too slow and improvement required)	Zero points

Distance Travelled to Muster at Correct Location (benchmark and threshold)

To score the participants' distance travelled to complete the scenarios, a benchmark distance of an experienced individual was used. A passing threshold or distance range was also used to identify the participants who were unsuccessful at meeting the benchmark distance and required improvement. The benchmark distance in each scenario represented the total distance travelled necessary to safely walk to the correct muster location within the temporary safe refuge area. A range was selected for the benchmark distance as the distance travelled to muster is route dependent and the benchmark range can account for the primary, secondary or tertiary routes selected. The threshold distance in each scenario represented an acceptable distance to complete the scenario, but where proficiency was not achieved (mainly in wayfinding). Total distances travelled of participants who completed the scenario incorrectly (e.g. mustered at the wrong location) did not receive a passing score.

Participants who completed the scenario at or near the benchmark distance received passing score. Participants who did not meet the benchmark distance but fell within the threshold passed the scenario but were flagged as requiring improvement. Participants whose total distance travelled fell outside the threshold distance received a failing score for the scenario and were flagged as requiring significant improvement.

Accommodation Scenarios:

The benchmark distance range used for the accommodation scenarios was 45 to 85 meters. Participants that completed the cabin egress scenarios in 45 to 85 meters received a passing score (full points). Participants who completed the scenario between 85-125 meters received an “improvement required” score and received half points. Participants who completed the scenario in more than 125 meters for the cabin scenarios are likely to not have a good understanding of the environment and received a failing score (zero points). Table 4-22 describes the behaviour observed and the scoring system applied to the distance travelled to muster from the cabin.

Table 4-22: Distance to muster for cabin scenarios and the associated score scheme.

Distance to Muster	Behaviour	Score
45 – 85m	Acceptable (meets benchmark range)	Full points
85 – 125m	Satisfactory (improvement required)	Half points
> 125m	Unacceptable (significant improvement required)	Zero points

Worksite Scenarios:

The benchmark distance range used for the worksite scenarios was 80 to 130 meters. Participants who completed the engine egress scenarios in 80 to 130 meters received a passing score (full points). Participants who completed the egress route in 130-

180 meters have performed satisfactorily but received an “improvement required” score and received half points. Participants who travelled more than 180 meters to complete the engine room scenarios are likely to not have a good understanding of the environment and received a failing score (zero points). Table 4-23 describes the behaviour observed and the scoring system applied to the distance travelled to muster from engine room.

Table 4-23: Distance to muster for the worksite scenarios and the associated score scheme.

Distance to Muster	Behaviour	Scoring
80 – 130m	Acceptable (meets benchmark range)	Full points
130 – 180m	Satisfactory (improvement required)	Half point
> 180m	Unacceptable (significant improvement required)	Zero points

2. Routes and Mapping

Route Selection

To be successful in the Routes and Mapping learning objective in the scenarios, participants were required to select the safest and most efficient route given the situation in the scenario. Participants were provided in the tutorials with two egress route options from the cabin and three egress route options from their worksite in the engine room.

The route selection scoring depended on the complexity of the test sessions. For this reason, two different scoring schemes were applied for the first two sessions and the third session.

Session 1 and 2: Route Deviations

Route choices were ranked based on the primary, secondary, tertiary routes available from the cabin and worksite. In sessions 1 and 2, the focus of the routes and mapping learning objective was on developing route knowledge of all routes available from the cabin and the worksite. Therefore, all route selections were scored equally. Participants were assessed on how well they could follow the particular route they had selected. If participants strayed off the route they lost points. Table 4-24 describes the route deviation behaviour observed and the scoring system applied to the route selection for each scenario.

Table 4-24: Route selection behaviour and associated score scheme for each scenario.

<table>
<tr><th>Behaviour</th><th>Scoring</th></tr>
<tr><td>Follows designated route (primary, secondary or tertiary)</td><td>Pass (Full points)</td></tr>
<tr><td>Minor off route (still room for improvement)</td><td>Half point</td></tr>
<tr><td>Major off route (improvement required)</td><td rowspan="2">Zero points</td></tr>
<tr><td>Not following any learned route (lost behaviour– improvement required)</td></tr>
</table>

Session 3: Route Selection

In session 3, route selection and subsequent re-routing were both more important due to the increased scenario complexity of the test scenarios. Each test scenario in session 3 provided information through PA announcements to help the participant select the most appropriate route to take given the situation. The most appropriate route to take in a scenario changed depending on the emergency circumstance. Participants were not told which route was the most appropriate for the situation but at the end of the scenario they were given the option to review the route they chose and where they went during the scenario on the path tracker map. Although participants could review the route they chose in the scenario, they were not given additional feedback on the quality of the route they selected. The scoring of route selection was scenario dependent because route selection varied from scenario to scenario.

Tables 4-25 and 4-26 describe the route selection behaviour observed and the scoring system applied for the accommodation and worksite egress scenarios.

Table 4-25: Behaviour for the accommodation scenarios and the associated score scheme.

Scenario	**Route Selected**	**Score**
TH1	Secondary Route or Tertiary Route	Pass (full points)
	Primary route (blocked by hazard)	Fail (zero points)
	Lost Behaviour	
TH2	Primary – full points	Pass (full points)
	Secondary or Tertiary Route (blocked by hazard)	Fail (zero points)
	Lost Behaviour	

Table 4-26: Behaviour for the worksite scenarios and the associated score scheme.

Scenario	**Route Selected**	**Score**
TH3	Secondary or Tertiary Routes	Pass (full points)
	Primary (blocked by hazard)	Fail (zero points)
	Lost Behaviour	
TH4	Primary – full points	Pass (full points)
	Secondary or Tertiary Routes (blocked by hazard)	Fail (zero points)
	Lost Behaviour	

The correct egress route option to select was specific to the context of the test scenario. In the cabin scenarios, no priority was assigned to the secondary or tertiary routes. Due to the layout of the accommodation block, the secondary and tertiary routes merge together along the egress path and are therefore very similar. For scoring purposes, the two routes were valued equally. For example, in the first hazard scenario (denoted TH1), the muster station and primary egress route were blocked by smoke and as a result the safest route option was to select either the secondary or tertiary routes. In this situation, both were equally correct. Similarly, in the worksite scenarios, no priority was assigned to the secondary or tertiary routes. The secondary and tertiary routes were two distinct routes from the worksite, but due to the layout of the engine room, they were situated in close proximity to one another. As a result, the hazard conditions in the scenarios could impact both routes. For scoring purposes, the two routes were valued equally. For example, in the final hazard scenario (denoted TH4), the explosion and fire rendered the secondary and tertiary egress routes inaccessible and as a result the only available route option was to select the primary egress route. In all scenarios, participants who selected any available clear route received a passing score and participants that selected the route blocked by the hazard received a failing score.

Backtracking Time:

The benchmark percent time backtracking range used for all scenarios was 0 to 15%. This time represents the mean percentage time spent backtracking that was acceptable to successfully complete the egress scenarios. Participants who completed the scenarios within 15% time spent backtracking received a passing score (full points). Participants who completed the scenarios with a total backtracking percentage time between 15-25% received an "improvement required" score and received half points. Participants who completed the scenarios with a total backtracking percentage time more than 25% are likely to not have a good understanding of the environment or poor proficiency with the controller interface and received a failing score (zero points). Table 4-27 describes the behaviour observed and the scoring system applied to the backtracking percentage time spent.

Table 4-27: Percent time spent backtracking and the associated score scheme.

Percent Time Backtracking	Behaviour	Score
0-15%	Acceptable (within passing threshold)	Full points
15-25%	Satisfactory (still room for improvement)	Half point
>25%	Unacceptable (improvement required)	Zero points

Backtracking Distance

The benchmark backtracking distance range used for all scenarios was 0 to 45 meters. Participants who completed the scenarios within 10 meters of backtracking distance received a passing score (full points). Participants who completed the scenarios with a total backtracking distance between 10-25 meters received an "improvement required" score and received half points. Participants who completed the scenarios with a total backtracking distance more than 25 meters are likely to not have a good understanding of the environment and received a failing score (zero points). Table 4-28 describes the behaviour observed and the scoring system applied to the backtracking distance travelled.

Table 4-28: Distance backtracking and the associated score scheme.

Distance Backtracking	Behaviour	Score
0-10m	Acceptable (within passing threshold)	Full points
10-25m	Satisfactory (still room for improvement)	Half points
> 25m	Unacceptable (improvement required)	Zero points

3. Assess Situation & Avoid Hazards:

Raising the Alarm (pass-fail):

Participants were tested on their ability to raise the alarm in the last test scenario (TH4). To be successful in raising the alarm the participants were required to activate the manual alarm call point (MAC) within a 20 second time limit. Points were also given to participants who recognized the need to raise the alarm but were unable to do so within the time limit. Zero points were given to those who did not attempt to raise the alarm and to those who attempted to raise the alarm but were unable to identify the correct location of the MAC to properly raise the alarm. Table 4-29 depicts the raising the alarm behaviour and associated scores.

Table 4-29: Raise the alarm behaviour and scoring scheme.

Behaviour	Scoring
Correctly raised alarm within time limit	Pass (Full points)
Attempted to raise alarm at correct location but were unable to do so within time limit	Half points
Attempted to raise alarm but at the wrong location	Fail (Zero points)
Did not attempt to raise the alarm	

<u>Re-route to avoid Hazards on Platform (pass-fail):</u>

To be successful in avoiding hazards on the platform, participants were required to avoid hazardous routes or re-route from hazards blocking their path. Re-routing behaviour was a measure of how well the participants remembered their route knowledge and were able to use their survey knowledge. Re-routing was assessed during the hazard test scenarios in session 3.

These test scenarios were designed to test the participants' route and survey knowledge by blocking the primary or secondary egress routes, the routes with which the participants to be were most likely familiar. The participant received a passing score if the initial route they selected was free of hazards. The participants also received a passing score if they re-routed immediately upon encountering a hazard along her/his selected route. If the participant did not re-route when encountering a hazard along her/his selected route then s/he received a failing score. Table 4-30 depicts the re-routing behaviour and associated scores.

Table 4-30: Re-routing behaviour and scoring scheme.

Behaviour		Scoring
Initial route selection resulting in no re-routing required	Initial route selection resulting in no re-routing required	Pass
Initial route selection resulting in re-routing required	Re-routing successfully executed (avoids hazard or area around hazard)	Pass
	No re-routing executed (did not re-route and went directly into hazard)	Fail

<u>Exposure time to Smoke and Fire Hazards (pass-fail):</u>

The exposure time to a hazard is a pass-fail measure. The participant received a passing score if their avatar did not come in contact with smoke or fire hazards during the scenarios. Any time recorded of the participant exposed to a smoke or fire hazard was considered a fail score. Table 4-31 depicts the smoke and fire hazard exposure behaviour and associated scores. Table 4-32 depicts the incurred injury due to hazard exposure and associated scores.

Table 4-31: Hazard exposure behaviour and scoring scheme.

Behaviour	Scoring
No exposure time to smoke or fire hazards (0 seconds)	Pass
Exposure time to smoke or fire hazards (> 0 seconds)	Fail

Table 4-32: Incurred injury (probability of 1st degree burns or death) and scoring scheme.

Behaviour	Scoring
No incurred injury due to fire hazards (zero probability of 1^{st} degree burns and/or death)	Pass
Incurred injury due to fire hazards (> 0 probability of 1^{st} degree burns and/or death)	Fail

4. Register at Temporary Safe Refuge

<u>Correctly Moving T-Card at Correct Muster Station (pass-fail):</u>

To be successful at Registering at Temporary Safe Refuge learning objective, participants must first reach the correct location and then correctly move their own T-card at the correct muster point. Participants were instructed in the tutorials on how to register at the muster station and lifeboat station and received feedback after each scenario on whether they went to the correct location. Table 4-33 describes the behaviour observed and the scoring system applied to the Register at TSR learning objective.

Table 4-33: Registering at the TSR behaviour and scoring scheme.

<table>
<tr><th>Behaviour</th><th>Scoring</th></tr>
<tr><td>Successfully moving his/her own T-card at right location for situation</td><td>Pass</td></tr>
<tr><td>Unsuccessfully interacting with T-card (i.e. not moving card) at right location for situation</td><td rowspan="3">Fail</td></tr>
<tr><td>Moving T-card at wrong location for situation</td></tr>
<tr><td>Looking for T-card at wrong location for situation and then readjusting to correct location</td></tr>
</table>

5. Safe Practices

<u>Percentage Running:</u>

To be successful in this Safe Practice learning objective, participants must spend 0% of the scenario time running. Participants were instructed in the tutorials not to run on the platform and received feedback after each scenario showing the time spent stationary, walking and running in the scenario. However, participants still ran on average 15% of the time spent in the scenarios across all three sessions. To instil the importance of this learning objective, participants that never ran received a passing score, participants that occasionally ran received an "improvement required" score and participants frequently ran received a failing score. Table 4-34 describes the behaviour observed and the scoring system applied to the no running on platform safety practice.

Table 4-34: Running behaviour and scoring scheme.

Percent Running	**Behaviour**	**Scoring**
0%	Never running (safe)	Full points
0-10%	Occasionally breaking out into a run (improvement required)	Half points
> 10%	Frequently running (unsafe and improvement required)	Zero points

Closing Fire and Watertight Doors (pass-fail):

To be successful in this safe practice learning objective, participants must close all fire and watertight doors while evacuating the platform in the scenarios. Participants were instructed in the tutorials not to leave fire or watertight doors open and received feedback after each scenario showing the number of doors left open in the scenario. On average, 56% of participants closed all important doors in the scenarios across all three sessions. To instil the importance of this learning objective, participants who closed all doors received a passing score and participants that left one or more doors opened received a failing score. Table 4-35 describes the behaviour observed and the scoring system applied to the door-closing task.

Table 4-35: Closing doors behaviour and scoring scheme.

# of Doors Left Open	Behaviour	Scoring
0	Closing all doors (safe)	Pass
1 or more	Leaving doors opened (unacceptable and needs improvement)	Fail

4.2.2 Weighting of Learning Objectives for Each Session

Following the step-by-step approach to teaching style (Grantcharov and Reznick, 2008), the AVERT training targeted specific learning objectives for each session. The focus of the first session was to establish spatial awareness and route knowledge. Additional learning objectives were incrementally added for each session. The alarm recognition and hazard avoidance learning objectives were incrementally added in order to help build a core spatial knowledge and procedural knowledge. The participants' task performance for each learning objective was measured repeatedly across the three sessions. This allowed for repeated measures to compare performance. Table 4-36 describes which learning objectives were tested during each session.

Table 4-36: AVERT learning objectives tested for each session.

Session 1	Session 2	Session 3
• Spatial Awareness • Routes and Mapping • Register at TSR • Safe Practices	• Spatial Awareness • Alarms Recognition • Routes and Mapping • Register at TSR • Safe Practices	• Spatial Awareness • Alarms Recognition • Routes and Mapping • Hazard Avoidance • Register at TSR • Safe Practices

The above learning objectives were targeted based on subject matter expert guidance and AVERT capabilities. They were gradually assessed over the course of three test sessions to allow participants time to establish spatial learning and gain knowledge on an unfamiliar topic. A prior study suggested that active exploration time with the simulator was important for participants to develop spatial awareness and for the success of virtual environment training (Bradbury-Squires, 2013). Therefore, the training and testing structure was developed to allow participants more time to develop wayfinding in the virtual environment. The weighting of the learning objectives for each session followed the content and learning emphasis for each session: spatial awareness, alarms and hazard avoidance.

A weighting scheme was developed to emphasize the importance of each task. The weighting system was used to grade all scenarios for each session. Tables 4-37, 4-38, and 4-39 show the weighting scheme used for session 1, session 2 and session 3 respectively. Each session was weighted out of 100%.

Session 1: Spatial Awareness

The goal of session 1 was mainly to establish spatial knowledge while introducing the basic procedural knowledge required for emergency egress offshore (instilling the value of the temporary safe refuge (TSR) and general safe practices). This session focused on learning objectives: Spatial Awareness (LO1), Routes and Mapping (LO3), Register at TSR (LO5) and Safe Practices (LO6). For this reason 60% of the weighting was assigned to spatial knowledge acquisition (overall spatial awareness including landmark recognition, route mapping and developing the foundation for survey knowledge development). The remaining 40% involved the procedural knowledge acquisition, specifically TSR procedures and general safe practices. Table 4-37 describes the learning objectives, the associated performance measures and assigned weighting for session 1.

Table 4-37: Session 1 weighting scheme (out of a total of 100 points).

<table>
<tr><th>Learning Objectives</th><th>Performance Measure</th><th colspan="2">Weighting</th></tr>
<tr><td rowspan="3">LO1. Establish Spatial Awareness of Environment</td><td>Correct location</td><td>25</td><td rowspan="3">35</td></tr>
<tr><td>Total time to muster at correct location</td><td>5</td></tr>
<tr><td>Total distance travelled to correct location</td><td>5</td></tr>
<tr><td rowspan="2">LO3. Routes and Mapping: Determine <u>Primary and Alternative</u> Routes to Muster Stations</td><td>Route selected (prim, second, or other)</td><td>15</td><td rowspan="2">25</td></tr>
<tr><td>Off route:
Total backtracking time
Total backtracking distance</td><td>
5
5</td></tr>
<tr><td>LO5. Register at Temporary Safe Refuge (TSR)</td><td>Correct location + Move t-card correctly</td><td>15</td><td>15</td></tr>
<tr><td rowspan="2">LO6. Safe Practices</td><td>Speed of trainee (% running)</td><td>10</td><td rowspan="2">25</td></tr>
<tr><td>Number of fire/watertight doors left open (closed)</td><td>15</td></tr>
</table>

Note: A passing competence score is 80%. If a participant failed to achieve any of the above learning objectives individually it would result in a failing grade.

Spatial Awareness and Route Mapping (60%):

Spatial Awareness and Route Mapping was broken down into reaching the correct location, time and distance spent getting to correct location, route selected to get to the location, and any backtracking that took place. This learning objective was aimed at helping the participants establish spatial orientation and wayfinding as described by Waller et al. (1999), Darken and Peterson (2001) and Farrell et al. (2003). To assist with this task, participants were provided time to familiarize themselves with accommodations, engine room machinery spaces, the temporary safe refuge (TSR) and the route options from accommodations and worksite to the TSR. This performance measure assessed how well the participants remembered:

- the assigned muster stations (landmarks),
- the egress routes (route knowledge),
- and overall wayfinding in the environment (survey knowledge).

The main focus of session 1 was to help establish spatial awareness of the environment (both route memory and survey knowledge). For this reason the spatial awareness and route mapping performance measures were allocated a 60% weighting for session 1.

Register at TSR (15%):

The second focus of session 1 was to establish the correct muster locations and the procedures necessary to correctly muster (i.e. what to do once one reached the correct location?). This learning objective is dependent on the previous objective. To properly muster at the TSR, the participant required both spatial and procedural knowledge to recognize the muster and lifeboat stations (landmarks) and knowledge of procedures to identify the correct muster location. The performance measures associated with Muster Station Protocol were allocated a 15% weighting for session 1.

Safe Practices (25%):

The first focus of session 1 was on emphasizing the importance of safe practices. As the participants in the study had no prior work experience in an offshore environment, they were generally unaware (prior to the study) of general safety practices that are necessary while working in these environments. Two main elements that were emphasized in AVERT were: not to run on the platform and to remember to close all doors. These tasks are important and need to be emphasized in this form of training. For this reason, a weighting of 25% of the overall competency score in session 1 was assigned to this learning objective.

The scoring scheme was not presented to the participants in the feedback portion of the testing and thus there was not sufficient reinforcement to help correct this behaviour for some participants in the study. This is a lesson learned to apply to future programs.

Session 2: Alarm Recognition

The goal of session 2 was to expand the participants' procedural knowledge specifically in alarm recognition. Therefore, the Alarm Recognition learning objective

(LO2) was added to the assessment criteria in session 2. This change is also reflected in the weighting scheme of session 2. The overall focus transferred from spatial awareness to alarm recognition. For this reason, 35% of the weighting was assigned to spatial knowledge. The alarm recognition category was assigned 25% of the weighting (for the reinforcing of the alarm recognition element of the procedural knowledge). The remaining 40% involved the procedural knowledge acquisition again for registering at the TSR procedures and general safe practices. The weightings were rebalanced to 100% for the session. Table 4-38 describes the learning objectives, the associated performance measures and assigned weighting for session 2.

Spatial Awareness and Route Mapping (35%) and Alarm Recognition (25%):

Once spatial awareness and route mapping was established in session 1, session 2 added the alarm recognition learning objective to the participants' responsibilities for responding to the scenarios. This resulted in a shift of weighting from the spatial awareness objective to the alarm recognition learning objective. Both spatial awareness and alarm recognition learning objectives are measured in AVERT with the same metric and can be considered as one and the same from session 2 onward.

Register at TSR (15%) & Safe Practices (25%):

These learning objective weightings remained the same in session 2.

Table 4-38: Session 2 weighting scheme (out of a total of 100 points).

Learning Objectives	Performance Measure	Weighting	
LO1. Establish Spatial Awareness of Environment	Correct location	See LO2	**10**
	Total time to muster at **correct location**	5	
	Total distance travelled to **correct location**	5	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Correct location (GPA = Mess Hall, PAPA = Lifeboat)	25	**25**
LO3. Routes and Mapping: Determine Primary and Alternative Routes to Muster Stations	Route selected (prim, second, or other)	15	**25**
	Off route:		
	Total backtracking time	5	
	Total backtracking distance	5	
LO5. Register at Temporary Safe Refuge (TSR)	Correct location + Move t-card correctly	15	**15**
LO6. Safe Practices	Speed of trainee (% running)	10	**25**
	Number of fire/watertight doors left open (closed)	15	

Session 3: Assess Situation and Avoid Hazards

The goal of session 3 was to evaluate the participants' procedural knowledge specifically in assessing the emergency situation and avoiding hazards on route. Therefore the Assess Emergency Situation and Avoid Hazards learning objective (LO4) was added to the assessment criteria in session 3. The weighting scheme of session 3 was adjusted to reflect the focus of the assessment. The overall focus transferred from spatial awareness and alarm recognition to assessing the situation and avoiding hazards. For this reason 30% of the weighting was reassigned to spatial knowledge (awareness and routes and mapping). The alarm recognition category was reassigned 10% of the weighting. Assessment of the situation category was assigned 25% (for reinforcing the situation assessment element of the procedural knowledge). Registering at the TSR was adjusted to 10% and the remaining 25% involved safe practices. The weightings were rebalanced to 100% for the session. Table 4-39 describes the learning objectives, the associated performance measures and assigned weightings for session 3.

Assess Situation and Avoid Hazards (25%):

The main focus of session 3 was establishing the importance of assessing the situation and avoiding hazards along the egress route. For this reason the performance measures associated with Assess Situation were allocated a 15% weighting for session 3.

Spatial Awareness & Route Mapping (30%), Alarm Recognition (10%), Register at TSR (10%)

Spatial awareness, route mapping, alarm recognition and register at TSR learning objectives were established and tested in session 1 and session 2. In session 3, the participants' responsibilities transitioned from demonstrating knowledge of these learning objectives to their application in emergency situations. The assess situation and avoid hazards (LO4) learning objective encompasses the application of the first four learning objectives. As a result, these learning objective weightings had a reduced emphasis in session 3.

Safe Practices (25%):

This learning objective weighting remained the same in session 3.

Table 4-39: Session 3 weighting scheme (out of a total of 100 points).

Learning Objectives	Performance Measure	Weighting: Scenarios TH1-3	Weighting: Scenarios TH4	Weighting: Total
LO1. Establish Spatial Awareness of Environment	Correct location	See LO2	See LO2	**10**
	Total time to muster at correct location	5	5	
	Total distance travelled to correct location	5	5	
LO2. Alarms Recognition: Understand role of alarms and urgency of situation	Correct location (GPA = Mess Hall, PAPA = Lifeboat)	10	10	**10**
LO3. Routes and Mapping: Determine <u>Primary and Alternative</u> Routes to Muster Stations	Route selected (prim, second, or other)	10	10	**20**
	Off route:			
	Total backtracking time	5	5	
	Total backtracking distance	5	5	
LO4. Assess Emergency Situation and <u>Avoid Hazards on Route</u>	Successful Interaction with MAC (TH4 only)	n/a	5	**25**
	Route selected (prim, second, or other) & re-route in event of alarm change/ PA update	5	n/a	
	Re-route in event of encounter Hazard (Most efficient route selected when re-routing)	10	5	
	Exposure time to hazard – smoke	10	5	
	Exposure time to hazard – fire (TH4 only)	n/a	5	
	Incurred injury – probability of 1st degree burns/death (TH4 only)	n/a	5	
LO5. Register at Temporary Safe Refuge (TSR)	Correct location + Move t-card correctly	10	10	**10**
LO6. Safe Practices	Speed of trainee (% running)	10	10	**25**
	Number of fire/watertight doors left open (closed)	15	15	

4.2.3 Aggregated Competence Score Results

The competence score provides a good indication of how closely the participants in both groups were able to reach competence in offshore egress using AVERT. The competence score for each scenario represents the individual participant's ability to demonstrate her/his understanding of all six learning objectives in the virtual offshore egress situation. As described in the previous section, to receive a passing competence score the participant was required to achieve an 80% or higher in the aggregated competence score. This (high) percentage passing score was chosen to ensure that the participants were meeting all learning objective requirements adequately. Table 4-40 summarizes the mean competence scores for the entire sample.

