PEDAGOGICAL STUDIES IN VIRTUAL OFFSHORE SAFETY TRAINING

by

© Jennifer Joan Elizabeth Smith

A Thesis submitted to the

School of Graduate Studies

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

January 2020

St. John's Newfoundland and Labrador

ABSTRACT

To better prepare the offshore workforce for emergencies, operators and regulators need to use evidence-based safety training. This research aims to provide such evidence by employing an experimental program to evaluate virtual environment (VE) training as a plausible means to enhance mandatory offshore egress training. Combining VE technology with a well-designed, pedagogically informed training program, and carefully selected data-mining tools, can support the development of trainee competence in emergencies by providing artificial experience in credible situations and tracking trainee performance throughout the VE training.

Evidence from this research supports the use of VE training to address pedagogical gaps in the training. Key gaps include the following: 1) conventional training is predominantly provided by fixed-time instruction, which results in crews with nominal competence, 2) the frequency of recurrency training is not informed by evidence on crews' susceptibility to forget training, 3) crews' learning outcomes are not measured or monitored, which results in no information to assess training transfer, and 4) due to safety constraints, muster drills lack the realism of how emergency situations unfold in real life.

Lessons from pedagogical theory and data-mining methodology were used to provide empirical and modeling evidence to inform offshore and maritime domains on the application of VE training. The scope of the research involved using the VE training as a human behaviour laboratory during a longitudinal study. The context of the study was to teach the necessary egress skills to evacuate a virtual oil platform during an emergency.

To address the pedagogical gaps and evaluate VE training, this thesis is comprised of four research papers. The first paper investigates the influence of the simulation-based mastery learning (SBML) pedagogical framework on the development of competence at the different learning phases, specifically the acquisition, retention, and transfer of egress skills. The second paper uses human performance data from the VE training to develop a decision tree (DT) diagnostic tool to compare the efficacy of different delivery methods for VE training. The third paper evaluates the retention and maintenance of the VE training after a period of 6 to 9-months. The fourth paper uses DT modeling to evaluate skill transfer and develop a predictive tool to analyze the efficacy of VE training on a systemic level to support future adaptive training programs.

The overall contribution of this research is the use of pedagogical frameworks and data-mining tools to improve the design, delivery, and assessment, of VE training. The concept of this work is established in the context of offshore and maritime safety, however the approaches are generalizable to many virtual training applications in other domains.

ACKNOWLEDGEMENTS

I would like to acknowledge with many thanks the support of my supervisors, Dr. Brian Veitch, Dr. Sarah Power, and Dr. Scott MacKinnon. They have provided invaluable guidance, thought-provoking questions, and helped produce impactful research.

I am grateful for the financial support provided by the Natural Sciences and Engineering Research Council of Canada - Husky Energy Industrial Research Chair in Safety at Sea.

I would also like to recognize the hard work and support of the Safety at Sea team. In particular, thanks to Sean Jessome, Robert Rees, Kyle Doody, Allison Blundon, Sinh Duc Bui, Norafneeza Norazahar, Syed Danial, Randy Billard, and Erik Veitch. Special thanks to Mashrura Musharraf for being a true inspiration, courageous leader, and friend throughout this research.

Thanks to all the participants who volunteered for the longitudinal experimental study. Without their interest and time, this research would not have been possible.

Many thanks to my family and friends for their support over the years, without their encouragement I would not have finished. Finally, I owe so much to my husband, Geoffrey Hoffe. After I suffered a stroke (of neither luck nor genius) that nearly ended my PhD journey, Geoff helped me pick up the pieces and finish what I had started. I am eternally grateful for his unwavering support and encouragement.

Table of Contents

List of Tables

List of Figures

Nomenclature

ACQ	Acquisition
AVERT	All-hands virtual emergency response trainer
DM	Data Matrix
DT	Decision Tree
EDM	Educational Data Mining
GPA	General Platform Alarm
HRA	Human Reliability Analysis
KB	Knowledge Base
LA	Learning Analytics
LO	Learning Objective
LB	Lifeboat station
LBT	Lecture Based Teaching
MS	Muster Station
PA	Public Address
PAPA	Prepare to Abandon Platform Alarm
PSF	Performance Shaping Functions
RET	Retention
RPDM	Recognition Primed Decision Making
SBML	Simulation Based Mastery Learning
SME	Subject Matter Expert
VE	Virtual Environment

1. INTRODUCTION

1.1. Problem & Purpose Statement

Pedagogical gaps in offshore safety training bring to question the petroleum industry's preparedness for major emergencies. Emergencies offshore present complex, time-sensitive situations that can make evacuations difficult to execute. These safety-critical egress situations do not afford second chances and require the prompt response of prepared crews to ensure that all personnel onboard have been accounted for in an emergency. Conventional emergency egress training is often participatory in nature and is provided using fixed-time instruction, resulting in a workforce with nominal certification and variable competence. Thus, depending on the training received, how people respond to emergencies can vary. This variability in procedure compliance is a safety concern for the offshore workplace, especially during critical operations or emergencies. One of the reasons conventional training is not properly preparing crews for emergency response is that this form of training lacks the pedagogical structure to assure crew competence is acquired and maintained. There are four problems with conventional training methods:

(1) the training does not address individual variability in learning skills,

(2) the training is often forgotten before the mandated recurrency training is scheduled,

(3) the training does not measure learning outcomes and as a result does not inform the transfer of training, and

(4) the training is not representative of the conditions in real emergencies.

The purpose of this research is to address the pedagogical gaps in conventional training by investigating the efficacy of virtual environment (VE) training in the offshore safety domain. More specifically, this research draws on pedagogical frameworks and data-mining methodologies to empirically evaluate the efficacy of VE training for offshore emergency egress, and to provide a data-driven diagnostic lens for a more thorough assessment of different VE training interventions. A human behaviour laboratory, in the form of a VE training setting, is used to measure trainee competence and evaluate new training interventions. From a pedagogical perspective, a simulation-based mastery learning (SBML) framework is explored. This research provides empirical evidence of VE training efficacy at three stages of learning to improve the delivery (acquisition), maintenance (retention), and application (transfer) of emergency egress skills. From a data-mining perspective, a method called decision tree (DT) modeling is applied and evaluated. This research provides modeling evidence to assess the diagnostic and predictive capabilities of DT tools for individual competence assessment and to diagnose the strengths and weaknesses of the VE training at a systemic level. This work is a precursor to the development of future adaptive VE training programs in the offshore safety domain.

This chapter describes the rationale for the research (Section 1.2), explains the VE training applied for offshore emergency egress (Section 1.3), and outlines the scope of work including the research questions, hypotheses, and methods used to test the hypotheses (Section 1.4). This chapter also presents relevant literature on pedagogical theories and data-mining methodologies and outlines the knowledge gaps related to these domains (Section 1.5). The rest of the chapter discusses the contributions of the work (Section 1.6), and explains the organization of the thesis (Section 1.7).

1.2. Gaps in Current Training and the Potential of Virtual Training

Conventional training in many industries tends to set fixed instructional time due to regulatory, logistical, and financial constraints. Fixed-time instruction contributes to learning outcome variance because individual learning needs are not adequately addressed (Cook et al., 2013). In the context of offshore safety training, new personnel or short-term contractors who arrive on an offshore platform are provided with conventional training (e.g. videos and walk-throughs). As a result, the competence of each individual participating in the same training program can be very different. The initial variability in acquired competence can have long lasting repercussions on maintaining egress skills.

In a review of multi-day safety training courses in medical, military, and marine domains, Sanli and Carnahan (2018) concluded that complex skills could be remembered for at most a six-month period following training. Knowing that egress skills deteriorate over time without practice, offshore industry standards require personnel to undergo recurrency training if they have been away from the platform for an extended period. For example, regulations for the Canadian offshore industry dictate mandatory retraining if personnel have been away from the platform for a period of 6 months or more (CAPP, 2015). However, these regulations are not evidence-based and do not provide recommendations on how training maintenance schedules should account for individual differences in relearning forgotten egress skills. The expectation is that any existing training gaps will be addressed through on-the-job training. Regulations require crews working offshore to practice egress skills in the form of routine muster and evacuation drills offshore. The drills typically occur at a regularly scheduled time each week and are usually performed in calm, non-risky situations that do not mimic real emergencies. Over time, this form of routine drill in benign conditions can cause negative learning and complacency amongst the crew.

Experience from the medical domain indicates that well-designed training programs using simulation and virtual environment (VE) technologies can address pedagogical gaps and provide competence assurance (McGaghie et al., 2006; Barsuk et al., 2010; Cohen et al., 2013; Moazed et al., 2013; McGaghie et al., 2014; Barsuk et al., 2016). For the offshore industry, VE training has been proposed as a training solution because it can provide a safe means to practice emergency evacuation exercises and help prepare personnel to respond effectively to realistic emergencies. VE technology can enhance conventional classroom and on-the-job training by teaching basic offshore safety protocols, such as onboard familiarization and emergency evacuation, before crews have been deployed offshore. This technology can be used to test new training protocols to determine if the intervention in training can improve competence and compliance (e.g. to improve safety behaviours in offshore emergencies). However, before offshore and maritime industries can adopt simulation-based training, its training utility should be evaluated from a pedagogical perspective to ensure it is properly teaching workers.

Although simulation and VE training technologies are being used in offshore and maritime education, few pedagogical studies have been conducted to examine the training efficacy (Sellberg et al., 2017). This shows that maritime and offshore sectors are gradually transitioning, shifting from passive training to simulation-based training methods. While this transition is positive because it recognizes the need to improve how training is provided in these domains, the lack of evidence-based pedagogical studies using simulation in offshore safety training presents a lingering problem that must be addressed. There is a need in offshore and maritime domains to provide guidance on how to deliver training and how to assess performance using simulation technology (Sellberg et al., 2017).

To optimize VE training for offshore egress, operators and regulators need to determine how comprehensive VE training should be to prepare crews for the multitude of emergencies that could arise. Virtual practice exercises should match the conditions (i.e. high risk, time pressure, and complexity) crews would be expected to experience in real emergencies. This would enhance routine drills that cannot otherwise replicate these conditions due to the ethical, logistical, and financial constraints. However, it is impracticable to rehearse for all possible situations in VE training. Virtual training should focus on supporting the transfer ability of training, specifically, the application, generalization, and maintenance of skills learned in one training context to new situations (Blume et al., 2010). Therefore, VE training should measure learning outcomes as a way to determine when the trainee has achieved competence.

Unlike conventional training, VE technology can collect human performance data during the virtual training (e.g. how people responded to cues from a situation). VE technology combined with data-mining methods can be used to recognize behavioural patterns and inform how people make decisions (Musharraf et al., 2018). This information can be used on an individual level to diagnose a person's strengths or weaknesses in performing a particular task in the VE training. It can also be used to detect when the person has achieved competence and predict when that person is sufficiently equipped to apply their skills to new situations (e.g. predicting their future decisions based on the value of the attributes in a given VE scenario). At a systemic level, data-mining methods can be used to diagnose the strengths and weaknesses of the VE training as a whole. This analysis could help inform instructional designers on how to design, deliver, evaluate and improve VE training.

1.3. Overview of Virtual Environment Training for Offshore Egress

To investigate the efficacy of VE training in the offshore domain, this research uses the All-hands Virtual Emergency Response Trainer (AVERT) developed at Memorial University. AVERT is a first-person perspective desktop VE that provides participants with a naturalistic representation of an offshore Floating Production Storage and Offloading (FPSO) vessel (House et al., 2014). AVERT provides a variety of emergency preparedness training exercises, from basic muster drills to full emergency evacuation scenarios (e.g. on board fires, blackouts, and explosions). Participants can move within the FPSO by controlling a first-person perspective avatar of an offshore worker using a gamepad controller (Xbox). AVERT is intended to train personnel in safe work practices and how to muster at their designated muster stations in the event of an emergency. The training curriculum for AVERT is based on subject matter expert guidance and industry regulations (Transport Canada 2007; International Maritime Organization 2001; Canadian Association of Petroleum Producers 2015; International Association of Drilling Contractors 2009; International Association of Oil and Gas Procedures 2010). The core learning objectives include familiarity with the platform layout, emergency alarms, egress routes, safety protocols, and mustering procedures.

1.4. Scope of Work and Research Objectives

This doctoral thesis encompasses four papers to address the following research objectives: (1) determine how to design and deliver VE training to assure demonstrable competence in the workforce, (2) assess data modeling tools to improve the assessment of different VE training interventions, (3) evaluate the retention and maintenance of VE training, and (4) improve VE training to foster skill transfer and prepare the workforce for a wide variety of emergencies. The following questions directed the research for each paper:

Q1. <u>How to design and deliver VE training:</u> Can the simulation-based mastery learning (SBML) framework be used to effectively deliver virtual offshore emergency egress training? Does the SBML approach yield quantitatively better results when compared to the lecture-based teaching (LBT) alternative in terms of bringing participants to competence?

Q2. <u>How to assess VE training using data-mining tools:</u> Can decision trees (DTs) be used to evaluate the efficacy of virtual training at a systemic level (e.g. diagnose the strengths and weaknesses of different pedagogical approaches to VE training)? Can DT modeling provide improved diagnostics compared to traditional performance scores on participants' competence and compliance?

Q3. <u>How to evaluate the retention of VE training:</u> Can egress skills acquired using SBML training be remembered after a period of 6 to 9-months (without any other form of training intervention)? Can the VE-based retraining bring participants back up to competence in all learning objectives?

Q4. How to improve VE training to foster skill transfer: Can the SBML training prepare participants to demonstrate the transfer of their egress skills to novel emergency scenarios? Can DTs predict when participants have sufficiently learned enough to apply their skills to new emergencies within the scope of the training? Can the collection of DTs from a group of participants' data be used to identify group patterns in performance and predict the preparedness of the participants (e.g. assessing the VE training at a systemic level)?

The following hypotheses were posed to inform the research questions:

H1. The SBML framework can be used to provide virtual offshore emergency egress training. The SBML approach can improve learning outcomes (e.g. competence and compliance) when compared with the LBT approach.

H2. DT modeling is a useful data classification tool to evaluate the efficacy of a VE training program. The visual representation of the DTs can identify gaps in the training, which traditional performance metrics lack, by comparing the DT depiction of the participant's understanding of emergency egress to the intended learning objectives of the VE training program.

H3. The SBML trained participants can retain egress skills over a period of 6 to 9-months. The retraining matrix can address participants' deficiencies and return them to demonstrable competence in all learning objectives.

H4. The SBML training adequately prepares participants to demonstrate competence in novel emergencies. DTs can help predict participants' future decisions in scenarios by comparing the DT depiction of the participants' understanding of an emergency to the value of the attributes in a given new scenario. DTs are useful for evaluating training transfer (e.g. detecting when participants have achieved competence and predicting when they are prepared to apply their skills to new situations).

Several methods were used to test these hypotheses. The first and third goals of this research were to determine how to design and deliver VE training and how to evaluate the retention and maintenance of VE training in order to address variability and assure demonstrable competence in the workforce. To investigate the efficacy of VE training, the pedagogical theory of simulation-based mastery learning (SBML) is explored. SBML is an established framework from medical education (McGaghie et al. 2014) and is grounded in Benjamin Bloom's (1971) mastery learning theory. Mastery learning is an approach that tailors the instruction and pace to each individual learner (Bloom 1971; Gusky, 2007). This is achieved by using variable instruction time to accommodate different learning paces (e.g. allowing everyone to reach competence by meeting their individual needs). This framework is a suitable for behavioural learning and procedural skills (e.g. safety training tasks). The VE training is used as a human behaviour laboratory in order to determine the impact of the different pedagogical frameworks on competences.

The evaluation of VE training was divided into four parts. The first part focuses on comparing the SBML approach to a conventional lecture-based teaching (LBT). Both the LBT and SBML approaches are applied to the same virtual environment to teach participants evacuation procedures for an offshore petroleum installation. Chapter 2 reports the results of this comparison.

A longitudinal study encompasses the second, third, and fourth parts of this research as depicted in Figure 1.1. The longitudinal study investigates the training efficacy of VE training empirically at three critical learning stages: 1) a skill acquisition phase; 2) a skill retention assessment and retraining phase; and 3) a transfer of training to novel situations phase. The scope of the work involves collecting performance metrics from the VE training at each of the learning phases. The empirical evidence reported from this portion of the research offers one approach to evaluate the training efficacy. Chapter 2 reports on the skill acquisition phase, chapter 3 reports the skill retention and retraining phase, and chapter 5 reports the training transfer phase.

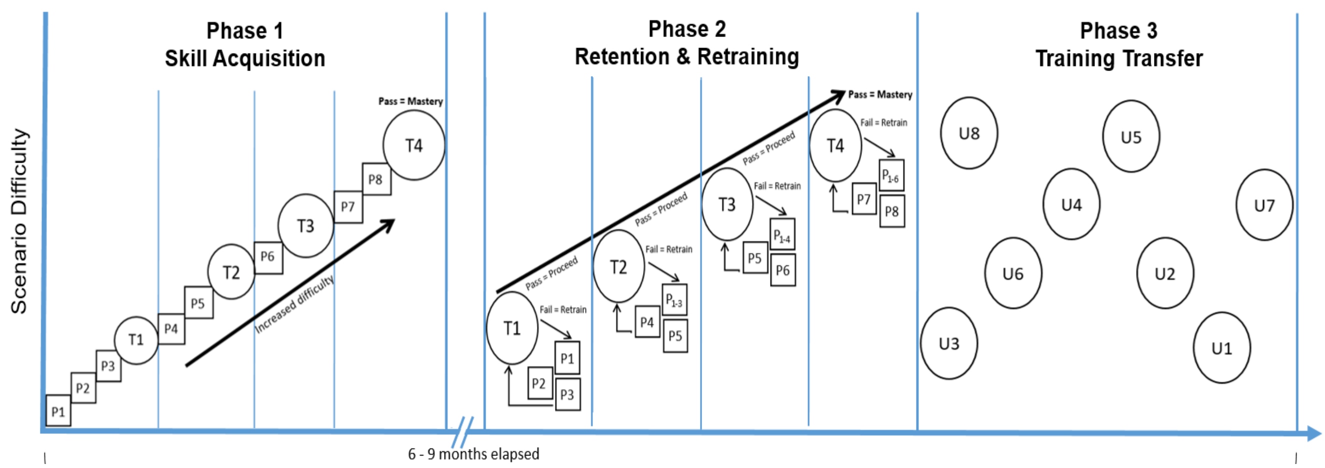


Figure 1.1: Longitudinal Experiment Timeline

The second goal of this research is to assess data modeling tools to improve the VE training capabilities to assess competence in the offshore safety domain. To determine ways to improve competence assessment, a data-mining method called decision tree (DT) modeling is used. DTs offer a pattern recognition lens to assess the training efficacy that goes beyond performance metrics. Unlike conventional training, VE technology can easily collect human performance data during the virtual training (e.g. how people responded to cues from a situation). DTs are based on supervised learning theory and use classification to visualize data patterns from data collected from VE training. Figure 1.2 depicts the process used to develop decision trees from the human performance data recorded from the VE training.

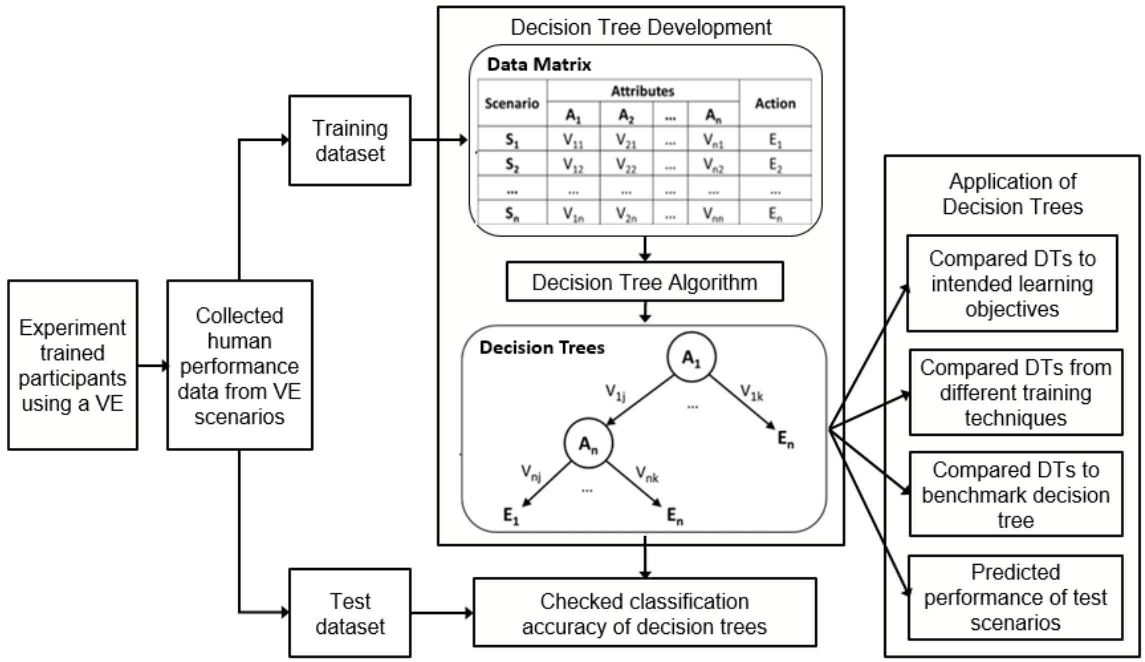


Figure 1.2: Process used to develop DTs and assess training efficacy.

The data-driven behavioural or learning patterns from the DTs can be used to inform how people make decisions. Musharraf et al. (2018) used DTs to model participants' route selection strategies from recorded evidence in the VE training. DT modeling in educational contexts often provide methods to diagnose the strengths and weaknesses of individuals. This research extends Musharraf et al. (2018) methodology beyond assessing individual learners' strengths and weaknesses, to focus on evaluating the effectiveness of an entire VE training program. The intention is to use DT modeling to develop adaptive training tools. This research investigates DT modeling as a complementary measure to conventional empirical evidence to improve both the delivery and assessment of competence in VE training. The diagnostic and predictive capabilities of the DTs are used to systemically assess the strengths and weaknesses of the two pedagogical approaches, SBML and LBT. Chapter 3 reports the results of this analysis.

The fourth goal of this research is to improve VE training from the perspective of fostering training transfer. DTs have the potential to predict when a trainee is sufficiently equipped to apply their skills to new situations. Following the findings from Musharraf et al. (2018), DTs constructed from information collected during VE training exercises can predict how a trainee will perform in similar circumstances. Similarly, visually representing trainees' decision strategies based on prior performance in VE training can inform the effectiveness of the training program (e.g. did participants develop strategies that match the intentions of the training). This research investigates the utility of DTs to diagnose the strengths and weaknesses of VE training at a systemic level. This is accomplished by using the DTs to evaluate the efficacy of VE training at the three learning phases in the longitudinal study: skill acquisition, skill retention and skill transfer. The results of this work can inform the development of future adaptive training programs in the offshore safety domain. Chapter 5 reports the results of this work.

1.5. State of Knowledge and Gaps

Two theoretical frameworks constitute the foundation of this doctoral research: pedagogical theory of mastery learning (Bloom, 1971), and data-mining methodology of decision trees (Quinlan, 1986; Han et al., 2012). The literature review below provides an overview of the pedagogy framework, and the data-mining methodology used to investigate the efficacy of training from training design, delivery, and assessment perspectives.

1.5.1. Pedagogical Theory

This section describes three pillars for assessing training: 1) acquiring skills – learning, 2) retaining skills and fending off forgetting, and 3) transferring skills – the versatility of training to apply lessons learned to new situations.

1.5.1.1. Skill Acquisition (Learning)

Benjamin Bloom’s learning for mastery is an instructional paradigm that ensures all learners achieve competence by providing them with individualized feedback and corrective measures (Bloom, 1971; Gusky, 2007). Mastery learning, by design, has a fixed competence threshold for learners to achieve, and the instruction time required to achieve that threshold usually varies due to individual variability (Cook et al., 2013). Mastery learning adapts the pace of instruction for each learner by monitoring learner progress through formative assessment and feedback (see Figure 1.3). Formative assessments help determine if learners have mastered the material sufficiently to move on to enrichment activities, or if they require corrective exercises before advancing to more difficult concepts.

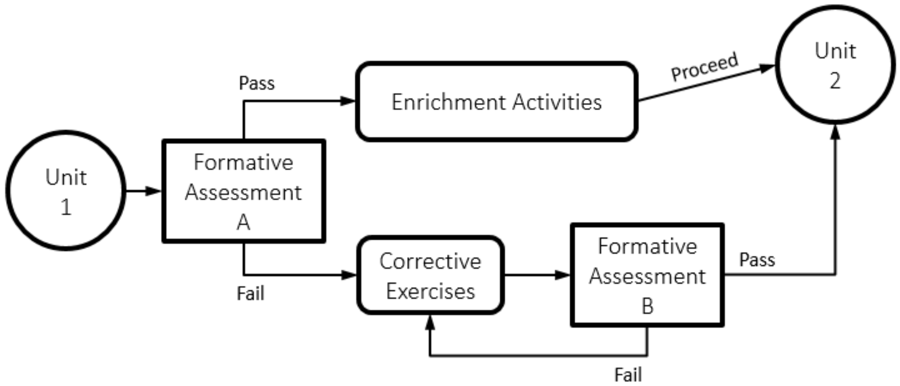


Figure 1.3: Bloom's mastery learning framework (after Gusky, 2007)

Simulation based mastery learning (SBML) is a pedagogical approach developed within the medical education field (McGaghie et al., 2006; Barsuk et al., 2010; Cohen et al., 2013; Moazed et al., 2013; Barsuk et al., 2016) and is based on Bloom's competence-based theory of learning for mastery. The SBML protocol gradually guides learners through the learning objectives, assesses their performance, provides corrective feedback, and allows the learners to proceed to the next level of training once they have demonstrated proficiency in the basic concepts. McGaghie et al. (2014) described seven aspects of SBML as follows:

(1) Assess the learner's entry level;

(2) Provide clear learning objectives that are arranged in increasing difficulty;

(3) Provide training exercises that are engaging and focused on each learning objective;

(4) Set a minimum pass standard for each unit;

(5) Provide a formative assessment of the learner's progress with feedback so they can gauge their progress and recognize where their proficiency level is in relation to the standard;

(6) Allow learners to advance to the next unit once they have reached the mastery standard (i.e. met or exceeded the mastery competence level); and

(7) Encourage learners to continue their practice on the unit until they have achieved the mastery standard.

The advantages of SBML are illustrated in the Dreyfus/Benner five-staged skill acquisition framework (see Figure 1.4). Dreyfus (1980), Benner (1982), and Griswold-Theodorson et al. (2015) have described the five-staged framework of training: 1) novice, 2) advanced beginner, 3) competent, 4) proficient, and 5) expert. According to Griswold-Theodorson et al. (2015), SBML training brings learners from stage 2 to stage 3 on the skill acquisition model (advanced beginner to competent).

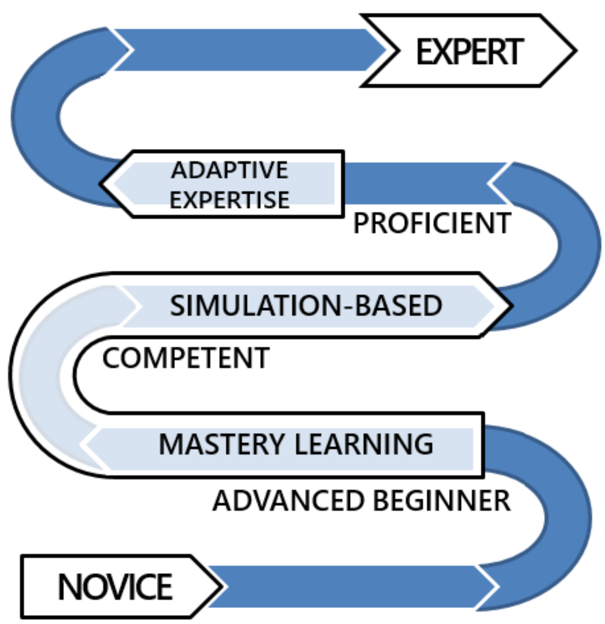


Figure 1.4: Stages of skill acquisition (after Griswold-Theodorson et al., 2015)

This framework describes how that SBML training can bring all learners to a standardized competence, especially for procedural tasks (e.g. where rules must be remembered and followed).

The SBML approach has been tested in the medical field with generally positive results (McGaghie et al., 2014; Griswold-Theodorson et al., 2015), although some evidence is mixed (Cook et al., 2013). Cook et al. (2013) concluded in their systematic review and meta-analysis that mastery learning in simulation-based medical education showed limited evidence of improved outcomes compared to non-mastery instruction, and that the mastery learning method takes longer than other training methods.

1.5.1.2. Retention & Forgetting

Many factors influence how well skills are remembered. Arthur et al. (1998) performed a meta-analysis of skill retention literature and described seven factors that influence skill decay and retention: i) length of time elapsed of non skill use during retention interval, ii) the quality of the original skill acquisition and the amount of overlearning that occurred; iii) skill type and task characteristics (e.g. physical versus cognitive tasks); iv) the methods used to test learning and retention; v) conditions of retrieval or specificity of training (i.e. the similarity of learning and testing contexts); vi) the instructional strategies or methods used to teach the skills; and vii) individual differences in acquiring and retaining skills.

Sanli and Carnahan (2018) in their review of multi-day training courses in medical, military, marine, and offshore safety fields discussed five similar factors that influence skill retention. According to Sanli and Carnahan (2018), the factors that influence skill and knowledge retention in these safety-critical domains include: a) type of skill (e.g. practical and declarative knowledge); b) task complexity and difficulty (e.g. number of steps and order of tasks); c) individual differences and the experience of the learner; d) specificity of training (i.e. closeness of the learning and testing contexts); e) the amount of practice and on the job exposure provided; and f) the frequency that refresher interventions are delivered.

Two main topics will be discussed in the context of virtual offshore egress training: (1) the impact of skill type, such as declarative and procedural knowledge, on forgetting, and (2) the frequency of practice, that is, how often recurrency training is provided and the length of time that passes between the training sessions (Schmidt & Lee, 2005; Wickens et al., 2013).

Skill Type:

Kim et al.'s (2013) learning theories model provides a framework to help explain how skills are learned and forgotten. Kim et al. (2013) integrated four learning theories (Fitts, 1964; Anderson, 1982; Rasmussen, 1986; and VanLehn, 1996) into a three-staged skill acquisition process: 1) declarative stage, 2) mixed stage, and 3) procedural stage. The declarative stage involves learning information or facts. Declarative knowledge can be attained through rote memorization. However, declarative knowledge will degrade with the lack of use (e.g. information will no longer be available in memory for retrieval). The mixed stage involves consolidating information related to the task into a mix of declarative and procedural knowledge. This mix of knowledge occurs when declarative knowledge is transformed into procedural knowledge over time (e.g. gradually associating knowledge, transforming it into rules, and developing heuristics and biases). Therefore, with time and practice, the procedural stage transforms the mixed knowledge into predominantely procedural knowledge. This final procedural knowledge stage is often achieved through overlearning.

According to Kim et al.'s (2013) learning theories model, frequent practice and contextual experience allow experts to proceduralize skills so that they rely less on declarative knowledge and are able to perform the task automatically in response to a situation. Procedural knowledge is implicit as experts possessing the knowledge are able to perform the actions without effort, but are unable to verbalize the knowledge (Wickens et al., 2012). Siu et al. (2016) suggest that learners should be provided with sufficient practice to allow them to reach the proceduralization stage, thereby increasing likelihood of skill retention.

Frequency of Retraining:

The amount of time that elapses between retraining sessions is an important factor to investigate in order to ensure safety-critical skills are maintained. Predicting the rate at which skills will be forgotten can help inform the frequency with which recurrency training should be provided (Wickens et al., 2012).

In a review of multi-day safety training courses, Sanli and Carnahan (2018) concluded that complex skills could be remembered for at most a six-month period without any form of training interventions. Dunlosky et al. (2013) reviewed the common learning techniques in education and found the most effective techniques for retention were self-testing and distributed practice. Similarly, Atesok et al. (2016) reviewed literature on the retention of SBML trained orthopaedic surgery skills and found that repetitive practicing of skills learned in a simulator helped mitigate skill decay even after some time had elapsed (these studies ranged in time elapsed; e.g. follow-up retention assessments occurred at 1 month, 3 months, 6 months, to a maximum of 30 months).

1.5.1.3. Training Transfer

Training should support transferability, specifically, the application, generalization, and maintenance of knowledge and skills learned in one training context to new contexts or situations (Blume et al., 2010). Ideally, the transfer of training is measured by first training skills using a VE or simulator, and then evaluating the skills in the real environment. For example, Magee et al. (2012) conducted a forward and reverse transfer of training experiment to evaluate VE training in the context of developing spatial knowledge for emergency drills on a submarine. The results found positive training benefits from the VE training. However, this form of training transfer study is logistically challenging because of the same issues that make the real world training difficult: limited access, ethical and safety concerns, logistical and financial constraints, in addition to difficulties in ensuring experimental control.

The purpose of training transfer studies is to assess the learners' performance of a practiced task in a new context, or how learning the practiced task helped improve their performance of a new version of the task (Sanli and Carnahan, 2018). Therefore, training transfer can also be evaluated by comparing how knowledge learned in one context or environment can help when applying the skills (or even learning new skills) in a novel or unforeseen context (Wickens et al. 2012). This logic justifies evaluating training transfer using the same environment. Therefore, a VE or simulator can be used as a substitute to the real environment and be used as a human behaviour laboratory to investigate the versatility and transfer of training. However, it is important to consider transfer proximity as it relates to how far reaching the scope of the training is to different contexts. Training transfer is considered near transfer when the test setting is very similar (or has a close proximity) to the knowledge covered in the training. Far transfer occurs when the test setting is very different from the training context (Barnett and Ceci, 2002; Ford et al., 2018).

According to Grossman and Salas (2011), training organizations consider three factors that influence the transfer of training:

(1) Trainee characteristics, such as the learner's cognitive ability, self-efficacy, motivation, and their perceived utility of training;

(2) Training design, such as behavioural modeling, error management, and realism of training environment; and

(3) The work environment, such as transfer climate, as well as the opportunity and support from management to allow workers to apply their training.

1.5.2. Data-Mining (Classification and Visualization)

This section (1) provides an overview of data-mining and learning analytics and their educational applications, (2) describes supervised machine learning, and (3) explains the application of decision tree modeling for VE training data.

1.5.2.1. Data-Mining in Education

Conventional data-mining (DM) tools were not designed for educational purposes and as a result, DM methods are often too technical for non-experts (Romero and Ventura, 2010). The concept of applying data-mining to provide insight into learning processes stems from two domains: learning analytics and educational data-mining. Learning analytics (LA) focuses on the use of data collection and analysis to understand and optimize learning (Papamitsiou and Economides, 2014). Educational data-mining (EDM) takes this a step further and involves research and innovation in applying computerized methods to detect patterns in large collections of educational data (Romero and Ventura, 2013). Baker and Yacef (2009) summarize the scope of EDM in a taxonomy, which includes using data-mining for prediction, clustering, relationship mining, refinement of data for human judgement, and discovery with models. Both domains have grown in the last decade in response to the increased use of technology in education and the availability of big data from the online learning sector. In general, data-mining for learning-specific applications can assist with pedagogical decisions and improve the overall instructional design of training (Romero et al., 2010).

The scope of this literature review will focus on EDM methodology because it is more applicable to assessing the efficacy of virtual training. Researchers Aldowah et al.,

(2019) and Papamitsiou et al. (2014) recommended supporting virtual learning environments with EDM tools to broaden the technologies' capabilities of addressing learning-related issues. Such EDM capabilities include the ability to assess the teaching and learning effectiveness and to optimize the training through mapping the student and instructor performance (Aldowah et al., 2019). The following is a list of relevant applications selected from Romero and Ventura's (2010) review of EDM applications, which illustrates the link between EDM methods and VE training: 1) visualizing and analyzing data, 2) student modeling, 3) providing recommendations for students, 4) predicting students' performance, 5) detecting undesirable student behaviours, 6) providing feedback for instructors, and 7) constructing course material. Augmenting virtual learning environments with these EDM applications could lead to making VE training more interactive, adaptive, and personalized (Papamitsiou et al., 2014).

1.5.2.2. Supervised Learning

Within the scope of data-mining, there are two machine-learning categories applicable to the analysis of human performance data collected from VE training: 1) supervised learning, and 2) unsupervised learning (Kotsiantis, 2007). Both machine-learning methods use a repository of data to form classifications, although they employ different techniques to classify the data. Supervised learning is a classifier that uses a repository of solved problems to draw inferences. This means the data in the repository has known class labels (or attributes) that help the classifier identify to what class the data belongs (Han et al., 2012). Unsupervised learning uses a cluster analysis on repository data that is not labeled (or when the class labels are unknown) in order to partition data into similar groups, and in doing so discovers new classifications of the data (Kotsiantis, 2007; Han et al., 2012).

The focus of this research is to apply DM methods to help instructors interpret the trainee performance data from the VE training. Supervised learning is the most suitable for this application, thus the remainder of this review will focus on supervised learning methods. Supervised learning is a two-step process: (1) learning – building the model or classifier using a training data set; and (2) classification accuracy – determining the accuracy of the model by testing the decision tree's classification rules on test data (Han et al., 2012). Both steps will be discussed in detail.

Within supervised learning, there are three techniques: logic, perceptron, and statistics (Kotsiantis, 2007). Logic-based algorithms group instances by sorting them based on class labels (or attributes). Common types of logic-based supervised learning include rule-based classifiers (e.g. IF-THEN rules) and decision trees. Perceptron-based techniques iteratively run through batches of training datasets to define a prediction vector or rule (Kotsiantis, 2007). Examples of perceptron-based supervised learning include artificial neural networks (ANN) and support vector machines (SVM). Statistical learning algorithms focus on modeling the probability relationship among the attributes of the dataset. An example of statistical supervised learning is Bayesian networks (BN).

Each technique has its own benefits and drawbacks. Selecting a DM technique largely depends on the intended application and the characteristics of the dataset. For assessing VE training, the goal of using supervised learning methods is to assist in understanding what trainees learned from the VE and to determine how to modify the training curriculum based on trainee behaviours. VE technology already tracks and records in-simulation performance. Decision trees (DT) are an example of logic-based supervised learning that is particularly well suited for VE training applications. DTs are commonly used for their visual simplicity and diagnostic capabilities. In addition, DTs can be constructed quickly and do not require prior assumptions about the data, particularly when compared to other methods, such as artificial neural network or support vector machines (Liu, 2009).

1.5.2.3. Decision Trees

Decision trees (DT) are well suited to VE training applications because of the characteristics of the VE training datasets. For instance, VE training can record each user's in-simulation performance during practice exercises and store this data in a user specific database (e.g. a repository of solved problems with labelled data). DT modeling applies an algorithm to the observed performance data in order to draw inferences and develop generalized decision rules (Badino, 2004). These generalized rules can be used for many applications. For example, researchers have used DTs to inform learning analytics and to develop artificially intelligent (AI) agents. Musharraf et al. (2018) used DTs to model participants' route selection strategies from recorded evidence in VE training in the context of selecting safe egress routes in virtual offshore emergencies. The main benefits of DTs for this application were that the DTs were easy to interpret, useful in identifying patterns in participants' performance, and had diagnostic potential for determining the strengths and weaknesses of different decision-making strategies.

From a diagnostic perspective, the DT model of participants' decision strategies can determine whether participants have achieved competence. DTs provide the transparency and high interpretability (Romero and Ventura, 2010) needed for instructional designers to understand the VE data and make informed decisions about the design of the training curriculum. DTs can model behavioural patterns, which is a capability that goes beyond conventional methods of tracking trainee progress and performance outcomes, and offers another lens to assess training efficacy. DT modeling allows instructional designers to observe patterns in peoples' data to identify their learning heuristics. These learning heuristics are otherwise not easy to identify using conventional performance assessment methods. In Romero and Ventura's (2010) review of EDM methods, DTs were commonly used for the following applications: visualizing data, modeling student behaviours, detecting undesirable behaviours, and predicting performance. These applications are especially important for helping instructional designers to assess the efficacy of different training methods.

The following sections describe the two-step process used to develop the decision trees: the DT algorithm and the classification accuracy.

1.5.2.1. Step 1: Modeling Decision Trees

Figure 1.5 shows the supervised learning process for developing decision trees. This process starts with a database, known as the training dataset, that consists of a list of labeled attributes and actions. This database is created following the rule-based methodology (Cacciabue et al., 1992) using the human performance data from the VE training. Each row represents a record (e.g. an individual's characteristics or performance in specific situations). In the case of VE training data, information from each participant's performance in VE scenarios is used to populate a database consisting of scenarios (S_1-S_n), attributes (A_1-A_n), values (V_{11}-V_{nn}), and actions (E_1-E_n). The scenarios and attributes are labelled inputs to the database and the participants' corresponding actions in the scenarios are known as classes.

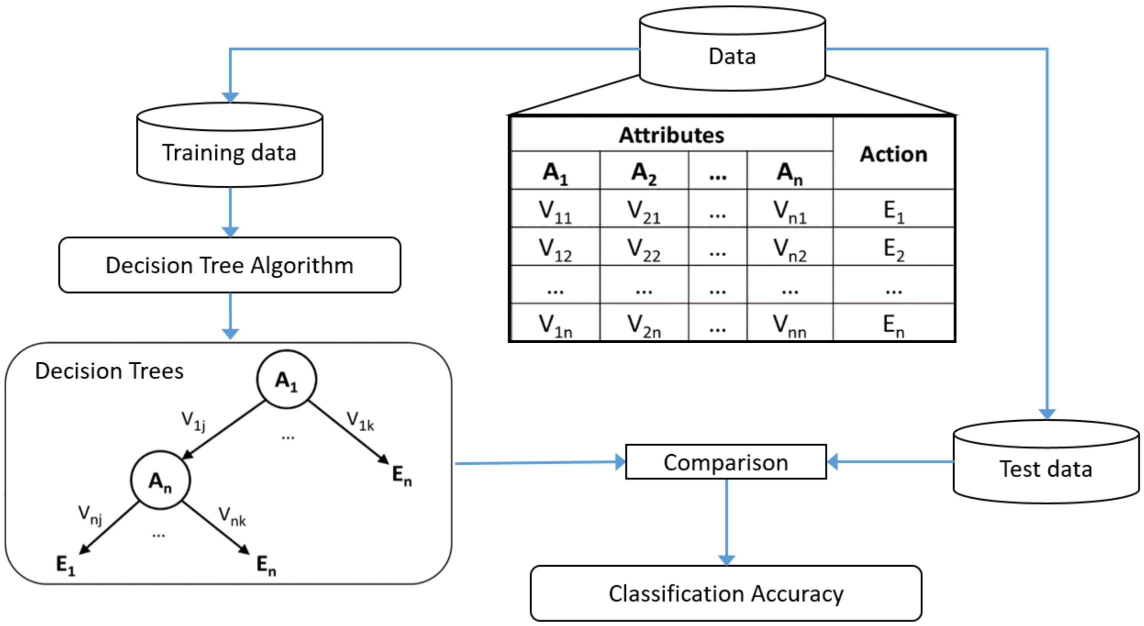


Figure 1.5: Supervised learning process for decision trees (after Han et al., 2012).

The DT algorithm uses induction to create generalized decision rules by classifying the information in the database into groups such that the dataset in each group belongs to the same class. The output is a DT that visually describes the individual's decision rules based on the content in the database. This DT can be used to predict their future decisions based on the value of the attributes in a given scenario.

Decision Tree Algorithm:

The DT algorithm is applied to observed performance data in order to develop generalized decision rules (Badino, 2004). This decision rule classification is based on the attribute selection method. Han et al. (2012) describes three DT algorithms and each uses a different attribute selection measure: *ID3* (Iterative Dichotomiser uses *information gain*), *C4.5 (*a revised version of ID3 uses a *gain ratio*), and *CART* (Classification and Regression Trees uses G*ini index*). All three use a non-backtracking method to construct the decision trees (Han et al., 2012). According to Kotsiantis' (2007) review of classification techniques, the C4.5 algorithm appears to be the most commonly used DT algorithm.

Musharraf et al. (2018) used the *ID3* algorithm (Quinlan, 1986) to develop AI agents for VE training and to investigate how changing attributes of virtual emergency scenarios influenced participants' decisions on egress routes. This demonstrated the diagnostic utility of DTs in determining the strengths and weaknesses of participants' performance during emergency scenarios (Musharraf et al., 2018). Musharraf et al. (2018) recommended that the diagnostic and predictive capabilities of DTs can be used to evaluate the effectiveness of a training program and to develop adaptive training tools.

The *ID3* decision tree algorithm takes two basic inputs: the performance database from the VE scenarios, and the list of attributes that were varied in each scenario. During the DT induction, data are iteratively classified using the attribute that has the highest information gain. The *ID3* algorithm calculates the highest information gain using three main calculations: 1) the entropy of the dataset, 2) the average information entropy of attributes, and 3) the information gain for each attribute.

First, the entropy of the entire dataset is calculated as a measure of the uncertainty of the data (Liu, 2009). This is achieved by defining the training set as S, where S contains m class labels and S_i is a subset of scenarios within the training set S. Then the entropy of S is calculated using Eq. 1.

$$Entropy(S) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{1}$$

Second, the training set S is partitioned using attribute A, where A has k distinct outcomes. This partition will result in subset S_j with j to k values. The average information entropy for all attributes (A_1-A_n) in S_j is calculated using Eq.2.

$$Entropy(A) = \sum_{j=1}^{k} \frac{|S_j|}{|S|} Entropy\ (S_j) \tag{2}$$

Finally, the *information gain*, which is the difference in entropy before and after splitting the dataset on the attribute A, is calculated for each attribute in the database using Eq. 3.

$$Gain\ (A) = Entropy\ (S) - Entropy\ (A) \tag{3}$$

The attribute with the highest *information gain* is selected as the root node, which begins the partition of the dataset. The root node represents the attribute that minimizes the information needed and reduces the randomness of the partitions (Han et al. 2012). This process repeatedly splits the data subsets at each internal node until no attributes are left for classification, or the data set is empty, or data in each group belong to the same class and no further classification is needed (Musharraf et al., 2018). A complete tree has branches to leaf nodes (that represent the class label or final action of the participant). Algorithm 1 describes the iterative steps used to develop a decision tree.

Algorithm 1. General algorithm to generate a DT from the database (Han et al., 2012)

Inputs: database; attribute list; information gain attribute selection method

Output: a decision tree

Method:

Start

(1) Create a node, A_i

(2) If all scenario examples at the current node are of the same class, then label the leaf nodes with the class labels and stop (e.g. branch, V_n; leaf node, E_n).

(3) If the data subset at the current node is empty then label the node with the majority class label in its parent data set (e.g. branch, V_n; internal node, A_n).

(4) If no attributes are left for further classification, then label the leaf node with the majority class label in the current data subset and stop (e.g. branch, V_n; leaf node, E_n).

(5) For each remaining attribute A_n, compute the value of information gain $Gain(A_n)$

(6) Choose the attribute with the highest $Gain(A_n)$ to branch the current node.

(7) For each branch node, go to step 2.

End

The output of this process is a DT that visually describes the VE user's decision preferences based on their performance data. The resulting DT can be used to diagnose the person's strengths or weaknesses in performing a particular task in the VE training, as well as for predicting their future decisions based on the value of the attributes in a given scenario. Instructional designers can use the collection of DTs from a cohort of trainees' data to identify group patterns in the performance outcomes and use this information to diagnose the strengths and weaknesses of the VE training at a systemic level.

1.5.2.2. Step 2: Calculating the Classification Accuracy:

Before DTs can be applied to make decisions on new data, the fit of the classification method needs to be verified. This can be done by calculating the DT classifier accuracy (i.e. prediction accuracy). Kotsiantis (2007) describes three methods to calculate the classifier accuracy: 1) split training sets, 2) cross validation, and 3) leave-one-out validation. This research applied *split training set* approach to calculate the classification or prediction accuracy. The *split training set* approach divides the dataset and uses 2/3 of the dataset for training and the other 1/3 of the dataset to test the classifier's performance. The classification accuracy of the DTs is the percentage of test sets that are correctly classified using the DT. The accuracy is calculated by comparing the DT prediction to the test data set as depicted in Figure 1.5. If the accuracy is considered acceptable, then the DT can be used to make decisions on new data (Han et al., 2012).

1.6. Novelty and Contribution

There are three main contributions of this research:

1. The longitudinal pedagogical study provides empirical evidence on learning, retention, and transfer to support a shift in the offshore safety domain, specifically in: (a) how VE training is designed and delivered; (b) how VE training interventions are assessed using data modeling tools; (c) how VE training is evaluated from a retention and maintenance perspective; (d) how VE training is improved to prepare all personnel to respond to a wide variety of emergencies.

2. Without a supporting pedagogical framework, simulation-based training is not being used to its full potential. Combining VE technology with a well-designed training approach, like SBML, can support the development of trainee competence

in emergencies. The SBML framework can also provide standardization and accountability to offshore safety training. Although this work demonstrated the applicability of the SBML pedagogical framework in the context of offshore oil and gas safety training, these methods are generalizable to other tasks and domains.

3. VE and simulation-based training should move beyond traditional forms of performance assessment and employ data driven diagnostics and learning analytics. This work aims to demonstrate that decision trees are useful for evaluating training efficacy. Data classification methods will be the foundation of future adaptive training tools for simulation-based training. The diagnostic capabilities of decision trees can be used to find systemic gaps in future automated VE training. Similarly, the predictive capabilities of decision trees can be used to predict how people will perform in unforeseen situations, offering further potential to optimize how recurrency training is delivered and monitored.

1.7. Organization of the Thesis

The PhD thesis is written in manuscript format and includes the following four journal papers as chapters:

Ch 2. Smith, J., and Veitch, B. (2018). A better way to train personnel to be safe in emergencies. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B*, 5(1), 011003.

Ch 3. Smith, J., Musharraf, M., Veitch, B., and Khan, F. Diagnosing the efficacy of virtual offshore egress training using decision trees. (Unpublished Manuscript) Submitted to *IEEE Transactions on Learning Technologies*.

Ch 4. Smith, J., Doody, K., and Veitch, B. (2019). Being prepared for emergencies: A virtual environment experiment on the retention and maintenance of egress skills. *WMU Journal of Maritime Affairs*, 18(3), 425-449.

Ch 5. Smith, J., Musharraf, M., Blundon, A., and Veitch, B. Preparing for skill transfer: a decision tree tool for curriculum design and assessment of virtual offshore emergency egress training. (Unpublished Manuscript) Submitted to the *International Journal of Training Research*.

Figure 1.6 shows the organization of research. Table 1.1 connects the papers, specific research objectives, and task descriptions for each paper related to this research.

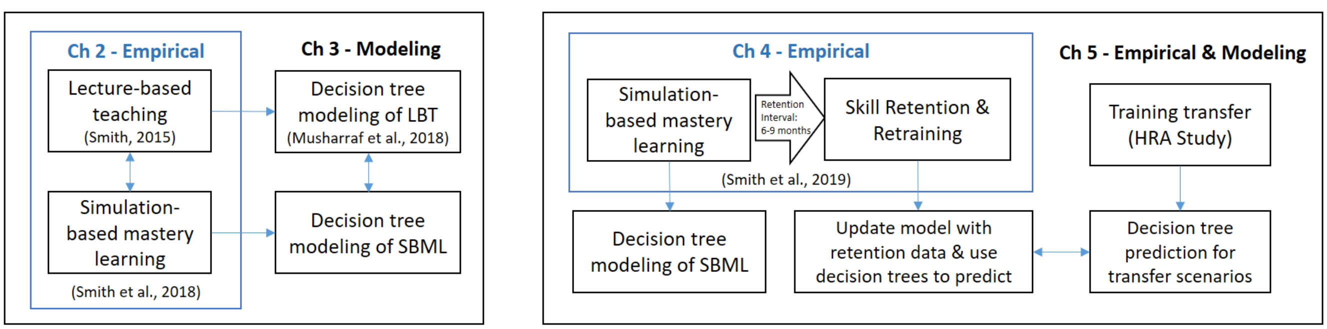


Figure 1.6: Organization of Research into Four Manuscripts

Table 1.1: Organization of manuscript thesis

Papers as Chapters	Research Objectives	Task Descriptions
Chapter 2: A better way to train personnel to be safe in emergencies	• Determine suitability of Simulation-Based Mastery Learning (SBML) pedagogical approach for delivering virtual egress training from the perspective of skill acquisition. • Compare the performance outcomes of the SBML approach with the outcomes from the early Lecture-Based Teaching (LBT) study.	• Consult with subject matter experts (SMEs) on training scope, learning objectives, and development of credible test scenarios in the virtual environment (VE). • Implement the SBML framework by creating training scenarios with built in guidance and feedback in the VE. • Verify test scenarios match earlier experiment to allow for comparison between the two experiments. • Conduct SBML experiment with human subjects. • Collect data on participants' performance in the scenarios. • Discuss the efficacy of the SBML approach.
Chapter 3: Diagnosing the efficacy of virtual offshore egress training using decisions trees	• Understand learning behavior during VE-based training. • Determine utility of decision tree (DT) modeling for diagnosing the strengths and weaknesses of the SBML training program.	• Select scenario attributes and populate the SBML knowledge base with data from the acquisition phase. • Apply DT algorithm to the SBML dataset. • Use DTs to identify SBML group's egress strategies • Compare DTs from the SBML and LBT datasets. • Use DTs to diagnose the strengths and weaknesses of the SBML training. • Discuss the value of DTs at diagnosing training efficacy.

Chapter 4: Being prepared for emergencies: a virtual environment experiment on the retention and maintenance of egress skills	• Measure the retention of egress skills after a period of 6 to 9-months (without any training interventions). • Assess the utility of the adaptive retraining matrix at bringing participants back to competence in all learning objectives.	• Verify test scenarios match earlier experiment phase to allow for comparison between phases. • Collect data on participants' performance in the scenarios • Conduct retention and retraining phase with same human subjects and collect data on participants' performance in the scenarios. • Discuss the results of the retention assessment and the efficacy of the retraining.
Chapter 5: Preparing for skill transfer: a decision tree tool for curriculum design and assessment of virtual offshore emergency egress training	• Empirically measure the transfer of egress skills to novel emergency scenarios. • Model participants DTs to determine capabilities to predict training transfer. • Assess the change in participants DTs as they transition from the skill acquisition, retention and transfer phases of the experiment.	• Populate the knowledge base with SBML data from the retention and transfer phases. • Apply DT algorithm to new dataset and use output DTs to observe changes in the SBML group's egress strategies. • Compare DTs from the SBML dataset at the skill acquisition, retention, and transfer phases of the experiment. • Use DTs to predict participants' performance in novel emergency scenarios. • Discuss this methods efficacy at predicting training transfer.

1.7.1. Chapter Descriptions

A statement of co-authorship is provided at the beginning of each chapter to describe the contribution of the authors throughout the stages of the research.

Chapter 2 presents the empirical results of two experiments that investigated the delivery methods of VE training for offshore emergency egress. The first experiment (Smith, 2015) used the LBT approach and the second experiment (Smith & Veitch, 2019) investigated the utility the SBML pedagogical framework. Efficacy of both training methods was measured by comparing the time spent training and the performance achieved by each training group (SBML & LBT). The results from this comparison corroborate the findings in the literature, that SBML training can address individual variability in competence. VE training using the SBML approach reinforced the learning objectives through guided practice and feedback. The SBML framework also provided the tools to standardize competence assessment and ensure that all participants of the VE training program reached the intended demonstrable competence.

Chapter 3 examines the efficacy of the SBML approach to VE training by using decision tree modeling. The SBML training was evaluated in two ways: 1) by comparing the decision tree depiction of the participants' understanding of emergency egress to the intended learning objectives, and 2) by comparing the SBML decision strategies with those developed under lecture-based teaching (LBT). The decision tree analysis identified deficiencies in the VE training. The comparison of the decision trees generated from SBML and LBT performance data showed that different training methods influenced the participants' egress strategies. The SBML approach resulted in decision trees with better route selection strategies compared to the LBT approach. This work demonstrated the diagnostic capabilities of decision trees and highlighted the value of integrating decision trees into the VE training as built-in tools to support adaptive training programs that could better meet the training needs of individuals.

Chapter 4 investigates the long-term retention and maintenance of emergency egress competence obtained through SBML training (Smith, Doody, & Veitch 2019). In particular, this chapter focuses on answering two questions: 1) what egress skills can be remembered after a period of 6-months? and 2) how effective is a VE-based retraining matrix at maintaining egress skills? Two main performance metrics were used to investigate retention and impact of retraining: 1) the overall competence (performance scores) demonstrated after the retention period, and 2) the performance by each learning objective after the retention period. The results of the experiment indicated that emergency egress skills (both spatial and procedural knowledge) are susceptible to skill decay over a period of 6 to 9-months. The overall performance scores in the participants' first test attempt showed an initial skill fade in the first two test scenarios and provided less evidence of skill fade in the latter two test scenarios. Participants' performance in terms of learning objective showed that most of the participants (89%) did not retain the full requisite skill set over the retention interval. Although skill decay occurred, the adaptive retraining matrix employed in the study was successful in bringing all participants back to demonstrable competence at the end of the experiment.