Table 4-40: The descriptive statistics of the aggregated competency score for each scenario.

Scenario	N	Mean Score	Std. Dev	Minimum	Maximum
Normal (TE1)	36	58	26	0	95
Normal (TE2)	36	65	26	0	98
Blackout (TE3)	36	73	17	8	95
Blackout (TE4)	36	67	19	13	95
Alarm (TA1)	36	79	14	25	100
Alarm (TA2)	36	76	20	3	100
Alarm + Blackout (TA3)	36	76	19	10	100
Alarm + Blackout (TA4)	36	77	16	33	100
Hazard situation (TH1)	36	61	23	15	95
Hazard situation (TH2)	36	67	24	5	100
Hazard situation (TH3)	36	62	23	5	98
Hazard situation (TH4)	36	67	23	5	95

As shown in Table 4-40, neither group achieved a mean competence score of 80% or higher in any of the test scenarios. Thus, as a collective, neither group met the expected passing competence score. It is notable that the entire sample shows signs of improvement between scenarios in session 1 and further improvement across session 1 to session 2. However, there appears to be an overall degradation of performance in session 3 which involved the most challenging scenarios. To compare competence scores in relation to scenario complexity, the competency scores have been organized by accommodation and worksite egress scenarios. Figures 4-14 and 4-15 depict the mean competence scores for each group in the accommodation and worksite scenarios, respectively.

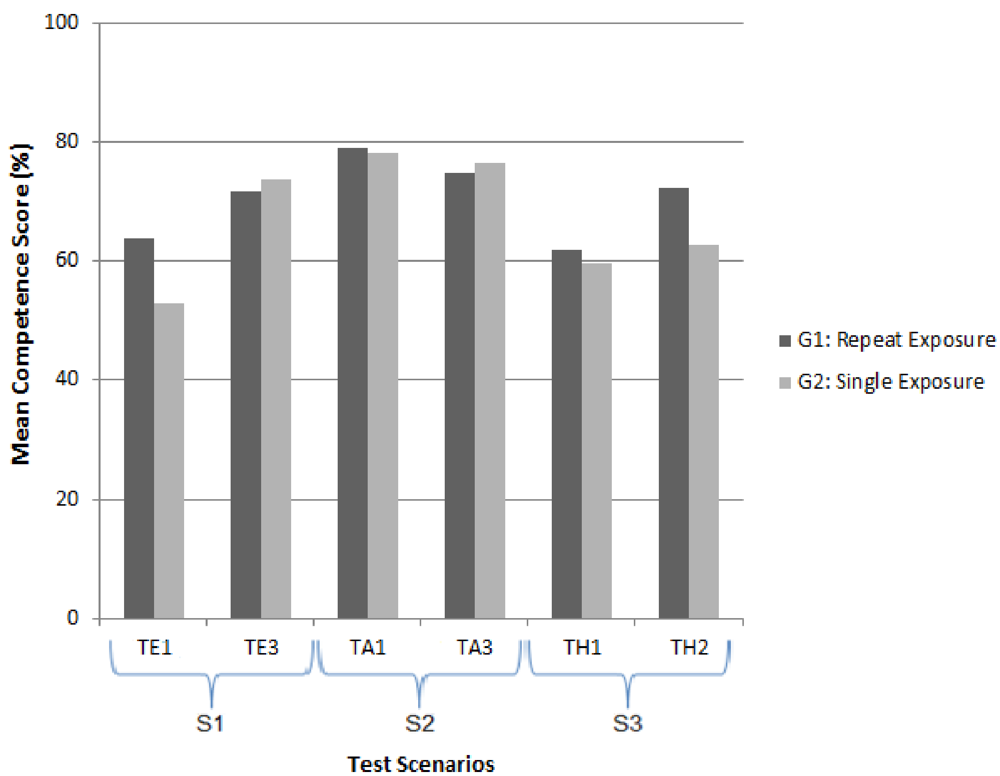

Figure 4-14: Cabin Egress Competence Scores for both groups.

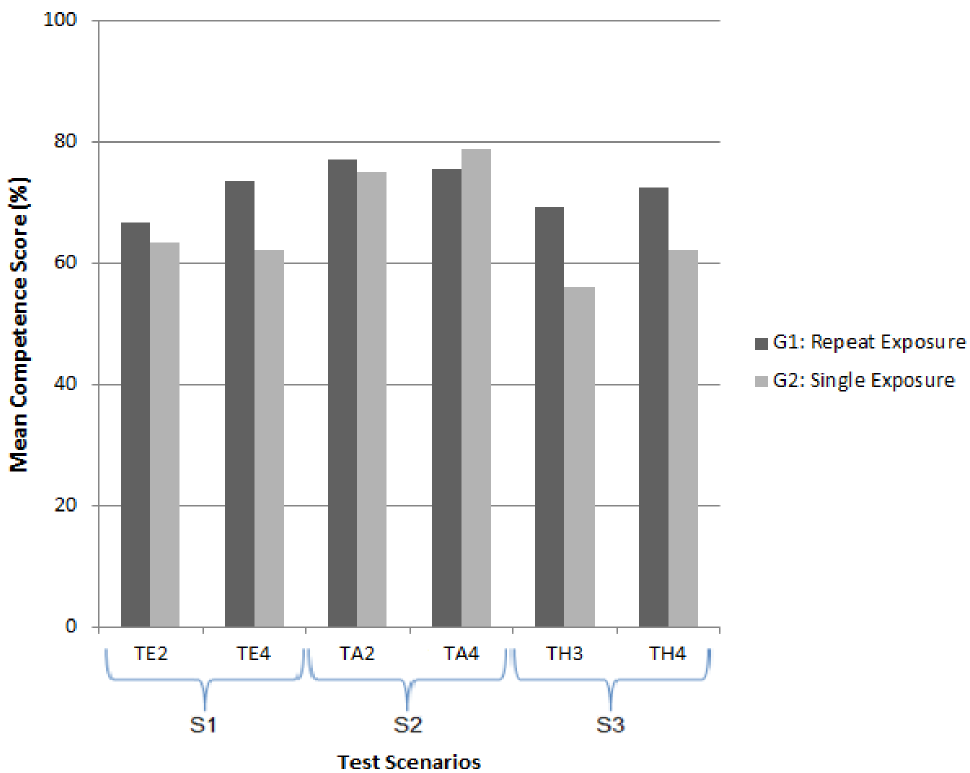

Figure 4-15: Worksite Egress Competence Scores for both groups.

Figure 4-14 shows a similar trend as the overall competence score in the accommodation egress scenarios. In session 1 there is a notable improvement for both groups in the overall competence score from TE1 to TE3 in the spatial awareness scenarios. Both groups remain consistent in competence score throughout session 2 during the alarm recognition scenarios. The competence score begins to drop in session 3 when the advanced emergency scenarios are introduced. The mean competence scores for both groups do not reach the passing score, however there are individuals in both groups that were successful in reaching the 80% passing score. In the first scenario (TE1), 24% of group 1 and 21% of group 2 achieved the passing score. In the blackout condition

(TE3), 35% of group 1 and 47% of group 2 achieved 80% or higher. Similarly, in the alarm recognition scenarios (TA1), 65% of group 1 and 47% of group 2 achieved the passing score. In TA3, 41% of group 1 and 63% of group 2 were successful. In the advanced emergency situations (TH1), 35% of group 1 and 26% of group 2 achieved 80% or higher. In TH2, 53% of group 1 and 32% of group 2 were successful at reaching the passing competence score.

Figure 4-15 shows the competence score in the worksite egress scenarios. In session 1, there was a small improvement for group 1 and no notable improvement for group 2 in the overall competence score from TE1 to TE3 in the spatial awareness scenarios. Both groups reached the highest competence score in session 2 during the alarm recognition scenarios. Similarly to the accommodation egress scenarios, the mean competence scores for both groups in the worksite scenarios do not reach the passing score. There were, however, several individuals in both groups that were successful in reaching the 80% passing score. In the first engine room scenario (TE2), 41% of group 1 and 42% of group 2 achieved the passing score. In the blackout condition (TE4), 35% of group 1 and 26% of group 2 achieved 80% or higher. In the alarm recognition scenarios (TA2), 71% of group 1 and 53% of group 2 achieved the passing score. In TA4, 59% of group 1 and 74% of group 2 were successful. In the advanced emergency situations (TH3), 29% of group 1 and 21% of group 2 achieved 80% or higher. In TH4, 41% of group 1 and 42% of group 2 were successful at reaching the passing competence score.

4.2.4 Competence Score Comparison between Groups

Fisher's Exact Test was used to compare the group proportions to determine if the level of training exposure between group 1 and group 2 impacted the participants' overall competence score. Tables 4-41, 4-42 and 4-43 summarize the competence scores by group for the three sessions.

Table 4-41: The mean competency scores for each scenario in session 1 by group.

Scenario Condition	Group	N	Mean	SE	p - value Fisher's Exact Test
Normal (TE1)	G1	17	64	5.44	0.314
	G2	19	53	6.62	
Normal (TE2)	G1	17	67	6.13	0.344
	G2	19	63	6.36	
Blackout (TE3)	G1	17	72	4.66	0.825
	G2	19	74	3.27	
Blackout (TE4)	G1	17	73	2.94	0.454
	G2	19	62	5.29	

In session 1, group 1 outscored group 2 in all scenarios except the cabin blackout scenario (TE3). As indicated in Table 4-41, no statistical difference was found between group participants in competence score for all session 1 scenarios. Overall, the competence scores for both groups were similar (mean 6% difference). The biggest difference was observed in the first (TE1) and last (TE4) scenarios, where the mean competence score of group 1 was 11% higher than group 2.

Both groups show improvement in competence score from session 1 to session 2. On average, group 1 improved by 6% while group 2 improved by 14% from session 1 to session 2. The competence scores are more balanced in session 2 across the two groups (mean 2% difference). However, in session 2, group 2 outscored group 1 in all scenarios except the normal condition engine room scenario. As indicated in Table 4-42, no statistical difference was found between group participants in competence score for all session 2 scenarios.

Table 4-42: Session 2 mean competency scores by group.

Scenario Condition	Group	N	Mean	SE	p - value Fisher's Exact Test
Alarm (TA1)	G1	17	79	4.14	0.735
	G2	19	78	2.28	
Alarm (TA2)	G1	17	77	4.77	0.565
	G2	19	75	4.75	
Alarm + Blackout (TA3)	G1	17	75	4.99	0.241
	G2	19	76	4.30	
Alarm + Blackout (TA4)	G1	17	75	3.86	0.488
	G2	19	79	3.78	

The competence score performance of both groups degraded from session 2 to session 3. On average, group 1 participants performance decreased by 6% while group 2 decreased by 17% from session 2 to 3. Session 3 resulted in the biggest difference in competence scores between the two groups. In all session 3 scenarios, group 1 outscored group 2 (mean 9% difference). As indicated in Table 4-43, one scenario resulted in a statistical difference between the group participants (TH2, $p = 0.02$). This scenario involved cabin egress where the secondary egress route was blocked by heavy smoke. Group 1 scored 13% higher than group 2 in this scenario. No other statistically significant difference was found between groups in competence score for the remaining scenarios.

Table 4-43: Session 3 mean competency scores by group.

Scenario Condition	Group	N	Mean	SE	p - value Fisher's Exact Test
(TH1)	G1	17	62	5.39	0.818
	G2	19	60	5.35	
(TH2)	G1	17	72	6.14	**0.020**
	G2	19	63	5.00	
(TH3)	G1	17	69	4.45	0.661
	G2	19	56	5.63	
(TH4)	G1	17	72	5.17	0.902
	G2	19	62	5.65	

4.3: Learning Across Sessions

The aggregated competence score provided a measure of the competence reached by both groups. A closer look at each learning objective is required to determine how well the participants learned the training content and improved by gaining experience in the test scenarios. To measure learning, each participant's task performance was measured repeatedly across the three sessions. For the purpose of this study, learning is defined as having occurred if the participant's task performance improves on one or more of the learning objectives across the three test sessions. See Table 3-3 for a list of the learning objectives that were assessed for each session.

This section will first identify any performance improvements measured within sessions and across sessions and secondly summarize the major learning outcomes and common mistakes observed during the virtual egress scenarios. Non-parametric statistics were used to compare within session and across session differences on task performance. Separate Friedman's tests were used for each group to determine within-group differences between normal (control), blackout (visibility) and hazard (complexity) conditions. Wilcoxon Signed Ranks test were also used to check for within-group differences across two scenario conditions to identify any statistical significance.

To compare task performance to measure improvements (learning), the results were again split into the accommodation scenarios and the worksite scenarios. The following comparisons were made for cabin egress scenarios:

- To compare learning within the session on cabin egress scenarios, the following scenarios were compared: TE1-TE3 (normal to blackout conditions), TA1-TA3 (normal to blackout conditions), and TH1 and TH2 (hazard to hazard conditions).
- To compare learning across sessions on cabin egress scenarios, the following scenarios were compared: TE1-TA1 (normal to normal condition); and TE3-TA3 (blackout to blackout condition). No data were collected to compare hazard conditions across sessions.

The following comparisons were made for engine room egress scenarios:

- To compare learning within the session on engine room egress scenarios, the following scenarios were compared: TE2-TE4 (normal to blackout conditions) and TA2-TA4 (normal to blackout conditions) and TH3 and TH4 (hazard to hazard conditions).
- To compare learning across sessions on engine room egress scenarios, the following scenarios were compared: TE2-TA2 (normal to normal condition); and TE4-TA4 (blackout to blackout condition). No data were collected to compare hazard conditions across sessions.

1. Establishing Spatial Awareness of Environment

Correct Muster Location – Cabin

Within Sessions

For group 1, no significant effect was identified between normal, blackout and hazard conditions on the correct muster location. For group 2 there was a significant difference in reaching the correct location between the normal to blackout conditions (TE1 to TE3 scenarios) in session 1. Figure 4-16 shows the percent competence in reaching the correct location in the cabin scenarios for both groups. The task performance of group 2 in reaching the correct location improved by 37% from the normal condition to the blackout condition in the accommodation scenarios (TE1 to TE3, $p = 0.016$). There were no other differences between normal to blackout conditions for group 2 in the correct muster location task.

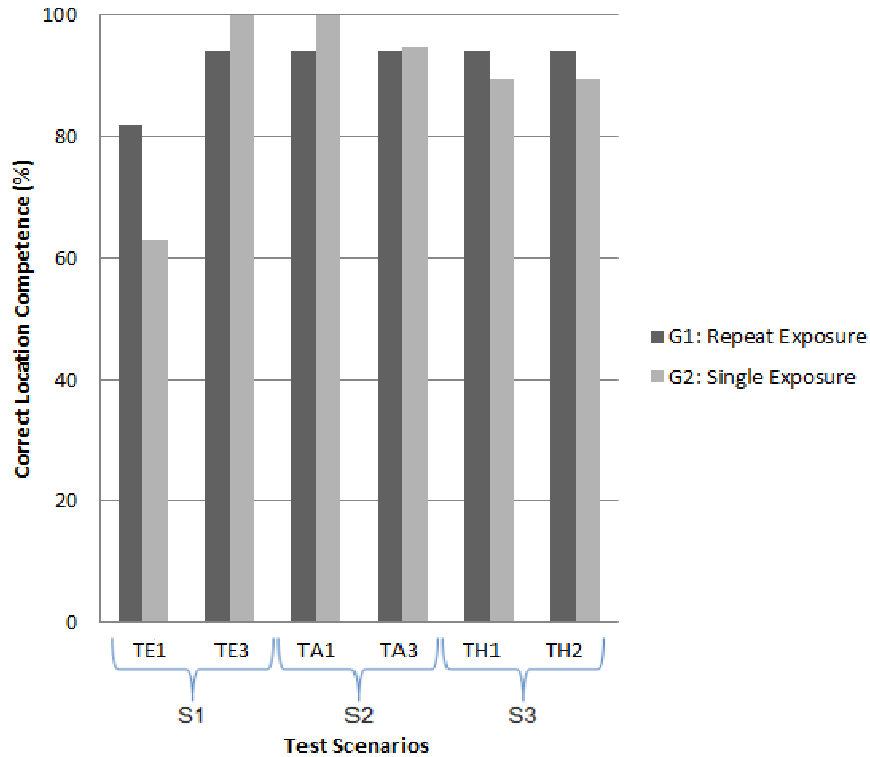

Figure 4-16: Cabin egress correct location and register at TSR competence scores.

Across Sessions

Comparing the correct muster location performance across sessions, there was no significant difference identified between session 1, session 2 and session 3 for group 1. For group 2 there was a significant difference ($p = 0.016$) in reaching the correct location between the normal condition scenario in session 1 and the normal condition scenario in session 2 (TE1-TA1). The task performance of group 2 in reaching the correct location improved by 37% from the normal condition in session 1 to the normal condition in session 2 in the accommodation scenarios. For group 2 there was also a significant difference between the first scenario in session 1 (TE1) and the hazard condition in session 3 (TH1) with a difference of 26% improvement ($p = 0.025$). There were no other differences between sessions for both groups in the correct muster location task.

Overall, the improvement in performance experienced by group 2 from the first scenario to scenarios in session 2 and 3 is likely due to learning the correct muster locations during the test scenarios. The lack of significant difference in performance in group 1 is in part because group 1 was able to practice similar scenarios prior to the test sessions. Thus there was less room for improvement for group 1 (as the group 1 participants' improvement in identifying the correct muster location occurred in the practice scenarios). There was also some degradation in performance for both groups from session 2 to session 3. This decrease in reaching the correct muster location was not statistically significant.

Correct Muster Location – Engine Room

Within Sessions

No significant difference was found for both groups in reaching the correct muster location for the engine room scenarios. Figure 4-17 shows the percent competence in reaching the correct location in the engine room scenarios for both groups

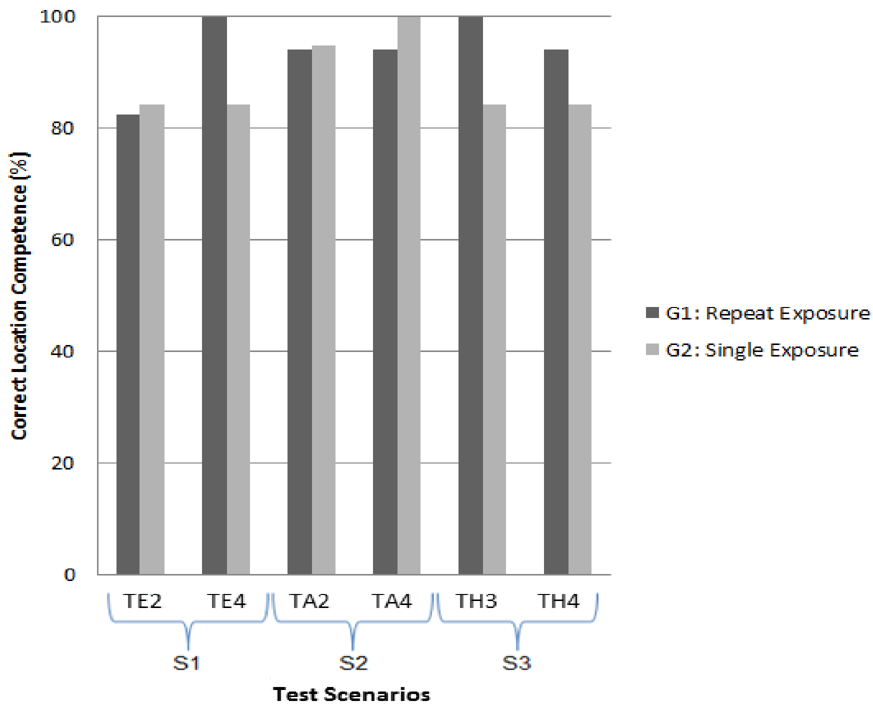

Figure 4-17: Worksite egress correct location and register at TSR competence scores.

Across Sessions

Comparing the correct muster location performance across sessions, there was no significant difference identified between session 1, session 2 and session 3 for group 1 or group 2.

<u>Time and Distance to Muster – Cabin Egress Scenarios</u>

Within Sessions

Figures 4-18 and 4-19 show the percent competence for time and distance to muster in the cabin scenarios for both groups. In comparing normal to blackout conditions in cabin scenarios, there was no significant within-group difference between normal to blackout conditions (TE1 to TE3) for time to complete and distance travelled to complete the scenarios for both groups in session 1. However, there was a significant within-group difference between normal and blackout conditions for group 1 (found between scenarios TA1-TA3) for time to complete scenario (difference = 15.73 seconds, $p = 0.031$) and distance to complete scenario (difference = 17.69 meters, $p = 0.022$) in session 2. There was no significant within-group difference for group 2 between (TA1-TA3) for time or distance to complete scenario ($p = 0.196$, $p = 0.291$).

Across Sessions

In comparing normal to hazard conditions in cabin scenarios, there was no significant difference in time and distance found for both groups in session 1. There was a trending difference in distance to complete scenario between normal to hazard (TE1 and TH1) conditions for group 1 (difference = 31.78m, $p = 0.062$) and group 2 (difference = 22.03m, $p = 0.070$). There was also a trending difference for group 1 between TE1 and TH2 for distance to complete scenario (difference = 61.97m, $p = 0.049$).

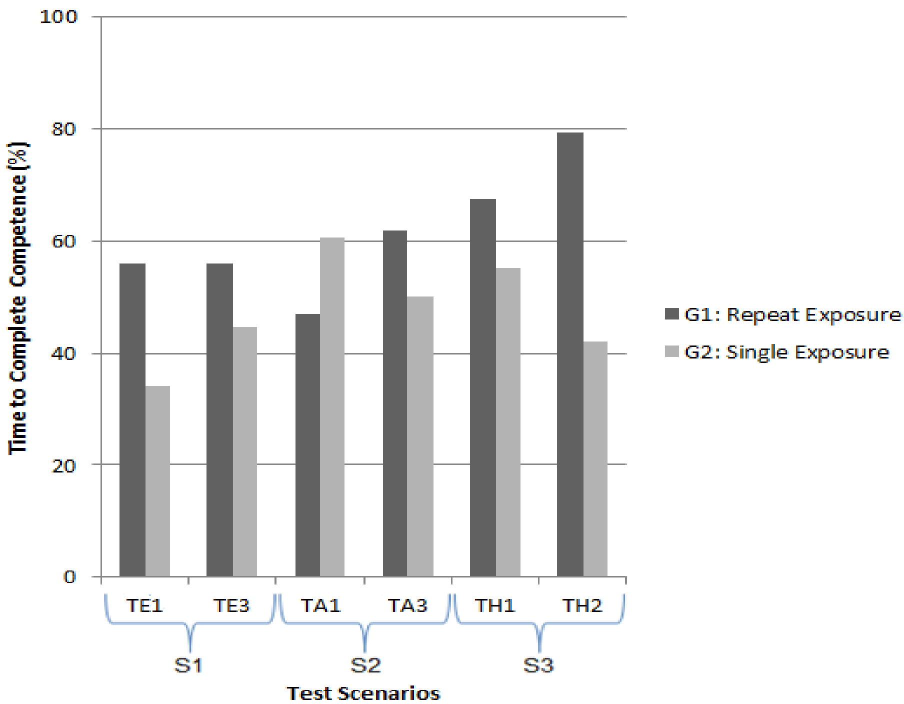

Figure 4-18: Cabin egress time to complete competence scores for both groups.

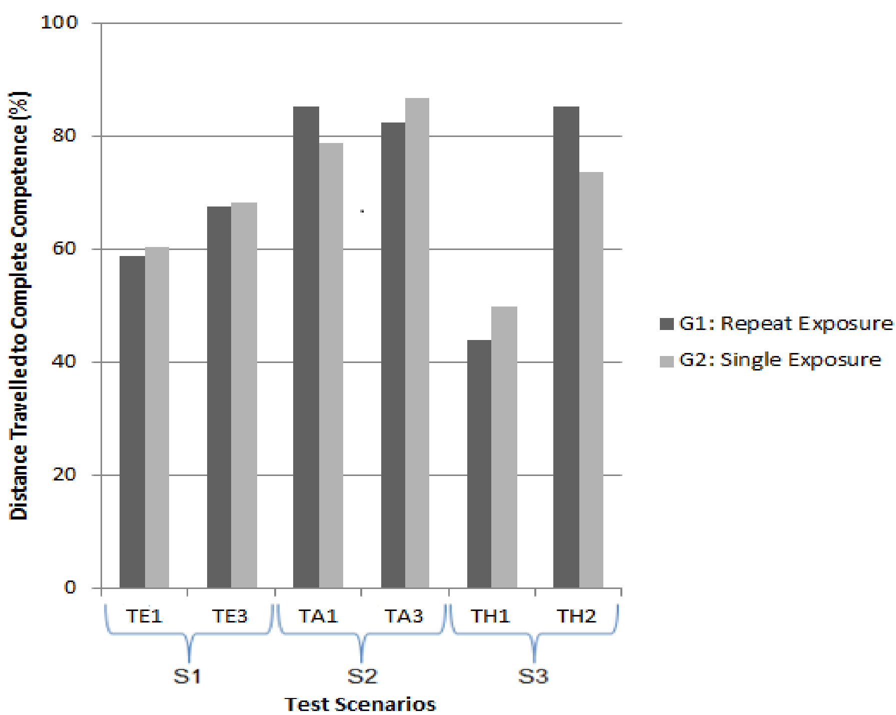

Figure 4-19: Cabin egress distance travelled competence score for both groups.