Chapter 5 describes the process of using decision trees as a data-informed curriculum design and assessment tool to evaluate the transferability of virtual offshore training. Visualizing participants' decision strategies can help instructional designers determine if participants are adequately prepared for new training transfer situations. In particular, this chapter extends the capabilities of DTs beyond individual performance assessment and demonstrates the use of DTs as an indicator of curriculum suitability. At a systemic level, decision trees can identify emerging group patterns in performance that help evaluate the efficacy of the training curriculum (e.g. diagnose the strengths and weaknesses of the VE training). The results of this work showed that DTs can model participants' decision-making strategies throughout the acquisition, retention and retraining, and transfer phases of the experiment. Modeling participants' DTs throughout the VE training helped to determine when the right amount of training had been achieved for each participant, or if further training was required. The results showed the potential of DTs to predict when a person is capable of transferring their skills to a variety of emergencies using inference. Overall, this work demonstrated that these DT tools play an important role in learning analytics but they have some limitations (e.g. DTs can only infer and not project the transfer of skills beyond the context of the training). This work further presents the DTs' potential as a tool for future learning applications, specifically to support intelligent or adaptive training programs.

Chapter 6 summarizes and concludes the thesis. It emphasizes the significance of the work, discusses the technical challenges and limitations, and offers some lines of inquiry for future research.

1.8. References

Aldowah, H., Al-Samarraie, H., and Fauzy, W. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*. 37, 13-49.

Anderson, J. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
Arthur, W., Bennett, W., Stanush, P., & McNelly, T. (1998). Factors that Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance,* 11(1), 57-101.

Atesok, K., Satava, R., Heest, A., Hogan, M., Pedowitz, R., Fu, F., Sitnikov, I., Marsh, J., & Hurwitz, S. (2016). Retention of Skills After Simulation-based Training in Orthopaedic Surgery. *Journal of American Academy of Orthopaedic Surgeons*, 2016(24), 505-514.

Badino, M., 2004. An application of information theory to the problem of the scientific experiment. *Synthese*, 140(3), 355-389.

Baker, R., and Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining*. 1(1), 3-17.

Barnett, S. & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.

Barsuk, J., Cohen, E., McGaghie, W., and Wayne, D. (2010). Long-Term Retention of Central Venous Catheter Insertion Skills After Simulation-Based Mastery Learning, *Academic Medicine*, 85(10), pp. 9–12.

Barsuk, J., Cohen, E., Wayne, D., Siddall, V., & McGaghie, W. (2016). Developing a Simulation-Based Mastery Learning Curriculum – Learning From 11 Years of Advanced Cardiac Life Support, *Simulation in Healthcare*, 11(1), pp. 52–59.

Benner P. (1982). From novice to expert, American Journal of Nursing. 82, pp. 402–407.

Bloom, B., S (1971). Mastery Learning. In J. H. Block (Ed.), Mastery Learning: Theory and practice (pp.47-63). New York: Holt, Rinehart & Winston.

Blume, B., Ford, J., Baldwin, T., & Huang, J. (2010). Transfer of training: a meta-analytic review. *Journal of Management*, 36(4), 1065–1105.

Canadian Association of Petroleum Producers (2015) Atlantic Canada offshore petroleum industry standard practice for the training and qualification of offshore personnel. CAPP. https://www.capp.ca/~/media/capp/customer-portal/documents/272259.pdf Accessed 16 March 2015

Cohen, E., Barsuk, J., Moazed, F., Caprio, T., Didwania, A., McGaghie, W., and Wayne, D., 2013. Making July safer: Simulation-based mastery learning during intern boot camp. *Academic Medicine*, 88(2), 233–239.

Cook, D., Brydges, R., Zendejas, B., Hamstra, S., Hatala, R. (2013). Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. *Academic Medicine*, 88:1178–1186.

Dreyfus, S., & Dreyfus, H. (1980). A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Fort Belvoir, Va: Defense Technical Information Center; 1980.

Liu, Y. (2009). Data mining and its applications in digital human modeling. In V. Duffy (Eds), *Handbook of digital human modeling: research for applied ergonomics and human factors engineering* (pp 37-1 – 37-44). Boca Raton, FL: CRC Press Taylor & Francis Group.

Dunlosky, J., Rawson, K., Marsh, E., Nathan, M., & Willingham, D. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology, *Psychological Science*, 14(1), 4-58.

Fitts, P. (1964). Perceptual-motor skill learning. *Categories of human learning*. A.W. Melton (Ed.), New York: Academic Press, 243-285.

Ford, J., Baldwin, T., & Prasad, J. (2018) Transfer of training: The known and the unknown. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 201–225.

Griswold-Theodorson, S., Ponnuru, S., Dong, C., Szyld, D., Reed, T., and McGaghie, W. (2015). Beyond the Simulation Laboratory: A Realist Synthesis Review of Clinical Outcomes of Simulation-Based Mastery Learning, Academic Medicine, 90(11), pp. 1553-1560.

Grossman, R., & Salas, E. (2011). The transfer of training: what really matters. *International Journal of Training and Development*, 15(2), 103–120.

Gusky, T., 2007. Closing achievement gaps: revisiting Benjamin S. Bloom's "Learning for Mastery". *Journal of Advanced Academics*, 19, 8-31.

Han, J., Kamber, M., & Pei, J., 2011. Data mining: concepts and techniques. Elsevier.

House, A., Smith, J., MacKinnon, S., and Veitch, B., 2014. Interactive simulation for training offshore workers. Proceedings, Oceans, St. John's.

International Association of Oil and Gas Procedures (2010) Evacuation, escape and rescue OGP risk assessment data directory. http://www.ogp.org.uk/pubs/434-19.pdf Accessed 16 March 2015

International Association of Drilling Contractors (2009) Health, safety and environmental case guidelines for mobile offshore drilling units. http://www.iadc.org/iadc-hse-case-guidelines/ Accessed 16 March 2015

International Maritime Organization (2001) The standards of training, certification and watchkeeping code (STCW code) convention for seafarers. International Maritime Organization, London.

Kim, J., Ritter, F., & Koubek, R. (2013). An integrated theory for improved skill acquisition and retention in the three stages of learning. *Theoretical Issues in Ergonomics Science*. 14(1), p22-37.

Kotsiantis, S. (2007). Supervised machine learning: a review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Edited by I. Maglogiannis, et al., IOS Press, 2007.

Magee, L., Thompson, A., Cain, B., & Kwan, C. (2012). From a submarine to a virtual environment and vice versa. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, 1-8.

McGaghie, W., Issenberg, S., Petrusa, E., and Scalese, R., (2006). Effect of Practice on Standardised Learning Outcomes in Simulation-Based Medical Education, *Medical Education*, 40(8), pp. 792-797.

McGaghie, W., Issenberg, S., Barsuk, J., and Wayne, D. (2014). A Critical Review of Simulation-Based Mastery Learning with Translational Outcomes, *Medical Education*, 48(4), pp. 375-385.

Moazed, F., Cohen, E., Furiasse, N., Singer, B., Corbridge, T., McGaghie, W., and Wayne, D. (2013). Retention of Critical Care Skills After Simulation-Based Mastery Learning, *Journal of Graduate Medical Education*, 5(3), pp. 458–463.

Musharraf, M., Smith, J., Khan, F., & Veitch, B., 2018. Identifying route selection strategies in offshore emergency situations using decision trees. Journal of Reliability Engineering & System Safety, In Press.

Papamitsiou, Z., and Economides, A. (2014). Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. *Educational Technology and Society*, 17(4), 49-64.

Quinlan, J. (1986). Induction to decision trees. *Machine Learning*, 1(1), 81-106.

Rasmussen, J. (1986). *Information processing and human-machine interaction: an approach to cognitive engineering*. New York: Elsevier.

Romero, C., and Ventura, S. (2010). Educational data mining: a review of the state-of-the-art. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews.* 40(6), 601-618.

Romero, C., and Ventura, S. (2013). Data mining in education. *WIREs Data Mining Knowledge Discovery*, 3: 12-27.

Sanli, E., & Carnahan, H. (2017). Long-term retention of skills in multi-day training contexts: A review of the literature. *International Journal of Industrial Ergonomics*, 66, 10-17.

Schmidt, R., & Lee, T. (2005) Motor control and learning: A behavioral emphasis (4th ed.). Champaign, IL, US: Human Kinetics.

Sellberg, C., Lindmark, O., & Rystedt, H. (2017). Learning to navigate: the centrality of instructions and assessments for developing students' professional competencies in simulation-based training. *WMU Journal of Maritime Affairs*. 17, pp. 249-265.

Siu, K., Best, B., Kim, J., Oleynikov, D., & Ritter, F. (2016). Adaptive virtual reality training to optimize military medical skills acquisition and retention. *Military Medicine*, 181, 214- 220.

Transport Canada (2007) Practical skills training for marine engineers (TP 13720E). Transport Canada. http://www.tc.gc.ca/eng/marinesafety/mpsp-training-examination-certification-publications-125.htm Accessed 26 June 2019

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology*, 47, 513-539.

Wickens, C., Holland, J., Banbury, S., & Parasuraman, R. (2013). Chapter 7: Memory and Training, *Engineering Psychology and Human Performance*, Fourth Edition, Pearson Education Inc. p234.

2. A BETTER WAY TO TRAIN PERSONNEL TO BE SAFE IN EMERGENCIES

Jennifer Smith* and Brian Veitch

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

St. John's, NL, Canada, A1B 3X5

*Correspondence author: jennifersmith@mun.ca

2.1. Co-authorship Statement

This manuscript was published in the ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering in July 2018. Author Jennifer Smith led the writing of this manuscript and co-author, Dr. Brian Veitch, provided guidance, editorial changes, and recommendations for improvements throughout the paper drafting.

2.2. Abstract

Offshore petroleum platforms present complex, time-sensitive situations that can make emergency evacuations difficult to manage. Virtual environments (VE) can train safety critical tasks and help prepare personnel to respond to real-world offshore emergencies. Before industries can adopt VE training, its utility must be established to ensure the technology provides effective training. This paper presents the results of two experiments that investigated the training utility of VE training. The experiments focused particularly on determining the most appropriate method to deliver offshore emergency egress training using a virtual environment. The first experiment used lecture-based teaching (LBT). The second experiment investigated the utility of a simulation-based mastery learning (SBML) pedagogical method from the medical field to address offshore emergency egress training. Both training programs (LBT and SBML) were used to train naive participants in basic onboard familiarization and emergency evacuation procedures. This paper discusses the training efficacy of the SBML method in this context and compares the results of the SBML experimental study to the results of the LBT training experiment. Efficacy of the training methods is measured by a combination of time spent training and performance achieved by each of the training groups. Results show that the SBML approach to VE training was more time effective and produced better performance in the emergency scenarios. SBML training can help address individual variability in competence. Limitations to the SBML training are discussed and recommendations to improve the delivery of SBML training are presented. Overall, the results indicate that employing SBML training in industry can improve human reliability during emergencies through increased competence and compliance.

2.3. Introduction

Offshore petroleum platforms are complex, safety-critical working environments. These platforms are characterized by their isolation and harsh marine weather. Emergencies on offshore petroleum platforms are time-sensitive situations and the safe evacuation of the platform can be challenging to manage. Offshore management teams rely on personnel to follow emergency protocols to help account for everyone in the emergency. During an emergency, all personnel onboard are required to muster at a temporary safe refuge (TSR) area to be accounted for. Assuring the compliance and competence of general personnel in responding to emergencies can help improve the overall emergency preparedness of the platform. There are two problems with conventional training methods: (1) they do not address individual variability and (2) they are not representative of the conditions in real emergencies. New personnel or short-term contractors who arrive on an offshore platform are provided with conventional safety training. Operational regulations typically require that these individuals complete an orientation period where they are supervised or accompanied by a full-time crew member for their own safety. This orientation period depends on the jurisdiction. For example, the orientation period can extend over two shift-turn overs, or six weeks, before the personnel can be unaccompanied on the platform. The amount of training time is usually fixed according to regulatory, logistical, and cost constraints. As a result, the competence of each individual participating in the same training program can be very different. This highlights a key problem: conventional training tends to fix the instructional time and as a result learning outcomes vary because individual learning needs are not adequately addressed (Cook et al., 2013).

Routine muster and evacuation drills offshore are required by regulations. The drills typically occur at a regularly scheduled time each week and require all personnel to muster at their designated muster stations in response to emergency alarms. The drills are usually performed in calm, nonrisky situations, and do not mimic real emergency situations. Practicing emergency evacuations in conditions that replicate real emergencies (high risk, stress) is not practicable due to the ethical, logical, and financial constraints. As a result, there is little variation in the muster drills and they do not represent the conditions that exist in real emergencies. Over time, this form of routine drill can cause complacency as personnel practice in benign conditions, rather than in the emergency conditions that the drills are intended to practice. This gap in the conditions of conventional training is concerning and highlights a second key problem: the gap between real emergency conditions and the conditions in which conventional training is performed undermines the contextual validity of conventional training. Virtual environment (VE) training has the potential of providing personnel with artificial experience in how to respond to situations in emergency conditions. VE training discussed in this paper focuses on two main areas: onboard familiarization and emergency egress.

2.3.1. Onboard Familiarization

VEs can be used to train personnel onshore before they are deployed offshore. Training with a VE can focus on "know your workplace training" and can familiarize personnel with the work environment and the safety procedures in a virtual setting before physically stepping foot offshore.

2.3.2. Emergency Egress Training

VEs are a safe means to practice emergency evacuation exercises and can help prepare personnel to respond effectively to realistic offshore emergencies. They can also be used to test new training protocols to determine if the interventions in a training program improve crew compliance and competence.

Before industries can adopt VE training, its utility must be established. The present work investigates the training utility of two different training methods: a conventional lecture based training (LBT) approach, and a simulation based mastery learning (SBML) approach. Both training methods were developed in the same VE and for the same learning objectives: basic onboard familiarization and emergency evacuation procedures.

2.4. Background

In general, conventional training tends to fix instructional time, allowing the learning outcomes to vary. Mastery learning, by design, has a fixed competence threshold for learners to achieve, and the instructional time required to achieve that threshold usually varies due to individual variability (Cook et al., 2013). Therefore, to determine the more appropriate method to deliver offshore emergency egress training using a virtual environment, two experiments were conducted. The first experiment looked at using LBT to deliver the VE training (Smith, 2015). The second experiment applied the SBML pedagogical approach to the VE training.

2.4.1. Simulation-Based Mastery Learning

Simulation-based mastery learning is a pedagogical approach developed in the medical field (McGaghie, et al., 2006; Barsuk et al., 2010; Cohen et al., 2013; Moazed et al., 2013; Barsuk et al., 2016). The SBML protocol gradually guides trainees through the learning objectives, assesses their performance, provides corrective feedback, and allows the trainees to proceed to the next level of training once they have demonstrated proficiency in the basic concepts. McGaghie et al. (2014) describe seven aspects of mastery learning as follows:

(1) Baseline or diagnostic assessment;

(2) Clear learning objectives, sequenced as units in increasing difficulty;

(3) Engagement in powerful and sustained educational activities focused on reaching the objectives (e.g., deliberate skills practice);

(4) A fixed minimum pass standard for each unit (e.g., test score, checklist percentage);

(5) Formative assessment with specific feedback to gauge unit completion at the minimum passing standard for mastery;

(6) Advancement to the next educational unit once competence is achieved at or above the mastery standard; and

(7) Continued practice or study on an educational unit until mastery standard is reached.

The SBML approach has been tested in the medical field with generally positive results (McGaghie et al., 2014; Griswold-Theodorson et al., 2015), although some evidence is mixed (Cook et al., 2013). For example, Cook et al. (2013) in their systematic review and meta-analysis concluded that mastery learning in simulation based medical education showed limited evidence of improved outcomes compared to nonmastery instruction, and that the mastery learning method takes longer than other training methods. The focus of this paper is to test the efficacy of the SBML pedagogical approach as applied to offshore safety training, particularly in the context of general personnel working offshore (i.e., individuals whose responsibility during an emergency is to muster at their designated muster stations).

2.5. Methodology

Two experiments were conducted to evaluate training methods using virtual environments to train offshore emergency egress. The first experiment evaluated LBT; the second experiment investigated SBML. This section will first explain the common elements of both experiments and then describe the methods used to deliver the training for each experiment.

2.5.1. All-Hands Virtual Emergency Response Trainer.

An emergency preparedness training simulator called all-hands virtual emergency response trainer (AVERT) was used in both experiments. AVERT is a first person perspective virtual environment that was developed to train personnel in basic offshore emergency duties within a naturalistic representation of the offshore work setting (House et al., 2014). The current configuration of AVERT is intended to train personnel in safe work practices and how to muster at their designated muster stations in the event of an emergency.

2.5.2. Task: Training Objectives and Test Scenarios

2.5.2.1. Learning Objectives.

Both experiments taught the same training content (basic offshore safety practices) using the AVERT virtual environment, but used different training delivery methods. Both training programs were designed to train and test six learning objectives in AVERT. The learning objectives were developed to address two knowledge dimensions: spatial knowledge and procedural knowledge. The spatial learning objectives included familiarity with the platform layout, and knowledge of the egress route options. The procedural learning objectives included recognizing emergency alarms, assessing the emergency situation, avoiding hazards, safety protocols, and mustering procedures.

2.5.2.2. Test Scenarios.

After completing the training, participants were tested in AVERT on their ability to respond to a variety of emergency preparedness exercises, from basic muster drills to full emergency evacuations. Table 2.1 provides a description of a selection of the test scenarios discussed in this paper. Participants were required to recognize the emergency situation based on immediate hazards, alarm type, and public address (PA) announcements (House et al., 2014). Depending on the alarm type that was sounded, personnel were required to either: (a) gather at the primary muster point, or (b) go to their secondary muster point, the lifeboat station. Following the alarm, participants were required to respond to the situation by selecting the safest route available to the TSR. The test scenarios gradually increased in difficulty based on how much training the participant had completed and the overall number of learning objectives being assessed.

Table 2.1 Scenarios used for the experiment

Test Scenario	Scenario Description
S1 (Wayfinding Drill)	Meet their supervisor at their designated lifeboat station by following their primary or secondary egress routes.
S2 (Muster Drill)	Respond to a muster drill (General Platform Alarm). During this alarm all personnel are required to muster at their primary muster station.
S3 (Emergency)	Respond to an emergency situation involving a General Platform Alarm (GPA) due to fire in the galley compromising the primary muster station with smoke. The situation escalates to a Prepare to Abandon Platform Alarm (PAPA) alarm. All personnel must head to the primary muster station but are forced to re-route to the lifeboat station as a result of the alarm change and compromised muster point.
S4 (Emergency)	Respond to an emergency situation involving a General Platform Alarm (GPA) due to an explosion and fire on the helideck. Smoke from heavy winds is blocking access to the secondary egress route. All personnel must go to the primary muster station. The situation escalates to a Prepare to Abandon Platform Alarm (PAPA) alarm, requiring all personnel to re-route and muster at their lifeboat station.

2.5.3. Experiment 1: Lecture Based Teaching (LBT) in AVERT

2.5.3.1. Participants

Forty participants were recruited for the study. Thirty-six participants completed the LBT experiment and four people withdrew from the study due to scheduling conflicts and symptoms of simulator sickness. Of the participants that completed the experiment, 27 participants were male and 9 participants were female. The participants' ages ranged from 19 to 39 years with the mean age of 26 years, (standard deviation (SD) of ± 4.4 years).

Participants were divided into two groups: LBT1 and LBT2. Due to participant withdrawals, the two training groups had an uneven number of participants. LBT1 had 17 participants and LBT2 had 19 participants.

2.5.3.2. LBT Training

The training scenarios in AVERT were designed to be similar to the orientation received by offshore personnel during their first experience onboard an offshore platform. The scope of the training included safety protocols for the offshore platform's Accommodation Block and the Engine Room (which was designated as the participant's worksite). The training and testing took place in three separate sessions: S1 – basic safety induction training, S2 – advanced alarm recognition, S3 – advanced hazard awareness. The participants were assigned to one of two groups, differentiated by the amount of exposure the participants were given to training in AVERT. One group of participants (LBT2) was given a single exposure to training. The other group (LBT1) was given multiple exposures. Both groups received initial training consisting of a training tutorial and an orientation scenario in AVERT that encouraged each participant to navigate areas of the vessel including their cabin, common rooms in the Accommodation Block, exterior vessel decks, and an assigned worksite in the Engine Room. LBT1 received repeated exposure to the training tutorials and were provided with practice scenarios in AVERT. LBT2 received one exposure to the initial training and did not receive any practice in AVERT. After the LBT1 group completed the training, participants performed a series of test scenarios to determine how well they were able to demonstrate their competence in AVERT. The LBT2 group did not receive any other form of training and returned for each session to perform the test scenarios to determine how much of the initial training they were able to retain and demonstrate in AVERT. The full experimental design and results of the LBT study are reported (Smith, 2015).

2.5.4. Experiment 2: Simulation Based Mastery Learning (SBML) in AVERT

2.5.4.1. Participants

Sixty participants were recruited in the study based on a priori power analysis to address participant attrition for a planned longitudinal study. Fifty-five people participated in the SBML experiment and five people withdrew from the study due to symptoms of simulator sickness and difficulty learning the control interface. Of the participants that completed the experiment, 42 participants were male and 13 participants were female. The participants' ages ranged from 18-54 years with a mean age of 27 years (SD of ± 7.9 years). The majority of participants for both experiments were undergraduate and graduate students. All volunteers who participated were naïve subjects with no prior offshore experience and no exposure to the AVERT simulator prior to the study.

2.5.4.2. SBML applied to AVERT

The SBML approach involved a series of training modules as depicted in Figure 2.1. All modules were completed during one session in the lab. Each module was designed to train specific learning objectives and gradually taught participants the platform layout, how to recognize alarms, what to do in the event of blocked routes, as well as how to assess the situation and avoid hazards while evacuating the platform.

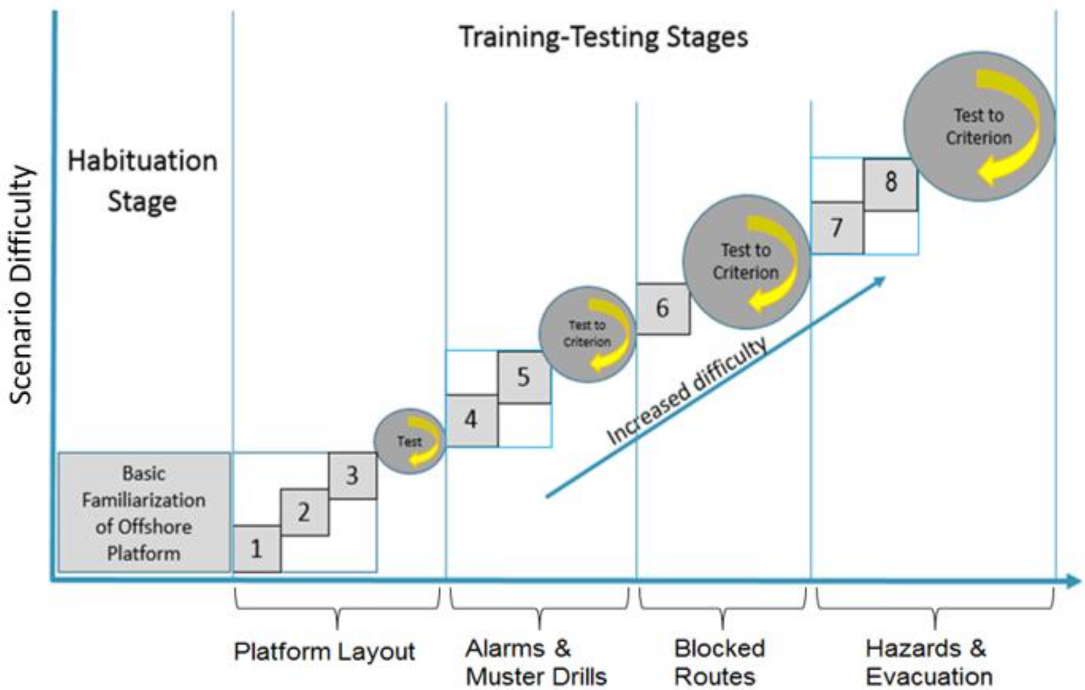


Figure 2.1: SBML training and testing stages

As depicted in Figure 2.1, the SBML experiment involved an habituation stage and four training and testing modules. The habituation stage was designed to provide participants with an initial exposure to the AVERT controls and introduced participants to the offshore platform. The SBML training and testing involved 8 training scenarios distributed across the four modules and test scenarios. Module 1 was designed to teach participants the platform layout. Module 2 provided lessons on different alarm types, muster locations, and muster procedures at the temporary safe refuge. Module 3 provided lessons on assessing the emergency situation and being prepared to re-route in the event that an egress route is blocked or muster point is compromised. Module 4 focused on situation assessment, hazard avoidance, and re-routing in the event that the primary or secondary egress route is obstructed due to poor lighting, or barriers. In this case, if the participant selected a route that was obstructed, then they were required to find an alternate route to their muster location.

As part of the SBML training, participants were required to demonstrate competence in all training and testing scenarios. After each training module in AVERT, the participant's performance was assessed using test scenarios. Participants received detailed feedback on their performance immediately after each attempt of a scenario. To achieve demonstrated competence, some participants required multiple attempts at the training scenarios.

2.6. Results & Discussion

To compare the efficacy of the training delivery methods, three measures were used: 1) the competence achieved, 2) the overall time spent training, and 3) the performance scores for each learning objective. This section presents the results of each measure.

2.6.1. Competence

The SBML training was successful in bringing all participants to competence in the targeted learning objectives. SBML participants were required to achieve a performance score of 100% in each test scenario (i.e. a passing score in all learning objectives) in order to continue with the training program. Seventy-one percent of participants successfully completed all the test scenarios on their first attempt. Some participants required multiple attempts to achieve competence. Conversely, only one participant demonstrated competence using the LBT training method. Forty-four percent of participants in the LBT training were able to achieve a score of 80% (or higher) in two or more of the test scenarios. The results of the LBT study also showed few differences in performance between the single and repeated exposure training groups. The results of the LBT study are reported in (Smith, 2015; Smith et al., 2015).

The efficacy of training was measured by comparing the level of competence achieved in terms of the participants' performance scores in the test scenarios by the LBT training and the SBML training. To compare the methods, the scoring rubric from the SBML study was applied to the LBT data. Figure 2.2 depicts the mean performance scores for the SBML, LBT1, and LBT2 groups. Participants in the LBT training were only tested once on their performance of the test scenarios. Therefore, the SBML participants' first attempt at the test scenarios is compared with the performance of the LBT participants' in Figure 2.2.

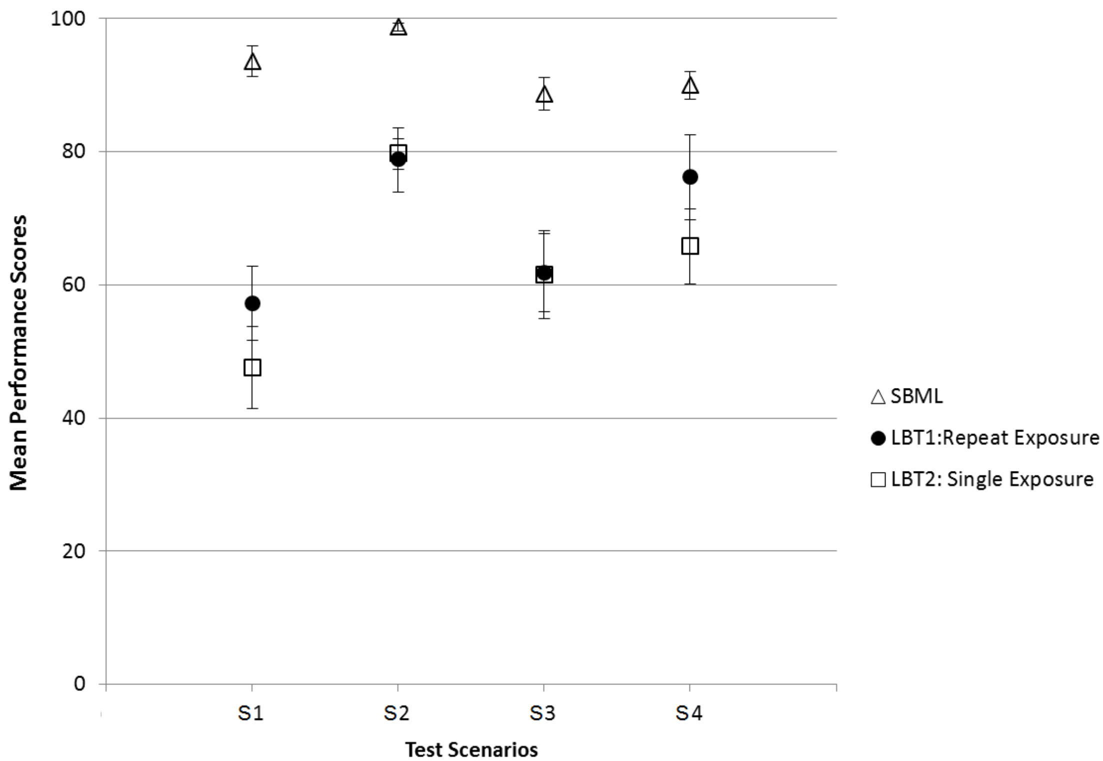


Figure 2.2: Comparison of the mean performance scores of SBML & LBT groups

The first two test scenarios focused on the basic procedures to perform muster drills. Test scenario S1 assessed participants on their knowledge of the platform layout. For S1, the average score of the SBML participants on their first attempt was 93.6% (SD = 16.6%). In comparison, the average score of the LBT groups (repeated and single exposure) were 57.2% and 47.6% (SD = 23.0% and 26.6%), respectively. The second test scenario, S2, assessed participants on their understanding of alarms and mustering at their designated muster stations. For S2, the average score of the SBML participants was 98.8% (SD = 4.4%). In comparison, the average scores of the LBT groups (repeated and single exposure) were 78.8 % and 79.7% (SD = 19.8% and 10.1%), respectively. The last two test scenarios focused on emergency situations. Participants were tested on their ability to assess the emergency situation and avoid hazards along their path. For S3, the average score of the SBML participants was 88.7% (SD = 17.8%). In comparison, the average scores of the LBT groups were 61.8% and 61.5% (SD = 24.2% and 28.7%), respectively. For S4, the average score of the SBML participants was 90.0% (SD = 15.1%). In comparison, the average scores of the LBT groups were 76.2% and 65.8% (SD = 26.3% and 24.8%), respectively.

The mean performance scores of the subgroups of LBT training were compared by (Smith, 2015). The results showed that there were no statistical differences between the two subgroups of LBT training for all four test scenarios. To compare the performance of the SBML-trained group to each of the LBT-trained subgroups, a Mann Whitney U test was used. The Mann-Whitney U test compares the median scores of two independent samples and is the non-parametric equivalent of a t-test (Corder & Foreman, 2014). A p-value of less than 0.05 was used to signify a statistical significance between the groups (i.e. the probability that the performance scores of the training groups were different). The Mann Whitney U tests showed significant statistical differences between the performance of the SBML group and the LBT subgroups for all four test scenarios ($p < 0.05$).

2.6.2. Time Spent Training

Not only did the SBML training successfully bring all participants to competence, it did so in less time. Table 2.2 summarizes the average total time spent by participants for each training method (including reviewing tutorials, performing practice scenarios, and completing test scenarios). The SBML training focused on teaching egress routes from the trainee's cabin in the accommodation block. The LBT training involved teaching egress routes from the cabin and worksite. For comparison purposes, only the time the LBT groups spent on learning the cabin egress routes are included in the calculations of the time spent training.

Table 2.2: Mean total time spent training by each group

Category	Mean Time Spent Training (minutes)		
	SBML	**LBT1**	**LBT2**
Tutorials	20.0	125.3	75.4
Practice Scenarios	63.6	42.5	30.0
Evaluation Scenarios	12.6	11.4	13.2
Total Training Time	96.3	179.2	118.6

On average, the SBML group required less time than the LBT trained groups (approximately 40% less time than LBT1 and 20% less time for LBT2). This difference in time shows that SBML training uses the available training time more effectively.

The LBT training focused on computer based training tutorials, so the majority of the training time (125 minutes for LBT1 and 75 minutes for LBT2) was spent on viewing lectures and videos. Consequently, a relatively short amount of LBT training time was spent on actual practice of egress tasks in AVERT. The LBT participants completed a 30 minute practice scenario designed to orientate them to the platform. The LBT2 group received initial tutorial training and was subsequently repeatedly tested. The LBT1 group received practice opportunities to help prepare for the test scenarios, but was not required to repeat scenarios until they reached competence. Therefore, there was very little time provided to the LBT1 group (repeated training exposure) for practice.

The SBML training was structured to provide participants with practice and feedback in AVERT. The bulk of the training time (64 minutes on average) was spent on these activities. The tutorial material was embedded into the training scenarios so more time was spent learning the task in AVERT. The practice scenarios were designed to gradually teach participants the learning objectives through in-simulation instructions and feedback. The SBML training required participants to demonstrate a minimum passing standard in the training and testing scenarios before progressing to the next training scenario or testing block. This approach resulted in participants taking multiple attempts to reach competence.

Even though the SBML participants repeated practice scenarios until competence was achieved, the overall combined training time was still shorter than LBT training time. This is due to the fact that SBML training focused less on instructing participants what to do in the event of emergency and instead offered opportunities to practice emergency egress procedures in a variety of situations. From this perspective, the SBML training was a more effective use of training time in comparison to the LBT training.

2.6.3. Comparing SBML and LBT Performance by Learning Objective

Tables 2.3 and 2.4 show the percentage of participants by training group who were successful in demonstrating the learning objectives on their first attempt at the test scenarios (drill and emergency scenarios, respectively). The learning objectives were categorized as spatial performance and procedural performance; each will be discussed separately.

2.6.3.1. Spatial Performance

Effective wayfinding in emergencies depends on an individual's spatial knowledge of the platform layout. Seigel and White (1975) describe a spatial knowledge acquisition model known as Landmark-Route-Survey. This model provides one explanation for how individuals develop a spatial understanding of an environment. Spatial knowledge usually starts with the recognition of salient landmarks, followed by connecting the landmarks with learned routes. Over time, individuals develop a map-like representation of an environment (e.g. learning how landmarks and routes are interconnected) known as survey knowledge.

The basic muster drill scenarios tested the participants' recognition of landmarks and their ability to follow designated routes. Table 2.3 shows the percentage of participants who reached competence for each learning objective in the two muster drill scenarios. A clear difference between SBML and LBT groups is seen in the first test scenario (S1) for reaching the correct location and correctly following the egress routes. This difference is less prominent in the second test scenario (S2). The LBT trained groups matched the performance of the SBML group in reaching the correct location and showed improvements in following the designated egress route during the alarm recognition muster drill scenario.

Table 2.3: Percentage of successful participants for drill scenarios

Performance Measures	**Basic Wayfinding (S1)**			**Alarm Recognition (S2)**		
	SBML	**LBT1**	**LBT2**	**SBML**	**LBT1**	**LBT2**
Spatial Performance:						
1. Reached correct location	93%	82%	63%	100%	94%	100%
2. Correctly following egress route	85%	65%	53%	98%	88%	74%
Procedural Performance:						
3. Recognized alarm & registered at TSR	n/a	n/a	n/a	98%	94%	100%
4. Avoided running	100%	12%	21%	100%	18%	16%
5. Closed all fire and watertight doors	95%	35%	21%	96%	59%	53%

n/a = not applicable. Some performance metrics are not applicable for all test scenarios.

The participants' route selection and re-routing in the emergency scenarios are good measures of how effective the training was in preparing participants for emergency situations. The emergency scenarios tested the participants' route and survey knowledge by forcing them to reroute after they encountered an obstructed path. Table 2.4 shows the percentage of participants who were successful at each learning objective in the two emergency scenarios. For these test scenarios, the route selection learning objective was divided into four subcategories, which are listed as items 2 to 5 in Table 2.4. The correct behaviours for the emergency scenarios were to select the safest route or re-route based on the PA information. Individuals who did not re-route until they encountered the hazard and those who did not re-route at all failed the test scenario. This categorization was designed to determine: 1) what information the participants were using to select their egress route, 2) the level of risk the participants were willing to take, and 3) if the participants had sufficient survey knowledge of the platform to re-route if their designated route was compromised by a hazard.

Table 2.4: Percentage of successful participants for emergency scenarios

Performance Measures	Emergency Situation (S3)			Emergency Situation (S4)		
	SBML	LBT1	LBT2	SBML	LBT1	LBT2
Spatial Performance:						
1. Reached correct location	93%	94%	89%	91%	94%	89%
2. Selected safest route available	55%	35%	47%	62%	76%	47%
3. Re-routed based on PA information	22%	0%	0%	22%	0%	0%
4. Re-routed if path blocked (encountered hazard)	16%	18%	16%	14%	12%	21%
5. Did not re-route (opened door to hazard and/or went through the hazard)	7%	47%	37%	2%	12%	32%
Procedural Performance:						
6. Recognized alarm & registered at TSR	100%	94%	89%	95%	94%	89%
7. Avoided Hazard Exposure	93%	53%	63%	98%	76%	47%
8. Avoided Running	100%	24%	5%	100%	24%	11%
9. Closed all fire and watertight doors	93%	65%	47%	100%	47%	79%

The overall percentage of participants who selected the safest route for a given scenario is one indicator of the training efficacy and how well the training helped participants develop egress strategies. As shown in Table 2.4, the SBML trained group was more successful at selecting the safest route at the onset of the emergency situation in scenario S3. The LBT1-trained group was more successful at selecting the safest route in scenario S4. In this case, the SBML trained group did not outperform the LBT1-trained group on their first attempt at the scenario in terms of reaching the correct location and selecting the safest egress route (item 1 and 2 in Table 2.4). This highlights some limitations in the SBML training for developing spatial knowledge. Two possible reasons for the performance variability are: 1) developing survey knowledge takes time and this process is subject to individual variability regardless of training method, and 2) the different training methods (SBML and LBT) resulted in different decision making strategies.

To better understand the limitations of the SBML and LBT training methods, Smith et al. (2017) investigated participants' decision making strategies in the virtual emergency scenarios using decision tree modeling. This analysis showed that the SBML-trained group tended to employ route selection strategies that involved listening to the PA announcement and taking into consideration information from the announcement. While these route strategies were more successful in general, this does highlight a limitation of the delivery of the training. The SBML training could be improved by focusing more attention to the development of survey knowledge and teaching participants on how to select routes and how to reroute due to blocked routes in the absence of PA announcements. Implementing improved performance assessment and built-in diagnostic tools into the VE, such as decision tree modeling, could help improve the implementation of the SBML training by addressing each individual's learning needs.

2.6.3.2. Procedural Performance

The biggest difference between the SBML and LBT training groups was the participants' compliance (or lack thereof) with safe practices. The SBML group was more compliant with safe practices and showed more risk adverse behaviours than the LBT groups. The majority of SBML participants were able to successfully avoid hazard exposure, close fire doors, and avoid running on the platform. These behaviours were not demonstrated by the LBT training groups.

Overall, the difference in competence and compliance between the SBML and LBT groups is attributed to the delivery of the training. There are two main reasons why the SBML participants outperformed the LBT trained participants: 1) SBML participants received an informative assessment with specific corrective feedback in the training scenarios, and 2) SBML training's fixed minimum passing standard for each scenario forced the participants to repeat scenarios until competence was demonstrated.

With regards to informative assessment, the SBML training group completed practice scenarios and had in-simulation instructions and feedback. During the SBML training scenarios, participants received immediate in-scenario feedback if they performed the safe practices incorrectly. This allowed the participants to recognize what was correct or incorrect performance. The LBT training groups both received instructions through lecture-style tutorials, followed by after scenario feedback. In the absence of corrective feedback, the LBT participants repeated the same errors throughout the training and testing scenarios.

Another aspect of SBML training that impacted performance on compliance to safety procedures was the use of a fixed minimum passing score for the training and testing scenarios. The SBML trained participants were only able to proceed to the next training scenario if they reached a minimum passing score in the learning objectives. This restriction in the training ensured all the participants complied with the safety procedures. The SBML participants had more opportunities to learn from their mistakes and adjust their strategy to respond to the emergency situations.

2.7. Conclusion

As shown by the results of these two experiments, how people respond to emergency situations can vary depending on the training they have received. Human variability in the workplace, especially during an emergency or other safety critical operation, is a safety concern. Virtual environment training using the SBML approach has been shown to address human variability and provide a means to reach demonstrable competence. Further, in the cases of LBT and SBML presented here, the SBML approach was more time effective. SBML training addresses individual differences and reinforces the learning objectives through feedback and practice. It also provides the tools to standardize competence assessment to ensure that all participants of a training program reach the intended competence. Evidence from this study indicates that SBML training is a suitable training method to provide structure, standardization, and accountability to offshore egress training.

Employing SBML training in industry has the potential to lower risks and improve the overall safety of operations.

2.8. Acknowledgments

The authors acknowledge with gratitude the support of the NSERC /Husky Energy Industrial Research Chair in Safety at Sea.

2.9. References

Barsuk, J., Cohen, E., McGaghie, W., and Wayne, D., 2010, “Long-Term Retention of Central Venous Catheter Insertion Skills After Simulation-Based Mastery Learning’” *Academic Medicine*, **85**(10), pp. 9–12. DOI: 10.1097/ACM.0b013e3181ed436c

Barsuk, J., Cohen, E., Wayne, D., Siddall, V., & McGaghie, W., 2016, “Developing a Simulation-Based Mastery Learning Curriculum – Learning From 11 Years of Advanced Cardiac Life Support,” *Simulation in Healthcare*, **11**(1), pp. 52–59. DOI: 10.1097/SIH.0000000000000120.

Cohen, E., Barsuk, J., Moazed, F., Caprio, T., Didwania, A., McGaghie, W., and Wayne, D., 2013, “Making July Safer: Simulation-Based Mastery Learning During Intern Boot Camp,” *Academic Medicine*, **88**(2), pp. 233–239. DOI: 10.1097/ACM.0b013e31827bfc0a

Cook, D., Brydges, R., Zendejas, B., Hamstra, S., and Hatala, R., 2013, “Mastery Learning for Health Professionals Using Technology-Enhanced Simulation: A Systematic Review and Meta-Analysis,” *Academic Medicine*, **88**(8), pp. 1178-86. DOI: 10.1097/ACM.0b013e31829a365d

Corder, G., and Foreman, D., 2014, *Nonparametric Statistics: A Step-by-Step Approach*, Wiley, New Jersey, Chap. 4. ISBN: 978-1-118-84031-3

Griswold-Theodorson, S., Ponnuru, S., Dong, C., Szyld, D., Reed, T., and McGaghie, W., 2015, “Beyond the Simulation Laboratory: A Realist Synthesis Review of Clinical Outcomes of Simulation-Based Mastery Learning,” *Academic Medicine*, **90**(11), pp. 1553-1560. DOI: 10.1097/ACM.0000000000000938.

House, A., Smith, J., MacKinnon, S., and Veitch, B., 2014, "Interactive Simulation for Training Offshore Workers," *Proc. Oceans'14 MTS/IEEE Conference*, St. John's, NL. DOI: 10.1109/OCEANS.2014.7003211

McGaghie, W., Issenberg, S., Barsuk, J., and Wayne, D., 2014, "A Critical Review of Simulation-Based Mastery Learning with Translational Outcomes," *Medical Education*, **48**(4), pp. 375-385. DOI: 10.1111/medu.12391.

McGaghie, W., Issenberg, S., Petrusa, E., and Scalese, R., 2006, "Effect of Practice on Standardised Learning Outcomes in Simulation-Based Medical Education," *Medical Education,* **40** (8), pp. 792-797. DOI: 10.1111/j.1365-2929.2006.02528.x

Moazed, F., Cohen, E., Furiasse, N., Singer, B., Corbridge, T., McGaghie, W., and Wayne, D., 2013, "Retention of Critical Care Skills After Simulation-Based Mastery Learning," Journal of Graduate Medical Education, **5**(3), pp. 458–463. DOI: 10.4300/JGME-D-13-00033.1

Seigel, A., and White, S., 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, pp. 9-55. DOI:10.1016/S0065-2407(08)60007-5

Smith, J., 2015, "The Effect of Virtual Environment Training on Participant Competence and Learning in Offshore Emergency Egress Scenarios," Masters thesis, Memorial University of Newfoundland, St. John's, NL. http://research.library.mun.ca/8401/1/thesis.pdf

Smith, J., Musharraf, M., and Veitch, B., 2017, "Data informed cognitive modelling of offshore emergency egress behaviour," *Proc. International Conference on Cognitive Modeling 2017* (pp. 1-6), Coventry, UK.

Smith, J., Veitch, B., & MacKinnon, S., 2015, "Achieving Competence in Offshore Emergency Egress Using Virtual Environment Training," *Proceedings of ASME 34th International Conference on Ocean, Offshore, and Arctic Engineering OMAE 2015* (pp. 1-8). St. John's, NL: ASME. DOI:10.1115/OMAE2015-41132.

3. DIAGNOSING THE EFFICACY OF VIRTUAL OFFSHORE EGRESS TRAINING USING DECISION TREES

Jennifer Smith*, Mashrura Musharraf, Brian Veitch, and Faisal Khan

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

St. John's, NL, Canada, A1B 3X5

*Correspondence author: jennifersmith@mun.ca

3.1. Co-authorship Statement

A version of this manuscript has been submitted for publication in the IEEE Transactions on Learning Technologies and is still under review. Author Jennifer Smith led the writing of this manuscript. Dr. Mashrura Musharraf verified the suitability of the methods used for this application and assisted in the analysis. Dr. Brian Veitch provided guidance and assisted in the interpretation of the results. All co-authors revised, edited, and made recommendations for improvements to drafts of this paper.

3.2. Abstract

For the offshore energy industry, virtual environment (VE) technology can enhance conventional training by teaching basic offshore safety protocols such as onboard familiarization and emergency evacuation. Combining VE technology with well-designed training can support the development of trainee competence in simulated emergencies. VEs can also act as human behavior laboratories to investigate the impact of different pedagogical approaches on competence. This paper examines the training efficacy of the simulation-based mastery learning (SBML) approach in a VE using decision tree modeling. Decision trees (DTs) are an inductive reasoning approach that can identify participants' egress strategies in offshore emergencies subsequent to training. The efficacy of the SBML training program is evaluated in three ways: 1) analyzing participants' performance scores in test scenarios, 2) comparing the DT depiction of participant's understanding of emergency egress to the intended learning objectives, and 3) comparing the SBML decision strategies with those developed under a different pedagogical approach - lecture based teaching (LBT). The results from the empirical study show that the SBML pedagogical approach was successful in bringing all participants to competence and this training resulted in concise DTs. A comparison of the resulting SBML training DTs with trees generated from a LBT approach show that the different training methods influenced the participants' egress strategies. The SBML approach resulted in better route selection strategies compared to the LBT approach. This paper demonstrates the diagnostic capabilities of decision trees as training assessment tools and recommends integrating DT tools into adaptive VE training programs to better support the training needs of individuals.

3.3. Introduction

Virtual environments (VE) can enhance conventional training for offshore energy and maritime industry personnel by providing crews with worksite familiarization, practice with safety-critical operations, and experience in responding to emergencies. To assess whether crews are adequately prepared for real-life emergencies, VE technology can track in-simulation performance metrics, provide corrective feedback, and deliver adaptive training scenarios.

Decision trees (DT) are an example of data-mining tools with behavioural pattern recognition capabilities that go beyond conventional methods of tracking trainee progress and performance outcomes, and offer another lens to assess training efficacy. DT modeling is a classification method that is particularly well suited for VE training applications because DTs employ supervised learning, which requires a repository of solved problems to draw inferences. For instance, VE training can record each user's in-simulation performance data during practice exercises and store this data in a user specific data repository. DT modeling applies an algorithm to the observed performance data (i.e. collected during VE training) in order to develop generalized decision rules (Badino, 2004). These generalized rules can be used for many applications. As an example, Musharraf et al. (2018) used DTs to identify individuals' decision-making strategies in the context of selecting safe egress routes in virtual offshore emergencies. The main benefits of DTs for the offshore emergency egress application were that the DTs were easy to interpret, useful in identifying patterns in participants' performance, and had diagnostic potential for determining the strengths and weaknesses of different decision-making strategies. This paper builds on the research from Musharraf et al. (2018) by using the diagnostic capabilities of DT modeling, as a complement to conventional performance metrics, to investigate the efficacy of pedagogical approaches applied to VE training.

3.3.1. Related Work

An earlier experiment by Smith et al. (2015) investigated the efficacy of VE training on competence using conventional lecture-based teaching (LBT). LBT is a passive learning approach that is instructor-centered and in line with traditional lecture-style instruction (Wingfield, 2005). This LBT method is a participatory form of training in which learners are exposed to the content through video instructions, demonstrations, and practice exercises (i.e. similar to conventional orientation training). Following this setup, the LBT training exposed participants to the emergency egress training content using video tutorials, platform walkthroughs, practice scenarios, and test scenarios in a first-person perspective virtual environment. This method does not have a formative assessment or a fixed minimum passing component, thereby leaving the participants with no means to assess their comprehension or gauge their progress. As a result, it was observed that the majority of participants in the LBT training failed to learn successful problem-solving strategies for emergency situations, which cast doubt on the efficacy of the LBT training approach, and its suitability to VE modes of training (Smith, 2015). Data collected during the study was used by Musharraf et al. (2018) to identify the problem-solving strategies of general personnel in emergency egress situations. The results showed that given the same training, people employed different learning strategies and developed their understanding of emergency protocols differently. In particular, decision-making in high-stress emergencies varied from person to person. These results coincide with those from (Smith, 2015), which found that the LBT training did not provide adequate assessment in the form of practice and feedback to ensure all participants gained competence.

Based on those findings, a new experimental study was conducted to assess the relative merits of another pedagogical approach: simulation-based mastery learning (SBML). The SBML framework was selected due to its reported effectiveness by the medical education field (Barsuk et al. 2010; Cohen et al. 2013; Moazed et al. 2013; McGaghie et al. 2014; Barsuk et al. 2016). SBML is a method designed to meet the needs and pace of the individual learner. The SBML approach gradually walks learners through the content and requires that they practice the exercises until they demonstrate competence. Learners are provided with formative assessments throughout the SBML training, which provides constructive feedback for them to improve or correct their performance. Once learners have demonstrated their understanding in test exercises, they are able to move on to more advanced content. Applying the SBML framework in this experiment, the SBML training repeatedly exposed participants to the emergency egress training content using platform walkthroughs, practice scenarios, and test scenarios in the same first-person perspective virtual environment. The same learning objectives, testing scenarios and performance metrics were used in both the SBML and the earlier LBT studies. The key difference between the two was the pedagogical approach, which included the delivery framework, formative assessment, and feedback method. This was intentionally controlled so the results from both experiments could be compared in terms of pedagogy.

3.3.2. Objectives

This paper uses DTs to evaluate how well the SBML training prepared participants for emergency scenarios. We apply the decision tree induction approach to the SBML dataset and compare the resulting DTs against the intended learning objectives. We compare participants' performance in test scenarios with the performance predicted by their DTs, which illustrates the utility of DTs as a predictive tool. We also compare the DTs generated from the SBML study with those from the earlier LBT study as a way of assessing the relative merits of the two pedagogical approaches in terms of improving the performance of participants. Partial results of this comparison were presented by Smith et al. (2017).

3.3.3. Organization

Section 2 presents the theoretical framework of SBML and the decision tree induction process. Section 3 explains the experimental design and the application of decision tree modeling to the SBML dataset. Sections 4 and 5 present the performance results and subsequent decision trees from the SBML and LBT training and discuss the strengths and weaknesses of the SBML training.

3.4. Theoretical Background

3.4.1. Simulation-Based Mastery Learning

SBML is a pedagogical approach developed in the medical education field (McGaghie et al. 2006; Barsuk et al. 2010; Cohen et al. 2013; Moazed et al. 2013; Barsuk et al. 2016) and is based on Bloom's competence-based theory of learning for mastery. Bloom's mastery learning is an instructional strategy that ensures all learners achieve competence by providing them with formative assessment, individualized feedback, and corrective measures (Bloom 1971; Gusky 2007). The SBML protocol builds on Bloom's framework and uses simulation to provide instruction at the learner's pace by gradually guiding them through the learning objectives, assessing their performance, providing corrective feedback, and allowing them to proceed to the next level of training once they have demonstrated proficiency in the basic concepts. McGaghie et al. (2014, p.376) describe seven aspects of simulation-based mastery learning as follows:

(1) Assess the learner’s entry level;
(2) Provide clear learning objectives that are arranged in increasing difficulty;
(3) Provide training exercises that are engaging and focused on each learning objective;
(4) Set a minimum pass standard for each unit;
(5) Provide a formative assessment of the learner’s progress with feedback so they can gauge their progress and recognize where their proficiency level is in relation to the standard;
(6) Allow learners to advance to the next unit once they have reached the mastery standard (i.e. met or exceeded the mastery competence level); and
(7) Encourage learners to continue their practice on the unit until they have achieved the mastery standard.

The advantages of SBML are illustrated in the Dreyfus/Benner five-staged skill acquisition framework. Dreyfus (1980) and Benner (1982) described the five-staged framework of training as: 1) novice, 2) advanced beginner, 3) competent, 4) proficient, and 5) expert. According to Griswold-Theodorson et al. (2015), SBML training brings learners from stage 2 to stage 3 on the Dreyfus/Benner skill acquisition model (i.e. from advanced beginner to competent). To illustrate this improvement in learning, advanced beginners tend to see their actions as a series of steps and are able to achieve some steps using their own judgement. However, advanced beginners usually require supervision for most tasks. Conversely, competent learners have established a working knowledge, they are able to use their own judgement when responding to a problem and can usually foresee the impact of their actions on long-term goals (Griswold-Theodorson et al. 2015). This is an important distinction for instructional designers aiming to establish competent workers.

Although the SBML pedagogical approach was been predominately reported from the medical education field (Cook et al. 2013; McGaghie et al. 2014; Griswold-Theodorson et al. 2015; Barsuk et al. 2016), it has the potential to be applied to any discipline to ensure learners are trained to competence. For this work, the SBML approach was applied to train participants in offshore emergency egress in a VE. Section 3 provides a detailed description of the SBML offshore egress training.

3.4.2. Decision Tree Induction

Among the different supervised machine learning techniques, this paper uses decision trees. DTs can be constructed quickly and do not require prior assumptions about the data, particularly when compared to other methods, such as artificial neural network or support vector machines (Duffy 2009). DTs were selected for their visual simplicity and diagnostic capabilities. From a diagnostic perspective, the DT model of participants' decision strategies can determine whether participants have achieved competence. This was especially important for assessing the efficacy of different training methods because the goal of this research was to provide a training diagnostic lens for instructional designers who do not have domain expertise in data-mining.

The decision tree algorithm is based on an induction process whereby generalizations are made based on observed phenomena (Badino 2004). Following the rule-based methodology (Cacciabue et al. 1992), a data matrix is first created using human performance data from simulation training. In this paper, information from each participant's performance in VE scenarios is used to populate a data matrix consisting of scenarios (*S_1-S_n*), attributes (*A_1-A_n*), values (*V_{11}-V_{nn}*), and actions (*E_1-E_n*). The scenarios and attributes are labelled inputs to the matrix and the participants' corresponding actions in the scenarios are known as classes. As depicted in Figure 3.1, the induction process creates generalized decision rules based on the content of the data matrix. The goal of the induction process is to classify the data in the matrix into groups such that the dataset in each group belongs to the same class. This paper uses the ID3 decision tree algorithm, which uses *information gain* as an attribute selection method, the means to classify the data into groups (Han et al. 2011).

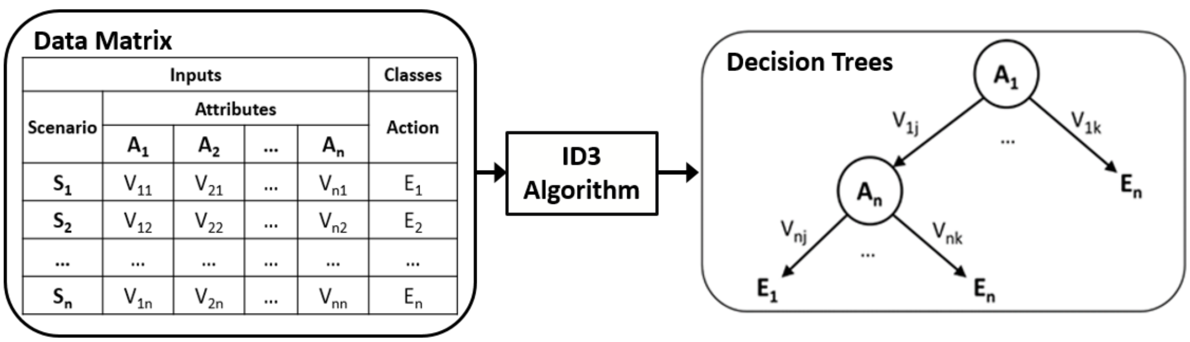


Figure 3.1: Decision tree development framework.

The ID3 decision tree algorithm takes two basic inputs: the performance data matrix from the VE scenarios, and the list of attributes that were varied in each scenario. The output is a decision tree that describes a participant's decision preferences and can also be used to predict their future decisions based on the value of the attributes in a given scenario. During the decision tree induction, data are iteratively classified using the attribute that has the highest information gain. The ID3 algorithm calculates the highest information gain using three main calculations: 1) the entropy of the dataset, 2) the average information entropy of attributes, and 3) the information gain for each attribute.