For group 1, a significant within-group difference was found between normal and hazard (TA1-TH1) conditions for time and distance to complete scenario (time difference = 40.3s, $p = 0.000$ and distance difference = 49.51m, $p = 0.001$). For group 2, a significant within group difference was found between normal and hazard (TA1-TH1) conditions for time and distance to complete scenario (time difference = 43.90s, $p = 0.002$ and distance difference = 46.31m, $p = 0.007$). In addition for group 1 a significant within-group difference were found between TA1-TH2 (normal to hazard) for time and distance to complete scenario (time difference = 36.01seconds, $p = 0.000$ and distance difference = 19.33m, $p = 0.025$). For group 2 there was no significant within-group difference found between normal to hazard (TA1-TH2) conditions for time and distance to complete scenario ($p = 0.225$ and $p = 0.227$ respectively).

In comparing blackout to hazard conditions in cabin scenarios, for group 1, there was a significant difference in distance to complete scenario between the blackout and hazard condition (TE3-TH1 distance difference = 18.40m, $p = 0.049$). For group 2, there was no significant within-group difference between blackout and hazard conditions (TE3-TH1). For group 1 there was a significant within-group difference between the blackout to hazard conditions (TA3-TH1) for time and distance to complete the scenario (time difference = 24.57s, $p = 0.023$ and distance difference = 31.81m, $p = 0.013$). For group 2, there was a significant difference between the blackout and hazard condition (TA3-TH1) for time to complete scenario (time difference = 49.91s, $p = 0.001$ and distance difference = 47.62m, $p = 0.002$).

In comparing TA3-TH2 there was a significant difference for time to complete scenario for group 1 (difference = 20.28seconds, p = 0.023), but no difference for distance to complete scenario (p = 0.943). For group 2 there was no significant within-group difference between TA3-TH2 for time to complete scenario (p = 0.113), but there was a significant difference in distance to complete scenario (difference = 12.69m. p = 0.016).

Time and Distance to Muster – Engine Room Egress Scenarios

Within Sessions

Figures 4-20 and 4-21 show the percent competence for time and distance to muster in the engine room scenarios for both groups. In comparing normal to blackout conditions in engine room scenarios, there was no difference for time and distance between normal to blackout conditions (for TE2-TE4 and TA2-TA4) for both groups for session 1 and session 2.

Across Sessions

In comparing normal to hazard conditions in engine room scenarios, for group 1, there is a trending difference for time to complete between normal to hazard (TE2-TH3) conditions (p = 0.064), but not for distance (p = 0.204). For group 2, there was a significant difference for distance (difference = 52.77m, p = 0.007) and a trending difference for time between TE2 to TH3 (time = 51.9s, p = 0.060). Comparing normal to hazard conditions (TA2-TH3), for group 1, there was a significant difference for time and distance for normal to hazard conditions in the engine room scenarios (time difference =

110.88s p = 0.000 and distance difference = 87.68m, p = 0.000). As well for group 2, there was a significant within-group difference in time and distance for normal to hazard (TA2-TH3) conditions in the engine room scenarios (time difference = 125.86s, p = 0.000 and distance difference = 100.64m, p = 0.000). For group 1, there was also a significant difference between TA2-TH4 for distance (difference = 39.87m, p =0.017) and a trending difference for time (t = 37.4s; p = 0.071).

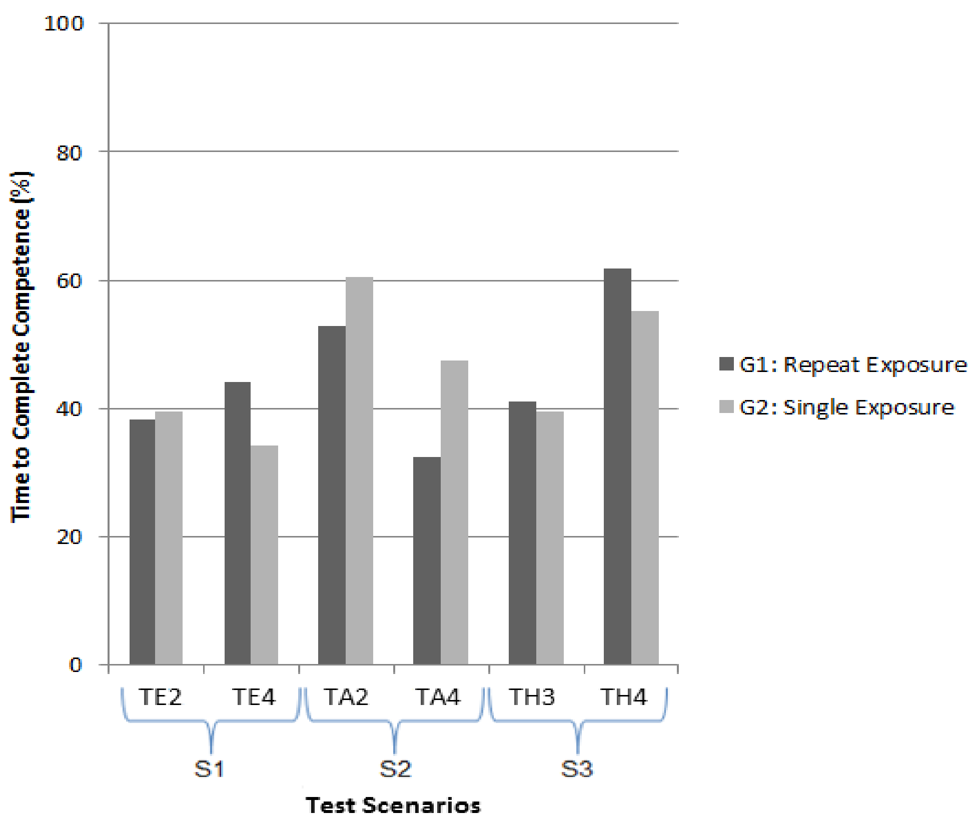

Figure 4-20: Worksite egress time to complete competence score for both groups.

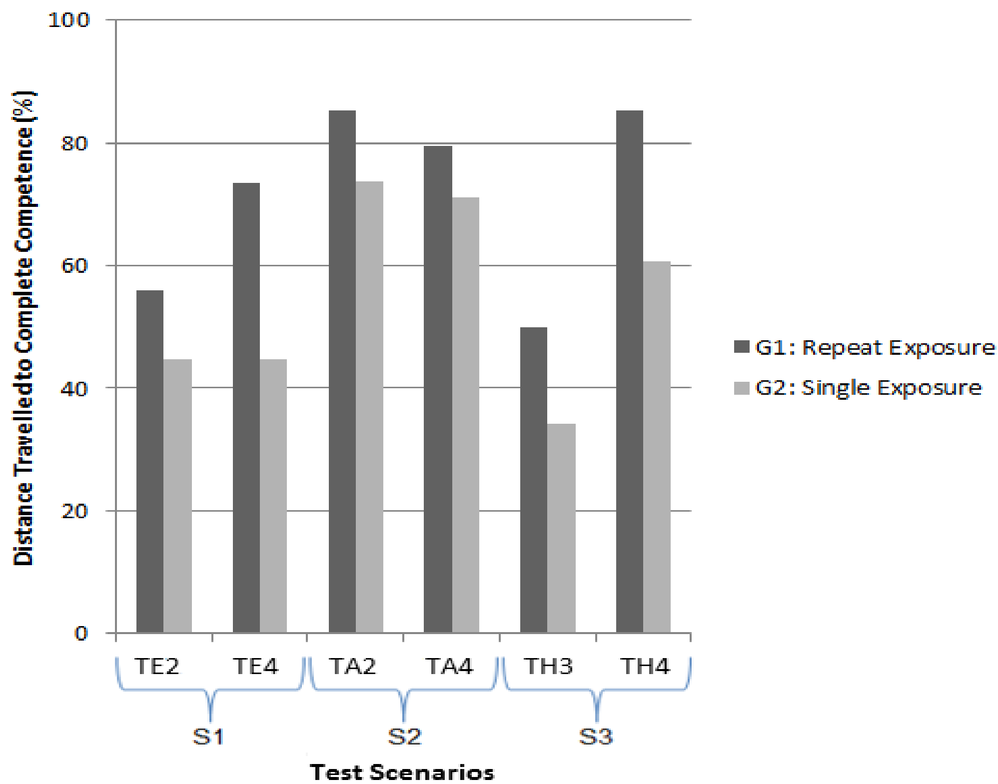

Figure 4-21: Worksite egress distance travelled competence score for both groups.

In comparing blackout to hazard conditions in engine room scenarios (TE4-TH3/TH4 or TA4-TH3/TH4), comparing blackout to hazard conditions (TE4-TH3), there was a significant difference in time to complete for group 2 (difference = 56.70s, $p = 0.045$) and a trending difference for group 1 (difference = 47.89s, $p = 0.064$). There was no within group difference for distance between TE4-TH3 scenarios. Comparing TE4-TH4 there was a significant difference in time to complete and distance to complete for group 2 between blackout to hazard conditions (time difference = -47.61s, $p = 0.012$ and distance difference = 75.93m, $p = 0.001$). There was also a trending difference for time to complete for group 1 (difference = -25.57s, $p = 0.071$). For group 1, there was a significant difference between blackout to hazard condition (TA4-TH3) for time and distance to complete (time difference = 100.01s, p = 0.000 and distance difference = 80.19m, p =0.001). Likewise, there was a significant within-group difference for group 2 between TA4-TH3 (time difference = 126.95s, p = 0.000 and distance difference = 99.56m, p =0.001). Finally, there was a significant difference in distance to complete between TA4-TH4 for group 2 (distance difference = 65.28m, p =0.027).

2. Routes and Mapping

Backtracking Time and Distance – Cabin Egress Scenarios

Figures 4-22 and 4-23 show the percent competence for backtracking time and distance to muster in the cabin scenarios for both groups. There was no significant difference between normal and blackout conditions for both groups in session 1 cabin scenarios. In session 2, there was a significant difference in backtracking time between normal and blackout conditions (TA1-TA3) for group 1 only with a backtracking time difference of 9.24s (p = 0.000).

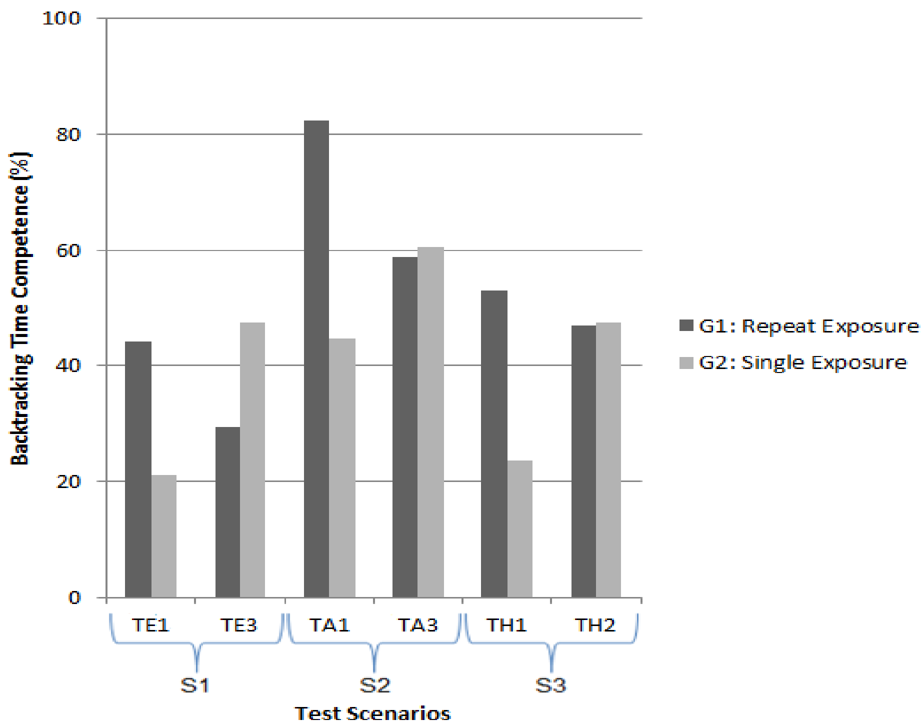

Figure 4-22: Cabin scenario backtracking time competence scores for both groups.

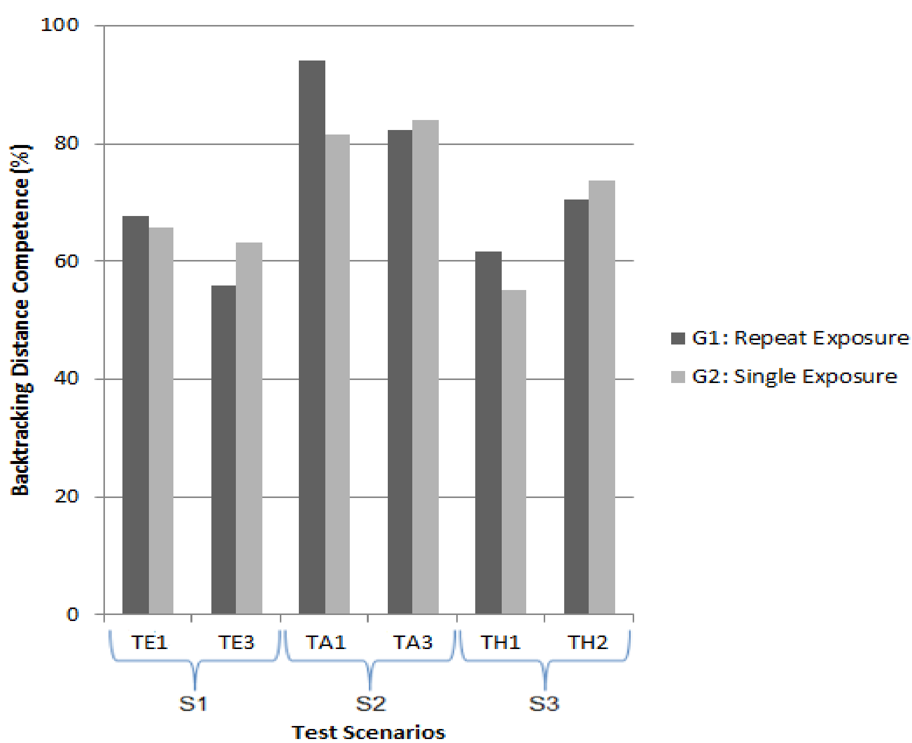

Figure 4-23: Cabin scenario backtracking distance competence scores for both groups.

In comparing normal to hazard conditions there was a significant difference for both groups (TA1-TH1) for both backtracking time and distance. For group 1 there was a significant backtracking time difference of 12.30s ($p = 0.002$) and backtracking distance difference of 10.87m ($p = 0.001$). For group 2 there was a significant backtracking time difference of 27.52s ($p = 0.006$) and backtracking distance difference of 16.55m ($p = 0.007$). For group 1 there was also a significant difference in normal to hazard conditions (TA1-TH2) for both backtracking time and distance with a difference of 19.83s ($p = 0.001$) and 13.20m ($p = 0.003$).

In comparing blackout to hazard conditions there was a significant difference for group 2 (TA1-TH1) for both backtracking time and distance. In group 2 there was a significant backtracking time difference of 30.46s ($p = 0.004$) and a backtracking distance difference of 17.78m ($p = 0.008$). No other statistical significance was found for backtracking time and distance for the cabin scenarios.

Backtracking Time and Distance - Engine Room Egress Scenarios

Figures 4-24 and 4-25 show the percent competence for backtracking time and distance to muster in the engine room scenarios for both groups. There was no significant difference between normal and blackout conditions for both groups in session 1 and session 2 engine room scenarios. In session 2, there was a trending difference in backtracking distance between normal and blackout conditions (TA2-TA4) for group 2 only ($p = 0.091$).

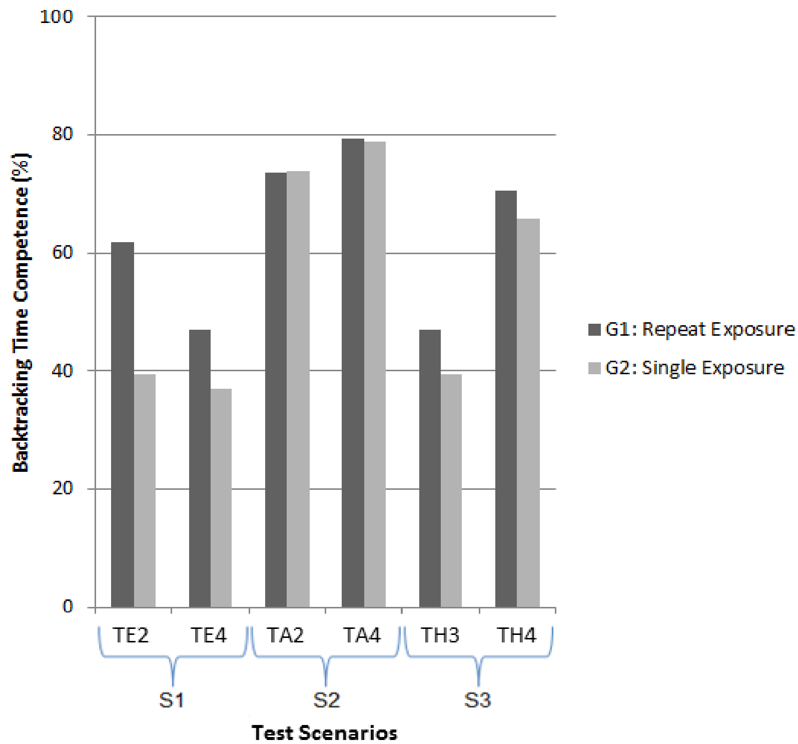

Figure 4-24: Engine room scenario backtracking time competence scores

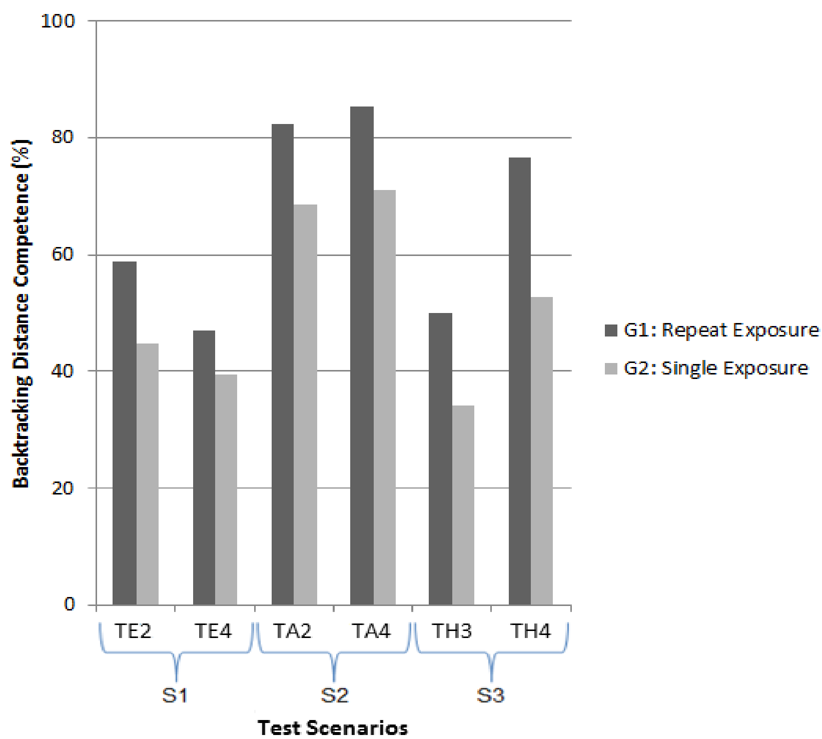

Figure 4-25: Engine room scenario backtracking distance competence score.

In comparing normal to hazard conditions there was a significant difference for both groups (TA2-TH3) for both backtracking time and distance. For group 1 there was a significant backtracking time difference of 56.94s ($p = 0.002$) and backtracking distance difference of 43.52m ($p = 0.002$). For group 2 there was a significant backtracking time difference of 75.86s ($p = 0.000$) and backtracking distance difference of 45.18m ($p = 0.001$).

In comparing blackout to hazard conditions there was a significant difference for both groups (TA4-TH3) for both backtracking time and distance. For group 1 there was a significant backtracking time difference of 52.58s ($p = 0.006$) and backtracking distance difference of 38.59m ($p = 0.002$). In group 2 there was a significant backtracking time difference of 70.65s ($p = 0.000$) and a backtracking distance difference of 44.10m ($p = 0.000$). For group 2 there was also a significant difference in blackout to hazard conditions (TE4-TH4) for backtracking time with a difference of -27.09s ($p = 0.014$). No other statistical significance was found for backtracking time and distance for the engine room egress scenarios.

3. Safe Practices:

Percentage Running:

Figures 4-26 and 4-27 show the percent competence for running in the cabin and engine room scenarios for both groups. There was no significant difference between normal and blackout conditions for both groups in session 1 and session 2 cabin and engine room scenarios.

In comparing normal to hazard conditions for the cabin scenarios there was a significant difference in percentage running for group 1. For group 2, there was a significant difference in percent running across all cabin scenarios. For group 1 there was a significant percent running difference between scenarios (TE1-TH2) resulting in a decrease of running by 8% (p = 0.007). There was also a trending difference between normal and hazard conditions for group 1 in scenarios (TA1-TH1, p = 0.055) and TA1-TH2, p = 0.081).

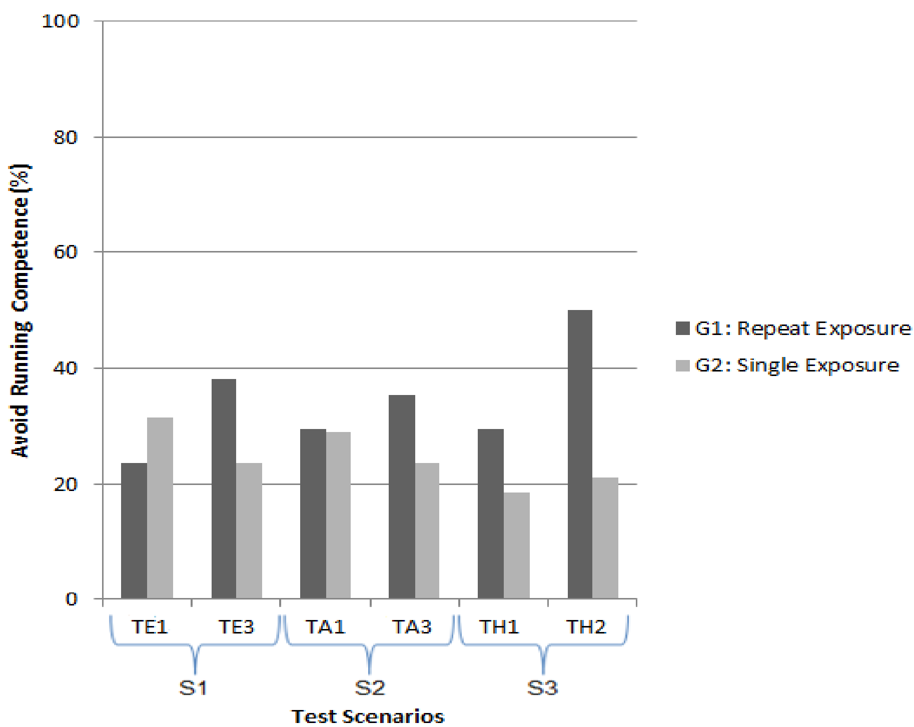

Figure 4-26: Cabin egress scenarios not running competence scores.

In comparing blackout to hazard conditions for the cabin scenarios there was a significant difference for group1 (TA3-TH2) for percentage running. For group 1 there was a significant percent running difference resulting in a decrease of running by 7%

(p = 0.035). No other statistical significance was found for percent running for the cabin egress scenarios.