First, the entropy of the entire dataset is calculated as a measure of the uncertainty of the data (Duffy 2009). This is achieved by defining the data matrix training set as *S*, where *S* contains *m* class labels and S_i is a subset of scenarios within the training set *S*. Then the entropy of *S* is calculated using Eq. 1.

$$Entropy(S) = -\sum_{i=1}^{m} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \tag{1}$$

Second, the training set, *S* is partitioned using attribute *A*, where *A* has *k* distinct outcomes. This partition will result in subset S_j with *j* to *k* values. The average information entropy for all attributes (A_1-A_n) in S_j are calculated using Eq.2.

$$Entropy(A) = \sum_{j=1}^{k} \frac{|S_j|}{|S|} Entropy\ (S_j) \tag{2}$$

Finally, the *information gain*, which is the difference in entropy before and after splitting the dataset on the attribute *A* is calculated for each attribute in the data matrix using Eq. 3.

$$Gain\ (A) = Entropy\ (S) - Entropy\ (A) \tag{3}$$

The attribute with the highest *information gain* is selected as the root node, which begins the partition of the dataset. The root node represents the attribute that minimizes the information needed and reduces the randomness of the partitions (Han, et al. 2011). This process repeatedly splits the data subsets at each internal node until no attributes are left for classification, or the data set is empty, or data in each group belong to the same class and no further classification is needed (Musharraf et al. 2018). A complete tree has branches to leaf nodes (that represent the class label or final action of the participant). Algorithm 1 describes the iterative steps used to develop a decision tree.

Algorithm 1. General algorithm to generate DT from data matrix (Han et al. 2011)

Inputs: data matrix; attribute list; information gain attribute selection method

Output: a decision tree

Method:

Start

(8) Create a node, A_i

(9) If all scenario examples at the current node are of the same class, then label the leaf nodes with the class labels and stop (e.g. branch, V_n; leaf node, E_n).

(10) If the data subset at the current node is empty then label the node with the majority class label in its parent data set (e.g. branch, V_n; internal node, A_n).

(11) If no attributes are left for further classification, then label the leaf node with the majority class label in the current data subset and stop (e.g. branch, V_n; leaf node, E_n).

(12) For each remaining attribute A_n, compute the value of information gain $Gain(A_n)$

(13) Choose the attribute with the highest $Gain(A_n)$ to branch the current node.

(14) For each branch node, go to step 2.

End

3.5. Methodology

The decision tree development and analysis framework used in this paper is depicted in Figure 3.2. First, a pedagogical experiment was conducted in the VE with 55 participants. These participants were trained using the SBML approach. The participants' performance data was collected and divided into two datasets: a training and a testing dataset. The training data was stored in a repository in the form of a data matrix. The test scenarios were set aside to form the testing dataset. The data matrix was used to train the DT algorithm and form the decision trees, which represent participants' behavioural pattern for route selection (Musharraf et al. 2018). The testing dataset was used to calculate the prediction accuracy of the newly formed DTs. The resulting DTs were used to compare participants' understanding of the training with the intended learning objectives and to assess the efficacy of different training techniques. Section 3.1 describes the experimental design, including a description of the participants, the AVERT simulator, and how the SBML training was applied to VE. Section 3.2 describes the decision tree modeling from the SBML data, including the development of the data matrices, how scenario frames were used from dynamic scenarios, and an illustration of the DT development.

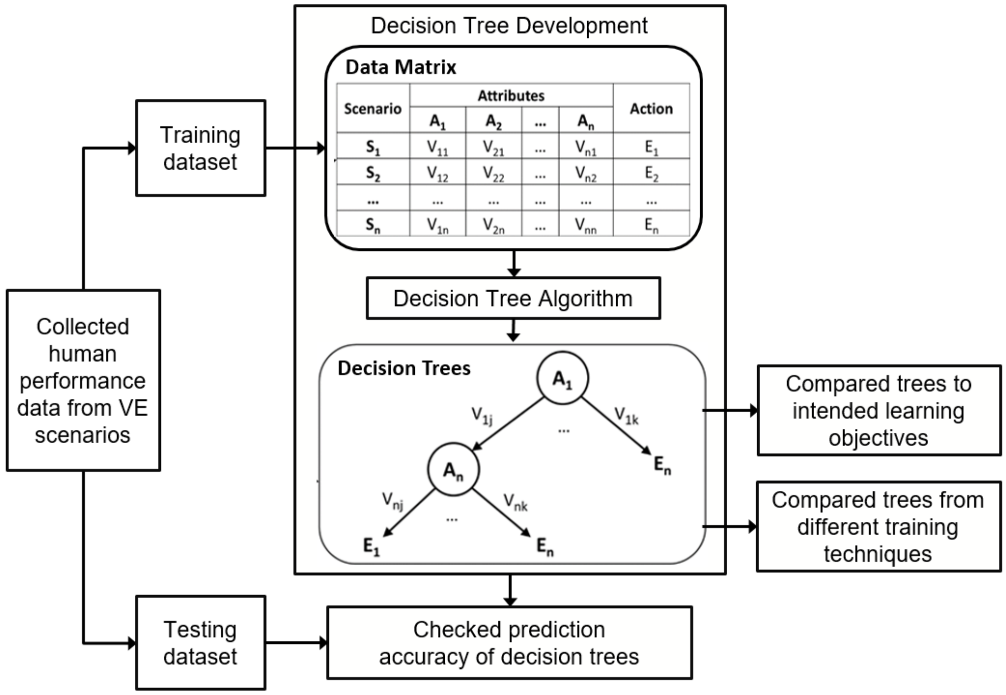


Figure 3.2: Process used to develop decision trees and assess training efficacy (after Musharraf et al. 2018).

3.5.1. Experimental Design

The DT methodology was applied to the new SBML dataset as a means to evaluate the effectiveness of this training approach. The efficacy of the SBML training program can be determined in three ways: 1) analyzing the SBML trained participants' performance scores in the test scenarios, 2) by comparing the decision tree depiction of the participant's understanding of emergency egress to the intended learning objectives, and 3) by comparing the SBML decision strategies with those identified by Musharraf et al. (2018) for the LBT experiment. The two decision tree comparisons are explored in this paper. This section describes the participants, AVERT simulator, how the SBML training was applied to AVERT, and the process used to develop the DTs.

3.5.1.1. Participants

Two separate experiments were used to test the SBML and LBT pedagogical approaches. Fifty-five naïve participants were trained using the SBML approach (42 participants were male and 13 participants were female). The SBML participants' ages ranged from 18-54 years (M = 27 years, SD = ± 7.9 years). The LBT experiment had 36 participants. These participants were divided into two treatment groups for different training exposures: LBT1, which represented multiple training exposures, and LBT2, which represented a single training exposure. This paper includes the results of 17 participants from LBT1 (13 participants were male and 4 participants were female). The LBT participants' ages ranged from 19-39 years (M = 27 years, SD = ± 5.0 years). All participants had no prior offshore experience and no exposure to the simulator prior to the study. The majority of participants for both experiments were undergraduate and graduate students.

3.5.1.2. AVERT Simulator

Both the SBML and LBT experiments trained participants in offshore emergency egress using the All-hands Virtual Emergency Response Trainer (AVERT). AVERT is a first-person perspective desktop VE that provides participants with a naturalistic representation of an offshore Floating Production Storage and Offloading (FPSO) vessel (House et al., 2014). Participants use a gamepad controller (Xbox) to control their avatar of an offshore worker and interact with the virtual FPSO platform. Participants were provided with general instructions by reading short tutorial slides before starting the in-simulation training scenarios. The current configuration of AVERT is intended to train general personnel in basic offshore emergency egress duties. General personnel are individuals whose responsibility during an emergency is to muster at their designated muster stations. The core learning objectives include familiarity with the platform layout, emergency alarms, egress routes, safety protocols, and mustering procedures.

3.5.1.3. SBML applied to AVERT

The SBML training involved an initial habituation stage and four training and testing modules as depicted in Figure 3.3. The habituation stage trained participants on how to use the AVERT controls and introduced participants to the offshore platform. Subsequent to the habituation stage, participants proceeded to the training and testing modules. Each module was designed to train specific learning objectives and gradually taught participants the platform layout, how to recognize alarms, what to do in the event of blocked routes, and how to assess the situation and avoid hazards while evacuating the platform. The learning objectives were developed with guidance from subject matter experts to address both spatial and procedural knowledge.

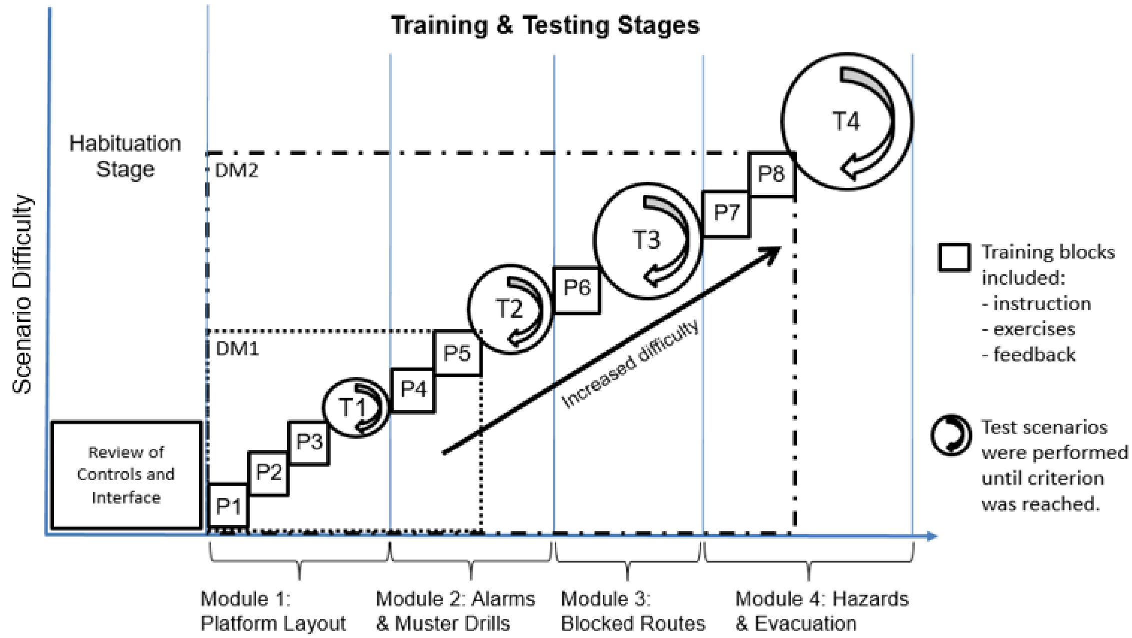


Figure 3.3: SBML training and testing stages

Each training and testing module involved 1 to 3 training scenarios and 1 test scenario. As shown in Figure 3.3, the SBML training consisted of 12 scenarios in total (8 practice and 4 testing scenarios). As part of the SBML training, participants were required to demonstrate competence in each scenario before they could advance to scenarios that were more complex. Module 1 taught participants the platform layout and all the available egress routes from the cabin. Participants were tested on their spatial knowledge in scenario T1 by asking them to meet their supervisor at their designated lifeboat station. Module 2 taught participants the different alarm types on the platform: general platform alarm (GPA), prepare to abandon platform alarm (PAPA), and mustering procedures. Participants were tested on their spatial and procedural knowledge in scenario T2 by asking them to safely respond to a muster drill. Module 3 reminded participants of the alternative routes from the cabin to ensure they knew the available route options in the event their egress route was obstructed. Module 4 taught participants the emergency protocols necessary to respond to emergency scenarios with hazards such as smoke and fire. Test scenarios T3 and T4 tested the participants' ability to respond to emergency conditions and re-route if their planned route was blocked by hazards.

After each training module in AVERT, the participants' performance was assessed using test scenarios. In the scenarios, participants were tasked with responding to muster drills or emergency alarms and selecting the safest egress route from their cabin. There were two main routes for participants to choose from: primary or secondary. Each route had multiple decision points along the path. Participants were instructed to listen to the alarm, pay attention to the public address (PA) announcements, and follow the safest route to their muster or lifeboat station. Participants were assessed on their ability to recognize the alarm, take their safety equipment, follow the safest egress route, avoid exposure to hazards, reach the correct muster location, and register at the temporary safe refuge area. Participants received corrective feedback on their performance after each scenario attempt. To achieve demonstrated competence, some participants required multiple attempts at the scenarios.

3.5.2. Decision Tree Modeling for SBML Data

The DT development and analysis, as depicted in Figure 3.2, involves six steps. First, the human performance data from the SBML virtual environment scenarios were separated into training and testing data sets. Second, the training data set was used to develop a data matrix consisting of scenarios, attributes, values, and actions. Third, the decision tree algorithm was applied to the data matrix to identify each participant's problem-solving strategies.

Once the DTs were generated, the fourth step involved using the testing data set to check the prediction accuracy of the DTs. In step five, the resulting SBML decision trees of each participant were compared to the intended learning objectives for each test scenario. Finally, in step six, the SBML decision strategies were compared with the DTs generated using data from the earlier LBT experiment (Musharraf et al., 2018).

3.5.2.1. Data Collected from the VE Scenarios

As participants completed the VE scenarios, their performance data was collected. Participants' performance during the scenarios was recorded in AVERT report files for each scenario. Observation logs were kept by the researcher to note any details that were not recorded in the automated report files. The participants' data was organized into training and testing datasets. Of the twelve scenarios, 11 were used for the decision tree development. One training scenario was an orientation scenario and was not used in the analysis. Among the remaining 11 scenarios, 9 were used to populate the data matrix that was used to train the decision tree algorithm and form the DTs. These scenarios are referred to as the training dataset. Two test scenarios were set aside to form the testing dataset. The testing dataset was used to calculate the prediction accuracy of the DTs.

3.5.2.2. Data Matrix

A two-dimensional data matrix (DM) was created using each participant's performance in the training scenarios. To populate the matrix, data was collected from the AVERT report files and from observations logged in-situ. The data matrix consisted of a combination of programmed attributes and the participants' actions. The programmed scenario attributes were varied for each scenario, such as the end location, alarm type, information presented in the PA announcements, presence of hazards, and location of obstructed routes. For each scenario, the data matrix included a record of the participants' actions, such as their route choices in the current and previous scenarios. Table 3.1 lists the attributes varied in the scenarios and their possible values.

Table 3.1: Description of scenario attributes.

Attribute	**Possible Values**
End location	Muster, Lifeboat
Alarm type	None, General Platform Alarm (GPA), Prepare to Abandon Platform Alarm (PAPA)
Hazard presence	No, Yes
Route directed by PA	None, Primary route, Secondary route
Obstructed route	None, Primary route, Secondary route
Previous route taken	N/A, Primary route, Secondary route

The data matrices were developed to correspond with the training scenarios that were completed in two training modules: module 2 (Alarm Recognition) and module 4 (Assessing the Situation). The first half of the scenarios were used to generate the data matrix for module 2 (denoted DM1). The full suite of training scenarios was used for the data matrix to represent module 4 (denoted DM2). Two scenarios, T2 and T4, were selected to test the classification accuracy of the DTs for these modules and are described in Table 3.2. Test scenario T2 occurred at the halfway mark of the SBML training. By T2, participants had familiarized themselves with the platform layout, the different alarm types, and the mustering procedures at the temporary safe refuge area. Test scenario T4 occurred at the end of the SBML training. By T4, participants were able to assess the emergency, listen to cues in the PA announcement, recognize the tenability of the egress routes, and re-route if the primary or secondary egress route was obstructed due to poor lighting or other barriers.

Table 3.2: Description of the test scenarios (After Smith & Veitch, 2018)

Test Scenario	Scenario Description
T2 Muster Drill	This scenario assessed the participants' understanding of alarms and muster procedures. Participants responded to a muster drill (General Platform Alarm). During this alarm, all personnel were required to collect their safety equipment and muster at their primary muster station.
T4 Emergency Situation	This scenario assessed the participants' ability to avoid hazards and follow the safest route to their lifeboat station. Participants responded to an emergency involving a General Platform Alarm due to fire in the galley. The fire compromised the muster station with smoke and the situation escalated to a Prepare to Abandon Platform Alarm. Initially all personnel were required to go to the muster station but were forced to re-route to the lifeboat station because of the compromised muster station.

Based on the value of the scenario attributes, the participant's goal was to select a safe egress route. Since the SBML training required participants to reattempt the scenarios until they demonstrated competence, the data matrix was updated after each attempt and only the participant's successful final attempt was stored in the data matrix. Table 3.3 shows the state of the data matrix for a sample participant after finishing all of the training modules. Each row in the matrix contains the different attribute values for the scenario and the corresponding route choice.

Table 3.3: Sample data matrix for training modules 2 (DM1) and 4 (DM1 and DM2).

Category	Scenario	Attributes						Route choice
		End Location	Alarm	Route by PA	Hazard	Blocked Route	Previous Route	
DM1	P1	Muster	None	Primary	No	None	N/A	Primary
	P3 (F1)	Lifeboat	None	Primary	No	None	Primary	Primary
	P3 (F2)	Muster	None	Secondary	No	None	Primary	Secondary
	T1	Lifeboat	None	None	No	None	Secondary	Primary
	P4	Muster	GPA	Primary	No	None	Primary	Primary
	P5	Muster	GPA	None	No	None	Primary	Primary
Test 1	T2	Muster	GPA	None	No	None	Primary	Primary
DM2	P6	Lifeboat	PAPA	Primary	No	Secondary	Primary	Primary
	T3	Muster	GPA	Secondary	No	Primary	Primary	Secondary
	P7	Lifeboat	PAPA	Secondary	Yes	Primary	Secondary	Secondary
	P8 (F1)	Muster	GPA	Primary	Yes	Secondary	Secondary	Primary
	P8 (F2)	Lifeboat	PAPA	Primary	Yes	Secondary	Secondary	Primary
Test 2	T4 (F1)	Muster	GPA	Secondary	Yes	Primary	Primary	Secondary
	T4 (F2)	Lifeboat	GPA	Secondary	Yes	Primary	Primary	Secondary
	T4 (F3)	Lifeboat	PAPA	Secondary	Yes	Primary	Primary	Secondary

* Highlighted rows represent the scenarios used to test the DTs classification accuracy.

As a basic example, scenario P4 from Table 3.3 is a muster situation in which participants practiced their egress routes and muster procedures. For a sample participant (A45), the scenario attributes were recorded as: End location = Muster; Alarm type = GPA; Route directed by PA = Primary, Hazard presence = No, Blocked route = none, and Previous route = Primary. In this case, the participant’s choice of route was the primary route.

3.5.2.3. Scenario Frames

Complex emergency scenarios were dynamic in the sense that the value of some attributes changed during the scenarios. To capture the dynamic aspect, these scenarios were divided into multiple frames so that the value of the attributes in each frame remained static (e.g. the first frame F1 depicted the initial conditions and the second frame F2 depicted the changed conditions of the scenario). Consequently, many of the emergency scenarios were multi-frames scenarios. Figure 3.4 shows an example of a two-framed training scenario (P8) and how the data matrix was updated to reflect the change in scenario attributes.

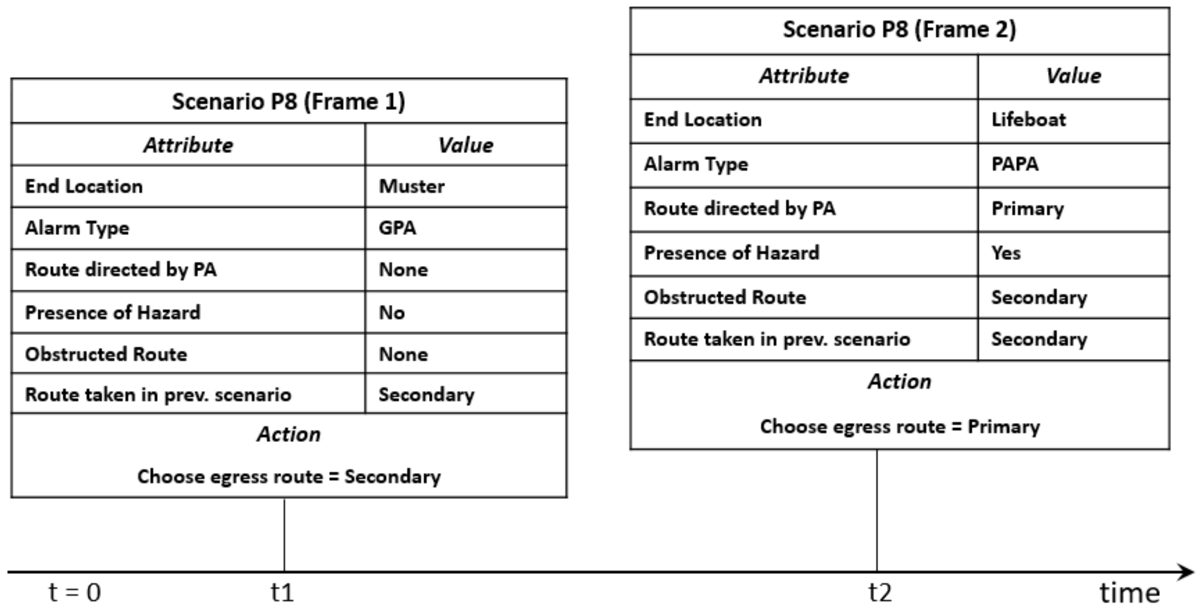


Figure 3.4: Example of scenario frames 1 and 2 for P8.

As a dynamic example, scenario P8 is an emergency in which participants responded to changing conditions. For a sample participant (A16), the scenario attributes in F1 were initially recorded as: End location = Muster; Alarm type = GPA; Route directed by PA = None, Hazard presence = No, Blocked route = None, and Previous route = Secondary. However, the severity of the situation escalated in F2 and some attributes changed: End location = Lifeboat; Alarm type = PAPA, Route directed by PA = Primary, Hazard presence = Yes, and Blocked route = Secondary. In this case, the participant's choice of route was originally the primary route, but they re-routed to the secondary route when the value of the attributes changed.

3.5.2.4. Decision Trees

The data matrix generated in the previous step was used as an input for the decision tree algorithm. The resulting DTs were used to visualize how participants formed emergency egress decision rules based on the content in the data matrix. The DTs provided insight as to which attributes had the biggest impact on participants' decision-making. Figure 3.5 shows a decision tree based on the matrix in Table 3.3, sample participant A45.

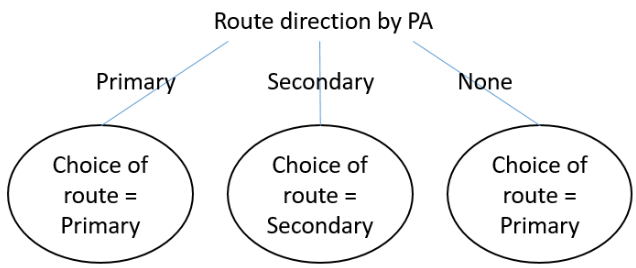


Figure 3.5: Example decision tree developed after DM1 and DM2 in Table 3.

The DT is based on evidence from the participant's performance in a series of virtual scenarios. The DT can be used to predict a participant's choice of route for a given future scenario. In this case, the participant's route selection was decided based on their understanding of the PA announcement. In future scenarios, if the PA directs them to a safest route then the participant will likely take that route. If the PA does not provide any information regarding the safest route, then the participant's choice will likely default to their primary egress route.

3.5.2.5. Testing the DTs Classification Accuracy

To evaluate the classification accuracy of the DT models, the predicted routes of the DTs for the two test scenarios (T2 and T4) were compared to the routes actually taken by the participants. The classification accuracy was calculated based on the average number of successful matches between the outcomes predicted by the decision tree and the observed outcomes. For the muster drill scenario (T2), the decision tree accurately predicted all the participants' behaviour. Errors in classification occurred for the emergency evacuation scenario (T4), where the DTs were able to predict the route selection with an accuracy of 96% on average. This classification accuracy shows the predictive potential of the DTs.

3.6. Results & Discussion

The efficacy of the SBML training was assessed using three measures: 1) analyzing the SBML trained participants' performance scores in the test scenarios, 2) comparing the participants' DTs to the intended learning objectives, and 3) comparing the SBML groups' decision strategies with those developed by the LBT group. The following subsections summarize the findings.

3.6.1. Empirical Results of SBML Training

Table 3.4 shows the percentage of participants who successfully completed each learning objective for test scenarios T2 and T4. These results compare the SBML trained participants' ability to respond to emergencies against the LBT trained participants.

As shown in Table 3.4, there were differences in the SBML participants' compliance for both the spatial and procedural learning objectives when compared to the LBT participants' performance in the same scenarios. Overall, the training that SBML participants received helped improve their spatial competence (LO2, LO3, and LO4) and their procedural safety compliance (LO8 and LO9).

Table 3.4: Percentage of successful participants by learning objective (after Smith & Veitch, 2018).

Performance Measures	T2 Muster Drill SBML n = 55	T2 Muster Drill LBT1 n = 17	T4 Emergency Situation SBML n = 55	T4 Emergency Situation LBT1 n = 17
Spatial Performance:				
LO1. Reached correct location	100%	94%	93%	94%
LO2. Correctly selected and followed safest egress route	98%	88%	55%	35%
LO3. Re-routed based on PA information	n/a	n/a	22%	0%
LO4. Re-routed if path blocked (encountered hazard)	n/a	n/a	16%	18%
LO5. Did not re-route (opened door to hazard and/or went through the hazard)	n/a	n/a	7%	47%
Procedural Performance:				
LO6. Recognized alarm & registered at TSR	98%	94%	100%	94%
LO7. Avoided Hazard Exposure	n/a	n/a	93%	53%
LO8. Avoided Running	100%	18%	100%	24%
LO9. Closed all fire and watertight doors	96%	59%	93%	65%

n/a = not applicable. Some performance metrics are not applicable for all test scenarios.

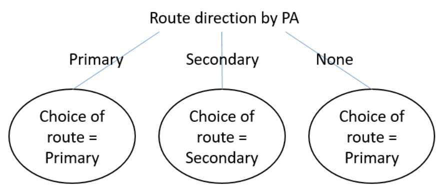
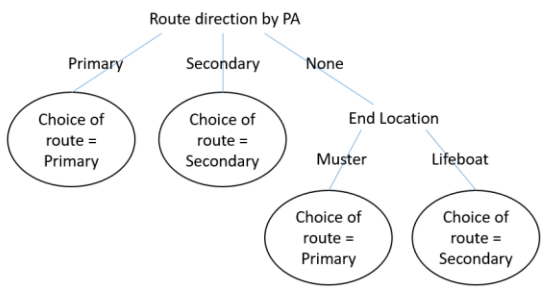
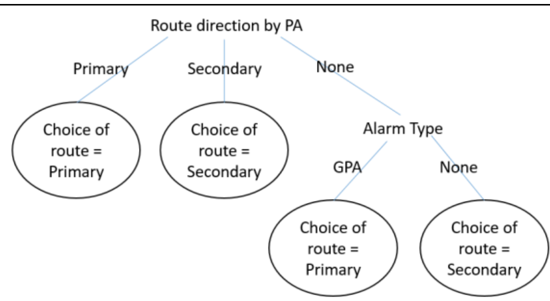
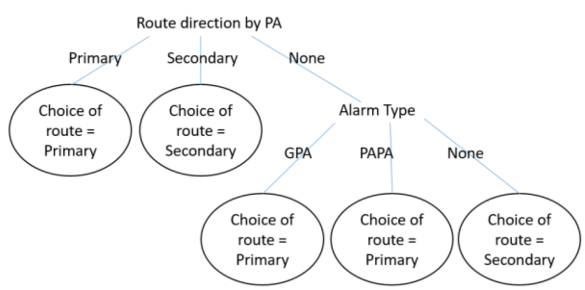
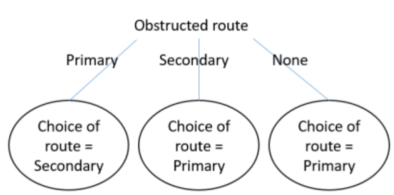
From a spatial competence perspective, both the SBML and LBT trained groups were able to locate the correct muster location (LO1) and follow the egress routes (LO2) in benign conditions. This is shown in the results from the muster drill scenario (T2 in Table 3.4). In the T2 scenario, 55 participants in the SBML group reached the correct muster station and 54 participants (representing 98%) were able to follow the safest egress route. For the LBT group, 16 participants reached the correct location and 15 (representing 88%) were able to follow the safest route. The main spatial competence differences between the SBML and LBT groups were observed in the emergency scenario (T4 in Table 3.4), specifically in route selection (LO2) and rerouting when the egress path was blocked by hazards (LO3, LO4, and LO5). Thirty participants (representing 55%) in the SBML group selected the safest route while only six participants (representing 35%) in the LBT group selected the safest route from onset of the emergency. Twelve participants (representing 22%) in the SBML group were able to avoid hazards by using information from PA announcement to re-route to the safest egress route. Nine participants (representing 16%) in the SBML group were forced to re-route when their path was blocked while three participants (representing 18%) in the LBT group were forced to re-route when they encountered the hazard blocking their path. Four participants from the SBML group (representing 7%) and eight participants from the LBT group (representing 47%) continued on the unsafe route and went directly through the smoke hazard.