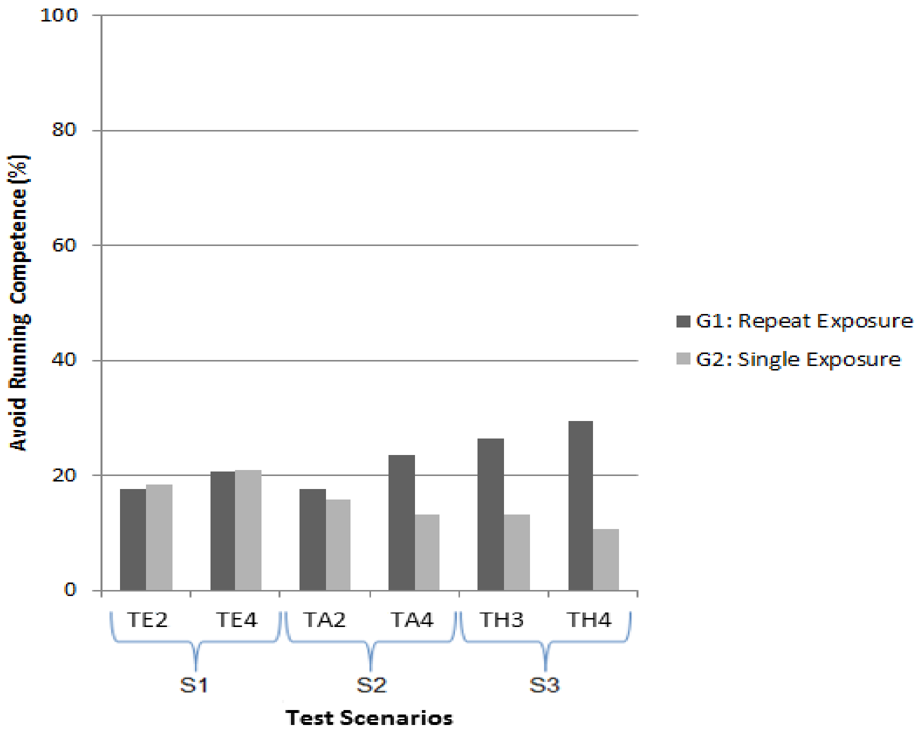

Figure 4-27: Engine room egress scenarios not running competence scores.

In comparing normal to hazard conditions for the engine room scenarios there was a significant difference for both groups (TE2-TH4 and TA2-TH4) for percent running. For group 1 there was a significant percent running difference of 4% (TA2-TH4, p = 0.026). For group 2 there was a significant percent running difference of 4% (TE2-TH4, p = 0.027).

In comparing blackout to hazard conditions for the cabin scenarios there was a significant difference for group2 (TE4-TH4) for percentage running. For group 2, there was a significant percent running difference resulting in an increase of running by 6%

(p = 0.024). No other statistical significance was found for percent running for the engine room egress scenarios.

Overall, group 1 percent running decreased across each session (slight performance improvement). Group 2 percent running increasing across each session (decrease in performance).

Closing Doors:

Figures 4-28 and 4-29 show the percent competence for closing doors in the cabin and engine room scenarios for both groups. No statistical significant difference was found for the door closing task in the engine room scenarios. Similarly for the cabin scenarios no significant difference was identified for the door closing percentage for both group 1 and group 2.

In session 1, 50% of group 1 closed the doors while 41% of group 2 closed the doors. In session 2, 54% of group 1 and 62% of group 2 closed the doors (improvement). In session 3, 63% of group 1 and 62% of group 2 closed the doors. There was an improvement in both groups for door closing from session 1 to session 2. For group 1 there was a slight performance improvement for closing doors across session 1 and session 2 from 50% to 54%. A more substantial performance improvement for group 1 occurred from session 2 to session 3 from 54% to 63%. For group 2 there was a substantial performance improvement for closing doors across session 1 and session 2 from 41% to 62%. However there was no further improvement for group 2 from session 2 to session 3.

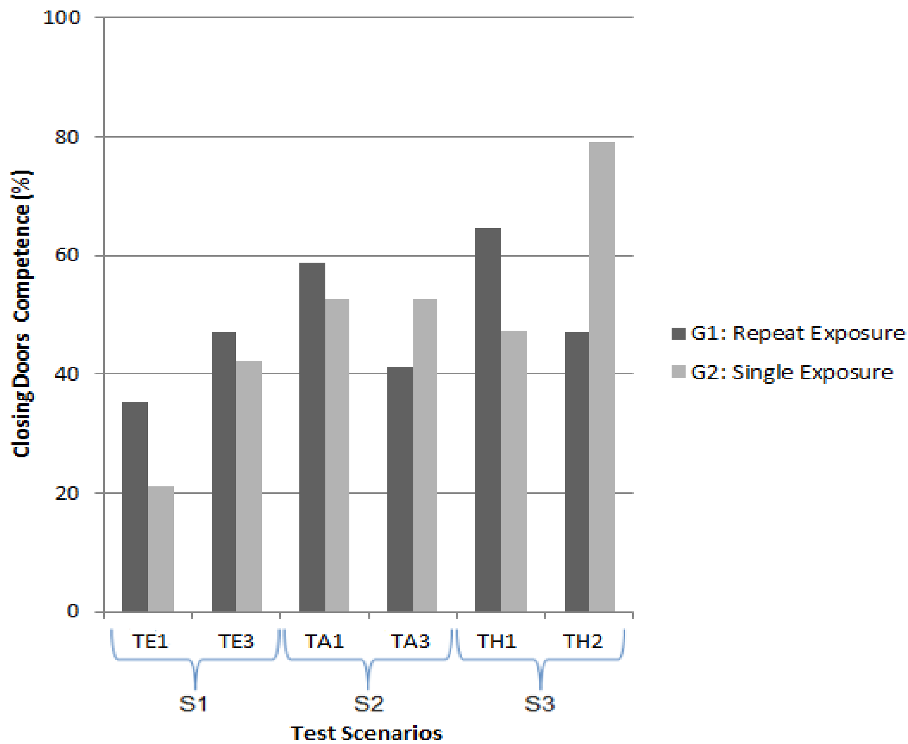

Figure 4-28: Cabin egress scenarios closing doors competence scores.

Overall, there was a significant difference between the first test scenario in session 1 and the two hazard scenarios in session 3. For group 1 there was a 29% increase in door closing (p = 0.025) between TE1 and TH1 (normal to hazard). While for group 2 there was a 58 % increase in door closing (p = 0.002) between TE1 and TH2 (normal to hazard) and a 37% increase in door closing (p = 0.020) between TE3 and TH2 (blackout to hazard). There was also a trending difference between the session 2 and session 3 door closing for group 2 (when comparing TA1 to TH2 and TA3 to TH2, p = 0.059). In session 3, there was also a trending difference of 32% (p = 0.082) between groups in scenario TH2. Group 2 received 32% higher competence in door closing than group 1. This is result is due the two groups route selection.

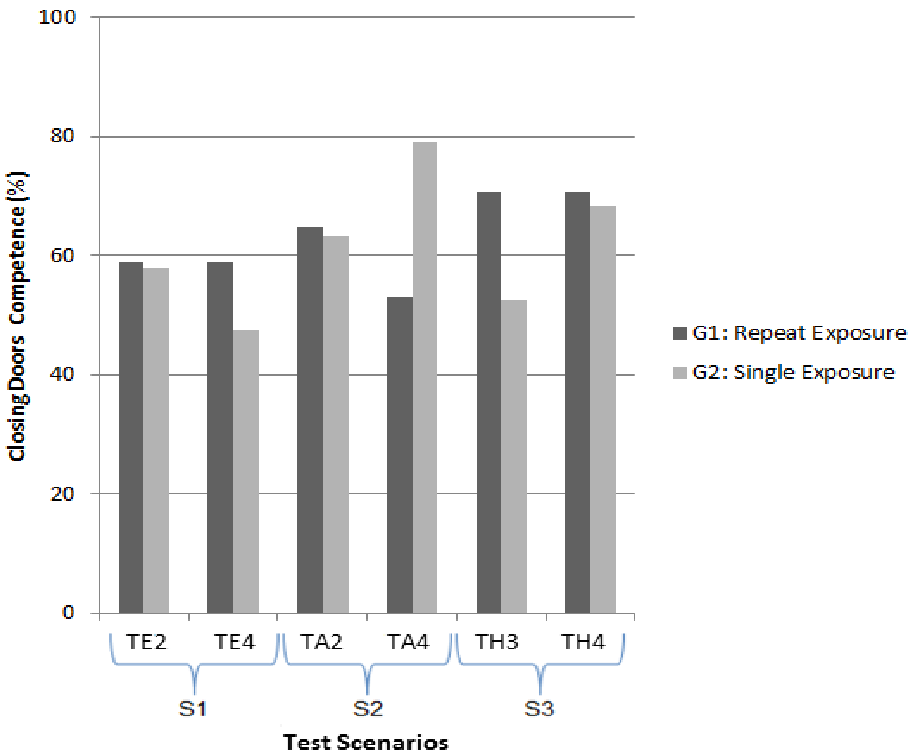

Figure 4-29: Engine room egress scenarios closing doors competence scores.

Overall, there was a significant difference between group 1 and group 2 in scenario TA4. Group 2 received a 26 % higher competence than group 1 in door closing.

Chapter 5 : Discussion

The goals of this study were to develop a strategy to assess performance in offshore emergency situations using virtual environments, measure time to competence, establish a baseline in task performance for novice individuals in simulated emergency situations and determine attributes of the virtual environment that activate learning. The main findings of this research were as follows: 1) performance in the test scenarios was lower than expected; 2) no major statistical significance was found in task performance between the two training exposure groups (with the exception of two scenarios out of a total of twelve); 3) basic offshore safety competence was not demonstrated for all six learning objectives; 4) time to competence for some learning objectives was achieved within the first session but for other learning objectives the time to competence is still unknown; and 5) learning occurred but at different rates for each group and each individual participant.

5.1 Achieving Competence in Basic Egress Safety Offshore

The main objective of the study was to develop a strategy to assess participant competence in offshore emergency situations using virtual environments. To measure competence and to understand the impact of virtual environment training on acquiring competence, two levels of training exposure were used. One group received repeated exposure while the other group received a single exposure to the training material. The participants in both groups were tested using the same test scenarios over the course of three sessions. The test sessions increased in difficulty across each session and ranged from finding a specific muster location to alarm recognition drills to full emergency evacuation scenarios. Multiple task performance measures were recorded and compiled to assess competence for six learning objectives. The learning objectives were categorized into two knowledge dimensions: spatial knowledge and procedural knowledge.

The spatial knowledge objectives included:

- Establishing Spatial Awareness (LO1)
- Routes and Mapping (LO3)

The procedural knowledge objectives included:

- Alarm Recognition (LO2)
- Assessing Situation and Avoiding Hazards (LO4)
- Registering at the Temporary Safe Refuge Area (LO5)
- General Safe Practices (LO6)

An aggregate score of the task performance parameters was developed using the performance data and a weighting scheme. Participants had to receive an 80% or higher in the aggregated competence score to pass.

5.1.1 Competence Was Not Demonstrated in All Six Learning Objectives

The results of the aggregate competency score show that participants were not able to demonstrate 100% competence. In fact, neither group was successful in demonstrating competence for all six learning objectives. As depicted in Figure 5-1, the mean scores for both groups did not reach the 80% competence standard for all three testing sessions.

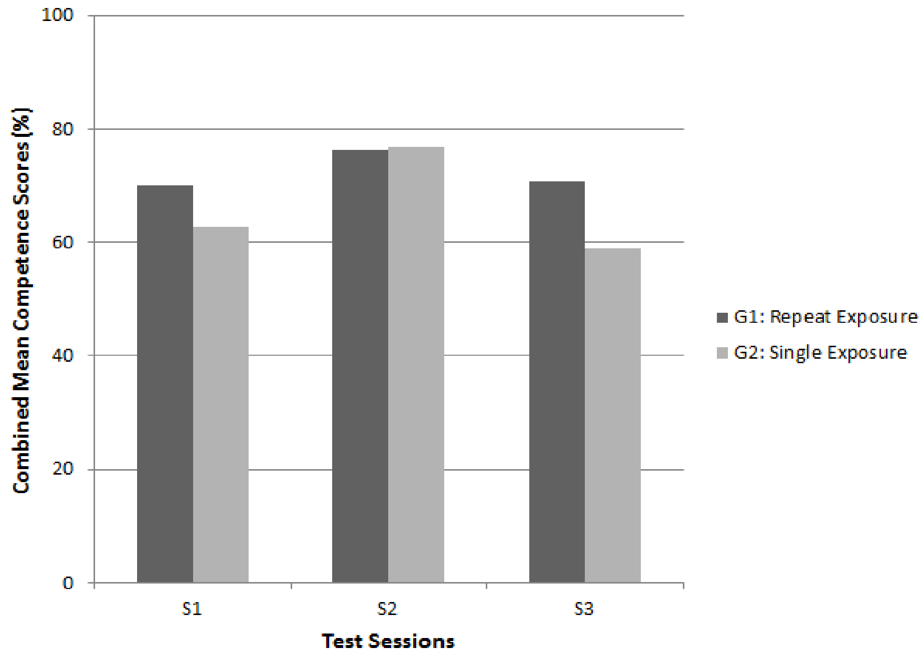

Figure 5 - 1: Combined mean competence scores for each group across sessions 1-3.

The overall mean competence reached for each session was 66%, 77% and 64% respectively. In session 1, the mean competence was 69% for group 1 and 63% for group 2. In session 2, group 1 and group 2 both scored a mean of 77%. In session 3, the mean competence scores dropped to 69% for group 1 and 60% for group 2. This shows that competence generally improved as participants progressed through each session from spatial awareness to alarm recognition. Substantial deterioration in task performance occurred when participants transitioned from session 2, alarm recognition to session 3, hazard avoidance. This result was unexpected. It was hypothesized that participants would be able to successfully demonstrate all learning objectives by the end of the third test session; however, this was the case for only a small minority of participants. Five people in total successfully achieved the minimum competence requirement in the final test session, four from group 1 and one from group 2. Three of the five participants demonstrated a mean competence score of 80% or above consistently across all sessions. However, only one participant was successful in achieving the 80% or above in all twelve test scenarios.

The step increase in scenario difficulty was used to help establish spatial learning and add procedural tasks step by step. However, only 4 participants in session 1 received a mean of 80% on the scenarios. This early poor performance suggests that few participants were grasping the basics, which included correct location, follow egress route, register at TSR and comply with basic safety practices. In session 2, 18 participants received a mean of 80% on the test scenarios. This increase in mean competence score showed a promising sign that learning had occurred and that the participants were grasping spatial and procedural knowledge required for basic emergency egress. It also is an indication that participants were starting to demonstrate their competence in relatively normal conditions (less stressful). By sessions 3, only 5 participants received a mean of 80% on the test scenarios. This indicates that the participants were unable to keep up with the increases in complexity from session 2 to session 3. There are several reasons that could contribute to the decline in performance:

1. The increase in difficulty was too abrupt for participants to demonstrate competence (steeper testing curve). The final test scenarios in session 3 were more challenging for the participants than anticipated and resulted in a larger step increase in minimum competence requirements than the participants were able to demonstrate.
2. Participants may have had difficulties retaining the initial training or participants may not have had a strong foundational knowledge both spatially and procedurally to handle the emergency egress scenarios in session 3.
3. Participants may not have been able to demonstrate spatial or procedural tasks in high stress emergency situations.
4. Due to passive reporting of feedback, participants may not have been adequately corrected for poor task performance that likely propagated into larger mistakes in the final test scenarios.
5. Participants may have become complacent with the study and not taken the final situations seriously.

5.1.2 Competence Achieved by Learning Objective

Overall, participants were not successful in demonstrating competence in all performance tasks in the study. The majority of participants were able to successfully demonstrate the following: reaching the correct muster location (LO1 – Spatial Awareness), recognizing the alarm type (LO2 – Alarm Recognition), registering at the temporary safe refuge area (LO5 – Register at TSR), and avoiding hazards blocking their path (LO4 – Assess Situation). However, participants were unable to successfully demonstrate the remaining competence: recognizing the hazardous situation and raising the alarm (LO4 – Assess Situation), resisting running on the platform and ensuring all fire or watertight doors were closed (LO6 – Safe Practices).

Spatial Knowledge (LO1 & LO3)

Overall, participants were able to reach the correct location in the cabin and engine room scenarios by the second session, with the exception of two outliers. This result indicates that participants were able to establish the landmark knowledge of the environment from Seigel and White's (1975) Landmark-Route-Survey spatial knowledge acquisition model. Some participants were unable to demonstrate the same level of spatial knowledge in the emergency situations in session 3. This degradation in performance raises doubts about whether foundational spatial knowledge was sufficiently established for emergency egress situations.

Deviations from egress routes were reduced from session 1 to session 2 as participants became more familiar with their primary and secondary routes from their cabin and worksite. This result indicates that participants were developing route knowledge of the platform. Route knowledge seemed to be successfully demonstrated for at least one egress route in the cabin scenarios, which involved a simple building environment structure. However, based on participant behaviours in the cabin scenarios, 53% of participants selected the same preferred route throughout all cabin test scenarios.

This result indicates that fewer than half of participants demonstrated that they were able to egress from their cabin using both their primary and secondary egress routes.

In the engine room scenarios, which represent a more complex spatial environment, the route knowledge was harder to establish for all egress options. Based on participant behaviours in the engine room scenarios, 11% of participants selected the same preferred route throughout all the worksite test scenarios. The small percentage selecting the same route throughout the scenarios may be an indication that participants were more active in learning multiple egress options from their worksite. In some cases it is more likely that the smaller percentage of participants selecting the same route is due to participants not having a firm grasp on even one preferred egress route. Some participants were prone to becoming lost in the engine room scenarios.

Route and survey knowledge was tested in session 3, with final test scenarios that were designed to be complex emergency situations to investigate how people behave in emergency situations. These test scenarios were designed to block primary or preferred routes at both the cabin and engine room locations on the platform. To test route knowledge the preferred route was blocked at the start of the route forcing the participant to select the next available alternative. The participants' route selection in these emergency situations gave an indication of their route knowledge and how well they remembered their route options. The route selection and re-routing that took place in the final test scenarios suggest that some participants did not have a firm grasp on the egress routes available and did not have a strong understanding of the survey knowledge of the platform.

To test survey knowledge, the preferred or primary route was blocked mid route (TH1 & TH3). The extent to which the participant's re-routing deviated from the ideal re-routing when encountering the hazard informed how well the participants understood the layout of the environment and how it interconnected on a survey knowledge level. The most efficient re-routing would reflect a thorough understanding of survey knowledge. In these scenarios there were several participants in both groups who demonstrated a lack of survey knowledge development. These behaviours included going up or down 4 to 5 decks to retrace steps to the starting point, choosing alternate route options, and passing alternative connecting routes that were not necessarily defined or designated routes. These results are consistent with Kobes et al. (2010) findings of participants passing the closest fire exits to take the main route they knew well.

Kobes et al. (2010), citing others (Graham and Roberts, 2000; Sandberg, 1997), found that in emergency situations people tend to evacuate buildings from the main entrance or by taking their preferred known route. In Kobes et al.'s (2010) study some participants passed available and closer fire exits, opting to travel through the smoke hazard to take the route they knew well. This behaviour seems to occur when there is poor spatial knowledge, especially survey knowledge. Instances of these behaviours were observed in the final evacuation scenarios. In situations where spatial knowledge was limited or when stress took over, participants who were faced with blockages on all known egress options took on more risk and entered into the hazard in the direction of their preferred route. Individuals who preferred to follow to the same route in every scenario showed a behaviour associated with low route and survey knowledge in emergency situations. Participants who varied their route selection between scenarios seemed to be more comfortable with their spatial representation of the virtual environment and were successful in re-routing when encountering hazards blocking their path.

The aforementioned behaviours are indicators of spatial understanding and should be tracked automatically in a virtual environment in order to provide more comprehensive feedback and adjust the training scenarios to help expedite the development of route and survey knowledge. In the study, tracking and performance assessment of route selection and re-routing were performed by researcher observation and post scenario analysis. Further improvements to AVERT and the automated brief/debrief system are required to automate route tracking and performance assessment. Built in algorithms are required to track and compare route selection and re-routing in real-time to the most efficient route available. This integration is important in order to evaluate the participant’s route and survey knowledge and to provide them with the necessary prompt and relevant feedback.

Procedural Knowledge (LO2, LO4, LO5, & LO6)

Competence in alarm recognition and registering at the TSR are linked to the correct location learning objective. Overall, participants were able to recognize the alarm type and register at the correct muster location within the TSR for both the cabin and engine room scenarios. The outliers showed several instances where participants went to the wrong location or forgot to move their T-card correctly in the final test scenarios. It is possible that these behaviours are linked to participants forgetting procedural steps in stressful situations.

With regards to the fourth learning objective - Assess Emergency Situation and Avoid Hazards, participants were largely unsuccessful in raising the alarm. Only one participant was able to apply sufficient procedural and spatial knowledge, assess the situation and remember where the manual alarm call point was located to raise the alarm. This poor result is likely due to the steep increase in difficulty from session 2 to session 3. The increase in procedural tasks required of them to perform is likely the cause of participants not being able to properly raise the alarm. Participants were more successful in navigating away from hazards blocking the path. The outliers in this case exhausted all known route options and entered into known hazardous areas to take the only route they knew. This behaviour is a symptom of the participant not having a firm understanding of survey knowledge of their environment.

1) The Safe Practices learning objective was the poorest demonstrated of all the competencies. Participants showed improvement (20-30% increase in performance) across all sessions in remembering to close watertight and fire doors. Participants were consistently unable to demonstrate even 40% competence for not running on the platform. There are several reasons that could contribute to the behaviour observed:Participants reported that the interface was challenging to use when opening and closing doors resulting in some participants becoming frustrated with doors.
2) Participants reported that the walking speed seemed too slow and resulted in participants preferring to run in the scenarios.
3) Due to the stress of emergency situations, some participants may have forgotten the procedural task of closing doors or in some cases may have started running due to the sound of the PAPA alarm.
4) The after-action review feedback that participants received was an optional and passive form of feedback. As a result, participants were not receiving a firm message that running was not acceptable and these behaviours persisted throughout the test scenarios.

To address some of these issues in future virtual environment training, more emphasis on performance feedback is required. The procedural learning objectives require more stringent reinforcement during training scenarios to ensure knowledge acquisition and to ensure mistakes are not propagated throughout the performance assessment. The delivery of virtual environment training and assessment can also be improved to promote spatial and procedural learning. A more gradual addition of procedural tasks between training sessions may be required. Instead of progressing from drills in session 2 to major accidents in session 3, a proposed new incremental step would go from alarm recognition to minor incidents and better equip trainees for major accidents scenarios.

For future studies, it is important to be able to identify which participants are having trouble and what their main areas of difficulty are in order to provide these individuals with specially designed feedback and practice scenarios. Given that helping crews train and overcome stressful situations in a safe virtual setting is the main goal of AVERT, further development of features to determine when trainees are struggling and to automate scenarios to target specific deficiencies is necessary for future virtual environment development and research.

5.1.3 Time to Competence:

Another objective of the study was to investigate the training time required to reach a competence in basic safety in offshore emergency egress using AVERT. The hypothesis was that participants would reach basic competency in offshore emergency egress using AVERT within the allocated time of the study, after completing all three sessions. This hypothesis assumed that the time required for the majority of the participants to achieve competence would be measurable during the time allocated for the study.

Based on the time allocations used in other studies, the groups received 60-90 minutes of cumulative total exposure to the AVERT virtual environment. The difference in cumulative exposure between the two groups was an additional 30 minutes practice time received by group 1. Participants in group 1 were allocated 40 minutes in each session to practice, however most of group 1 did not make use of the full practice time. This exposure was in line with previous studies (Bradbury-Squires, 2013; Darken and Peterson, 2001) and provided a much longer exposure to the virtual environment than other spatial learning studies (Darken and Perterson, 2001 citing Witmer, Bailey, and Kerr, 1995; Koh, et al., 2000; and Waller et al., 1998). Even though this study allowed participants more time with the virtual environment than most presented in the literature review, measuring time to competence still fell beyond the time allocated for the study. The study was unable to definitively determine the time required to achieve competence in basic safety in offshore emergencies using AVERT for two main reasons: 1) neither group was successful in achieving competence in all learning objectives and 2) based on the behaviours observed by outliers in the study (e.g. some participants that were consistently unable to perform in the spatial or procedural tasks), individual differences in learning pace and style likely played a large role in participants not demonstrating minimum competence.

For these reasons, time to competence as whole is not yet known for all participants and exceeds the time allocated in this study. Looking at each learning objective individually, as was done in the previous section, helped understand which learning objectives were easier to achieve within the allocated time and which ones would require more time. As anticipated, spatial knowledge acquisition takes time. Researchers have indicated the potential for virtual environments to eventually surpass map accrued spatial knowledge with sufficient exposure to the virtual setting (Darken and Peterson, 2001; Waller et al., 1998). However, this study was unable to definitively support this theory because the training exposure participants received was standardized (e.g. participants received a set amount of time in the virtual environment regardless of their individual differences in spatial learning) and in some cases participants were not given sufficient time to reach spatial competence. Therefore, a limitation in this study is that spatial knowledge acquisition is heavily influenced by individual learning styles and pace. Although some participants were able to gain sufficient spatial knowledge to demonstrate competence in the test scenario, others were not able to do so in the time allocated. Individual learning styles should be taken into consideration in the design of future studies in order to determine the required exposure time to virtual environments necessary to reach competence. In addition to this, on an individual learning objective basis, competence that was demonstrated in normal conditions was not always demonstrated in full evacuation situations. The increase in scenario complexity is something that should also be addressed in future studies to help trainees reach spatial understanding more quickly regardless of the circumstances.

Procedural knowledge acquisition requires the trainee to understand and perform smaller components of a procedural separately followed by the entire procedure, and receive feedback throughout the process (Grantcharov and Reznick, 2008). This was generally the case for participants who received repeated exposure to training. However, the poor procedural performance observed in this study is primarily due to the quality of feedback the participants received and the absence of reflection exercises after each test scenario. This is something that should be addressed in future studies to help trainees reach and exceed the minimum level of competence in a short time frame with reinforcement of learning objectives.

5.2 Impact of Training Exposure on Task Performance

Two different exposure levels to training were used to assess the impact of the training factor on task performance. The hypothesis was that participants who spent more time training (G1 - repeated exposure to training) will perform better from a task performance perspective in the emergency response testing scenarios than participants who spent less time training (G2 - single exposure to training). In reviewing the results of the task performance it was found that there was no significant difference between the groups with the exception of two scenarios (denoted TA2 and TH2).

5.2.1 Statistically Different Groups Task Performance:

Only one scenario was found to be statistically different in competence score between groups, denoted as TH2. In this cabin egress scenario, group 1 achieved 9% higher competence score than group 2 ($p = 0.020$). This competence score difference is attributed to group 1 reaching the correct muster location in a shorter time to complete, while avoiding hazard exposure and running significantly less in comparison to group 2. Even though group 1 outperformed group 2 in the aforementioned task performance measures, group 1 did receive a lower score in route selection based on the emergency situation and in closing all fire and watertight doors.

Another scenario (denoted TA1 – the first scenario of session 2) was not statistically significant from a competence score perspective but was statistically different between groups on multiple parameters (specifically time to muster, distance travelled to muster, route selection, backtracking time and backtracking distance). The reason why scenario TA2 was statistically significant on multiple parameters and not statistically significant from competence score is due to the weightings of the tasks and their overall importance to competence in emergency egress.

In this cabin drill scenario, group 2 was a mean of 30 seconds slower ($p = 0.036$) and travelled 14 meters more ($p = 0.005$) than group 1 to complete the scenario. Group 1 selected the primary egress route 30% more than group 2 (trending $p = 0.065$) and as a result spent less time backtracking (12 seconds, $p = 0.004$) and travelled less distance backtracking (4 meters, $p = 0.006$). This difference in performance at the beginning of the second session may have been a sign of group 2 not being able to retain the information from the training session. In contrast, group 1 had the opportunity to review the information prior to testing. Group 2 was quick to learn from the first scenario and improved their overall performance to be statistically the same as group 1 in session 2. Group 2 was able to catch-up in performance just based on the scenario and feedback alone. The test scenarios themselves may have been a learning tactic for group 2 to help facilitate their comparable performance to group 1 in the scenarios that followed.

The two scenarios where statistical significance was found are cabin scenarios. There was no statistical difference found between groups for the engine room scenarios.

This is an indicator that both groups were equally prepared for the engine room scenarios and that the additional practice time provided no advantage to group 1 in gaining additional spatial knowledge of the engine room.

There are several possible reasons for observing little to no statistical difference between groups in spatial and procedural tasks:

1) The basic safety training provided to both groups in session 1 was sufficient to provide equal performance across groups.
2) The practice scenarios group 1 received were not sufficient to differentiate task performance during the test scenarios.
3) The difficulty of the test scenarios was an effective training or learning tactic for group 2 to catch up to group 1 in task performance.

5.2.2 Notable Group Differences in Spatial Task Performance:

Although the majority of scenarios were not statistically different in task performance between groups, there were other notable differences in behaviour. The following describes the group differences by spatial and procedural tasks. In some cases group 1 appeared to be a quicker than group 2 to demonstrate competence in the test scenarios (within the first session as opposed to by the time the participants perform the second session). This may be due to the practice scenarios they received, for two reasons: 1) the practice scenarios gave group 1 a more comprehensive spatial knowledge due to more exposure with the virtual environment, and 2) the practice scenarios allowed group 1 the opportunity to make mistakes prior to being tested. The advantage to the practice time in

AVERT appears to have its limits as group 2 in some cases was able to reach equal performance as they progressed through scenarios.

Reaching the Correct Muster Location

There was no statistically significant difference between groups in the task of reaching the correct muster location. In general, group 1 successfully demonstrated this task midway through session 1 and maintained their performance throughout session 2 and session 3. Group 2 successfully demonstrated this task by session 2 but it appears the group's task performance degraded in session 3. This lag in demonstrating performance is likely to do with training retention or differences in spatial understanding.

Proficiency in Reaching Correct Location

Total time and distance travelled are two measures of how proficient the participants were at reaching the correct muster location. In comparing group performance, in one cabin egress scenario there was a significant difference between groups (denoted as scenario code TA1). In the more complex environment (engine room egress) scenarios there was no statistical significance found between groups for the time to muster and distance travelled.

The fact that there was a difference in groups in the less complex environment and not in the more complex environment suggests that the two groups had an equal understanding of the engine room spatial layout. It also may be that group 2 had a significantly slower performance in the first scenario of the second session due to issues in remembering the environment and not so much due to their understanding of the accommodation block.

Route Selection and Re-routing

There was no statistical difference between groups in route selection. There was however a trending difference in groups for scenario TA1: in group 1, 88% selected the primary and 12% selected secondary egress routes; in group 2, 58% selected the primary and 42% selected secondary egress routes. In general, group 1 seems to be more flexible when selecting routes in the cabin test scenarios. For the cabin scenarios, 53% of group 1 and 26% of group 2 varied their route selection. In contrast, 41% of group 1 and 63% of group 2 participants selected their preferred route in every cabin test scenario. Others had a tendency to follow the same route (observed in 5 out of 6 scenarios). 6% of group 1 and 11% of group 2 followed the same route for most of the cabin scenarios. It is possible that group 1 participants were more adaptable to altering their route selection due to an increased knowledge of the environment. Conversely, group 2's fixed route selection might be due to a weaker knowledge of the environment and their lack of confidence in knowing the route options.

Although the groups were not statistically significantly different in their route selection, there were interesting behavioural tendencies by both groups. In general, both groups were equivalent in route flexibility in the more complex environment scenarios. For the engine room scenarios, 35% of group 1 and 47% of group 2 varied their route selection. Similarly, 29% of group 1 and 32% of group 2 selected the same route in every engine room scenario. In the engine room scenarios there were also participants who showed a tendency to follow the same route (observed 5 out of 6 scenarios): 29% of group 1 and 21% of group 2 followed the same route for most of the engine room scenarios. Group 1 also had 6% of participants lost in every engine room scenario. It appears both groups had less familiarity with the engine room than the accommodation block. More participants became lost in the engine room scenarios, especially in the complex emergency scenarios. Thus, the smaller percentage of participants selecting the same route each time may be due to both groups having trouble developing spatial knowledge of one preferred route. The variation in route selection for the engine room scenarios may be due to poor route and survey knowledge by both groups.

This difference in participant route preference is likely due to two factors: 1) group 1 may have established a better understanding of the route options for the cabin scenarios with additional exposure to the training material and practice scenarios and thus had more spatial knowledge to draw on, and 2) group 2 was not given sufficient time with the environment to establish sufficient spatial knowledge or to feel comfortable enough to take a less known route in the emergency situation. Due to the added complexity of the engine room scenarios, participants likely required more survey knowledge of the engine room in order to properly identify the route options. As survey knowledge takes longer to acquire, both groups likely did not have enough time to build the necessary spatial model of the engine room to help remember egress route options for their worksite scenarios. It is also possible that the lack of difference between groups in spatial knowledge of the worksite may be due to the repeated training group (group 1) not using the full amount of time available to them in the practice scenarios. If participants in group 1 had taken full advantage of the practice time available to them, it is possible that more survey knowledge of the worksite in the engine room could have been established.

5.2.3 Notable Group Differences in Procedural Task Performance

Assess Emergency Situation and Avoid Hazards on Route

There was no statistically significant difference between groups in raising the alarm. In general, the performance in this task was below expectations. The majority of participants were unable to recognize their procedural responsibility to raise the alarm in the scenario. Only one participant (from group 1) was able to successfully raise the alarm in scenario TH4. Other participants who attempted to raise the alarm had spatial issues in finding the location of the manual alarm call point. Two participants from each group attempted to raise the alarm but were unable to successfully raise the alarm within the time limit for the scenario. Three other participants from group 2 tried to raise the alarm but were not at a proper MAC station necessary to raise the alarm. This poor performance is a result of unclear test scenario goals in which participants were conditioned to listen for the alarm and PA announcement and then follow their egress route. When the scenario deviated from all previous scenarios to test them on raising the alarm, most participants mistook the cues of the situation requiring them to raise the alarm and proceeded to follow their egress route without alerting the emergency response team of the observed hazardous situation. Another possible reason for the poor performance in raising the alarm is due to weak spatial knowledge of the MAC stations. More training and clearer scenario instructions are required to assure competence in the more complex scenarios.

There was a trending statistical difference between groups in hazard exposure for one scenario (TH2). In this scenario, group 2 participants accepted more risk and were exposed to more smoke than group 1 ($p = 0.051$). This increase in smoke exposure may be a result of the preferred route behaviour observed in the test scenarios. Regardless of the hazardous situation blocking the secondary route, which was the preferred route for some participants, 42% of group 2 and 11% of group 1 proceeded to take the route most familiar to them. This is likely why there is a trending difference in hazard exposure because almost half of group 2 did not properly re-route and were subsequently exposed to smoke on their preferred path. This is again a symptom of poor route and survey knowledge. More exposure to the environment would likely benefit group 2 and help them acquire with route and survey knowledge.

Safe Practices

There was no statistically significant difference between groups in percent running with the exception of two scenarios (trending for TH2 and TH4). Although statistical significance was not found, the results show that group 1 participants' performance gradually improved and group 2 participants' performance degraded in both the cabin and engine room scenarios.

For the cabin scenarios, the amount of running by group 1 participants decreased across each session. For group 1, there was an 8% decrease ($p = 0.007$) in running from session 1 (TE1) to session 3 (TH2), and a 7% decrease ($p = 0.035$) from session 2 (TA3) to session 3 (TH2). For group 2, the participants increased the amount of running across each session, especially in the final emergency evacuation scenarios. For the engine room scenarios, the amount of running by group 2 participants increased across each session. For group 2, there was a 4% ($p = 0.027$) and 6% ($p = 0.024$) increase in running from session 1 (TE2 and TE4) to session 3 (TH4). For group 1, the participants decreased their running by 4% ($p = 0.026$) from session 2 (TA2) to session 3 (TH4). For the two emergency situations scenarios (TH2 and TH4) there was a trending difference between groups in percent running. It is possible that these emergency situations evoked an increased running behaviour in group 2. This opposite performance by the two groups is likely a result of the corrective feedback the groups received. Group 1 was reminded in the training tutorials not to run on the platform. Group 2 only received the after action review feedback indicating the time spent idle, walking, and running with no explicit feedback advising them not to run on the platform.

There was no statistical difference between groups in closing doors. There was a trending difference between groups for TH2 ($p = 0.082$). In this scenario, 47% of group 1 and 79% of group 2 correctly closed all fire and watertight doors. This difference in door closing is linked to the two group’s preferred route selection. In this scenario, group 1 preferred to take the primary route and group 2 was more likely to take the secondary route. The primary route in this case has more doors to open and close and subsequently more room for error in forgetting to close doors.

Overall, both groups gradually improved in closing doors across scenarios. For session 1 to session 3, group 1 participants improved by 13% and group 2 participants improved by 21%. Both groups were only able to reach 62% competence in closing doors by the end of the study. This is not entirely due to participants forgetting to close the doors and can be partly attributed to how participants were expected to interact with the doors in the virtual environment. Many participants reported difficulty interfacing with doors and frustration with the doors not closing when they were actively trying to close the doors. Future development of the virtual environment will need to address the deficiencies in interacting with objects within the environment.

5.3 Measuring Learning in a Virtual Environment:

The final objective of the study was to measure learning in a virtual environment and determine what attributes helped contribute to learning.

5.3.1 Learning Occurred Across Sessions

Although both groups were unable to archive 100% competence, the results show that learning occurred and performance improved. Most of the observed performance improvements occurred as participant progress from session 1 to session 2. An overall degradation of performance occurred from session 2 to session 3. These results indicate that the step progression from session 1 to session 2 facilitated learning and therefore an improvement in task performance. Conversely, as a result of the steep increase in scenario complexity and the increase in required emergency duties from session 2 to session 3, the participants did not demonstrate improvement. Any learning benefit from the test scenarios themselves was lost in the third session due to the difficulty of the scenarios. A more manageable increase in emergencies duties and scenario complexity may have benefited both groups. For example, adding another training and testing session between session 2 and session 3 on minor emergency situations with minimal hazards could help support better understanding of the hazard avoidance learning objectives. This approach would balance the required emergency duties between the two sessions while gradually increase the scenario complexity. The minor emergencies session would assess the trainees' ability to raise the alarm, follow mustering protocol and re-route in normal and blackout conditions. This smaller step in training would promote learning and allow the next training session to focus on major emergencies. The major emergencies session could remain the same as session 3 of the study and would assess the trainees' ability to identify hazards, respond to emergency situations and re-route in the event that the primary egress route is blocked by a hazard.

Another reason for the poor performance in the third session may be that the participants were able to demonstrate some spatial and procedural knowledge in less stressful situations but unable to do so in more complex emergency situations. For example, some participants forgot a procedure, suddenly lost spatial awareness or never had enough spatial awareness to manage the complexity of the evacuation situation. This reaction of some participants to the more complex emergency situations may be due to the participants' response to stress or the scenarios complexity exceeding their cognitive load. Virtual environments have the potential to train participants to be able to build-up exposure to stressful situations as required for real offshore emergency situations. Data analysis of the participants' physiological responses collected during the test scenarios is beyond the scope of this thesis. Future studies should take into consideration the impact of scenario complexity on trainee cognitive load and stress. Training and test scenarios should be designed to gradually build-up complexity and match trainees' expectations.

5.3.2 Attributes of a Virtual Environment that contribute to learning

Several benefits and disadvantages to the training and technology were identified after observing participants interface with the training material and virtual emergency scenarios using the learning management system. The benefits of the AVERT training included: 1) the overall training objectives were easy to understand; 2) participants were able to progress through the course material independently and at their own pace; 3) the platform walkthrough scenarios used navigation aids to help participants establish spatial knowledge; 4) the test scenarios were engaging and acted as a learning tactic for group 2; and 5) the after-action review helped correct some erroneous behaviours (although not all) such as reaching the correct location, and closing doors.

Guidance from Grantcharov and Reznick (2008) was followed in course design and feedback, however further improvements can be made to maximize the utility of virtual environment training. The following are areas where improvements can be made: organization and delivery of training; design and development of features in the technology to guide the trainee in-simulation; and overall experimental design improvements. Recommendations for improvement to training and virtual environment design include: 1) recognizing individual difference in skill acquisition, learning and retention and tailoring virtual environment features to individual learning styles; 2) engaging trainees and avoiding complacency; 3) designing training modules using mastery of learning approach; 4) designing research experiments for multiple applications; and 5) importance of feedback and time to reflect on performance.

Recognizing Individual Differences in Skill Acquisition, Learning and Retention

Spatial learning occurred at different rates for many participants in this study. As Weisberg et al., (2014) suggest each individual is different in the amount of time and effort required to acquire spatial knowledge. Consequently, the time allocated in this study may not have been sufficient to accommodate individual differences in acquiring spatial knowledge. As a result, it is possible that individual differences in spatial learning attributed to the low competence scores.

In general, some individuals required more exposure to virtual settings to ensure knowledge retention. Some participants regardless of the training exposure, had difficulty with spatial or procedural tasks, which required additional time and training to properly address. Some examples include: 1) the individual who was unable to correctly reach the muster location (in 8 of the 12 test scenarios); 2) the individual who chose to run through the hazard instead of finding an alternative route; and 3) the individual who forgot how to muster in the emergency situation. These cases are representative of the finding that the cumulative total exposure time in AVERT used in this study may not be sufficient to gain competence in all six learning objectives (for all participants or for the majority of participants). Therefore, the training model employed for both groups in this study did not accommodate individual learning needs.

Its important to recognize these individual differences in skill acquisition as this will impact how virtual environments are designed. To facilitate different learning styles, virtual environments should record the aforementioned types of behaviours in-game in order to customize training scenarios to support the individual needs. Virtual environment features that can support individual learning styles include: active walkthroughs with artificially intelligent guides (instead of orientation videos), in-game notifications when incorrect behaviours are happening, and allowing trainees to progress through the training content at their own pace by unlocking achievements and tailoring training to address individual behavioural tendencies.

Engaging Trainees and Avoiding Complacency

Trainee engagement is important for learning to occur using virtual environments. The decline in trainee engagement may have influenced the degradation in performance for some participants. Part of trainee engagement involves the overall attentional resources of the trainee. If the attentional resources are exceeded due to an overload of information, then loss of learning opportunities can occur. For the purposes of the study, much of the training was condensed into basic training tutorials at the onset of the first session. These training tutorials provided a large amount of information over the course of a two hour period. This may have exceeded some participants' attentional resources resulting in some loss in trainee engagement and subsequently a loss of learning or retention due to lack of interest. It is important to design the frequency and delivery of virtual environment training so as not to exceed the attentional resources of the trainee.

Trainee engagement can also be improved with the increase of AVERT's simulation fidelity, particularly from an interaction to the environment perspective. Improving the environment interaction realism would help address the difficulty some participants reported regarding opening and closing doors. From a procedural knowledge perspective, adding interaction functionality in AVERT that allows participants to interact and equip themselves with safety equipment should be the next step for software development. This would provide added contextual realism and improve trainee engagement for the participants.

An interesting theory that could impact both groups with regards to the decline in performance in session 3 is participant complacency. The participants in this study were novice (no prior experience in the offshore industry) volunteers. This study involved a considerable time commitment with little to no incentive aside from the participants learning something new in the field of emergency egress. It is possible that some participants by the third session were less engaged in the study and wanted simply to 'get through' the exercises and put less effort into their performance. It is possible that this phenomenon would be less likely to occur in participants who worked in the offshore or maritime industries as they would have a vested interest to succeed based on job safety. Although this is outside of scope of the current study, it would be interesting to see how differently (if at all) personnel from the offshore work environment background might perform in comparison to the baseline of novice individuals. As some researchers have already demonstrated (Magee et al., 2012), this question may be answered in a reverse transfer study using experienced personnel.

Designing Training Modules using Mastery of Learning Approach:

A critical process to help trainees reach the necessary level of competence is to improve training delivery by implementing a mastery of learning approach with reinforcement of learning in quality feedback. This mastery of learning concept is based on educational scaffolding principles (Sawyer, 2006). The mastery of learning approach would provide trainees with the foundational knowledge necessary to build on for more advanced learning objectives and support the development of competence in more complex situations. The master of learning approach would only allow the trainee to advance to the next learning objective once they are able to demonstrate the minimum acceptable competence of their current learning objective or task. This would help avoid trainees being tested on more advanced elements when they have not successfully demonstrated foundational requirements (such as spatial knowledge in this study).

It is also important to provide transparency of competence scoring so that the participants know what they are being graded on. The performance measures used in this study were described to participants at the beginning of the study. However, the scenarios did not adequately brief the participants on the targeted learning objectives of the scenario. The after-action review feedback only provided the participants with a subset of the overall performance measures used in the competence score. If the feedback had been more detailed to encompass all the performance measures, it is more likely that participants would have been aware of how well they were performing against the performance standards.

To support the mastery of learning teaching approach, in-simulation training aids and automated performance assessment and feedback should be integrated in to the virtual environment training program. This configuration would support spatial and procedural knowledge development by allowing a trainee to demonstrate his/her capabilities in one foundational competence before moving on to more advanced competence requirements. Pairing mastery-of-learning training with detailed and automated feedback is necessary in order to identify trainee deficiencies and customize training to address these deficiencies.

Designing Research Experiment for Multiple Applications

This study was designed to inform two distinct research projects: 1) assess competence in offshore emergency egress using virtual environments and 2) to collect human reliability data in offshore emergency situations. To accommodate both forms of data collection, compromises in the experimental design were made. To support learning, the order of the test scenarios was chronological and gradually added procedural tasks. This approach was used instead of randomizing the test scenarios the participants received. As a result, the design affected the analysis of how the performance shaping factors, such as visibility and complexity, impacted human behaviour in emergency egress. Conversely, to support a more naturalistic response to emergency situations, the two group's exposure to training from a human reliability analysis perspective was categorized as advanced training and minimal training. This design affected the amount of corrective instruction group 2 received in order to see how participants would react to emergency situations with little training. The human reliability analysis research required behavioural naturalness and the training assessment research required interruption and corrective measures to instil learning and behavioural conformance to protocols and procedures. As a result, the feedback provided to the trainees was designed to be more informative than corrective, which was one of the reasons participants continually made procedural mistakes.

Considerable planning and design iterations were used in the development of the experimental design. The results indicate areas for improvement. Future competence assessment and human reliability assessment experiments should be performed separately so that the data collection and analysis will not interfere with one another.

Importance of Feedback and Time to Reflect on Performance

The study employed a brief after-action review of the trainee's performance at the end of each practice and test scenario. This information included: correct location reached, percent walking, running and stationary, path tracker map, total time to complete, number of doors left open (if any) and degree of injury incurred if hazard exposure occurred. This information was presented passively in that it was optional to review. Although this model followed the active learning style in that it allowed the trainee to have control over their learning process, it resulted in trainees skipping the essential review of their performance due to over confidence in their performance. Participants who felt they performed adequately tended not to review their feedback in detail, which could account for poor performance in session 3.

The feedback was also presented to inform the participants of their actions, but not the quality or correctness of the actions. This is likely a flaw in the design of the experiment. It was done this way so as not to heavily impact the data for the Human Reliability Assessment (HRA). Specifically, the data represented a more natural behaviour in emergency situations but as a result negatively impacted the training quality as mistakes were not adequately addressed from one scenario to another. It was also done so that the act of testing group 2 (which is essentially recurrent training for the group) did not retrain the individual. Thus the feedback was informative but not corrective. As Klein (1997) cautioned, trainees who do not understand the seriousness of the errors made nor spend enough time reflecting on their performance, are unlikely to develop strategies to improve their performance in the next scenario or test session.

Therefore, the procedural learning objectives of this study require more stringent reinforcement of correct behaviours during training and testing scenarios to ensure knowledge acquisition. With immediate, relevant and corrective feedback in after action reviews of test scenarios, trainees would be able to correct their strategy and make greater improvement across test scenarios. If feedback is only informational and optional, like the feedback provided in this study, then participants may skip over the learning lessons and become overconfident in their performance.

It is also recommended that trainees receive feedback that is more immediate (e.g. in-simulation tips and help), direct (e.g. explicit correct and incorrect – not ambiguous) and relevant (e.g. feedback that reiterates learning objectives – route selected compared with the most efficient route instead of map of route taken). These forms of feedback can be automated in the learning management system to better inform the trainee of their errors and help mitigate any spread of potential errors. There should also be built in debrief exercises in the learning management system to encourage the trainee to reflect on their performance. Without these measures, participants will propagate mistakes throughout the training.

Chapter 6 : Conclusion and Recommendations

6.1 Summary of Findings

The following is a summary of the findings in this study:

1. Basic offshore safety competence was not demonstrated for all learning objectives. It was hypothesized that participants would be able to successfully demonstrate all learning objectives by the end of session 3. In practice this was the case for only a small minority of participants.
2. The majority of participants were able to successfully demonstrate the following competences:
   a. reach the correct muster location (LO1 – Spatial Awareness)
   b. recognize the alarm type (LO2 – Alarm Recognition)
   c. follow designated egress routes (LO3 – Routes and Mapping)
   d. avoid hazards blocking their path (LO4 – Assess Situation), and
   e. register at the temporary safe refuge area (LO5 – Register at TSR).
3. The majority of participants were unable to successfully demonstrate the following competences:
   a. select safest egress routes (LO3 – Routes and Mapping)
   b. recognize the hazardous situation, raise the alarm and re-route effectively if the path was blocked (LO4 – Assess Situation)
   c. resist running on the platform, and ensure all fire or water tight doors were closed (LO6 – Safe Practices).