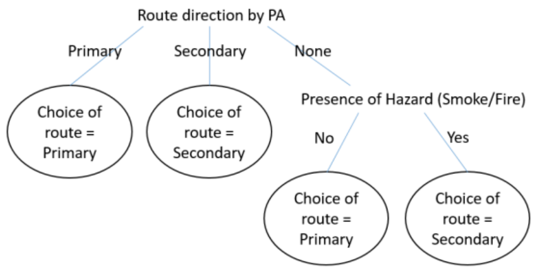
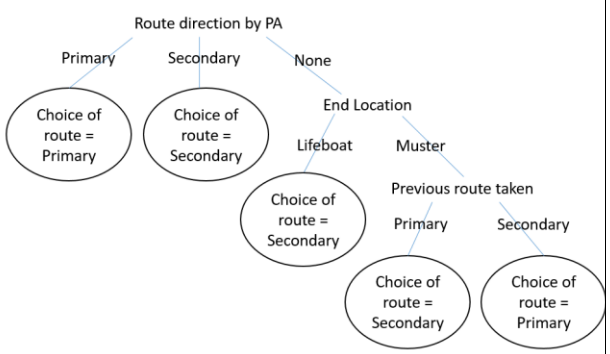
This work focused mainly on the spatial learning objectives LO1 to LO5 in Table 3.4. However, large differences were also observed between the trained groups in the procedural performance, specifically learning objectives LO7, LO8, and LO9 (i.e. avoiding hazards, refraining from running, and remembering to close the fire and watertight doors).

3.6.2. Comparing the SBML group's DTs to the intended learning objectives

The DTs were used to judge the efficacy of the SBML training by comparing the SBML trained group's DTs to the intended learning objectives at two stages of the training program. The SBML trained participants' DTs were developed using data from modules 2 and 4 to see how the trees evolved as more training content was added to the participants' data repository. The different DTs for the SBML training are summarized in Table 3.5.

Table 3.5: Types of DTs formed after SBML training modules 2 and 4 (DM1 & DM2).

Type	Decision Rules	% Participants DM1	% Participants DM2	Learning Objective Comparison
1		73% (40 participants)	64% (35 participants)	Correct
2		27% (15 participants)	16% (9 participants)	Correct
3.1		0%	5% (3 participants)	Correct
3.2	 *One participant A10 had a similar DT but reversed rules for PAPA & None	0%	5% (3 participants)	Correct
4		0%	2% (1 participant)	Incomplete

5		0%	5% (3 participants)	Incorrect
6		0%	2% (1 participant)	Incorrect

3.6.2.1. Alarm Recognition Decision Tree (DM1)

In the muster drill (T2) and the emergency (T4), the alarm type indicated the severity of the situation and dictated the final muster location (e.g. muster or lifeboat station). During the GPA alarm, personnel were required to gather at the muster station. During the PAPA alarm, personnel were required to muster at the lifeboat station. The main learning objective for module 2 was for participants to listen to the alarm and relevant instructions from the PA announcement and take the safest route available in response to the situation. Table 3.5, column DM1 shows the intended decision tree that was taught for a muster drill situation (denoted as Type 1). Seventy-three percent of participants achieved this type of decision tree before the test scenario (T2). The remaining 27% of participants also formed their route selection based on the PA announcement, but when the PA provided no route information, they relied on their intended end location, which was dictated by the alarm type. For these participants, the muster station meant taking the primary route and the lifeboat station meant taking their secondary route (denoted as Type 2). The formation of both DTs (Type 1 and 2) after module 2 demonstrated that all participants achieved the intended learning objectives and were adequately prepared to respond to the muster drill test scenario (T2).

3.6.2.2. Assess Emergency Situation Decision Tree (DM2)

Building on earlier learning objectives, module 4 trained participants how to assess the situation, avoid hazards, and follow the safest egress path to the designated muster or lifeboat station. In an emergency, if personnel encountered an obstructed route, they were required to re-route in response to the hazardous situation. A variety of DTs were developed after module 4. Table 3.5, column DM2 shows that there were six different strategies used by participants at the end of training. Sixty-four percent of participants continued to use the same decision tree in which they selected their egress route based on information from the PA (Type 1). Sixteen percent of participants continued to use the strategy in which the end location (dictated by the alarm type) indicated the route choice in the absence of a PA (Type 2). Ten percent of participants followed the alarm type and PA (Type 3). If no clear route direction was provided over the PA, the participants would link the alarm type to an egress route. For example, if the GPA or PAPA alarm sounded, the participants would take the primary route. However, in the event of no alarm, they would take their secondary route. The remaining 10% of participants demonstrated more varied behaviours. In these cases, when the PA did not provide a route direction, some individuals put emphasis on different attributes to make their decision. One participant's data (representing 2%) formed a correct but incomplete DT where the route decision was based solely on whether the route was obstructed or not (Type 4). The remaining four participants' data (representing 8%) formed incorrect DTs that wrongly considered the presence of hazards (Type 5) and the previous route taken (Type 6).

When comparing the DT variations with the learning objectives, some weaknesses in the training and the participants were identified. The formation of an incomplete DT (e.g. Type 4) suggests that this participant required more targeted scenarios to focus on the missing decision rules (e.g. additional practice variations for situations to create the intended PA decision rules). The incorrect DTs (e.g. Types 5 and 6) show that some participants (7%) require additional practice opportunities and feedback to ensure they reach the intended competence. If incorrect trees persist, then it is possible the participants are not suitable for VE training or are not taking the training seriously (e.g. Type 6 where the participant's decision involved their previous route taken).

3.6.2.3. DTs In-depth Analysis of SBML Training

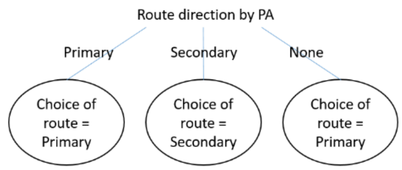
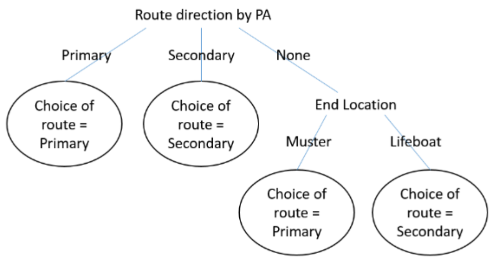
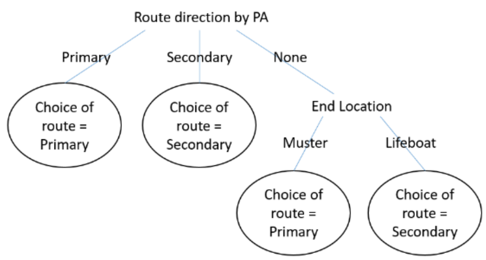
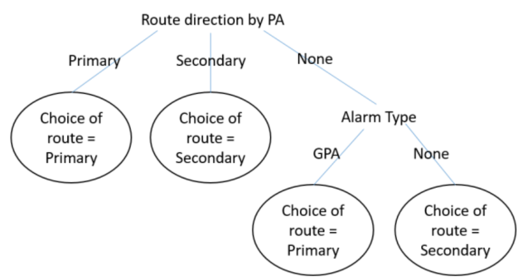
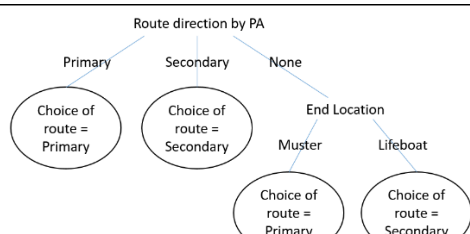
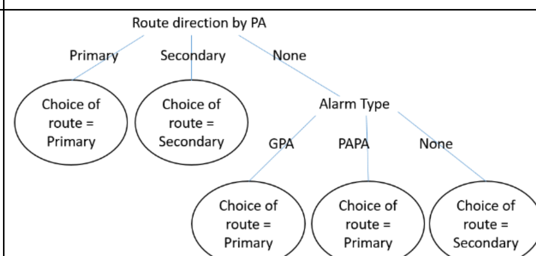
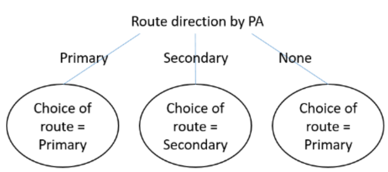
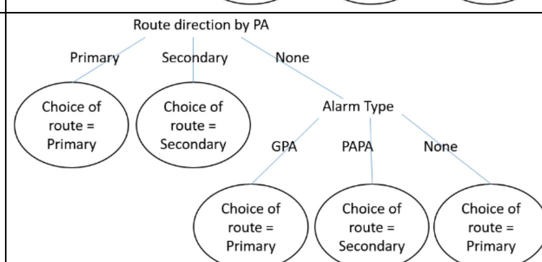
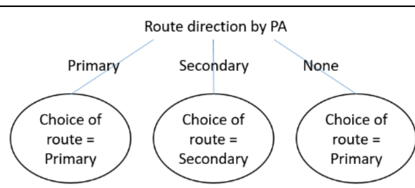
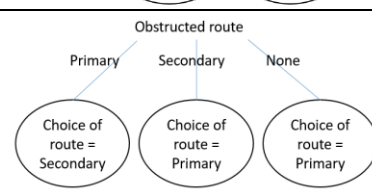
The decision tree analysis revealed information about the participants' performance that would otherwise not be apparent when looking solely at performance metrics in relation to the learning objectives. The diagnostic capabilities of DTs allowed for a more in-depth performance analysis because DTs can identify the strengths and weaknesses of participants' decision-making strategies.

The majority of participants' DTs matched the intended learning objectives (100% for DM1 and 90% for DM2). These participants, whose data formed DT types 1, 2, and 3, demonstrated the decision-making skills taught by the SBML training program. They were able to identify attributes that were critical to success and come up with strategies that led to safe egress. The DTs also provided indications of deficiencies in the training, such as the over reliance on PA announcements during emergencies, and the need to provide participants with sufficient spatial knowledge and re-routing strategies. For example, the DTs of some participants revealed that their decision strategies centered on the PA announcement (Types 1, 2, and 3). In the absence of a PA announcement, some participants focused their attention on a variety of different attributes (e.g. presence of hazards), which were useful in terms of their performance in making effective egress decisions. However, this variability in DTs formation due to missing or unclear PA announcements provides valuable information on whether the decision-making skills taught were sufficient for all emergencies. These are areas that could be improved in future iterations of the training. Adaptive training could recognize these deficiencies in real-time and focus further training on teaching participants what to do in the event that there is no PA announcement or instructions on what is happening during the emergency.

3.6.3. Comparison of SBML and LBT Trees

As another lens through which to observe the training efficacy of the SBML approach, the DT results from both experiments were compared directly. The DTs modelled from the SBML training data are summarized in Table 3.6. The DTs modelled from the LBT training data are summarized in Table 3.7.

Table 3.6: Resulting SBML DTs for all 55 participants after finishing modules 2 and 4 (DM1 & DM2).

<table>
<tr><th>% Participants</th><th>Decision rules from DM1 (until test scenario T2)</th><th>Decision rules from DM2 (until test scenario T4)</th></tr>
<tr><td>64% (35 participants)</td><td>
</td><td>Remained the same.</td></tr>
<tr><td>16% (9 participants)</td><td>
</td><td>Remained the same.</td></tr>
<tr><td>5% (3 participants)</td><td>
</td><td>
</td></tr>
<tr><td>4% (2 participants)</td><td>
</td><td>
</td></tr>
<tr><td>2% (1 participant)</td><td>
</td><td>
</td></tr>
<tr><td>2% (1 participant)</td><td>
</td><td>
</td></tr>
</table>

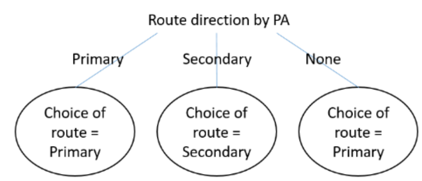
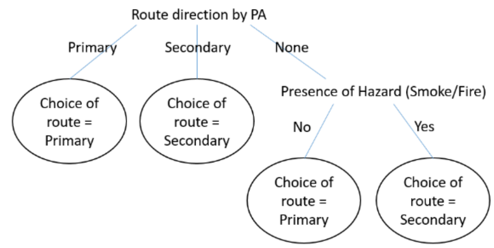
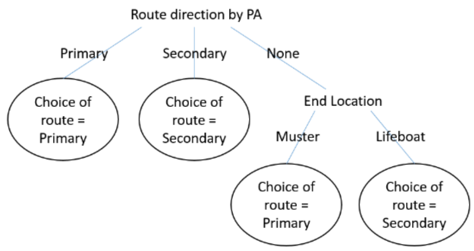
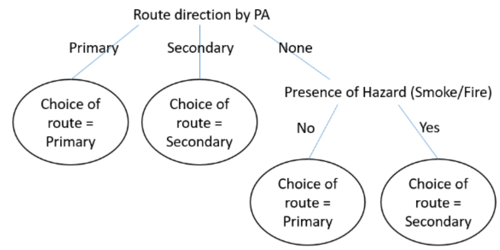
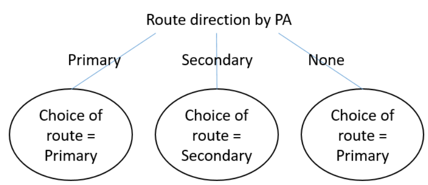
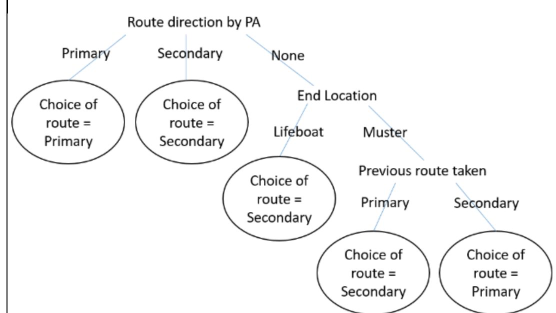
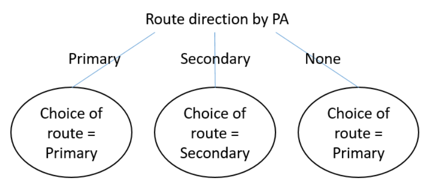
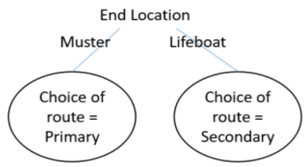
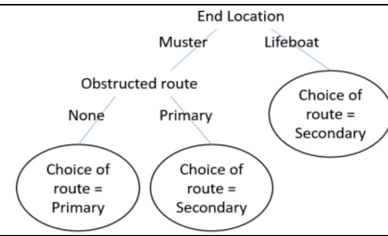
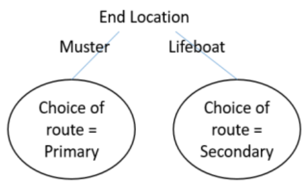
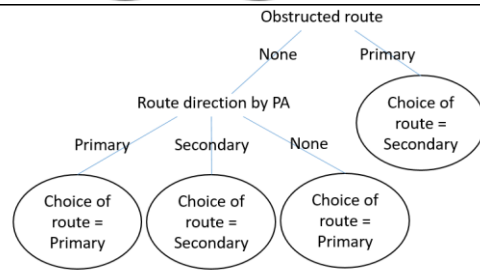
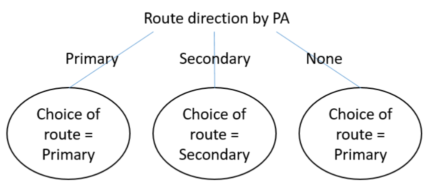
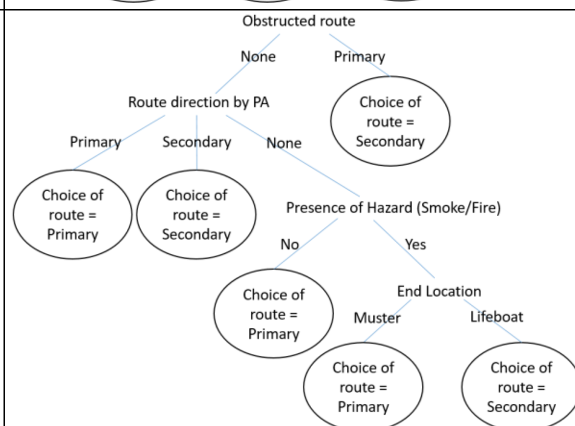
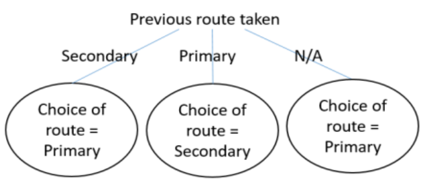
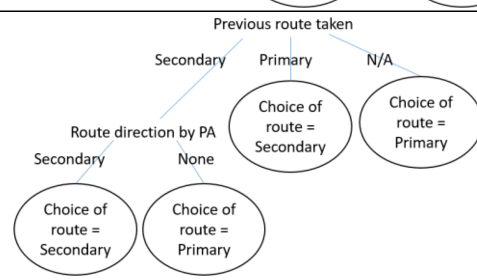
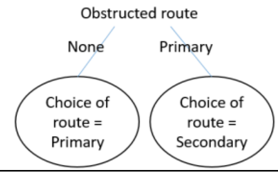
<table>
<tr>
<td>4%
(2
participants)</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2%
(1
participant)</td>
<td>
</td>
<td>
</td>
</tr>
<tr>
<td>2%
(1
participant)</td>
<td>
</td>
<td>
</td>
</tr>
</table>

Table 3.7: Resulting LBT DTs for all 17 participants for test scenarios T2 and T4 (Musharraf et al., 2018).

<table>
<tr><th>% Participants</th><th>Decision rules from DM1
(until test scenario T2)</th><th>Decision rules from DM2
(until test scenario T4)</th></tr>
<tr><td>*29%
(5 participants)</td><td>
</td><td>Remained the same.

*One participant had a slightly different DT for part 2. Depending on whether the participant understood the PA, they either followed the same DT or made a choice based on the obstructed route. If no route was obstructed, they would take the primary route. However, if the primary route was obstructed then they would take the secondary route.</td></tr>
<tr><td>17%
(3 participants)</td><td>
</td><td>
</td></tr>
<tr><td>6%
(1 participant)</td><td>
</td><td>
</td></tr>
<tr><td>6%
(1 participant)</td><td>
</td><td>
</td></tr>
<tr><td>6%
(1 participant)</td><td>
</td><td>
</td></tr>
</table>

6% (1 participant)	At any condition, the participant's choice of route was the primary route.	
6% (1 participant)	At any condition, the participant's choice of route was the secondary route.	Remained the same.
*24% (4 participants)	No behavioral pattern or strategy was identified. The participants' choice of route was random and as a result the decision tree could not provide any more generalization than the data matrix. *One participant had trouble with the controls to open/close doors and chose a route with fewer doors. This participant was excluded because the behaviour specific to AVERT and not realistic to real evacuations.	

Comparing the resulting DTs generated from the SBML and LBT data showed that the different training methods influenced the participants' egress strategies. Over the course of the SBML training, the SBML-trained participants' behaviours in responding to emergencies gradually converged to a few expected DTs (with the exception of a few participants). Ninety percent of SBML trained participants achieved the intended learning objectives as demonstrated by the DTs (Types 1, 2, and 3). Only 10% of SBML trained participants displayed varied behaviours that could be addressed with targeted training. Conversely, the DTs of the LBT-trained participants' behaviours for the emergency response scenarios diverged. Only twenty-nine percent of LBT trained participants achieved the intended learning objectives as demonstrated by the DTs (Type 1). Many of the remaining LBT participants had a poor understanding of the egress procedures and were not compliant. Thirty-five percent of the LBT participants' data presented DT strategies that included special conditions for PA announcements, alarm type, obstructed routes, and hazards. The DTs for two participants (representing 12% of LBT trained participants) showed how inflexible they were on their route choice. For example, their DT represented behaviours of taking the same route regardless of the emergency condition. For twentyfour percent of LBT trained participants, the choice of route was random and the DTs could not provide any more generalization than the data matrix. Overall, the LBT participants' DTs weighted attributes of the scenario that were not useful for making effective egress decisions (Musharraf et al., 2018). The variability and incorrect behaviours observed in the LBT decision trees show that this method of training was inadequate for preparing participants for emergency conditions.

The SBML approach resulted in better route selection strategies compared to the LBT approach. As shown in section 4.2, the majority of the observed route strategies for the SBML trained participants (representing 90%) led to the successful completion of the test emergency scenario. Conversely, the majority of LBT trained participants (representing 71%), displayed incomplete or incorrect DTs. Therefore, the SBML training resulted in higher safety compliance and more concise DTs than the LBT training. This indicates that participants from SBML training were generally better equipped for managing the emergency scenarios.

3.7. Conclusion

The training efficacy of two pedagogical approaches, SBML and LBT, were assessed experimentally using performance outcomes and decision tree modeling in the context of training naïve personnel for basic emergency duties. Overall, the decision tree modeling provided a more comprehensive analysis of the participants' route performance than the conventional performance outcomes. In terms of measured performance, the SBML pedagogical approach was clearly better than the alternative LBT approach. The comparison of performance metrics in both studies indicated that the SBML-trained participants performed better than the LBT-trained participants did; however, the performance metrics did not offer information as to why one group outperformed the other. Conversely, the DTs generated by the participants' data in both studies provided an explanation as to how the route selection performance in the SBML and LBT trained groups differed. The DTs showed that when selecting egress routes in virtual emergencies the decision-making strategies of the SBML-trained participants were more consistent with the intended learning objectives and represented safer behaviours than the DT strategies of LBT-trained participants.

This paper demonstrated the diagnostic capabilities of DTs as training assessment tools. In both training cases, the DTs provided a convenient visual representation of the individual strategies employed by participants. As illustrated in this work, this feature of the DTs can be useful for identifying systemic deficiencies in training (and even in how procedures are designed). They can also be used to diagnose the strengths and weaknesses of individual trainees, a capability that has additional value in terms of adapting training to meet the needs of individuals. This adaptive training potential could be realized by coupling the SBML approach to a virtual environment mode of training in which performance can be tracked and assessed automatically and in real-time, thereby providing the data required by a built-in decision tree diagnostic tool. For this to work in practice, the training scenarios must be carefully designed, as they are, in effect, experiment conditions for the diagnostic DTs. Additional training scenarios would also be required to provide sufficiently specific pathways for adaptive training.

Finally, the DTs were shown to have considerable predictive capability. This feature could also be useful in terms of pedagogical strategies, such as determining when personnel are likely to be sufficiently capable of responding to a wide variety of potential emergencies, without necessarily training them for all potential eventualities.

3.8. Acknowledgements

The authors acknowledge with gratitude the support of the NSERC-Husky Energy Industrial Research Chair in Safety at Sea.

3.9. References

Badino, M., 2004. An application of information theory to the problem of the scientific experiment. Synthese, 140(3), 355-389.

Barsuk, J., Cohen, E., McGaghie, W., and Wayne, D., 2010. Long-term retention of central venous catheter insertion skills after simulation-based mastery learning. Academic Medicine, 85(10), 9–12.

Barsuk, J., Cohen, E., Wayne, D., Siddall, V., & McGaghie, W., 2016. Developing a simulation-based mastery learning curriculum – Learning from 11 years of advanced cardiac life support. Simulation in Healthcare, 11(1), 52–59.

Benner P. (1982). From novice to expert, American Journal of Nursing. 82, pp. 402–407.

Bloom, B., S (1971). Mastery Learning. In J. H. Block (Ed.), Mastery Learning: Theory and practice (pp.47-63). New York: Holt, Rinehart & Winston.

Cacciabue, P., Decortis, F., Drozdowicz, B., Masson, M., & Nordvik, J., 1992. COSIMO: a cognitive simulation model of human decision making and behavior in accident management of complex plants. Systems, Man and Cybernetics, IEEE Transactions, 22(5), 1058-1074.

Cohen, E., Barsuk, J., Moazed, F., Caprio, T., Didwania, A., McGaghie, W., and Wayne, D., 2013. Making July safer: Simulation-based mastery learning during intern boot camp. Academic Medicine, 88(2), 233–239.

Cook D, Brydges R, Zendejas B, Hamstra S, Hatala R (2013) Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. Acad Med 88:1178–1186

Dreyfus, S., & Dreyfus, H. (1980). A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Fort Belvoir, Va: Defense Technical Information Center; 1980.

Duffy, V., 2009. Handbook of digital human modeling: research for applied ergonomics and human factors engineering. CRC Press Taylor & Francis Group.

Griswold-Theodorson, S., Ponnuru, S., Dong, C., Szyld, D., Reed, T., and McGaghie, W. (2015). Beyond the Simulation Laboratory: A Realist Synthesis Review of Clinical Outcomes of Simulation-Based Mastery Learning, Academic Medicine, 90(11), pp. 1553-1560. Doi: 10.1097/ACM.0000000000000938.

Gusky, T., 2007. Closing achievement gaps: revisiting Benjamin S. Bloom's "Learning for Mastery". Journal of Advanced Academics, 19, 8-31.

Han, J., Kamber, M., & Pei, J., 2011. Data mining: concepts and techniques. Elsevier.

House, A., Smith, J., MacKinnon, S., and Veitch, B., 2014. Interactive simulation for training offshore workers. Proceedings, Oceans, St. John's.

Moazed, F., Cohen, E., Furiasse, N., Singer, B., Corbridge, T., McGaghie, W., and Wayne, D., 2013. Retention of critical care skills after simulation-based mastery learning. Journal of Graduate Medical Education, 458–463.

McGaghie, W., Issenberg, S., Petrusa, E., and Scalese, R., 2006. Effect of practice on standardised learning outcomes in simulation-based medical education. Medical Education, 40, 792-797.

McGaghie, W., Issenberg, S., Barsuk, J., and Wayne, D., 2014. A critical review of simulation-based mastery learning with translational outcomes. Medical Education, 48(4), 375-385.

Musharraf, M., Smith, J., Khan, F., & Veitch, B., 2018. Identifying route selection strategies in offshore emergency situations using decision trees. Journal of Reliability Engineering & System Safety, In Press.

Smith, J., Veitch, B., & MacKinnon, S., 2015. Achieving competence in offshore emergency egress using virtual environment training. Proceedings, Offshore Mechanics and Arctic Engineering, OMAE-41132, St. John’s.

Smith, J., Musharraf, M., & Veitch, B., 2017. Data informed cognitive modelling of offshore emergency egress behaviour. Proceedings, International Conference on Cognitive Modeling 2017, Coventry.

Smith, J., & Veitch, B. 2018. A better way to train personnel to be safe in emergencies. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B, 5(1), 011003. https://doi.org/10.1115/1.4040660

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.

4. BEING PREPARED FOR EMERGENCIES: A VIRTUAL ENVIRONMENT EXPERIMENT ON THE RETENTION & MAINTENANCE OF EGRESS SKILLS

Jennifer Smith*, Kyle Doody, and Brian Veitch

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

St. John's, NL, Canada, A1B 3X5

*Correspondence author: jennifersmith@mun.ca

4.1. Co-authorship Statement

This manuscript was published in the WMU Journal of Maritime Affairs in August 2019 and partial results were first presented at the 9th annual Applied Human Factors and Ergonomics (AHFE) Conference in July 2018. Author Jennifer Smith led the writing of this manuscript. Kyle Doody performed the retention experiment and shared the data. Dr. Brian Veitch assisted in the interpretation of the results and provided guidance, editorial changes, and recommendations for improvements to several drafts of this paper.

4.2. Abstract

The retention of safety-critical egress skills is an essential part of emergency preparedness on offshore petroleum platforms. Virtual environment (VE) training has been shown to be an effective method for teaching basic onboard familiarization and offshore emergency evacuation procedures. This technology has the potential to train crews before they are deployed offshore. This paper investigates the long-term retention and maintenance of emergency egress competence obtained using a virtual offshore platform. In particular, the research aimed to answer two questions: 1) what egress skills can be remembered after a period of 6-months? and 2) how effective is a VE-based retraining program at maintaining egress skills? A two-phased experiment was designed to first teach basic egress skills and subsequently assess skill retention after a 6 to 9-month period. The first phase of the experiment used a simulation-based mastery learning (SBML) pedagogical approach to teach naïve subjects the necessary spatial and procedural skills to evacuate safely. In the second phase of the experiment, the same 36 participants were tested after the retention interval on their ability to respond to a series of egress test scenarios. Participants who had trouble remembering the egress procedures were provided retraining on deficient skills. The results of the experiment indicate that emergency egress skills (both spatial and procedural knowledge) are susceptible to skill decay. This paper will highlight the skills that were most susceptible to skill fade after a period of 6 to 9-months and discuss the efficacy of the retraining participants received to return to competence.