4. Time to competence for all learning objectives is still unknown. Individual differences in learning styles and pace may have impacted the amount of exposure to the AVERT training program required to ensure knowledge retention. This result highlights the importance of verifying that competence is achieved for each individual. In future studies, a shift in focus from time to competence to demonstrable competence is recommended. From a training utility perspective, the amount of time required to achieve competence is secondary to ensuring competence is achieved. Therefore a mastery of learning approach should be the focus of future AVERT training programs to ensure demonstrable competence.
5. Few instances of statistically significant difference were found between the two groups' task performance (both spatially and procedurally). It was hypothesized that participants who spent more time training (group 1) would outperform participants who spent less time training (group 2). However, this hypothesis was not supported for all learning objectives.
6. One scenario (denoted TA1 – the first scenario of session 2) was statistically different between groups on multiple parameters:
    a. time to muster ($p = 0.036$),
    b. distance travelled to muster ($p = 0.005$),
    c. route selection ($p = 0.065$)
    d. backtracking time ($p = 0.004$) and backtracking distance ($p = 0.006$).

7. One scenario (denoted TH2 – the second scenario of session 3) had a trending difference between groups on multiple parameters:
    a. exposure to smoke ($p = 0.051$),
    b. percentage running ($p = 0.086$), and
    c. number of doors left opened ($p = 0.082$).
8. Based on the mean competence scores, performance in the test scenarios improved from session 1 to session 2. However, a degradation in performance occurred between session 2 and session 3. It was hypothesized that task performance for both groups would improve as participants advanced through the training and testing sessions. This increase in average competence score between session 1 and session 2 showed a promising sign that learning had occurred and that the participants were grasping spatial and procedural knowledge required for basic egress conditions. The decrease in performance between session 2 and session 3 indicates that the participants were unable to keep up with the increases in complexity and that they may not have been adequately prepared for major emergency situations.

6.2 Concluding Remarks

This research studied the effect of the AVERT virtual environment training on participant competency and learning in emergency response evacuation scenarios. Overall, this baseline data informs future virtual environment studies on spatial and procedural knowledge transfer to real offshore environments.

Successful egress in an emergency situation requires both spatial and procedural knowledge. Realistic evacuation drills and emergency situations were recreated in the virtual environment to assess trainee competence in emergency offshore egress. The results from this study indicate that guided training or further training exposure to AVERT is required in order for the participants to be successful in demonstrating performance. When developing a virtual environment training program, careful consideration should be placed on individual differences in skill acquisition and retention. A mastery of learning approach should be considered to help all individuals with different learning needs to reach competence more quickly. The lackluster performance of participants in demonstrating competence in complex emergency situations (session 3) reflects more on the adequacy of the participants' preparation for the difficulty of those situations than the success of the technology in teaching them about offshore egress training. The study clearly indicates that performance in offshore egress can be systematically assessed using virtual environment technology. Overall, the results from this study indicate that there is still work to be done to validate the effectiveness and utility of the virtual environment technology on training offshore emergency egress.

6.3 Recommendations

The following are areas where improvements can be made in future research related to offshore safety training with virtual environments:

1. Training improvements to encourage skill acquisition and competence
    a. Apply a user-centered training program to virtual environments. Spend less time using traditional forms of teaching and focus more time training in the virtual environment, using engaging training scenarios.
    b. A mastery of learning approach can cater to the individual learning styles and paces to help improve overall trainee performance in the virtual setting. Better performance will occur if the trainee is able to focus specifically on the areas s/he finds more difficult and is only able to advance to new material once s/he has mastered the previous. To ensure better performance success, performance measures must be indicated to the trainee at the start of the virtual environment training program.
    c. After-action review feedback is most helpful when it is immediate, clear and relevant. This form of feedback should be supported with instantaneous in-simulation guidance and corrective reinforcement.
2. Simulation improvements to support learning
    a. Training scenario design should engage the trainee and enable them to actively focus on learning and mastering the performance tasks.

b. Where possible, scenarios should use aids to help solidify learning and reinforce relevant learning objectives. Further work needs to be completed to automate the assessment of performance measures in order to provide important in-simulation guidance and immediate reinforcement feedback.

3. Design of experiment improvements for multi-purpose data collection

   a. Human reliability assessment (HRA) experiments and training curriculum experiments should be done separately, otherwise as shown in this study, the experimental design, data collection and analysis may interfere with one another. Human reliability assessment requires behavioural naturalness in the virtual setting while training assessment requires interruption and corrective measures to instil learning and behavioural conformance to protocols and procedures.

4. Future studies should focus on:

   a. Long-term retention studies (greater than a week duration) to address remaining time to competence and retention questions.

   b. Reverse training transfer studies using experienced personnel. Some forms of behaviours and mistakes observed by the novice group may not be present in a group with offshore domain expertise.

References

Aldrich, C. (2009). *The Complete Guide to Simulations and Serious Games: How the Most Valuable Content Will be Created in the Age Beyond Gutenberg to Google.* San Francisco, CA: Pfeiffer

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Rarths, J. and Wittrock, M. (2001). *A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives.* New York NY: Addison Wesley Longman.

Assael, M. J., and Kakosimos, K. E. (2010). *Fires, explosions and toxic gas dispersions: effects calculations and risk analysis*. CRC Press, NW.

Babrauskas V., (1983). Estimating large pool fire burning rates, *Fire Technology*, 19(4), 251-261.

Bradbury-Squires, D. (2013). *Simulation training in a virtual environment of an offshore oil installation.* Published master's thesis, Memorial University of Newfoundland, St. John's, NL.

Brunye, T., Gardony, A., Mahoney, C., and Taylor, H. (2012). Going to town: visualized perspectives and navigation through virtual environments. *Computers in Human Behavior*. 28 (2012) 257-266.

Burgess D. S., and Herzberg, M. (1974). *Advances in thermal engineering*, Wiley and Sons, Chichester.

Cannon-Bowers, J.A., and Salas, E. (1998). Individual and team decision making under stress: theoretical underpinnings. *In*: J.A. Cannon-Bowers and E. Salas (eds). *Making decisions under stress: implications for individual and team training*, Washington, DC: American Psychological Association, 17–38.

Atlantic Canada Offshore Petroleum Agency: Standard practice for the training and qualifications of personnel. (2013). (No. 2013-007). Canadian Association of Petroleum Producers. Retrieved from http://www.capp.ca/library/publications/

Carassa, A., Geminiani, G., Morganti, F., and Varotto, D. (2002). Active and passive spatial learning in complex virtual environment: the effect of efficient exploration. *Cognitive Processing*, 3-4 (2002), 56-81.

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. Published online by *Psychonomic Society*, Inc. November 2011. 19 (2012), 1-23.

Clark, R. and Mayers, R. (2008). Learning by viewing versus learning by doing: evidence-based guidelines for principled learning environments. *Performance Improvement. International Society for Performance Improvement.* 47(9), 5-13.

Dahlstrom, N., Dekker, S., van Winsen, R., and Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science,* 10 (4), 305-314.

Dalgarno, B., and Lee, M. J. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology*. 41(1), 10-32.

Darken, R. P., and Peterson, B. (2001). Spatial Orientation, Wayfinding, and Representation. *Handbook of Virtual Environment Technology*. Stanney, K. Ed.

De Freitas, S., and Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers and Education*, 52 (2009), 343-352.

Dekker, S. (2006). *The Field Guide to Understanding Human Error.* Hampshire, England: Ashgate.

Det Nortas Veritas STCW Basic Safety for Seafarers. (2005). (No. 3.303). September 2005, Det Nortas Veritas (DNV). Retrieved from https://exchange.dnv.com/publishing/stdcert/2005-09/Standard3-303.pdf

Dewey, J. (1916). *Democracy and Education*, New York: The Free Press. (1966).

DiMattia, D., Khan, F., Amyotte, P. (2005). Determination of human error probabilities for offshore platform musters, *Journal of Loss Prevention in the Process Industries*, 18(4–6), July–November 2005, 488–501.

Farrell, M., Arnold, P., Pettifer, S., Adams, J., Graham, T., and MacManamon, M. (2003). Transfer of route learning from virtual to real environments. *Journal of Experimental Psychology: Applied.* 9(4), 219-227.

Flin, R., Slaven, G., and Stewart, K. (1996). Emergency decision making in the offshore oil and gas industry. *Human Factors and Ergonomics Society*, 38 (2), 262-277.

Foo, P., Warren, W. H., Duchon, A., and Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? Map- versus landmark-based navigation of novel shortcuts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, (31), 195–215.

Fowlkes, J., Schatz, S., and Stagl, K. (2010). Instructional strategies for scenario-based training: insights from applied research. *In Proceedings of SpringSim*. 2010, 32-32.

Fowlkes, J., Dwyer, D., Oser, R., and Salas, E. (1998). Event-based approach to training (EBAT), *The International Journal of Aviation Psychology*, 8(3), 209-221.

Gallagher, A., Ritter, M., Champion, H., Higgins, G., Fried, M., Moses, G., Smith, D., and Satava, R. (2005). Virtual reality simulation for the operation room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery,* 241(2) February 2005. 364-372.

Gaunet, F., Vidal, M., Kemeny, A., and Berthoz, A. (2001). Active, passive and snapshot exploration in a virtual environment: influence on scene memory, reorientation and path memory. *Cognitive Brain Research*, 11(2001), 409-420.

Gawron, V. (2008). *Human Performance, Workload, and Awareness Measures Handbook* (2nd edition), Boca Raton, FL: CRC Press.

Graham T. L., and Roberts D. J. (2000). Qualitative Overview of Some Important Factors Affecting the Egress of People in Hotel Fires. *Hospitality Management*, 19 (2000), 79-87.

Grantcharov, T. and Reznick, R. (2008). Teaching rounds: teaching procedural skills. *British Medical Journal*, 336 (2008), 1129- 1131.

House, A., Smith, J., MacKinnon, S., and Veitch, B. (2014). Interactive simulation for training offshore workers. *Oceans'14 MTS/IEEE Conference,* September 14-19, 2014. St. John's, NL.

Huang, H., Rauch, U., & Liaw, S. (2010) Investigating learners' attitudes toward virtual reality learning environments: based on a constructivist approach. *Computers and Education,* 55 (2010), 1171-1182.

IADC. (2009) *Health, safety and environmental case guidelines for mobile offshore drilling units, International Association of Drilling Contractors*, Issue 3.2.1(2009), 128.

International Maritime Organization. (2001). *STCW, The Standards of Training, Certification and Watchkeeping Code (STCW Code) Convention for Seafarers*. London: International Maritime Organisation.

Insko, B. (2003) Measuring presence: subjective, behavioral and physiological methods. Being there: concepts, effects and measurement of user presence in synthetic environments. G. Riva, F. Davide, W.A IJsselsteijn (Eds.) Ios Press, 2003, Amsterdam, The Netherlands.

International Association of Oil and Gas Procedures. (2010). *Evacuation, escape and rescue OGP Risk Assessment Data Directory* (Report No. 434 – 19), March 2010. Retrieved from http://www.ogp.org.uk/pubs/434-19.pdf

Ishikawa, T., & Montello, D. R. (2006). Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology*, 52, 93–129.

Kennedy, R., Lane, N., Berbaum, K., and Lilienthal, M. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*. 3(3), 203-220.

Klein, G. (1997). An overview of naturalistic decision making. In C. Zsambok & G. Klein (Eds.), *Natural Decision Making*, New York, NY: Psychology Press, 49-59.

Klein, G., (1999). *The Sources of Power: How People Make Decisions.* Massachusetts, NE: Institute of Technology Press.

Kobes, M., Helsloot, I., de Vries, B., and Post, J. (2010). Exit choice, (pre-)movement time and (pre-)evacuation behaviour in hotel fire evacuation – behavioural analysis and validation of the use of serious gaming in experimental research, *First International Conference on Evaluation Modelling and Management*, 3(2010), 37-51.

Koh, G., von Wiegnad, T., Garnett, R., Durlach, N., and Shinn-Cunningham, B. (2000). Use of virtual environments for acquiring configurational knowledge about specific real-world spaces: preliminary experiment, *Presence: Teleoperators and Virtual Environments*, 8(6), 632-656.

Kolb, D. (1984). *Experiential learning*, Upper Saddle River, NJ: Prentice-Hall.

Lang, K., Breitsprecher, K., Hughes, B., and McGrath, T. (2007). Use of a computer simulation to assess hazard and risk perception. *Society of Petroleum Engineers*. SPE 108904.

Lathrop, W. and Kaiser, M. (2005). Acquiring spatial knowledge while traveling simple and complex paths with immersive and nonimmersive interfaces, *Presence: Teleoperators and Virtual Environments*, 14(3), June 2005, 249-263.

Lee, E. A., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers and Education*, 55(2010), 1424-1442.

Levelt, W. J. M. (1982). Cognitive style in the use of spatial direction terms. In R. J. Jarvella & W. Klein (Eds), *Speech, place and action: Studies in deixis and related topics*, Chichester, England: Wiley. 251-268.

Linde, C., and Labov, W. (1975). Spatial networks as site for the study of language and thought. *Language*, 51(4), 924-939.

Linberg, E., and Garling, T. (1983). Acquisition of different types of locational information in cognitive maps: Automatic or effortful processing? *Psychological Research*, 45, 19-38.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-mediated communication*, 3(2).

Lynch, K. (1960). *The image of the city*, Cambridge, MA:MIT Press.

MacDonald, S., MacKinnon, S., Simões Ré, A., and Power, J. (2011). Performance measurements of the effect of simulator training on novice operators in simulated ice conditions, Proceedings of the *21st International Conference on Port and Ocean Engineering under Arctic Conditions*, July 10-14, 2011. Montreal, Canada.

MacKinnon, S., Patterson, A., Billard, R., Power, J., Veitch, B., Power-MacDonald, S., Baker, A., and Simões Ré, A. (2012) Designing simulators to maximize knowledge transfer. *Proceedings, MARSIM 2012,* Singapore.

Magee, L., Thompson, A., Cain, B. and Kwan, C., (2012). From submarine to a virtual environment and vice versa. *Interservice/Industry Training, Simulation and Education Conference (I/ITSEC)*. Paper No. 12071.

Mantovani, F., and Castelnuovo, G. (2003). Sense of presence in virtual training: enhancing skills acquitision and transfer of knowledge through learning experience in virtual environments. *There: concepts, effects and measurement of user presence in synthetic environments,* Chapter 11. G. Riva, F. Davide, W.A IJsselsteijn (Eds.) Ios Press, 2003, Amsterdam, Netherlands.

Meehan, M., Insko, B., Whitton, M., Brooks, F. (2002). Physiological measures of presence in stressful virtual environments, *Proceedings of the SIGGRAPH 2002*, San Antonio, TX, USA, 21(3), July 2002, 645-652.

Meehan, M., Razzaque, S., Whitton, M. and Brooks, F. (2003). Effect of latency on presence in stressful virtual environments. *Proceedings of the IEEE Virtual Reality 2003*, 22-26 March 2003, 141-48.

Melanson, B., Kelso, J. and Bowman, D. (2002). Effects of active exploration and passive observation on spatial learning in a CAVE. Retrieved December 2012, from http://vtechworks.lib.vt.edu/handle/10919/20082.

Michel, N., Cater, J. J., and Varela, O. (2009). Active versus passive teaching styles: an empirical study of student learning outcomes. Human Resource Development Quarterly. 20 (4), 397-418.

Montgomery, D. (2012). *Design and Analysis of Experiments*, 8th Edition (April 2012), Hoboken, NJ: Wiley.

Musharraf, M. (2014). *Bayesian network approach to human reliability analysis (HRA) at offshore operations.* Published master's thesis, Memorial University of Newfoundland, St. John's, NL.

Norsok. (2001). *Norsok Standard: Risk and emergency preparedness analysis,* (No. Z-013, Rev 2, 2001-09-01). Norwegian Technology Centre. Retrieved from http://www.standard.no/pagefiles/955/z-013.pdf

Peruch, P., Varcher, J., and Gauthier, G. (1995), Acquisition of spatial knowledge through visual exploration of simulated environments. *Ecological Psychology*, 7(1), 1-20.

Peruch, P., and Wilson, P. (2004). Active versus passive learning and testing in a complex outside built environment. *Cognitive Processing*, 5 (2004), 218-227.

Pula, R., Khan, F., Veitch, B., and Amyotte, P. (2005) Revised fire consequence models for offshore quantitative risk assessment. *Journal of Loss Prevention in the Process Industries*, 18 (4), 443-454.

Reason, J., (1990). *Human Error*, Cambridge, UK: Cambridge University Press.

Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, 21 (2011), 58-67.

Ruddle, R., and Lessels, S. (2006). Three levels of metric for evaluating wayfinding. *Presence: Teleoperators and Virtual Environments*, 15(6), December 2006, 637-654.

Ruddle, R., Payne, S., and Jones, D. (1998). Navigating large-scale "desktop" virtual buildings: effects of orientation aids and familiarity. *Presence: Teleoperators and Virtual Environments*, 7(2), April 1998, 179-192.

Sandberg A. (1997). Unannounced Evacuation of Large Retail-Stores. An Evaluation of Human Behaviour and the Computermodel Simulex. Lund Institute of Technology, Lund University, Sweden.

Salas, E., Rosen, M., Held, J., and Weissmuller, J. (2009). Performance measures in simulation-based training: a review and best practices. *Simulation & Gaming*, 40 (3), June 2009, 328-376.

Sawyer, R. K. (2006). *The Cambridge Handbook of the Learning Sciences*. New York: Cambridge University Press.

Scoresby, J., and Shelton, B. (2010). Visual perspectives within educational computer games: effects on presence and flow within virtual immersive learning environments. *Instructional Science*, (2011), 39, 227–254

Schinazi, V. R., Nardi, D., Newcombe, N. S., Shipley, T. F., and Epstein, R. A., (2013). Hippocampal size predicts rapid learning of a cognitive map in humans, *Hippocampus*, 23, 515–528.

Sebok, A., Nystad, E. and Helgar, S. (2004). Navigation in desktop virtual environments: an evaluation and recommendations for supporting usability, *Virtual Reality*, 8(2004), 26-40.

Shettleworth, S. J. (2009). *Cognition, Evolution, and Behavior* (2nd edition). New York, NY: Oxford University Press.

Siegel, A.W., and White, S.H., (1975). The development of spatial representations of large-scale environments, *Advances in child development and behavior*, H. W. Reese (Ed.), New York: Academic Press, Vol. 10, 9-55.

Skogdalen, J., Khorsandi, J. and Erik Vinnem, J. (2012). Evacuation, escape, and rescue experiences from offshore accidents including the Deepwater Horizon. *Journal of Loss Prevention in the Process Industries*. 25 (2012) 148-158.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire, *Presence: Teleoperators and Virtual Environments*, 8(5), October 1999, 560-565.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments, *Philosophical Transactions of the Royal Society of London: Series B, Biological* Sciences, 364(1535), 3549-3557.

Sneddon, A., Mearns, K. J. and Flin, R. (2013). Stress, fatigue, situation awareness and safety in offshore drilling crews. *Journal of Safety Science*. 56, 80-88.

Squire, K. (2006). From content to context: Digital games as designed experiences. *Educational Researcher*, 35 (8), 19–29.

Stanney, K., Kennedy, R., Drexler, J. and Harm, D. (1999). Motion sickness and proprioceptive aftereffects following virtual environment exposure. *Applied Ergonomics*. 30 (1999), 27-38.

Stewart-Wingfield, S., & Black, G. S. 2005. Active versus passive course designs: The impact on student outcomes. *Journal of Education for Business*, 81(2), 119-125.

Thorndyke, P., and Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, 14(4), 560-589.

Tichon, J. (2007). Training cognitive skills in virtual reality: measuring performance, *Cyber Psychology & Behavior*, 10(2).

Tichon, J., and Wallis. G. (2010). Stress training and simulator complexity: why sometimes more is less. *Behaviour & Information Technology*. 29(5), 459-466.

Transport Canada Transport Canada. (2007). Practical skills training for marine engineers. TP 13720E, Second Edition, July 2007.

Tversky, B. (1981). Distortions in memory for maps. *Cognitive Psychology*, 13(3), 407-433.

US Coast Guards Research and Development Center (2000) A method for developing mariner assessment. Chapter 3: a manual for assessment developers. Final Report No. CG-D-17-00, July 2000.

Van der Spek, E., Oostendorp, H. and Meyer, J. (2013). Introducing surprising events can stimulate deep learning in a serious game, *British Journal of Educational Technology*, 44(1) 2013, 156-169.

Veitch, B., Billard, R., and Patterson, A. (2008). Emergency response training using simulators. Paper OTC 19233-PP. *Proceedings of the 2008 Offshore Technology Conference*, 5-8 May 2008, Houston Texas.

Veitch, B., Billard, R., and Patterson, A. (2009) Evacuation training using lifeboat simulators – Bridging the Training Gap: Managing competence assurance and improving safety. *Sea-Technology*, April 2009, 50(4), 65.

Waller, D., Hunt. E, and Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and virtual environments*, 7(2), April 1998, 129-143.

Wallis, G., Tichon, J., and Mildred, T. (2007). Speed perception as an objective measure of presence in virtual environments. *In: Proceedings of SimTect2007*, Brisbane, Australia, 527–531.

Weisberg, S., Schinazi, V., Newcombe, N., Shipley, T. and Epstein, R. (2014). Variations in cognitive maps: understanding individual differences in navigation. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 40(3), 669-682.

Wickens, C., and Hollands, J. (2000). *Engineering Psychology and Human Performance*. Third edition. Prentice Hall. Upper Saddle River, New Jersey 07458.

Wilson, P., Foreman, N., Gillett, R., and Stanton, D. (1997). Active versus passive processing of spatial information in a computer-simulated environment, *Ecological Psychology*, 9(3), 207-222.

Wingfield, S. S., & Black, G. S. (2005). Active versus passive course designs: the impact on student outcomes. *Journal of Education for Business*. November/December 2005. 119-123.

Witmer, B. G., Sadowski, W. J., & Finkelstein, N. M. (2002). VE-based training strategies for acquiring survey knowledge, *Presence: Teleoperators and virtual environments*, 11(1), February 2002, 1-18.

Witmer, B. G., Bailey, J. H., & Knerr, B. W. (1995). Training Dismounted Soldiers in Virtual Environments: Route Learning and Transfer (Technical Report 1022): *U.S. Army Research Institute for the Behavioral and Social Sciences*.

Witmer, B. G., Bailey, J. H., Knerr, B. W. & Parsons, K. C. (1996). Virtual spaces and real-world places: Transfer of route knowledge, *International Journal of Human-Computer Studies*, Vol. 45, pp. 413-428.

Witmer, B. G., and Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and virtual environments,* 7(3). June 1998, 225-240.

Witmer, B., Jerome, C., and Singer, M. (2005). The factor structure of the presence questionnaire. *Presence: Teleoperators and virtual environments,* 14(3). June 2005, 298-312.

Woodward, J. L., and Pitblado M. (2010). *LNG risk based safety: modeling and consequence analysis*. John Wiley and Sons Inc., New Jersey.

Zapalska, A., Brozik, D., and Rudd, D. (2012). Development of active learning with simulations and games. *US-China Education Review*, A 2 (2012) 164-169.

Appendices

Appendix A: Presence Results to Virtual Environment Emergency Situations

Descriptives for Group 1 and Group 2 – Video Game Experience Scores, Immersive Tendencies Scores and Presence Scores.

Variables	Group 1	Group 2	All Participants (Total Sample)
N	17	19	36
Age	26.76 ± 1.22 (.823)	26.21 ± .91 (.823)	26.47 ± .74
VGEQ2	11.88 ± 1.88 (.427)	10.84 ± 2.07 (.427)	11.33 ± 1.39
VGEQ3	6.59 ± 1.85 (.654)	6.92 ± 1.98 (.654)	6.76 ± 1.35
VGEQ6	2.24 ± .22 (.871)	2.16 ± 0.23 (.871)	2.19 ± .16
VGEQ7	2.41 ± .23 (.672)	2.68 ± 0.19 (.672)	2.56 ± .15
ITQ Score	62.50 ± 2.02 (.609)	64.99 ± 2.53 (.609)	63.81 ± 1.63
PQ Score	79.49 ± 1.87 (.250)	77.80 ± 2.84 (.250)	78.59 ± 1.72

All values are reported as means ± standard error (*p* - value) (* indicates $p < 0.05$, ** indicates $p < 0.01$)

Group 1 - Mean Percentage Immersive Tendencies and Presence Scores

Participant	ITQ-%Score	PQ-%Score
G1-1	60	76
G1-2	54	69
G1-3	73	77
G1-4	48	87
G1-5	60	78
G1-6	72	87
G1-7	70	82
G1-8	52	88
G1-9	62	64
G1-10	60	76
G1-11	67	79
G1-12	69	71
G1-13	71	80
G1-14	75	93
G1-15	55	85
G1-16	54	73
G1-17	62	86

Group 2 - Mean Percentage Immersive Tendencies and Presence Scores

Participant	ITQ-%Score	PQ-%Score
G2-1	59	89
G2-2	64	63
G2-3	84	86
G2-4	71	77
G2-5	58	66
G2-6	51	60
G2-7	63	87
G2-8	63	87
G2-9	76	66
G2-10	71	92
G2-11	73	89
G2-12	67	87
G2-13	90	87
G2-14	63	87
G2-15	59	92
G2-16	68	76
G2-17	54	65
G2-18	50	67
G2-19	50	55

Appendix B: Physiological Response Results to Virtual Emergency Situations

Group 1 - Mean physiological measures for each test scenario

Participant	Session	GSR	ST	RR	HR
G1-1					
TE1	S1	9.46	94.12	21.52	68.75
TE2		9.66	93.87	17.57	76.48
TE3		9.42	94.63	17.16	69.79
TE4		6.21	94.12	18.17	67.03
TA1	S2	3.00	92.36	19.20	69.38
TA2		2.90	91.89	19.26	67.76
TA3		2.62	91.74	17.73	66.94
TA4		3.23	92.02	19.28	68.81
TH1	S3	1.66	87.88	20.22	68.54
TH2		2.04	87.02	15.84	72.70
TH3		2.70	86.10	18.28	70.42
TH4		3.10	85.18	17.71	72.67
G1-2					
TE1	S1	0.57	94.42	23.70	83.82
TE2		2.00	95.26	26.56	90.85
TE3		1.71	95.70	23.77	87.55
TE4		1.44	95.72	23.99	91.56
TA1	S2	0.46	87.63	21.12	76.52
TA2		0.42	85.91	21.78	74.57
TA3		0.42	85.60	22.14	69.68
TA4		0.49	84.14	22.75	71.37
TH1	S3	0.44	93.36	21.61	68.83
TH2		0.51	92.25	21.41	69.89
TH3		0.53	92.13	21.46	68.91
TH4		0.65	91.30	19.94	70.49
G1-3					
TE1	S1	4.73	82.31	25.65	97.48
TE2		6.97	95.47	26.43	108.17
TE3		8.81	96.26	25.35	106.98
TE4		8.01	96.09	26.24	106.19
TA1	S2	5.93	96.10	22.46	102.76

TA2		5.38	95.43	25.67	104.95
TA3		5.22	95.43	23.19	104.08
TA4		6.57	95.31	25.74	102.11
TH1	S3	2.95	94.77	25.48	97.94
TH2		2.72	95.63	25.22	103.13
TH3		3.47	95.82	22.15	104.77
TH4		4.21	95.60	24.77	103.62
G1-4					
TE1	S1	3.15	95.53	23.30	65.17
TE2		5.22	95.28	24.27	66.66
TE3		6.84	95.32	23.83	67.13
TE4		7.71	95.32	24.18	70.25
TA1	S2	4.08	94.81	23.34	70.02
TA2		5.72	94.84	24.28	62.69
TA3		5.28	94.81	22.65	66.44
TA4		7.31	95.14	22.29	64.60
TH1	S3	2.72	95.70	21.79	59.47
TH2		4.58	95.51	20.11	54.08
TH3		5.35	95.33	21.40	56.70
TH4		6.59	95.19	19.66	59.18
G1-5					
TE1	S1	1.67	85.11	23.73	79.20
TE2		1.53	84.73	21.65	81.49
TE3		1.35	84.71	22.93	77.05
TE4		1.91	84.28	24.44	75.64
TA1	S2	2.75	89.25	21.16	78.50
TA2		1.33	90.70	23.16	74.63
TA3		1.64	91.33	24.66	82.02
TA4		1.89	91.78	21.94	74.03
TH1	S3	1.95	88.15	24.05	80.16
TH2		2.36	87.38	24.08	73.66
TH3		2.64	86.00	24.47	74.02
TH4		2.87	84.86	26.27	79.17
G1-6					
TE1	S1	6.54	93.45	21.18	62.40
TE2		7.88	92.90	19.97	62.94

TE3		8.38	93.84	20.88	59.30
TE4		8.91	94.66	20.03	60.78
TA1	S2	5.04	93.77	19.18	59.62
TA2		5.80	94.22	19.22	56.30
TA3		6.68	94.01	17.55	59.60
TA4		6.80	93.71	18.90	53.60
TH1	S3	3.85	94.57	17.58	56.84
TH2		5.11	94.74	18.23	55.84
TH3		5.55	94.78	17.07	56.08
TH4		5.67	95.10	17.66	53.80
G1-7					
TE1	S1	1.31	88.40	22.57	70.80
TE2		1.54	87.88	22.79	73.44
TE3		1.71	88.12	23.26	72.65
TE4		1.76	88.48	21.50	72.70
TA1	S2	1.87	89.88	23.69	90.80
TA2		2.34	89.34	23.07	84.47
TA3		2.47	88.65	20.91	83.58
TA4		2.61	88.48	23.47	82.39
TH1	S3	1.59	91.30	25.19	80.97
TH2		1.77	91.91	23.17	80.83
TH3		1.95	91.93	24.14	79.09
TH4		2.09	92.06	24.49	81.10
G1-8					
TE1	S1	7.14	93.49	21.16	65.95
TE2		8.43	94.49	20.82	66.42
TE3		9.14	93.86	20.55	66.27
TE4		9.58	93.93	20.11	68.78
TA1	S2	7.05	95.09	22.02	82.63
TA2		7.53	94.90	20.05	82.94
TA3		7.97	95.33	22.13	80.48
TA4		8.18	94.91	21.44	78.38
TH1	S3	10.59	95.38	22.84	79.39
TH2		9.80	95.59	24.48	85.13
TH3		9.64	95.24	21.68	82.72
TH4		10.66	95.29	22.30	85.05

G1-9					
TE1	S1	12.09	91.23	18.88	55.98
TE2		13.12	90.86	19.54	54.77
TE3		13.89	91.16	18.67	56.99
TE4		14.21	91.15	18.42	54.01
TA1	S2	13.19	94.25	17.03	58.74
TA2		15.06	94.87	18.74	58.13
TA3		15.77	95.67	18.63	60.87
TA4		15.56	95.06	17.57	64.20
TH1	S3	8.70	96.58	18.57	67.37
TH2		10.46	96.63	18.56	62.68
TH3		11.40	96.88	18.72	65.03
TH4		10.81	96.45	19.56	60.79
G1-10					
TE1	S1	2.34	83.30	25.19	77.50
TE2		3.43	82.90	22.90	73.20
TE3		4.20	83.37	22.69	72.85
TE4		4.93	88.20	22.04	70.92
TA1	S2	2.71	93.85	21.56	86.68
TA2		3.70	92.69	22.82	76.36
TA3		4.47	92.88	23.24	80.43
TA4		5.27	93.02	20.38	75.81
TH1	S3	2.05	87.79	20.37	69.20
TH2		3.11	88.62	23.30	71.06
TH3		4.00	90.83	21.19	72.01
TH4		5.98	91.56	21.73	74.57
G1-11					
TE1	S1	4.82	89.78	19.34	78.83
TE2		5.70	90.89	20.98	78.94
TE3		5.87	91.80	22.46	81.04
TE4		5.90	92.32	22.45	81.91
TA1	S2	5.37	92.93	23.13	76.66
TA2		6.11	94.38	22.24	77.43
TA3		6.91	93.89	21.66	77.41
TA4		7.17	95.01	23.23	81.19
TH1	S3	5.11	93.47	21.89	78.21

TH2		6.51	93.89	21.22	75.93
TH3		7.06	94.28	21.38	83.64
TH4		7.07	94.03	21.45	82.06
G1-12					
TE1	S1	2.93	86.43	21.48	68.94
TE2		3.5	86.57	20.55	69.67
TE3		4.07	88.51	21.2	66.33
TE4		4.43	92.67	20.2	69.17
TA1	S2	3.92	94.74	22.59	76.25
TA2		4.91	94.64	21.61	75.86
TA3		5.21	94.98	21.43	72.39
TA4		5.4	95.13	21.32	73.05
TH1	S3	3.66	96.12	23.7	79.77
TH2		4.5	96.26	22.52	78.23
TH3		4.88	96.58	22.09	81.86
TH4		4.92	96.09	21.48	81.66
G1-13					
TE1	S1	2.72	85.30	24.77	74.85
TE2		3.75	84.98	25.14	79.79
TE3		5.00	86.63	22.33	74.49
TE4		5.56	89.14	23.85	72.66
TA1	S2	4.75	89.38	23.19	69.69
TA2		6.04	90.36	24.70	72.57
TA3		6.48	91.17	21.07	65.06
TA4		6.94	90.97	21.39	67.81
TH1	S3	3.35	92.31	23.69	69.87
TH2		4.71	93.33	25.18	71.49
TH3		5.34	93.98	23.73	67.85
TH4		5.76	93.91	23.41	72.00
G1-14					
TE1	S1	2.13	90.12	24.84	96.38
TE2		2.00	91.42	20.47	89.17
TE3		1.96	92.54	19.62	90.61
TE4		1.89	92.22	19.37	88.50
TA1	S2	1.36	90.05	22.07	87.17
TA2		1.38	91.61	18.84	86.99

TA3		1.99	92.08	20.20	85.02
TA4		1.90	92.17	18.29	86.13
TH1	S3	0.47	92.65	20.69	82.19
TH2		0.56	92.63	18.61	80.45
TH3		0.51	92.37	17.94	82.66
TH4		0.78	92.64	19.62	80.90
G1-15					
TE1	S1	0.94	89.37	19.85	67.60
TE2		1.03	88.79	20.66	66.36
TE3		0.95	88.49	18.94	66.01
TE4		1.43	88.54	19.97	66.65
TA1	S2	0.79	89.14	23.39	71.50
TA2		0.78	89.86	19.36	72.16
TA3		0.95	90.87	22.24	72.78
TA4		0.94	91.56	21.20	68.07
TH1	S3	0.67	90.02	21.23	72.37
TH2		1.02	90.38	21.79	68.03
TH3		1.19	90.09	21.15	70.95
TH4		1.50	90.16	21.43	68.71
G1-16					
TE1	S1	2.01	89.42	20.32	70.33
TE2		2.20	89.18	20.13	75.08
TE3		2.46	89.13	22.28	74.39
TE4		2.68	88.76	22.07	71.38
TA1	S2	4.42	92.62	20.11	77.42
TA2		5.41	93.57	19.65	71.48
TA3		5.92	93.80	21.58	75.88
TA4		6.15	93.78	20.89	75.59
TH1	S3	2.35	85.64	21.45	75.06
TH2		3.03	86.79	21.05	74.54
TH3		3.72	89.54	21.90	74.71
TH4		4.11	89.65	19.96	72.23
G1-17					
TE1	S1	3.52	80.82	21.32	70.52
TE2		4.21	81.35	21.11	69.20
TE3		4.56	82.44	21.73	71.24

TE4		4.90	83.71	21.64	70.45
TA1	S2	3.81	86.78	21.82	74.47
TA2		4.38	84.89	22.56	71.46
TA3		4.44	83.41	20.73	69.81
TA4		4.43	82.92	21.52	71.25
TH1	S3	3.16	86.15	19.90	64.35
TH2		3.73	85.99	21.72	62.04
TH3		3.93	87.65	21.39	65.05
TH4		4.31	90.00	20.46	64.60

Group 2 - Mean physiological measures for each test scenario

Participant	Session	GSR	ST	RR	HR
G2-1					
TE1	S1	2.99	91.44	32.82	120.23
TE2		4.45	91.90	31.50	117.71
TE3		4.98	92.04	32.98	117.52
TE4		4.43	91.57	32.85	120.22
TA1	S2	1.86	86.28	26.90	94.77
TA2		2.69	85.26	27.63	94.03
TA3		2.81	84.12	28.20	90.65
TA4		3.10	83.76	27.58	90.97
TH1	S3	2.47	93.59	29.94	103.74
TH2		3.94	94.12	29.28	102.06
TH3		3.98	93.96	29.70	105.10
TH4		4.91	94.32	30.30	103.32
G2-2					
TE1	S1	5.66	89.57	25.96	86.28
TE2		6.90	88.99	25.98	86.86
TE3		6.64	88.24	27.81	80.21
TE4		6.02	89.20	27.51	82.52
TA1	S2	7.86	95.68	27.36	92.65
TA2		9.12	95.47	27.51	90.17
TA3		9.79	95.61	27.00	86.39
TA4		8.68	95.60	25.18	87.49
TH1	S3	2.89	92.61	26.31	76.82
TH2		2.93	92.59	25.97	79.09
TH3		3.65	93.30	26.51	79.40
TH4		4.79	93.35	26.54	79.40
G2-3					
TE1	S1	5.91	95.57	23.92	87.21
TE2		6.19	94.92	23.11	88.16
TE3		7.48	94.60	22.11	87.20
TE4		7.53	94.93	23.87	91.05
TA1	S2	2.69	94.66	23.81	91.58
TA2		2.77	94.92	24.17	90.39

TA3		2.85	94.62	24.54	93.92
TA4		3.11	94.42	23.52	88.85
TH1	S3	2.78	94.04	24.99	79.49
TH2		2.76	93.34	23.49	79.87
TH3		3.32	92.18	25.19	88.29
TH4		3.67	91.51	24.71	81.56
G2-4					
TE1	S1	1.60	95.34	22.23	78.47
TE2		2.19	95.35	22.16	78.91
TE3		1.93	94.18	19.95	74.51
TE4		2.19	95.35	22.53	79.65
TA1	S2	1.05	94.89	21.93	73.85
TA2		1.77	95.32	21.00	72.18
TA3		1.99	95.11	20.72	71.91
TA4		1.85	95.12	20.68	74.38
TH1	S3	1.59	94.99	22.42	74.72
TH2		2.38	95.32	21.98	73.68
TH3		1.93	94.18	19.95	74.51
TH4		2.45	94.84	20.87	72.76
G2-5					
TE1	S1	5.14	93.97	28.25	78.44
TE2		6.68	94.81	30.12	83.65
TE3		7.52	94.56	27.01	81.92
TE4		8.15	95.34	30.61	83.36
TA1	S2	4.06	97.49	33.30	102.06
TA2		5.65	97.30	34.31	99.55
TA3		5.41	97.50	31.06	95.81
TA4		6.56	97.49	34.22	96.88
TH1	S3	3.94	96.11	29.93	89.18
TH2		5.30	96.41	27.66	87.71
TH3		6.27	96.17	30.63	91.90
TH4		6.55	96.35	30.01	93.21
G2-6					
TE1	S1	4.25	94.63	28.16	87.47
TE2		4.59	96.01	25.95	86.57
TE3		6.99	95.96	26.13	87.43

TE4		6.92	95.43	26.48	85.15
TA1	S2	3.88	88.48	25.13	80.39
TA2		4.67	86.56	25.86	82.96
TA3		5.19	86.48	24.16	79.67
TA4		4.83	86.78	24.22	81.60
TH1	S3	4.57	91.72	25.94	81.44
TH2		5.95	92.47	25.05	75.12
TH3		6.61	91.65	23.52	79.13
TH4		6.08	90.69	23.45	73.93
G2-7					
TE1	S1	3.60	90.17	22.96	74.74
TE2		5.20	89.57	20.11	74.84
TE3		5.66	89.87	20.93	72.49
TE4		5.97	89.77	18.74	70.76
TA1	S2	3.91	93.32	20.03	88.69
TA2		4.49	94.54	21.97	83.15
TA3		4.64	95.29	19.60	85.12
TA4		4.23	94.28	20.86	81.59
TH1	S3	4.17	91.95	19.90	75.21
TH2		4.93	92.84	19.77	77.81
TH3		5.33	92.28	19.25	77.03
TH4		5.33	91.39	21.83	80.16
G2-8					
TE1	S1	5.55	93.51	21.25	81.48
TE2		7.98	94.46	21.35	80.21
TE3		9.06	94.68	21.91	79.09
TE4		9.14	94.80	21.74	74.14
TA1	S2	5.93	95.83	22.36	85.46
TA2		7.41	95.93	20.43	83.64
TA3		7.99	95.77	21.01	84.11
TA4		7.99	95.62	20.51	83.39
TH1	S3	6.19	95.79	20.36	88.01
TH2		7.76	95.58	20.48	82.61
TH3		8.41	95.75	22.02	81.17
TH4		8.66	94.37	21.43	81.69
G2-9					

TE1	S1	6.27	89.77	19.59	89.72
TE2		7.28	87.47	19.62	85.61
TE3		8.05	87.45	19.09	83.66
TE4		7.90	88.24	18.55	85.87
TA1	S2	6.66	93.56	19.54	89.96
TA2		6.75	91.32	18.09	94.37
TA3		6.35	91.44	20.34	89.25
TA4		6.22	91.33	20.40	91.58
TH1	S3	3.43	89.91	18.55	86.88
TH2		4.18	87.02	18.78	85.00
TH3		4.51	86.75	19.10	90.39
TH4		5.06	86.34	17.82	82.52
G2-10					
TE1	S1	2.00	93.95	25.45	76.20
TE2		2.49	95.00	26.20	82.44
TE3		2.17	94.72	25.23	77.02
TE4		2.56	93.74	26.22	77.90
TA1	S2	1.24	88.96	25.04	73.30
TA2		1.59	89.18	25.03	72.79
TA3		1.67	90.41	25.78	78.03
TA4		1.97	89.65	24.03	71.45
TH1	S3	2.94	92.21	30.95	76.59
TH2		3.73	93.33	26.20	84.07
TH3		4.45	93.36	25.27	80.03
TH4		4.98	93.96	26.75	79.19
G2-11					
TE1	S1	1.55	89.12	22.85	61.17
TE2		1.31	90.48	20.33	60.46
TE3		2.07	91.42	19.86	60.89
TE4		1.82	92.33	22.11	65.28
TA1	S2	2.16	93.34	19.42	61.00
TA2		1.28	92.94	22.60	61.97
TA3		1.95	92.99	19.66	62.11
TA4		1.38	94.41	23.15	60.74
TH1	S3	1.30	93.73	20.86	70.01
TH2		1.55	95.08	20.26	68.13

TH3		1.71	93.62	22.31	64.85
TH4		1.99	93.55	19.98	61.96
G2-12					
TE1	S1	0.95	89.46	20.32	64.18
TE2		1.36	89.55	21.19	69.43
TE3		1.56	88.45	19.97	70.96
TE4		1.61	87.82	19.85	67.57
TA1	S2	1.10	91.52	20.10	64.05
TA2		1.53	90.93	21.31	68.35
TA3		1.67	90.17	20.35	65.26
TA4		1.78	90.21	21.43	65.34
TH1	S3	1.61	91.27	18.88	67.08
TH2		1.62	90.36	19.97	69.03
TH3		1.78	90.00	20.96	66.93
TH4		1.86	89.79	23.08	67.22
G2-13					
TE1	S1	4.02	94.70	25.90	88.81
TE2		4.82	95.69	26.05	90.92
TE3		4.97	95.89	19.50	90.91
TE4		5.13	95.20	22.18	90.59
TA1	S2	3.78	96.23	23.01	98.64
TA2		4.43	96.38	24.63	96.20
TA3		4.63	96.49	26.24	97.20
TA4		4.92	96.36	26.82	97.80
TH1	S3	4.02	95.20	27.87	105.35
TH2		4.62	94.94	26.27	105.62
TH3		4.81	94.57	24.56	109.04
TH4		5.20	93.92	23.05	100.57
G2-14					
TE1	S1	5.42	91.06	22.64	79.60
TE2		6.02	90.71	24.51	83.82
TE3		6.21	89.44	22.77	76.17
TE4		6.24	88.34	22.79	82.85
TA1	S2	4.20	95.94	24.48	91.63
TA2		4.56	95.82	24.16	89.00
TA3		4.60	95.59	24.11	85.80

TA4		4.47	95.71	24.62	89.31
TH1	S3	5.85	94.62	25.86	85.58
TH2		7.15	94.80	22.42	86.13
TH3		7.06	94.48	24.18	85.37
TH4		6.88	93.81	23.71	82.77
G2-15					
TE1	S1	1.76	87.78	25.79	93.63
TE2		2.12	85.98	26.28	89.05
TE3		2.38	84.63	26.19	88.36
TE4		2.40	83.72	24.54	85.83
TA1	S2	3.21	91.81	29.63	104.31
TA2		3.68	91.02	28.12	102.14
TA3		4.01	90.94	28.59	100.01
TA4		4.14	90.52	28.26	99.45
TH1	S3	3.63	90.91	27.11	103.78
TH2		4.73	89.51	27.24	98.80
TH3		5.01	89.01	28.36	106.19
TH4		4.91	88.06	27.86	104.93
G2-16					
TE1	S1	4.30	92.65	26.73	81.58
TE2		5.04	92.78	26.62	81.79
TE3		5.14	93.61	25.28	79.51
TE4		4.70	93.30	25.17	78.14
TA1	S2	1.96	94.89	24.62	76.81
TA2		1.00	95.27	24.53	77.42
TA3		1.13	95.03	25.29	76.13
TA4		1.62	95.06	25.42	77.32
TH1	S3	1.46	94.15	24.24	75.90
TH2		2.35	93.75	25.94	77.78
TH3		1.27	92.83	26.12	78.58
TH4		2.30	93.84	27.37	75.74
G2-17					
TE1	S1	5.29	92.31	23.90	61.69
TE2		6.06	93.70	21.48	56.92
TE3		5.57	94.05	21.46	61.70
TE4		6.18	94.41	19.96	61.16

TA1	S2	2.34	92.72	18.43	57.19
TA2		3.06	93.64	22.23	57.48
TA3		N/A	94.15	20.11	56.09
TA4		3.36	93.71	20.82	57.43
TH1	S3	1.79	90.57	22.21	59.56
TH2		1.97	91.64	20.30	57.02
TH3		1.91	91.64	20.30	56.55
TH4		2.16	92.47	20.69	55.65
G2-18					
TE1	S1	1.71	93.24	26.96	75.65
TE2		1.73	93.38	25.77	74.57
TE3		1.90	93.60	28.20	77.92
TE4		1.72	93.27	27.07	74.96
TA1	S2	0.45	93.98	25.91	78.81
TA2		0.58	93.33	25.56	80.24
TA3		0.65	93.75	21.64	78.39
TA4		0.79	93.58	24.88	82.33
TH1	S3	0.91	93.93	26.80	93.24
TH2		1.43	93.90	27.67	91.36
TH3		1.67	94.00	25.40	91.21
TH4		1.82	94.14	24.52	88.61
G2-19					
TE1	S1	2.40	86.26	21.04	74.26
TE2		2.79	85.65	21.13	76.57
TE3		3.09	85.44	19.35	77.51
TE4		3.50	86.02	20.52	72.16
TA1	S2	0.46	84.98	18.02	73.46
TA2		0.42	85.79	16.74	71.31
TA3		0.78	85.57	19.81	76.93
TA4		0.45	85.73	19.70	73.72
TH1	S3	2.00	78.89	19.59	78.14
TH2		1.42	79.50	20.05	77.61
TH3		1.71	79.27	18.73	75.20
TH4		2.34	79.17	20.08	81.26

Appendix C: Detailed Description of Test Scenarios and Required Actions.

Sessions	Scenario Code	Context and Required Actions
Set 1: Environment	TE1	From cabin, find your lifeboat station using primary or secondary egress routes (as quickly as possible).
	TE2	From worksite, find your primary muster station using primary or secondary egress routes (as quickly as possible).
	TE3	From cabin, find your primary muster station in a blackout situation using primary or secondary egress routes (as quickly as possible).
	TE4	From worksite, find your lifeboat station in a blackout situation using primary or secondary egress routes (as quickly as possible).
Set 2: Alarms	TA1	From cabin, respond to a muster drill (GPA alarm). Required action: muster at your primary muster station as quickly as possible.
	TA2	From worksite, respond to an evacuation drill (PAPA alarm). Required action: muster at your lifeboat station as quickly as possible.
	TA3	From cabin, respond to emergency situation (PAPA alarm caused by equipment failure and anomalies in process controls resulting in blackout) Required action: muster at lifeboat station as quickly as possible.
	TA4	From worksite, respond to emergency situation (GPA alarm vessel wide blackout) Required action: muster at primary muster station as quickly as possible.
Set 3: Hazard Avoidance	TH1	From cabin, respond to emergency situation (GPA alarm due to fire in galley and escalating to a PAPA alarm due to smoke in adjacent muster station). Required action: head to primary muster station then re-route to lifeboat station due to compromised primary muster point. NOTE: primary route and muster point blocked.
	TH2	From cabin, respond to emergency situation (GPA alarm due to fire on helideck and escalating to a PAPA alarm due to explosion on helideck. High winds resulting in heaving smoke affection port side and center stairwell outside). Required action: head to primary muster station then re-route to lifeboat station due to escalating situation. NOTE: secondary route blocked.
	TH3	From worksite, respond to emergency situation (GPA alarm due to electrical fire and smoke in upper deck and escalating to a PAPA alarm due a blackout and thick smoke blocking access to upper deck and muster stations). Required action: head to primary muster station then re-route to lifeboat station due to escalating situation. NOTE: primary route blocked.
	TH4	From worksite, respond to emergency situation (No initial alarm, fire and explosion at main engine resulting in blackout) Required action: raise alarm, once alarm raised head to primary muster station then re-route to lifeboat station due to escalating situation. NOTE: secondary route blocked.

Appendix D: Call for Participants Recruitment Email

Call for Subjects- Recruitment Email Information

Are you interested in taking part in a research study?