4.3. Introduction

Offshore emergencies require the prompt response of prepared crews. Emergencies do not afford second chances. Thus, the retention of safety-critical emergency response skills is an essential part of an offshore emergency preparedness plan. Offshore emergency response teams rely on individuals to follow egress protocols to ensure that all personnel onboard have been accounted for in an emergency.

Workers typically acquire egress skills through conventional safety induction training on their first deployment offshore. This training involves watching safety videos followed by a supervised orientation period for their own safety. Crew are typically required to participate in induction training for a designated time period, but no formal assessment is performed to assure competence has been achieved. This form of training does not address individual learning requirements and allows learning outcomes to vary amongst inductees. To maintain competence while offshore, workers are required to perform weekly muster drills and quarterly evacuation drills. Due to safety constraints, the drills are not representative of real emergency conditions. This disconnect between drills and emergencies can result in negative training transfer (Wickens et al. 2013).

According to industry standards (e.g. CAPP, 2015), personnel who return to work on a platform after an extended period (e.g. 6 months or more) are required to undergo safety training again, regardless of their previous experience. This requirement is based on the understanding that egress skills deteriorate over time without practice. The mandated recurrency schedule is not informed by the individual skill retention abilities. The lack of personalized training can result in a recurrency schedule that is too infrequent for some, causing training to be forgotten, or conversely, a recurrency cycle that is too frequent for others, which can undermine the training by causing worker complacency.

Virtual environment (VE) training can address weaknesses in conventional safety training and provide a way to practice emergency egress skills regularly, unconstrained by safety, logistical, or financial concerns. VEs are effective at teaching basic onboard familiarization and emergency evacuation procedures for offshore petroleum platforms (Smith and Veitch 2018). Specifically, VE training can provide assurance that all individuals have at least achieved the same minimum standard of competence. VE training also provides practice in situations that muster drills onboard cannot replicate, such as the high-stress, dynamic, and hazardous conditions of emergency situations.

For safety-critical skills, training is only effective if the skills can be recalled and used in a real emergency. This brings to question: 1) can egress skills acquired using VE training be remembered after a period of 6-months without any other form of training? 2) can a VE-based retraining program help maintain egress skills by returning participants to competence? To inform these questions, this paper presents a two-phased experiment to investigate the long-term retention of offshore egress skills attained using a virtual environment.

The first phase, skill acquisition, was conducted using the simulation-based mastery learning (SBML) pedagogical framework to teach virtual offshore emergency egress training (Smith and Veitch 2018). Fifty-five novice participants participated in the skill acquisition phase of the experiment. All participants who completed the SBML training achieved the targeted performance outcomes and demonstrated competence at the end of the program. Smith and Veitch (2018) compared the SBML approach to a benchmark training program called lecture-based teaching (LBT). The LBT training program represented the existing safety protocols used offshore for egress training (Smith 2015). The results of this comparison showed that SBML training was more effective at bringing all participants to competence and did so in less time than the LBT methods. This phase of the experiment established a benchmark of competent performance and corresponding times required to achieve competence (i.e. for comparison with the retention phase measurements). All fifty-five participants were invited to return to participate in the second phase, skill retention, of the experiment. The second phase evaluated the retention of skills attained by the same participants who completed virtual offshore egress training in the first phase SBML experiments. After the retention interval of 6 to 9-months, thirty-six participants returned to complete the same test scenarios used in the SBML experiment. The participants' performance in the test scenarios at the end of the skill acquisition phase was compared to their first attempt performance in the same test scenarios at the beginning of the retention phase. This comparison assesses the retention of egress skills required to evacuate an offshore platform in an emergency. Participants in the retention study who failed to complete the test scenarios were retrained using exercises that focused on the particular skills they failed to demonstrate. The impact of retraining was measured to determine how well retraining improved participants' performance in subsequent test scenarios. The goals of this research were to: 1) determine if egress skills were retained for a period of 6 months without other training interventions, 2) identify the learning objectives that were more susceptible to skill degradation, and 3) determine the efficacy of the retraining in bringing participants back to competence. All three aspects are discussed in this paper.

4.4. Overview of Factors that Influence Skill Retention

Many factors influence how well skills are remembered. Arthur et al. (1998) performed a meta-analysis of skill retention literature and described seven factors that influence skill decay and retention: i) length of time lapsed of non-skill use during retention interval, ii) the quality of the original skill acquisition and the amount of overlearning that occurred;

iii) skill type and task characteristics (e.g. physical versus cognitive tasks); iv) the methods used to test learning and retention; v) conditions of retrieval or specificity of training (i.e. the similarity of learning and testing contexts); vi) the instructional strategies or methods used to teach the skills; and vii) individual differences in acquiring and retaining skills.

Sanli and Carnahan (2018) in their review of multi-day training courses in medical, military, marine and offshore safety fields discussed similar factors that influence skill retention. According to Sanli and Carnahan (2018), the factors that influence skill and knowledge retention in these safety-critical domains include: a) type of skill (e.g. practical and declarative knowledge); b) task complexity and difficulty (e.g. number of steps and order of tasks); c) individual differences and the experience of the learner; d) specificity of training (i.e. closeness of the learning and testing contexts); e) the amount of practice and on the job exposure provided; and f) the frequency that refresher interventions are delivered.

Three main topics will be discussed in the context of virtual offshore egress training: 1) the influence of instructional or pedagogical strategies on skill acquisition, 2) the impact of skill type on forgetting (such as spatial, declarative, and procedural knowledge), and 3) the frequency of practice (e.g. how often recurrency training is provided and the length of time that passes between training sessions).

4.4.1. Strategies for Improved Skill Acquisition

In emergency response domains, the amount of training provided is typically dictated by fixed timelines and does not take into consideration individual differences in learning. Training dictated by a fixed timeline refers to limiting the training material and/or opportunities to practice to within the allocated time of the course. In the medical field, fixed training times have been found to cause performance outcomes to vary (Cook et al. 2013; Gallagher et al. 2005). To assure skills are properly acquired in the first place, training programs are shifting from time-based frameworks to competence-based models. Virtual environment training using the pedagogical frameworks developed by the medical education field can assist offshore operators in transitioning from a fixed-time training model to competence-based training. For example, McGaghie et al. (2014) developed the simulation-based mastery learning (SBML) pedagogical framework for the medical education field. The SBML method accommodates different learning styles and paces by ensuring all individuals reach a minimum competence standard by the end of the training program. It achieves this using two main features: 1) opportunities to practice and receive formative corrective feedback until competence is demonstrated, and 2) allowing trainees to advance to more complicated training material only once foundational skills are demonstrated.

4.4.2. Skill Type - Spatial and Procedural Skills for Retention

Offshore egress training is provided with the expectation that egress skills remain current in the event of an emergency so that individuals are prepared to take action. The type of task influences how well skills are retained after a period of non-use. The safe evacuation of an offshore platform requires two types of skills: spatial knowledge of the platform to assist in wayfinding, and procedural knowledge of the protocols in place to protect personnel from harm in an emergency. Thus, understanding the retention of spatial and procedural skills is important for providing adequate training for real emergency situations.

This section will discuss the differences in acquiring and retaining spatial, declarative, and procedural knowledge.

4.4.2.1. Spatial Knowledge

Landmark-Route-Survey (LRS) is a spatial knowledge acquisition model (Seigel and White 1975) that explains how people develop their understanding of an environment. People first recognize landmarks, then learn the routes that connect landmarks, and over time they develop survey knowledge of how the landmarks and routes are interconnected. Developing a spatial understanding of an environment on all three levels (landmark, route, and survey) can also develop concurrently (Taylor, Brunye, and Taylor 2008). However, survey knowledge often requires longer exposure to the environment to gain a map-like representation of the environment (e.g. learning how landmarks and routes are interconnected).

Survey knowledge is important for evacuating an offshore platform because a well-known route may not always be available in an emergency, leaving personnel to find a less traversed tenable path to their muster stations. For example, researchers observed that in emergency situations, people tend to evacuate buildings by taking the known main exit instead of the nearest fire exit (Kobes et al. 2010). This risky behaviour can be addressed by providing people with more time to learn survey knowledge of an environment.

4.4.2.2. Declarative & Procedural Knowledge

Kim, Ritter, and Koubek (2013) integrated four learning theories (Fitts 1964; Anderson 1982; Rasmussen 1986; VanLehn 1996) into a three-staged skill acquisition process: 1) declarative stage – learning declarative knowledge (i.e. information or facts), 2) mixed stage – consolidating the acquired task knowledge to form a mix of declarative and procedural knowledge, and 3) procedural stage – tuning the knowledge towards predominately procedural knowledge through overlearning. This model provides a framework to help explain how skills are learned and forgotten.

Declarative knowledge will degrade with the lack of use (e.g. information will no longer be available in memory for retrieval). Declarative knowledge can be transformed into procedural knowledge over time (e.g. gradually associating knowledge, transforming it into rules, and developing heuristics and biases). Frequent practice and contextual experience allow experts to proceduralize skills so that they rely less on declarative knowledge and are able to perform the task automatically in response to a situation (Kim et al. 2013). Procedural knowledge is implicit as experts possessing the knowledge are able to perform the actions without effort but are sometimes unable to verbalize the knowledge (Wickens et al. 2013). Siu et al. (2016) suggest that trainees should be provided with sufficient practice to allow them to reach the proceduralization stage, thereby increasing likelihood of skill retention.

4.4.3. Frequency of Retraining

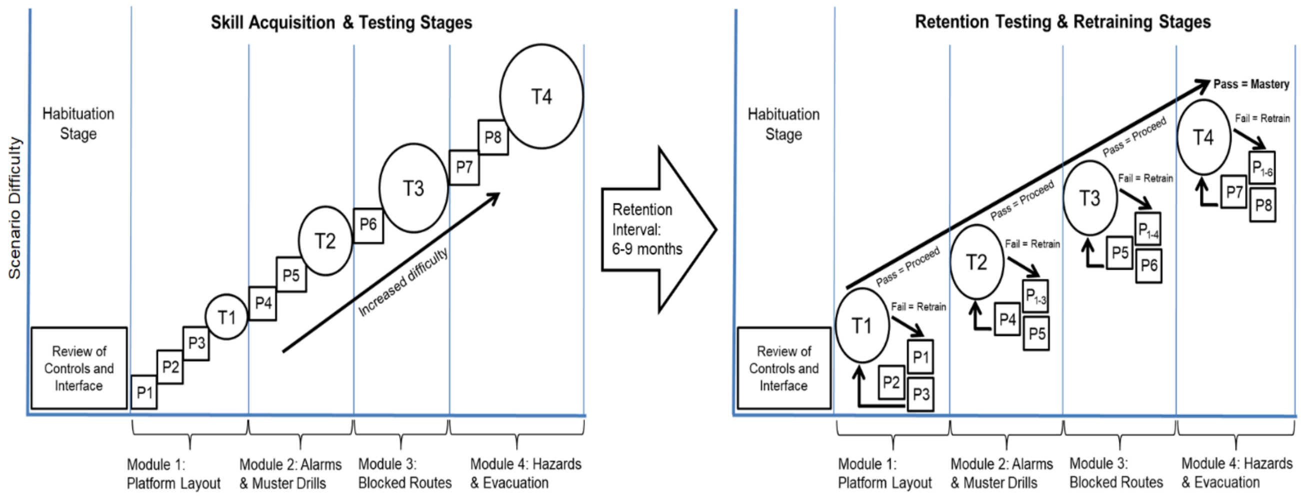
The amount of time that lapses between retraining sessions is an important factor to investigate in order to ensure safety-critical skills are maintained. Predicting the rate at which skills will be forgotten can help inform the frequency with which recurrency training should be provided (Wickens et al. 2013).

In the review of multi-day safety training courses, Sanli and Carnahan (2018) concluded that complex skills could be remembered for at most a six-month period without any form of training interventions. Atesok et al. (2016) reviewed literature on the retention of simulation-based trained orthopaedic surgery skills and found that repetitive practicing of skills learned in a simulator helped mitigate skill decay even after some time had lapsed (these studies ranged in amount of time lapsed; e.g. follow-up retention assessments occurred at 1 month, 3 months, 6 months, to a maximum of 30 months).

Knowing that egress skills (i.e. spatial, declarative, and procedural knowledge) deteriorate over time without practice, the offshore industry standards require personnel to undergo recurrency training if they have been away from the platform after an extended period (e.g. 6 months or more). This brings to question: how well are egress skills retained in a 6-month period? and how effective is a VE-based retraining program at maintaining egress skills? This paper presents results to answer these questions.

4.5. Methods

The experiment consisted of two phases: 1) a skill acquisition phase using the simulation-based mastery learning (SBML) approach, and 2) a skill retention assessment and retraining phase, which took place after a period of 6 to 9-months. Figure 4.1 depicts phases I and II of the experiment. Both phases of the experiment consisted of a habituation stage followed by a series of modules with practice scenarios and testing scenarios (denoted in Figure 4.1 as P1-P8 and T1-T4, respectively). In phase I, participants were tasked in each module with completing the practice scenario correctly (i.e. to criterion) before advancing to the test scenarios (Figure 4.1a). In phase II, participants were re-tested on the same test scenarios and were re-trained if they made any errors made in the test scenario. The retraining consisted of specific practice scenarios (Figure 4.1b).



(a) **(b)**

Figure 4.1: AVERT Skill Acquisition (phase I) and Retention & Retraining (phase II) (after Smith, Doody, and Veitch, 2018)

This section will briefly describe the effect size and power analysis, participants, the AVERT simulator, skill acquisition and test scenarios, and the retention assessment and retraining matrices. A detailed description of the methods used in phase I can be found in Smith and Veitch (2018). A description of the methods used in phase II can be found in Doody (2018).

4.5.1. Estimated Effect Size and Power Analysis

The effect of interest in this experiment was the change in performance score from the skill acquisition phase to the retention phase. The effect size was calculated based on an estimated drop in performance of 15% or greater due to skill fade and was informed by previous experiments (Smith 2015; Smith and Veitch 2018). Based on the estimated minimum amount of skill degradation to be detected, the effect size calculation resulted in an effect size $d = 0.6$. This is a large effect based on Cohen's convention for t-test on means of two dependent (or paired) samples (Cohen 1988).

A priori power analysis was performed using *G*Power3* (version 3.1.9.2) software (Faul et al. 2007) to determine the required sample size for the retention portion of the longitudinal experiment. For the repeated measures design, the following specifications were used: a matched pairs Wilcoxon signed-rank test (the non-parametric equivalent of a two dependent samples t-test), one-tailed (for directional hypothesis that some egress skills will be lost), with input parameters: significance level $\alpha = 0.05$, power level $(1 - \beta) = 0.95$, and effect size $d = 0.60$. For the Wilcoxon signed-rank test, *G*Power3* returned a sample size of N = 33 participants to achieve a power level of 0.95 with critical $t = 1.696$ and non-centrality parameter $\delta = 3.389$. This result indicated that the retention portion of the longitudinal study required at least 33 participants to return and complete the test scenarios in order to maintain a statistical power of 0.95 (i.e. 95% chance of the result was not due to a type II error).

4.5.2. Participants

Memorial University's Interdisciplinary Committee on Ethics in Human Research approved the experimental protocol. Following the approved research protocol, the recruitment strategy focused on naïve participants (to control for spatial knowledge and experience) and this translated into recruiting undergraduate and graduate students. Participants were recruited from the university's campus by email, posters, and by word of mouth. All volunteers who participated were naïve to the experimental design, had no prior experience working offshore and had no exposure to the simulator prior to the study. All volunteers provided their informed consent before participating in the experiment.

Sixty participants were recruited for the first phase of the study with an expectation of 25% attrition for the longitudinal portion of the study. Five participants withdrew at the onset, due to simulator sickness or difficulty with the controller. Fifty-five participants completed the skill acquisition training (phase I) and were invited to return after a period of six months to participate in the retention assessment (phase II). Seventeen participants opted out of the longitudinal study during the 6 to 9-month retention interval. The remaining 38 participants completed the retention phase. Two were identified as outliers (completed the retention assessment at 4 and 10 months) and were removed from the retention analysis. Thirty-six participants completed the retention phase within the designated 6 to 9-month period. Twenty-seven participants were male and nine participants were female. Participants ranged in age from 19 to 54 years (M = 29 years, SD = ±8.8 years).

4.5.3. AVERT Simulator

Emergency egress training was provided using the All-hands Virtual Emergency Response Trainer (AVERT). The AVERT simulator is a desktop virtual environment that allows participants to interact with the virtual offshore platform using a gamepad controller (Xbox). The virtual environment depicts a realistic representation of an offshore Floating Production Storage and Offloading (FPSO) vessel, similar to those used in the

Newfoundland offshore area. A generic virtual FPSO platform was chosen for its relevance to the local offshore industry.

Participants moved within the FPSO by controlling a first-person perspective avatar of an offshore worker. Participants were first provided with habituation scenarios for orientation with the simulator controls. The AVERT training provided a series of training scenarios with built-in guidance, multiple opportunities to practice, and test scenarios with after-action feedback. Participants were tasked with learning their way around the accommodation block of the platform, and the safety protocols for responding to emergency situations.

4.5.4. Skill Acquisition and Test Scenarios

A previous virtual environment training experiment by Smith (2015) used lecture-based teaching (LBT) methods with the AVERT simulator and found that fixed instructional time was ineffective at ensuring participants acquired the necessary skills to respond to virtual emergency situations. Individual learning differences, such as style and pace, were believed to contribute to the failure of participants to reach competence using conventional LBT training. The simulation-based mastery learning (SBML) pedagogical framework (McGaghie et al. 2014) was adopted for the first phase of the longitudinal experiment to accommodate for individual differences. The SBML framework was used to deliver offshore emergency egress training in the AVERT simulator.

The training curriculum and assessment criteria for the experiment were developed based on subject matter expert guidance and industry regulations (Transport Canada 2007; International Maritime Organization 2001; Canadian Association of Petroleum Producers

2015; International Association of Drilling Contractors 2009; International Association of Oil and Gas Procedures 2010). The learning objectives for the AVERT simulator were outlined by industry representatives in a workshop and the objectives were verified against the functionality and capabilities of the AVERT simulator. Table 4.1 provides a list of the learning objectives taught using AVERT.

Table 4.1: Learning objectives for AVERT (Smith, Doody, & Veitch 2018)

No.	Learning Objectives	Skill Type
LO1	Reach correct location	Spatial
LO2	Recognize alarm	Procedural
LO3	Select safest egress route	Spatial
LO4	Re-route based on PA information or if path blocked	Spatial
LO5	Avoid exposure to hazards	Procedural
LO6	Take safety equipment	Procedural
LO7	Register at the correct muster station	Procedural
LO8	Avoid running	Procedural
LO9	Close all fire and watertight doors	Procedural

The US Coast Guard's method for developing mariner assessments was used to develop the assessment criteria, proficiency standard, and performance scoring system (McCallum et al. 2000). 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 emergency scenarios for the research team to model in AVERT so that trainees could demonstrate their understanding of the 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. 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 to the participants for each scenario. The scenarios were tested and refined prior to starting the experiment.

This research looked at the retention of spatial and procedural skills in the context of emergency egress. The spatial learning objectives for this experiment included familiarity with the platform layout, and knowledge of the egress route options. The procedural skills were defined as a combination of declarative and procedural knowledge (e.g. remembering facts and formulating rules to follow). The training in the experiment aimed to teach personnel to comply with safety protocols. The procedural learning objectives included recognizing emergency alarms, assessing the emergency situation, avoiding hazards, following safety protocols, and mustering procedures.

As depicted in Figure 4.1a, participants were taught the learning objectives using four modules. Each module had training scenarios (depicted in Figure 4.1a, as P1, P2, P3, P4, P5, P6, P7 and P8) to teach participants how to accomplish the egress tasks, and subsequent test scenarios (depicted in Figure 4.1a, as T1, T2, T3, and T4) to assess participants' competence. The modules gradually increased in difficulty, building on previously presented learning objectives. Module 1 taught the spatial layout of the platform (LO1), the different egress routes available from the trainee's cabin (LO3), and how to safely move within the platform by avoiding running and remembering to close fire and watertight doors (LO8 & LO9). Module 2 taught trainees how to respond to different alarm types (LO2), and the mustering procedures at the temporary safe refuge (TSR) on the platform (LO6 & LO7). Module 3 taught trainees how to assess the emergency situation and to listen to the public address (PA) announcement for information on the tenability of the egress routes (LO4). Module 4 taught hazard avoidance and what to do when an egress route was obstructed (LO5).

Participants were required to complete each training scenario correctly before moving on to the next scenario. Participants who made errors in a particular training scenario were required to repeat the scenario until competence was demonstrated. This protocol of training until competent is referred to as trials to criterion. After each training module, the participants' performance was assessed using a test scenario. Table 4.2 provides a detailed description of the four test scenarios.

Table 4.2: Description of the test scenarios (Smith, Doody, & Veitch 2018)

Test Scenario	Scenario Description
T1 Wayfinding Drill	This scenario assessed the participants' spatial knowledge of the platform. Participants were asked to meet their supervisor at their assigned lifeboat station by following their primary or secondary egress routes.
T2 Muster Drill	This scenario assessed the participants' understanding of alarms and muster procedures. Participants were tasked with responding to a muster drill (General Platform Alarm). During this alarm, all personnel were required to collect their safety equipment and muster at their primary muster station.
T3 Blocked Route	This scenario assessed the participants' ability to deal with obstructions to their planned egress route. Participants were required to respond to the alarm, listen to the announcement, and follow the muster procedures. The PA announcements provided information to help the participants select the most effective route.
T4 Emergency	This scenario assessed the participants' ability to avoid hazards and follow the safest available route to their lifeboat station. Participants were tasked with responding to an emergency involving a General Platform Alarm due to fire in the galley. The fire compromised the muster station with smoke and the situation escalated to a Prepare to Abandon Platform Alarm. Initially all personnel were required to go to the muster station but were forced to re-route to the lifeboat station because of the compromised muster station.

The test scenarios required participants to use the knowledge learned in the module. Each module built upon the learning objectives taught in prior modules, and as a result, the corresponding test scenarios became more comprehensive. Following the same trials to criterion protocol as the training scenarios, participants who made errors in the test scenarios were required to repeat the scenarios until competence was demonstrated.

4.5.5. Retention Assessment and Adaptive Retraining Matrices

After a retention interval of 6 to 9-months, participants were given the opportunity to demonstrate their retention of offshore emergency egress skills by performing the same four test scenarios that they had successfully mastered in the skill acquisition phase. Figure 4.1b depicts the retention assessment and retraining phase in AVERT. The figure shows the test scenarios for each of the modules. Participants who were successful at completing a test scenario advanced to the next test scenario. This process continued until all the test scenarios were completed. A failure to complete a test scenario resulted in participants being required to do corrective training exercises. After successfully completing the retraining exercises, these participants were required to reattempt and pass the test scenario before moving on to subsequent test scenarios.

A series of adaptive training matrices were used to assign the participants the corrective training scenarios to address the specific errors they made in the test scenarios. Each learning objective that participants failed had a corresponding corrective training scenario. Figures 4.2 and 4.3 provide examples of the retraining matrices used for test scenarios T2 and T4, respectively.

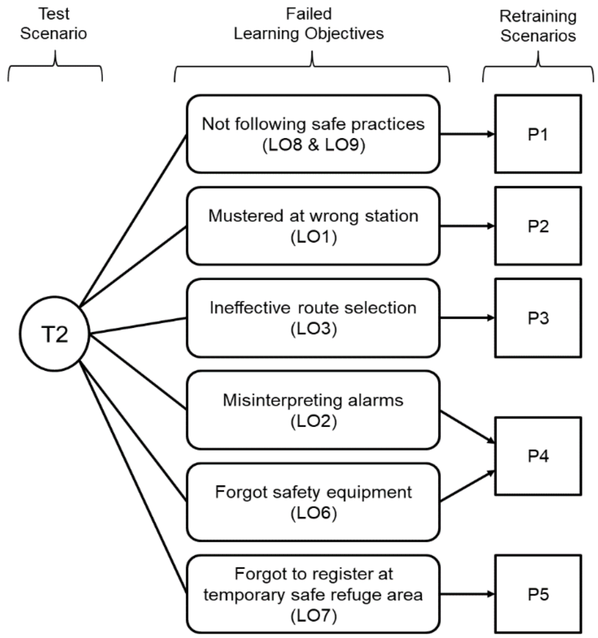


Figure 4.2: Example of the retraining matrix for test scenario T2

In this example of test scenario T2, a participant who failed to identify the alarm (LO2), and forgot how to register at the temporary safe refuge area (LO7), was required to complete one corrective training scenario focused on teaching the different alarms and the corresponding muster station for each alarm (P4), and another corrective training scenario that reinforced the importance of registering at your designated muster station to ensure all personnel onboard are accounted for during an emergency (P5). Once all prescribed retraining exercises were completed correctly, the participant was required to reattempt and pass test scenario T2 before moving on to the test scenario in module 3.

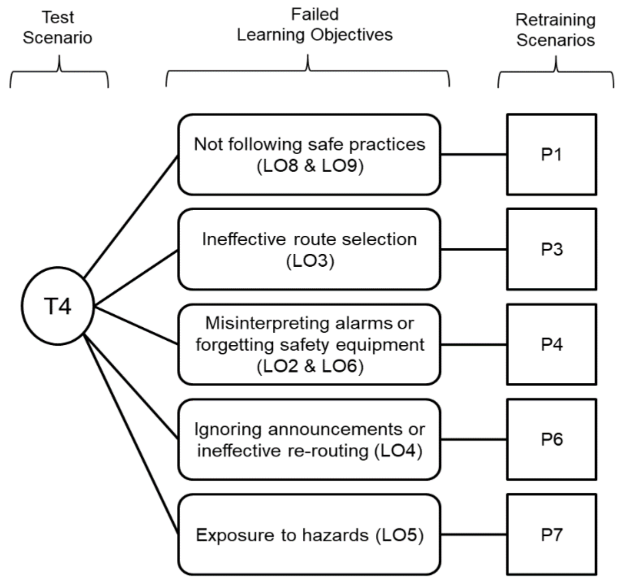


Figure 4.3: Example of the retraining matrix for T4, used for the final test scenario

As illustrated in Figure 4.3 for the final test scenario, a participant who failed to select the safest route (LO3), forgot to close fire doors (LO9), and encountered a smoke hazard (LO5), was required to complete three retraining scenarios. One corrective training scenario focused on teaching the available egress routes from the cabin to the muster station (P3). Another corrective training scenario reinforced the importance of keeping fire and water-tight doors closed (P1). The last corrective training scenario highlighted the importance of being aware of the surroundings and of avoiding exposure to hazards during emergencies (P7). Due to individual differences in the errors made in the test scenarios, participants received specific training scenarios to meet their individual needs.

4.6. Results & Discussion

To measure the retention of basic offshore emergency egress skills acquired using a virtual environment, the participants' performance in the skill acquisition phase was used as a benchmark to compare with the performance achieved after the retention interval. Several metrics were used to investigate retention and impact of retraining: 1) the overall competence demonstrated after the retention period, 2) the performance of each learning objective after the retention period, 3) the overall time spent retraining, and 4) the influence of the time lapsed on performance after the retention interval. This section presents the performance results as participants first encountered each test scenario. This section discusses which learning objectives were found to be more susceptible to degradation, and how quickly participants were able to return to competence following the retraining program.

4.6.1. Impact of Retention Period on Overall Competence Retention

To investigate the group's average competence after a period of 6 to 9-months, the final performance scores of the skill acquisition phase (Phase I) were compared with the first attempt performance scores of the retention phase (Phase II) for each of the four test scenarios. Table 4.3 shows the descriptive statistics for the performance in the skill acquisition and retention phases for all four test scenarios.

Table 4.3: Descriptive statistics of the performance scores for skill acquisition and retention phases

Test	n	Phase I: SBML Performance Scores (%) (Final attempt)					Phase II: Retention Performance Scores (%) (1st attempt)				
		Mean	St. Dev	Median	Min	Max	Mean	St. Dev	Median	Min	Max
T1	35	99.5	2.1	100	91.0	100	72.3	26.9	81.0	9.0	100
T2	35	100	-	100	100	100	83.7	18.0	91.0	35.0	100
T3	36	100	-	100	100	100	96.8	8.0	100	63.0	100
T4	34	97.3	4.3	100	89.0	100	95.7	13.4	100	29.0	100

Figure 4.4 provides a visual representation of the data in Table 4.3 using boxplots. The boxplots are grouped by the four test scenarios and experiment phases. Phase I data (skill acquisition) are denoted by ACQ, and phase II data (retention) are denoted by RET in the figure. Boxplots indicate the data distribution, including the median, first quartile, third quartile, minimum and maximum values of the data set, as well as outliers. The median is represented in the box, which is bounded by the first and third quartiles (interquartile range). The minimum and maximum values are represented by the whiskers. Outliers are represented as individual points and are defined as values outside the range of 1.5 times the first and third quartile.