Volunteers are needed to study the effect of virtual training systems on behaviour and learning in emergency response scenarios.

Brief Description of Experiment:

- Contribute to our understanding of how training protocols affects people's ability to train their memory and emergency response behaviors.
- You'll be using a simulator that is similar to a first person video game BUT for serious training situations. The simulation is controlled by an Xbox controller on a desktop computer.
- Volunteers will be asked to attend three (3) training and testing sessions.
- Each session could take upwards of four (4) hours to complete for a total of three (3) sessions. With no more than two (2) days between sessions.

Who can participate?

- Anyone between 19-55 years of age.
- No experience necessary to participate.

The study will be conducted in the Virtual Environments for Knowledge Mobilization laboratory located off-campus at 20 Hallett Crescent. Total time involvement will be approximately twelve hours over three visits to the lab.

Sources of data being collected:

- Performance during simulation scenarios;
- Sensors will be applied to each participant to measure:
  - Heart rate (2-Lead EKG);
  - Skin conductance (2 finger electrodes);
  - Peripheral skin temperature (one sensor applied to the hand), and;
  - Respiration (a band fixed around the torso),
- Subjective assessment of experience via questionnaires.

Please contact Jennifer Smith or Mashrura Musharraf for more information or to schedule a session. You can contact Mashrura by email (mm6414@mun.ca) or Jennifer by email (jennifersmith@mun.ca) or by telephone at (709) 864 - 6764.

Appendix E: Recruitment Poster

MEMORIAL UNIVERSITY

HOW WELL DO YOU COPE IN UNPREDICTABLE SITUATIONS?

PLAY A NEW VIRTUAL ENVIRONMENT PROTOTYPE BY PARTICIPATING IN A SIMULATION STUDY AT MUN

Description of Experiment:

- We are studying the effects of virtual training on human behaviour and learning in emergency response scenarios.
- The study consists of three training and testing sessions.
- Anyone between 19-55 years of age can participate.
- Volunteers will be tested on how well their training prepared them for unpredictable scenarios in the virtual world.

To find out more, contact:

Jennifer Smith – jennifersmith@mun.ca or (709) 864 6764

Appendix F: Free and Informed Consent Form

Title: The effect of virtual training systems on participant behaviour and learning in emergency response scenarios.

Researchers:

Dr. Scott MacKinnon
Principal Investigator
Memorial University of Newfoundland
School of Human Kinetics and Recreation
(709) 864-6936
smackinn@mun.ca

Ms. Jennifer Smith
Co-Investigator
Virtual Environments for Knowledge Mobilization Project
Human Factors Coordinator
(709) 864-6764
jennifersmith@mun.ca

Mr. David Bradbury-Squires
Co-Investigator
Virtual Environments for Knowledge Mobilization Project
Human Factors Graduate Student
(709) 728-5472
djbs32@mun.ca

Mr. Andrew Caines
Co-Investigator
Virtual Environments for Knowledge Mobilization Project
Human Factors Assistant – Cooperative Education
(709) 864-6764
a.caines@mun.ca

Mrs. Mashrura Musharraf
Co-Investigator
Virtual Environments for Knowledge Mobilization Project
Human Factors Graduate Student
(709) 769-6669
mm6414@mun.ca

You are invited to take part in a research project entitled *"The effect of virtual training systems on participant behaviour and learning in emergency response scenarios"*.

This form is part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. It also describes your right to withdraw from the study at any time. In order to decide whether you wish to participate in this research study, you should understand enough about its risks and benefits to be able to make an informed decision. This is the informed consent process. Take time to read this carefully and to understand the information given to you. Please contact the researchers *Dr. Scott MacKinnon or Jennifer Smith* if you have any questions about the study or for more information not included here before you consent.

It is entirely up to you to decide whether to take part in this research. If you choose not to take part in this research or if you decide to withdraw from the research once it has started, there will be no negative consequences for you, now or in the future.

Introduction

We are an interdisciplinary research team at Memorial University. The research is being conducted as a part of the Virtual Environments for Knowledge Mobilization Project. Disciplines collaborating on this project include engineers of various specialties (naval architecture, mechanical, software, electrical), human factors and ergonomics researchers, coop students from the faculty of engineering and kinesiology, and graduate students with related research interests.

There has recently been increased emphasis placed on ensuring the safety of crew on ships and offshore platforms. This has manifested in the form of improved safety management systems, more stringent regulations governing qualifications of seafarers and offshore crew, and requirements for more comprehensive training programs relating to all aspects of work and personal safety in maritime occupations.

The basic training related to personal safety and emergency response is most commonly delivered in lecture format, and routine drills. On offshore platforms there are weekly muster drills on a specific day and time where the onboard crew is required to proceed to their muster stations. Practically, this is their only opportunity to practice emergency response and muster procedures learned in lecture.

These weekly training drills are performed under optimal conditions. The crew knows to expect the drill at the same day and time, there is no threat of danger, and the route to the muster area will be free of obstacles. With no opportunity to experience an actual hazard or threat of danger, the value of this training may be lost.

This research project is evaluating naive participants training for emergency response and mustering on a virtual offshore oil rig. In this virtual environment there is the capability to implement hazards such as fires, explosions, oil leaks and toxic fumes. When practicing evacuation of an area of the vessel, doors can be barred, stairs can be rendered impassable, and the trainees will have to select alternate routes to avoid hazards, and safely make their way to their muster station or abandonment area.

Purpose of study:

This study will evaluate virtual training systems. Technology is advancing is such a way that training time may be increased without increasing the demand on instructors. In this study, minimal contact with the training facilitator is supplemented with training tools and artificial instruction. The research team is seeking to determine if there is a difference in performance between groups after single exposure training program and repeat exposure training program.

What you will do in this study:

Participants will attend three, four hour sessions in Virtual Environments Lab with no more than two days between sessions. Each session will consist of a training phase followed by virtual environment testing trials. Training will be an artificial-instructor led tutorial providing guidance on basic offshore safety practices including spatial awareness, alarm recognition, muster procedures and hazard avoidance. During the tutorials the instructor will provide a video walk-through tour of the virtual platform, outlining the platform floor plans, pointing out areas of access on the platform, important landmarks, muster locations and instructing on available escape routes. The instructor will also provide helpful lessons to facilitate learning of alarm types, mustering procedures and recognizing potential hazards. The instructions to the participant will be to learn, with the assistance of the familiarization materials, the alarm types, multiple escape routes, and muster procedures. Participants will have the opportunity to explore the virtual offshore oil rig in a thirty-minute self-guided walkthrough.

At the end of the training phase of each session, each participant will complete a quiz and perform four measured trials to assess the efficacy of each of the training protocols on learning alarm types, muster locations and escape routes. Each participant will perform the measured trials in the same order. The measured trial scenarios will increase in difficulty across each session. The trials will consist of scenarios, where the participant will encounter different alarm types as well as hazards such as poor lighting, barriers, fires, or explosions. Given these obstacles, the participant will have to respond appropriately to the alarm type and navigate from the starting point (cabin or workstation) to their muster location choosing alternative routes should they encounter an obstacle. After completing each trial scenario the participant will receive feedback on their performance.

Length of time:

Each participant will be required to attend three, four hour training and testing sessions, with no more than 2 days between sessions. Following the completion of the training the participant will be required to complete four testing trials for each session. The anticipated total time involvement of the participants is expected to be no more than twelve hours of the three visits to the Virtual Environments Lab.

Withdrawal from the study:

If you decide to withdraw from the study, the information collected up to that time will continue to be used by the research team. It may not be removed. This information will only be used for the purposes of this study.

Information collected and used by the research team will be stored by Scott MacKinnon and he is the person responsible for keeping it secure. Withdrawal from the study will not affect your standing with Memorial University, The School of Human Kinetics and Recreation, The School of Engineering and Applied Science, or the Virtual Environments for Knowledge Mobilization Project.

Possible benefits:

There are no known direct benefits to the participants of this study.

The knowledge gained from this study will support efforts to improve training in the maritime community.

Possible risks:

Participants will be equipped with electrodes on several locations on their body. While these self-adhesive electrodes are only applied to the skin, or are worn attached to a head cap, the adhesive gel and tape that is used to secure the wires may irritate sensitive skin. The application method employed in this study is common practice in both research and clinical applications. Skin sensitivity will be assessed prior to the application of the sensors and should the skin become irritated to a point of discomfort the participant retains the right to withdrawal. All efforts will be made to minimize the duration of skin exposure to the adhesive gel and tape.

Navigation through the virtual space may cause some to experience symptoms of motion (or simulator) sickness. Simulator sickness will be assessed prior to the study and will be monitored throughout. The third person view point that will be used in this study is not known to be highly provocative of simulator sickness symptoms in non- or minimally susceptible individuals.

Exposure to a computer screen may cause eye strain in some participants. Screen time exposure is minimal, and therefore there is minimal expected discomfort. The distance from the participant to the screen will be selected such that it reduces the potential for eye strain and discomfort. Eye strain is expected to be not more than would be experienced during normal computer usage of the same duration.

If at any time the participant experiences symptoms or discomfort which prevent them from continuing in this study they retain the right to withdrawal from the study.

Confidentiality vs. Anonymity

There is a difference between confidentiality and anonymity: Confidentiality is ensuring that identities of participants are accessible only to those authorized to have access. Anonymity is a result of not disclosing participant's identifying characteristics (such as name or description of physical appearance).

Confidentiality and Storage of Data:

Protecting your privacy and maintaining confidentiality is an important goal of the research team. Every effort to protect your privacy will be made. However it cannot be guaranteed. For example we may be required by law to allow access to research records.

When you sign this consent form you give us permission to

- Collect information from you
- Share information with the people conducting the study
- Share information with the people responsible for protecting your safety

The members of the research team will see study records that identify you by name. Other people may need to <u>look</u> at the study records that identify you by name. This might include the research ethics board. You may ask to see the list of these people. They can look at your records only when one of the research team is present.

<u>Use of records</u>

The research team will collect and use only the information they need for this research study. This information will include your:

- date of birth
- sex
- performance metrics
- physiological data
- subjective assessments

Your name and contact information will be kept in a locked office on a password protected PC by the research team at MUN. It will not be shared with others without your permission. Your name will not appear in any report or article published as a result of this study.

Information collected for this study will be kept for 5 years. Following this period, all electronic records of your participation will be permanently deleted and all paper files will be appropriately destroyed.

Anonymity:

Protecting your privacy and ensuring all personal data recorded during participation remains anonymous is an important goal for the research team. Every reasonable effort will be made to assure your anonymity. You will not be identified in any reports or publications without your explicit written permission.

Recording of Data:

As part of this study, we will be collecting various types of data. Performance metrics will be recorded electronically during computer-based activities: time, speed, and errors; physiological parameters will be collected to assess stress experienced during the test trials: heart rate (EKG), galvanic skin response, respiration rate, skin temperature, and electroencephalogram (EEG). Afterwards, you will also be asked to fill out a questionnaire to report perception of "presence" during the simulation, a questionnaire reporting symptoms of simulator sickness and a questionnaire reporting the overall utility the virtual environment training.

Reporting of Results:

The research team intends to publish the findings of this study in peer reviewed journals and academic conferences. Formal reports will be made available to funding agencies and industry partners. The data will be reported in statistical and descriptive form. Individual information or data will not be reported without your exclusive written consent.

Sharing of Results with Participants:

On completion of data analysis a report will be prepared for dissemination. Participants who wish to be informed of the results will have the opportunity to receive a copy of the final report.

Questions:

You are welcome to ask questions at any time during your participation in this research. If you would like more information about this study, please contact:
Dr. Scott MacKinnon
(709) 864-6936
smackinn@mun.ca

Jennifer Smith
(709) 864-6764
jennifersmith@mun.ca

ICEHR Statement:

The proposal for this research has been reviewed by the Interdisciplinary Committee on Ethics in Human Research and found to be in compliance with Memorial University's ethics policy. If you have ethical concerns about the research (such as the way you have been treated or your rights as a participant), you may contact the Chairperson of the ICEHR at icehr@mun.ca or by telephone at 709-864-2861.

Consent:

Your signature on this form means that:

- You have read the information about the research.
- You have been able to ask questions about this study.
- You are satisfied with the answers to all your questions.
- You understand what the study is about and what you will be doing.
- You understand that you are free to withdraw from the study at any time, without having to give a reason, and that doing so will not affect you now or in the future.
- You understand that any data collected from you up to the point of your withdrawal will be retained by the researcher for use in the research study.

If you sign this form, you do not give up your legal rights and do not release the researchers from their professional responsibilities.

Your signature:

I have read and understood what this study is about and appreciate the risks and benefits. I have had adequate time to think about this and had the opportunity to ask questions and my questions have been answered.

- ☐ I agree to participate in the research project understanding the risks and contributions of my participation, that my participation is voluntary, and that I may end my participation at any time.
- ☐ I agree to the use of quotations but do not want my name to be identified in any publications resulting from this study.
- ☐ I agree to having all of the following physiological parameters recorded during my participation in this study.

  - ☐ Heart Rate (EKG)
  - ☐ Galvanic Skin Response
  - ☐ Skin Temperature
  - ☐ Respiration
  - ☐ Electroencephalogram (EEG)

- ☐ I agree to the use of my responses to all questionnaires completed during my participation in this study

A copy of this Informed Consent Form has been given to me for my records.

______________________________ _________________________

Signature of Participant Date

Researcher's Signature:

I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

______________________________ _________________________

Signature of Principal Investigator Date

Appendix G: Video Game Experience Questionnaire

Participant Number: _______________

Gender (circle one): M / F

Age: __________

1) Have you ever played video or computer games? (circle one)
If no, then you do not need to fill in the rest of the questionnaire.

Yes / No

2) How long have you been playing video games?

Time (in years): ______________

3) How many hours, on average, do you spend playing videogames per week?

_________Hours per week

4) Which of the following video game genres do you prefer?

A) Action
B) Action-adventure
C) Adventure
D) Role-playing
F) Simulation
G) Strategy
H) Other (Please Specify): ________________________________

5) What is your preferred view angle in video games?

A) First person perspective (Player experiences the game from the eyes of the character)
B) Third person perspective (Player experiences the game from an over the shoulders view of the character)
C) Isometric perspective (Aerial viewpoint of the map)

6) How familiar are you using the controller/interface employed in this study?

A) Not at all
B) Somewhat proficient
C) Proficient
D) Expert

7) How proficient/experienced would you rate your level of skill, in first person vantage point games?

A) Never Played
B) Played but Not Experienced
C) Played With Some Experience
D) I Consider Myself an Expert

8) Which of the following gaming systems do you have experience using?

A) Xbox
B) Nintendo Wii
C) Play Station
D) iPad/Android/Mobile platform
D) PC Based
E) Other (Please Specify) _________________________________
F) Of the systems you have experience using, with which do you have the most experience? ________________________________

Appendix H: Offshore Experience Questionnaire

Participant Number: ______________

1) Have you worked in the marine or offshore industry? (i.e. ships, ports, oil platforms, etc.)

Yes / No

If yes, please list the marine environments you have worked:

____

____

2) Have you ever had emergency evacuation training?

Yes / No

If yes, please list the type of training you have:

____

____

3) Have you ever worked on an offshore platform/vessel?
If no, do not answer the rest of the questionnaire.

Yes / No

4) How long have you worked on an offshore platform/vessel?

________ In years

5) On what kind of an offshore platform or vessel have you worked on?

A) Fixed platform
B) Semi-submersible platform
C) Floating Production Storage and Offloading (FPSO)
D) Gravity Based Structure (GBS)
E) Spar platform

6) How many hours, on average, have you spent at sea?

________ In hours

7) Do you have any experience using the platform layout represented by AVERT in this study?

Yes / No

8) Have you ever participated in emergency preparedness training or fire drills?

Yes / No

Appendix I: Modified Witmer & Singer Immersive Tendencies Questionnaire

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators & Virtual Environments, 7(3), 225-240.

Participant Number: ____________________

1. Do you ever get extremely involved in projects that are assigned to you by your boss or your instructor, to the exclusion of other tasks?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Very Often

2. How easily can you switch your attention from the task in which you are currently involved to a new task?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all easily — Very Easily

3. How frequently do you get emotionally involved (angry, sad, happy) in the news stories that you read or hear?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never — Very Often

4. Do you easily become deeply involved in movies or TV dramas?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Very Easily

5. Do you ever become so involved in a television program or book that people have problems getting your attention?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all — Very Often

6. How mentally alert do you feel at the present time?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Alert Very Alert

7. How frequently do you find yourself closely identifying with the characters in a story line?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very frequently

8. Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick or watching the screen?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very Often

9. How good are you at blocking out external distractions when you are involved in something?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not good at all Very Good

10. Do you ever become so involved in a day dream that you are not aware of things happening around you?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very Often

11. How well do you concentrate on enjoyable activities?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not well at all Very Well

12. How well do you concentrate on disagreeable tasks?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not well at all Very well

13. Have you ever gotten excited during a chase or fight scene on TV or in the movies?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very Often

14. Have you ever gotten scared by something happening on a TV show or in a movie?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very Often

15. Have you ever remained apprehensive or fearful long after watching a scary movie?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Never Very Often

16. Do you ever become so involved in something that you lose all track of time?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very Often

Appendix J: Kennedy Simulator Sickness Questionnaire

Kennedy, R. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *International Journal Of Aviation Psychology*, *3*(3), 203

Participant Number: ____

Time:____________________
When: After / Before Testing

Symptom	0 No Symptoms	1 Minimal	2 Moderate	3 Severe
General Discomfort				
Fatigue				
Headache				
Eyestrain				
Difficulty Focusing				
Increased Salivation				
Sweating				
Nausea				
Difficulty Concentrating				
Fullness of Head				
Blurred Vision				
Dizzy (eyes open)				
Dizzy (eyes closed)				
Vertigo				
Stomach Awareness				
Burping				

Appendix K: Utility of Virtual Environment and Training Questionnaire

Post-Trial Questionnaire for Session 1 and Session 2:

Participant Number: _____________

1. What did you find most challenging in completing the scenarios?

____

2. What do you think are important factors for success in the scenarios?

____

3. Did you have a strategy to learn the environment and respond to scenarios? Y / N

If yes, please briefly describe your strategy.
____
____

4. Did you have enough time to complete the scenarios in the way you would have wanted?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not enough time Too much time

Post-Trial Questionnaire for Session 3

Participant Number: ___________

A. Tutorials:

1. How helpful was the basic offshore safety training material?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

2. How helpful was the walkthrough video?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

3. How helpful was the escape route videos?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

4. How helpful were the floor plans?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

5. How helpful was the 30-minute platform exploration time (free-roam around the platform)?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

6. Did you have enough time to review the material the way you would have wanted?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not enough time Too much time

7. Did the tutorial provide you with sufficient understanding to respond to the test scenarios?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Completely

8. How helpful were the quizzes?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

9. How helpful was the feedback you received?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

B. Scenarios:

4. Reflect on your performance and rate your overall performance in completing the scenarios.

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all successful Very successful

5. How realistic were the hazards/scenarios you experienced?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all compelling Very compelling

6. What went well during the scenarios?

____

7. What did you find most challenging in completing the scenarios?

____

8. What do you think are important factors for success in the scenarios?

____

9. Did you have a strategy to learn the environment and respond to scenarios? Y / N

If yes, please briefly describe your strategy.

____

____

10. How comfortable were you at navigating the platform?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very comfortable

11. How comfortable were you at responding to alarms?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very comfortable

12. How comfortable were you at avoiding hazards?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very comfortable

10. Did you have enough time to complete the scenarios in the way you would have wanted?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not enough time Too much time

C. Simulator:

1. Rate the ease of use of the controls:

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very Hard Very Easy

2. How helpful was the mini map?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

3. How helpful were the check box questions?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

4. How helpful was the helper text?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

5. How helpful were the signs on the platform?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very helpful

6. Rate the ease of use of interacting with things in the virtual environment (i.e. turning on flashlight, opening and closing doors, moving your T-card, etc):

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very Hard Very Easy

D. Comments and Suggestions:

Any Comments?

____

____

____

____

Appendix L: Modified Witmer and Singer Presence Questionnaire

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators & Virtual Environments, 7(3), 225-240.

Participant Number: _______

1. How responsive was the environment to actions that you initiated (or performed)?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all Responsive ... Very Responsive

2. How natural did your interactions with the environment seem?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all Natural ... Very Natural

3. How much did the visual aspects of the environment involve you?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all ... Completely

4. How much did the auditory aspects of the environment involve you?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all ... Completely

5. How natural was the mechanism which controlled movement through the environment?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all Natural ... Very Natural

6. How compelling was your sense of objects moving through space?

   ☐ ☐ ☐ ☐ ☐ ☐ ☐

   Not at all compelling ... Very Compelling

7. Were you able to anticipate what would happen next in response to the actions that you performed?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Completely

8. How completely were you able to actively survey or search the environment using vision?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Completely

9. How well could you identify sounds?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very well

10. How compelling was your sense of moving around inside the virtual environment?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all compelling Very Compelling

11. How involved were you in the virtual environment experience?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Involved Very Involved

12. How much delay did you experience between your actions and expected out-comes?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very delayed No delay at all

13. How quickly did you adjust to the virtual environment experience?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not quickly at all Very quickly

14. How proficient in moving and interacting with the virtual environment did you feel at the end of the experience?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all proficient Very Proficient

15. How much did the visual display quality interfere or distract you from performing assigned tasks or required activities?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very Distracting Not at all distracting

16. How much did the control devices interfere with the performance of assigned tasks or with other activities?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Very Interfering Not interfering at all

17. How well could you concentrate on the assigned tasks or required activities rather than on the mechanisms used to perform those tasks or activities?

☐ ☐ ☐ ☐ ☐ ☐ ☐

Not at all Very well

Appendix M: Multiple Choice Quiz Questions for Each Session.

Session 1 – Baseline Safety Induction Quiz

Learning Objectives	Quiz Questions
Spatial Awareness	1. What does TSR stand for? 2. On which deck is your cabin located? 3. Where is your worksite onboard the vessel? 4. Where is your primary muster station located? 5. Where is your secondary muster station located? 7. Which areas of the platform do you NOT have access to? 8. What does the blast wall do? 9. What is the purpose of a muster?
Cognitive Awareness	10. In what situation would you be required to go to your lifeboat station? 11. Who is in charge at the muster station? 17. What is the expected chain of events in an emergency situation?
Alarm Recognition	12. What does GPA stand for? 13. What does the PAPA stand for? 14. Where do you go after hearing the Prepare to Abandon Platform Alarm? 15. What information does the PA provide?
Routes	16. How many routes are available from your cabin to the lifeboat station?
Muster Procedures	18. What is the T-card for? 19. What is a clear call? 20. What is the purpose of the lifeboat station?
Safe Practices	6. Fire doors and watertight doors should always be?

Session 2 – Advanced Alarm Recognition Quiz

Learning Objectives	Quiz Questions
Cognitive Awareness	1. What are the emergency duties of general personnel? 2. Who is in charge at the lifeboat station? 3. What do you do in the event of a major incident and PAPA alarm? 4. What are the TWO most important pieces of information you need in order to respond to an emergency?
Alarm Recognition	5. When hearing a GPA alarm where are you being directed to go? 6. What alarm is this? [GPA] 7. What visual alarm is associated with the General Platform Alarm? 8. What alarm is this? [PAPA] 9. What do you do when you see a steady green light? 10. How do you raise the alarm? 11. What does MAC stand for? 12. Why is it important to listen for a PA? 13. When should you raise the alarm? 14. What is considered a reportable incident? 15. Where is the nearest manual alarm call point to your situation at your worksite?
Re-routing	16. List some major accidents offshore:
Muster Procedures	17. How do you correctly notify the muster checker that you've mustered? 18. What do you do after you've mustered? 19. Why is it important to report to your muster station immediately in an emergency situation?
Routes	20. In your opinion what is the most efficient route from your worksite to your muster station?

Session 3 – Advanced Hazard Awareness Quiz

Learning Objectives	Quiz Questions
Spatial Awareness	1. Where is the TSR on the platform? 2. The station bill provides what information? 3. What do you do in the event that your primary muster station is compromised? 4. If you can't remember how to get to your muster station what should you do?
Cognitive Awareness	5. What does the OIM do in the event of an emergency? 6. What do you do in the event of a minor incident?
Alarm Recognition	7. What would you do in the event of an alarm that wasn't followed by a PA announcement?
Routes	8. What is the safest exit to take given where the hazard is located?
Re-routing	9. What PPE should you wear in smoke environments? 10. What are some of the possible hazards of an offshore work environment? 11. What does EER stand for? 12. Why is it important to know more than one route? 13. What areas have increased risk on the platform? 14. What do you do when your primary muster route has been blocked? 15. What are some things that could cause a situation to escalate? 16. What is the expected chain of events in an emergency situation? 17. What is the safest exit to take given where the hazard is located? 18. What is the safest exit to take given where the hazard is located?
Muster Procedures	19. How do you indicate you've mustered at the lifeboat station? 20. Which of the following is a good reason to leave the muster station?