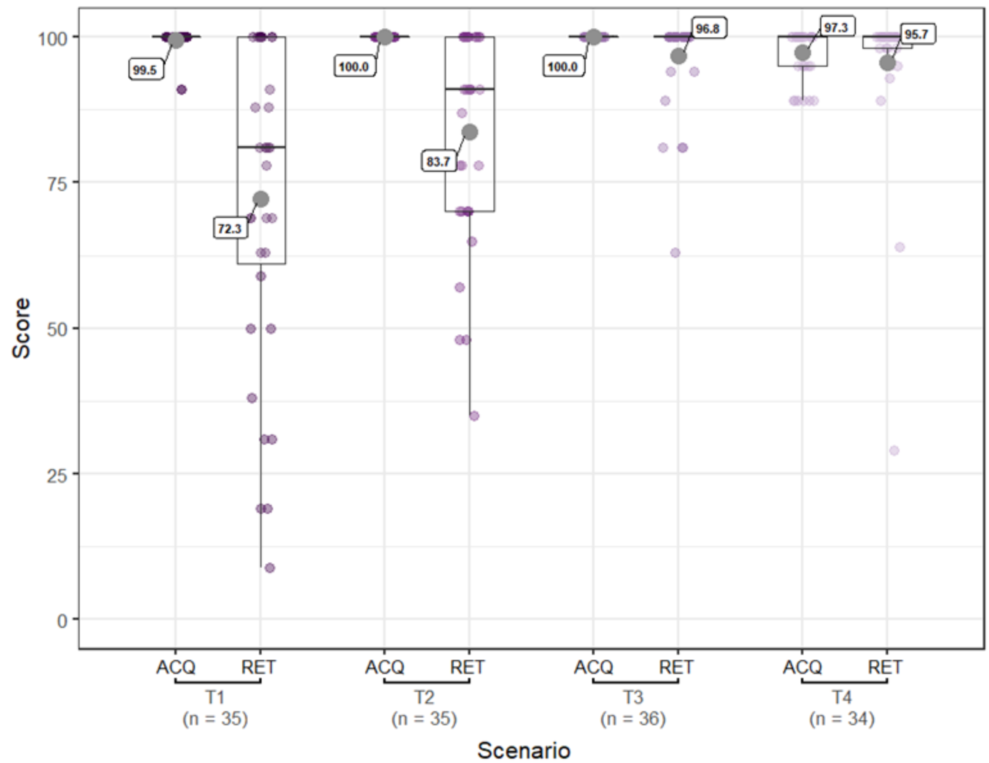


Figure 4.4: Boxplots of performance scores at the skill acquisition and retention phases for all test scenarios

As shown in Table 4.3 and Figure 4.4, the SBML training brought all participants to demonstrable competence at the end of the skill acquisition phase of the experiment. When 36 of the participants were reassessed after a period of 6 to 9-months, skill fade was observed. Only 4 participants (11%) were able to successfully complete all the test scenarios without making any errors. The average performance of participants drops from the skill acquisition phase in test scenarios T1 and T2 after the retention interval. Only 10 participants (28%) in T1, and 13 participants (36%) in T2, were successful in demonstrating competence in test scenarios T1 and T2, respectively. There appears to be little appreciable difference in average performance for test scenarios T3 and T4. Thirty-one participants

(86%) in T3, and 33 participants (92%) in T4 were successful in demonstrating competence in test scenarios T3 and T4, respectively, after the retention interval.

RStudio (version 3.5.0) software was used for statistical analysis (RStudio 2016). The data was tested for normality and found to be positively skewed, so non-parametric statistical tests were performed. The Wilcoxon signed-rank test is the non-parametric equivalent to a paired t-test and uses the median scores of two dependent samples (Corder and Foreman 2014). The Wilcoxon signed-rank test (using the Pratt method for pairs with ties) was used to compare the performance scores of the test scenarios (T1, T2, T3, and T4) before (pre-interval) and after the retention interval (post-interval). For each comparison, the statistical test (Z), p-value (p), and effect size (r) are reported.

The results showed significant differences between skill acquisition and retention phases. The output of the Wilcoxon-Pratt signed-rank indicated that the post-interval retention scores were statistically lower than the pre-interval acquisition scores, for three test scenarios, T1 ($Z = 4.67$, $p < .001$, $r = .79$), T2 ($Z = 4.55$, $p < .001$, $r = .77$), and T3 ($Z = 2.64$, $p = .008$, $r = .44$). No statistical differences between the acquisition and the retention scores were found for the final test scenario T4 ($Z = 0.05$, $p = .964$, $r = .008$).

These results indicate that participants had difficulties recalling the egress protocol, specifically the learning objectives that were tested in the first three scenarios (T1, T2, and T3). It also suggests that the combination of the retraining and exposure to the test scenarios helped the participants regain the competence required to correctly perform the final test scenario (T4). Further investigation into how participants performed when they first encountered each learning objective in the test scenarios provides more information on what skills were retained or lost during the 6 to 9-month period.

4.6.2. Performance by Learning Objective after Retention Period

The first time that the participants were tested on an individual learning objective after the retention interval is an important measure of how well the particular skill was retained for the learning objective. In the skill acquisition phase, participants were tested on an increasing number of learning objectives as they completed each additional module. In the retention phase, the learning objectives were again tested in a cascading format, each test scenario building on the previous scenario. In the first test scenario (T1 - Wayfinding), the retention of four learning objectives was assessed (LO1, LO3, LO8 and LO9). All three subsequent test scenarios tested these same learning objectives. The second test scenario (T2 – Muster drill), assessed the retention of three new learning objectives (LO2, LO6, and LO7). These learning objectives were tested again in the subsequent scenarios, T3 and T4. The third and fourth test scenarios assessed the retention of one more new learning objective each. The third test scenario (T3 – Blocked route) assessed the retention of learning objective LO4 for the first time in the retention phase. Learning objective LO4 was tested again in the final test scenario. The final test scenario (T4 - Emergency evacuation) assessed the retention of learning objective LO5 for the first time, as well as all the other learning objectives that had already been introduced in previous test scenarios.

Table 4.4 shows the percentage of participants who were successful at completing each learning objective for each of the test scenarios in the retention phase. The numbers in bold represent the first time the corresponding learning objective was assessed in the retention study.

Table 4.4: Percentage of participants who passed the learning objectives in each test scenario after the retention period

No.	Learning Objectives	Knowledge Type	Percentage of participants who passed at first attempt is shown in bold			
			T1	T2	T3	T4
LO1	Reach correct location	Spatial	**81%**	100%	97%	100%
LO2	Recognize alarm	Procedural	-	**92%**	100%	97%
LO3	Select safest egress route	Spatial	**42%**	94%	89%	92%
LO4	Re-route based on PA or if path blocked	Spatial	-	-	**92%**	92%
LO5	Avoid exposure to hazards	Procedural	-	-	-	**94%**
LO6	Take safety equipment	Procedural	-	**50%**	100%	97%
LO7	Register at the correct muster station	Procedural	-	**58%**	97%	97%
LO8	Avoid running	Procedural	**61%**	100%	100%	100%
LO9	Close all fire and watertight doors	Procedural	**86%**	94%	100%	100%

None of the nine learning objectives was successfully demonstrated by all 36 participants when first encountered in the test scenarios. All participants who were unsuccessful at completing a test scenario were re-trained using exercises that focused on the particular errors they made as prescribed by the adaptive training matrix (as described in section 3.4). Depending on the errors made, they received specific training scenarios to help improve their performance in subsequent attempts at the test scenarios.

In general, the retraining exercises were effective at returning participants to competence in specific skills. The skill retention of spatial (LO1, LO3 and LO4) and procedural (LO2, LO5, LO6, LO7, LO8, LO9) skills will be discussed separately by scenario.

4.6.2.1. Retention of Spatial Learning Objectives

The wayfinding test scenario (T1) focused on spatial knowledge and assessed participants on their ability to find their way around the platform (specifically testing spatial learning objectives LO1, and LO3). At the end of the skill acquisition phase, the participants were successful in all performance metrics. In the retention phase, not all participants retained the skills associated with the same learning objectives. Spatially, 81% of participants were able to recognize the correct muster location (LO1) and only 42% of participants were successful in following their egress route (LO3). In terms of the LRS model (Seigel and White 1975) this result suggests that participants remembered landmark knowledge, such as recognizing muster locations, better than route or survey knowledge of the environment after the retention period.

The muster drill test scenario (T2) retested participants' spatial knowledge of their designated muster locations (LO1) and the available egress routes from their cabin (LO3). The percentage of participants who were successful at the spatial aspects improved in this scenario: 100% of participants were able to recognize the correct muster location and 94% of participants were successful in following their egress route. This suggests that a combination of the testing that participants completed in the first scenario and the retraining received after the first scenario helped the majority of participants return to competence in route knowledge of the platform.

The final two test scenarios (T3 and T4) assessed participants on their survey knowledge of the platform by blocking known egress routes, requiring participants to reroute to find their muster stations. The blocked route test scenario (T3) assessed participants' ability to re-route in the event that their egress route was obstructed. This was the first-time participants were assessed on LO4 in the retention phase. In this scenario, 75% of participants selected the safest egress route and 8% of participants re-routed based on information from the PA. Some participants still experienced difficulties with route and survey knowledge; 11% of participants had trouble following their egress routes (they deviated from their routes) and 6% of participants had difficulties finding an alternate route when their path was disrupted, requiring them to re-route after encountering the blocked route.

The emergency test scenario (T4) assessed participants on their procedural knowledge to assess the situation, and on their survey knowledge of the platform to avoid hazardous egress routes and re-route effectively if their chosen egress route was obstructed. This scenario was the second time participants were assessed on LO4. In this scenario, 82% of participants took the safest route available for the situation. Another 9% of participants attempted to follow the safest route but had some difficulty and deviated at various points along the route. Only one participant followed the less optimal route, but re-routed effectively to avoid hazards by listening to the PA. Two participants (representing 6% of participants) followed an unsafe route and encountered the hazard (failing both spatial and procedural learning objectives).

4.6.2.2. Retention of Procedural Learning Objectives

The wayfinding test scenario (T1) focused on spatial knowledge, but also assessed participants on their ability to follow safety protocols on the platform (specifically testing procedural learning objectives LO8, and LO9). In this test scenario, 61% of participants remembered not to run on the platform (LO8) and 86% of participants remembered to close all the watertight doors (LO9).

The adaptive training matrix was used to retrain all participants who were unsuccessful at completing test scenarios (as described in section 3.4). The retraining for the procedural learning objectives appeared to correct participants' performance. For example, after the errors made in LO8 for T1, no one failed this learning objective in the three subsequent test scenarios.

The muster drill test scenario (T2) assessed participants on their understanding of alarms and basic muster procedures at their designated muster stations. The percentage of participants who were successful increased from scenario T1 to T2 for both the spatial (LO1, LO3) and procedural (LO8, LO9) elements that recurred in T2 (after being first assessed in T1). This may be a result of the re-training that took place after the first test scenario. However, there were still deficiencies in remembering procedural steps. The three procedural tasks (LO2, LO6, and LO7) that were first assessed in the muster drill scenario were forgotten by many participants. Some participants forgot that the alarm type dictated the muster location (8%), others forgot to take their personal protective equipment from their cabin (50%), and some forgot the mustering or unmuster procedures (42%). These skills were not practiced during scenario T1 or during the retraining associated with T1.

The blocked route test scenario (T3) did not assess any new procedural learning objectives. The emergency test scenario (T4) assessed participants on their procedural knowledge to assess the situation and avoid hazards (LO5) for the first time in the retention phase. The majority of participants did not make procedural errors in T3 and T4. Errors made by some individuals in these scenarios were in remembering safety equipment (LO6), registering at the TSR (LO7), and avoiding hazard exposure (LO5).

4.6.3. Time Spent Retraining

The retraining helped participants return quickly to competence. Table 4.5 provides a comparison between the mean time spent training in the skill acquisition phase and the mean time spent retraining in the retention assessment phase.

Table 4.5: Mean total time spent training and retraining by each group

Category	Time in minutes spent training	
	Skill Acquisition Mean (St. Dev)	Retention & Retraining Mean (St. Dev)
Habituation	16.7 (4.0)	7.9 (4.4) [a]
Tutorials	20.5 (8.9)	4.6 (3.8)
Practice Scenarios	45.6 (17.0)	15.9 (13.2)
Evaluation Scenarios	12.9 (4.3)	18.3 (6.3)
Total Training Time	95.6 (29.9)	46.9 (25.3) [a]

[a] Three participants' habituation scenario time was not recorded.

Overall, compared to the initial training, 47% less time was spent in the retention phase to return participants to demonstrable competence in egress skills. During the retraining, participants took less time reviewing tutorial material (4.6 minutes compared to 20.5 minutes) and training in AVERT (15.9 minutes compared to 45.6 minutes). However, participants spent more time demonstrating retention of competence in test scenarios (18.3 minutes compared to 12.9 minutes).

4.6.4. Influence of Retention Interval on Overall Competence Retention

The experiment was designed to evaluate retention after a 6-month interval (time lapsed). Due to logistical constraints, participants returned to complete the retention phase of the experiment after a period of 6 to 9-months. The average elapsed time between phases was 7.42 months (SD = 0.91 months). Participants were grouped based on the time that lapsed between phases (6, 7, 8 or 9 months) to determine if the difference in the retention interval impacted skill degradation. Five participants were assessed at 6 months, 16 participants were assessed at 7 months, 10 participants were assessed at 8 months, and 5 participants returned at 9 months. Of the four participants who were successful in all test scenarios, two completed the retention assessment at 6 months, and the remaining two participants completed the retention assessment at 7 months.

Due to the small sample size in months 6 and 9, the participants were grouped into two separate groups: group 1) time lapse of 6-7 months (n=21), and group 2) time lapse of 8-9 months (n=15). Figure 4.5 shows the distribution of performance by the participants in each retention interval for all test scenarios.

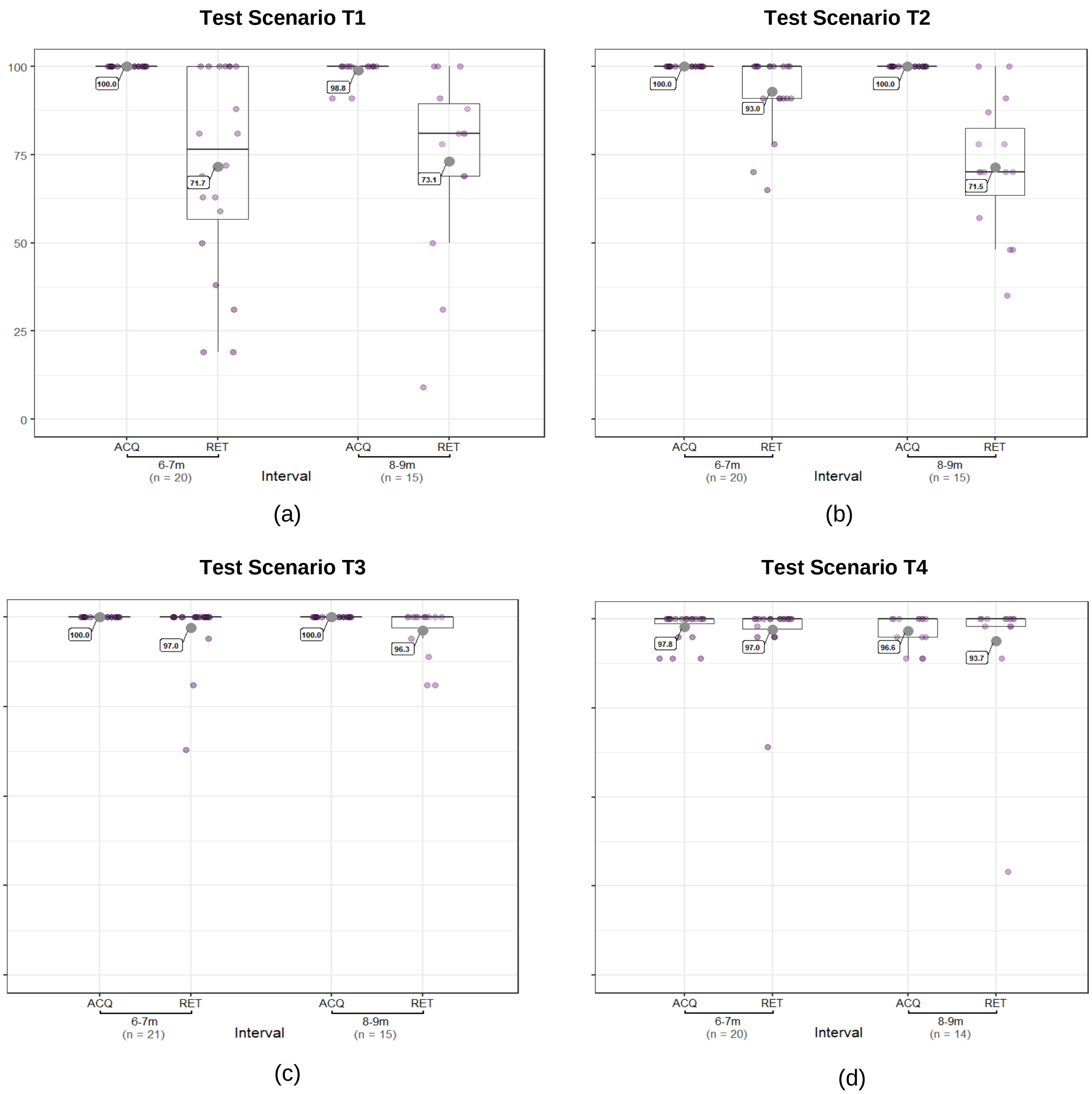


Figure 4.5 a, b, c, & d: Boxplots of the performance scores in all test scenarios for the skill acquisition and retention phases separated by retention interval

To compare the performance of the two groups' elapsed time, the Mann-Whitney *U*-test was performed in RStudio. Performing multiple tests on the same data inflates the type I error, therefore the Bonferroni procedure was used to correct $\alpha = 0.05$ for determining the significant differences between the samples (Corder and Foreman 2014). The adjusted significance level was determined to be $\alpha_B = 0.025$. For each comparison, the statistical test (U), p-value (p), and effect size (r) are reported.

No statistical difference was found for test scenarios T1 ($U = 150.5$, $p < .05$, $r = 0$), T3 ($U = 139.5$, $p = .416$, $r = .14$), and T4 ($U = 134.5$, $p = .833$, $r = .04$). A statistical difference in the performance between the groups was found for test scenario T2 ($U = 50.5$, $p = .0006$, $r = .58$). This result indicates that the difference in time that elapsed for the retention interval affected participants' performance in the second test scenario. Participants who completed the retention assessment at 8 and 9-months performed worse on T2 than those who completed the assessment at 6 and 7-months.

Tables 4.6-4.9 provide the distribution of spatial and procedural errors by learning objective for the four test scenarios for each group (6-7 months and 8-9 months). The majority of errors occurred in the first two test scenarios (T1 and T2). Spatial errors in test scenario T1 were equally distributed across all groups, regardless of the time lapse between skill acquisition and retention assessment. The majority of errors made in test scenario T2 were procedural errors (shown in Table 4.7), such as forgetting to take safety equipment (LO6) and difficulties remembering how to register at the designated muster station (LO7). This result indicates that the longer participants waited to be assessed, the more likely participants forgot these tasks. All other errors made throughout the retention phase were not found to be statistically different amongst the groups.

Table 4.6: Number of failed LOs after retention for T1 (by time lapsed in months)

No.	Learning Objectives	Number of participants who failed T1 after retention interval				
		6m	7m	8m	9m	Total Sample
		n=5	n=16	n=10	n=5	n=36
LO1	Reach correct location	**0**	**6**	**0**	**1**	**7**
LO2	Recognize alarm	-	-	-	-	-
LO3	Select safest egress route	**2**	**12**	**4**	**3**	**21**
LO4	Re-route based on PA or if path blocked	-	-	-	-	-
LO5	Avoid exposure to hazards	-	-	-	-	-
LO6	Take safety equipment	-	-	-	-	-
LO7	Register at the correct muster station	-	-	-	-	-
LO8	Avoid running	**0**	**7**	**4**	**2**	**14**
LO9	Close all fire and watertight doors	**0**	**2**	**2**	**1**	**5**

Table 4.7: Number of failed LOs after retention for T2 (by time lapsed in months)

No.	Learning Objectives	Number of participants who failed T1 after retention interval				
		6m	7m	8m	9m	Total Sample
		n=5	n=16	n=10	n=5	n=36
LO1	Reach correct location	0	0	0	0	0
LO2	Recognize alarm	**0**	**0**	**1**	**2**	**3**
LO3	Select safest egress route	0	1	1	1	2
LO4	Re-route based on PA or if path blocked	-	-	-	-	-
LO5	Avoid exposure to hazards	-	-	-	-	-
LO6	Take safety equipment	**2**	**6**	**5**	**5**	**18**
LO7	Register at the correct muster station	**0**	**3**	**7**	**5**	**15**
LO8	Avoid running	0	0	0	0	0
LO9	Close all fire and watertight doors	0	1	0	1	2

Table 4.8: Number of failed LOs after retention for T3 (by time lapsed in months)

No.	Learning Objectives	Number of participants who failed T1 after retention interval				
		6m	7m	8m	9m	Total Sample
		n=5	n=16	n=10	n=5	n=36
LO1	Reach correct location	0	1	0	0	1
LO2	Recognize alarm	0	0	0	0	0
LO3	Select safest egress route	0	1	2	1	4
LO4	Re-route based on PA or if path blocked	**0**	**1**	**1**	**1**	**3**
LO5	Avoid exposure to hazards	**-**	**-**	**-**	**-**	**-**
LO6	Take safety equipment	0	0	0	0	0
LO7	Register at the correct muster station	0	1	0	0	1
LO8	Avoid running	0	0	0	0	0
LO9	Close all fire and watertight doors	0	0	0	0	0

Table 4.9: Number of failed LOs after retention for T4 (by time lapsed in months)

No.	Learning Objectives	Number of participants who failed T1 after retention interval				
		6m	7m	8m	9m	Total Sample
		n=5	n=16	n=10	n=5	n=36
LO1	Reach correct location	0	0	0	0	0
LO2	Recognize alarm	0	0	1	0	1
LO3	Select safest egress route	0	1	1	1	3
LO4	Re-route based on PA or if path blocked	0	1	1	1	3
LO5	Avoid exposure to hazards	**0**	**1**	**1**	**0**	**2**
LO6	Take safety equipment	0	0	1	0	1
LO7	Register at the correct muster station	0	0	1	0	1
LO8	Avoid running	0	0	0	0	0
LO9	Close all fire and watertight doors	0	0	0	0	0

4.7. Conclusions

The results of this retention study show that emergency egress skills attained using a virtual environment are susceptible to skill decay over a period of 6 to 9-months. Although skill decay occurred, the adaptive retraining matrix employed in the study was successful in bringing all participants back to demonstrable competence at the end of the experiment. This section will discuss the answers to two questions: 1) what egress skills degraded after a period of 6 to 9-months? and 2) how effective was the VE-based retraining program at maintaining egress skills?

4.7.1. What Egress Skills Degraded After a Period of 6 to 9-Months?

Two indicators were used to understand the retention of egress skills acquired using a virtual environment: 1) the overall performance scores in each test scenario, and 2) the performance of participants in their first test attempt at each learning objective. The first indicator was useful in showing the initial skill fade in the first two test scenarios and provided less evidence of skill fade in the latter test scenarios. This result was interesting because the first two test scenarios were foundational scenarios that tested participants' basic knowledge of the platform and their understanding of the safety protocols in benign conditions. The latter test scenarios were more complex scenarios and assessed participants' survey knowledge of the platform and their ability to respond to emergency conditions, including blocked routes and hazards.

The first two test scenarios (T1 and T2) are the most important in terms of retention assessment because seven of the nine learning objectives were encountered for the first time in these test scenarios. The subsequent test scenarios (T3 and T4), although more complex emergency situations, built on the foundational learning objectives from early test scenarios and tested fewer new learning objectives. The results provided evidence that participants forgot foundational skills needed to perform in simple scenarios. Participants regained these skills and performed better in the more complex emergency test scenarios due to a combination of the exposure to the test scenarios and the corrective retraining they received.

The second indicator – participants’ performance in terms of learning objective – provided more practical information about the loss of specific egress skills (i.e. identifying which skills were most susceptible to skill fade). Participants’ performance in terms of learning objective showed that most of the participants (89%) did not retain the full requisite skill set over the study interval. It also identified which of the learning objectives were relatively more or less susceptible to skill fade, which is important in determining training interventions. This method is supported by other researchers, such as Atesok et al. (2016), who broke down orthopaedic surgery procedural skills into smaller components to investigate the decay of skills and identify how to improve long-term retention.

4.7.1.1. Spatial Skills

The learning objective that scored worst in terms of retention was remembering egress routes (LO3). The majority of participants failed to remember their egress routes when they first encountered this decision in the retention study (test scenario T1). Choosing the safest egress route also had the most persistent failures across test scenarios. Once the skills were forgotten, it took longer to retrain spatial skills than some of the procedural skills. This suggests that spatial competence needs relatively more training than the other skills, and that a shorter retraining interval is required to reduce spatial skill decay. There are some limitations of using a virtual setting to learn spatial skills when compared to the real environment. VEs often take longer to learn survey knowledge compared with the traditional maps or the real world for short-term exposures (Waller et al. 1998; Witmer et al. 2002; Darken and Peterson 2001). However, VEs are useful for longer-term exposures and for situations when the real-world environment is not easily accessible, which is the case for all offshore platforms.

4.7.1.2. Procedural Skills

Compliance with procedures was an issue for the first two test scenarios (T1 and T2) but was less so in the latter two test scenarios (T3 and T4). When the procedural learning objectives were first encountered, half the participants forgot to take their safety equipment, and about 40% did not follow the muster procedures, and about 40% forgot to refrain from running on the platform. As for retraining procedural skills, most participants only needed to be reminded of the protocols once in order to complete them successfully in the subsequent scenarios. In the cases where participants failed to complete procedural tasks, it is possible that the initial training did not provide adequate practice (or frequency of practice) for all participants to proceduralize the declarative knowledge related to these tasks. Since procedural knowledge takes time to develop, the retention assessment in this study may not have solely assessed participants' procedural knowledge, but also their ability to retrieve declarative knowledge (which is more susceptible to decay over time). Therefore, a shorter retraining interval would be beneficial for most participants. More frequent practice would help participants maintain declarative knowledge long enough to develop procedural skills.

4.7.2. How Effective was the Retraining at Maintaining Egress Skills?

This experiment demonstrated that the VE-based training can retrain participants who have lost egress skills after a period of 6 to 9-months. Only 11% of participants completed all test scenarios without errors and so did not require retraining. Eighty-nine percent of participants failed one or multiple test scenarios in this experiment and were required to complete corrective exercises. These participants regained foundational skills and performed better in the more complex emergency test scenarios due to a combination of the exposure to the test scenarios and the corrective retraining they received after the first test scenarios.

Most egress skills that were forgotten were quickly addressed with minimal retraining. A series of adaptive matrices were used to assign participants their corrective retraining scenarios based on their errors they made. Participants received a specific sequence of retraining scenarios based on the areas with which they had difficulty. For participants with spatial deficiencies, such as remembering their egress route (LO3), the spatial skills took the longest to relearn. Some participants required multiple iterations of the adaptive training matrix to re-acquire the forgotten spatial layout of the platform. For participants having difficulty with procedural skills, such as remembering not to run on the platform (LO8), and remembering to close fire doors (LO9), many participants only needed to be reminded of procedures in order to correct their behaviour. This finding is supported by other researchers (Hein et al., 2010) who found that minimal practice before performing a task can help improve performance after time has passed.

Overall, the adaptive retraining matrix was an integral part of accommodating different individual learning paths and an effective method to return all participants to demonstrable competence by the end of the experiment. This experiment demonstrated that VE-based training can help workers maintain their egress skills even if they have been absent from the platform for 6 months.

4.7.3. Limitations

The experiment was designed to combine the assessment of retention and retraining (a cascade format was used to measure the retention of targeted learning objectives for each test scenario and retraining was provided between test scenarios to address errors in prior tested learning objectives). This design was used to strike a balance between experimental control, ecological validity, and the practicality of training delivery. The authors recognize this design limits the conclusions that can be made from this work in the sense that it did not investigate only retention.

4.7.4. Future Work

Two questions arising from this research open new lines of inquiry. One question is: what is the most practical retention interval to maintain offshore egress skills? Industry standards require personnel to undergo safety training if they have been absent from the offshore platform for an extended period (e.g. 6 months or more). This experiment demonstrated that egress skills can be lost by the 6-month period. Therefore, a shorter retention interval is necessary to maintain egress skills, but what is the ideal frequency? Although this experiment did not examine the retention of skills between initial training and the 6-months interval, it would be interesting to look at the rate of decay of skills at shorter retention intervals (e.g. 1 month, 3 months). Investigating a shorter retention interval would help determine the appropriate recurrency training interval to maintain egress competence. This information would help inform the retention rate (e.g. predict the rate of decay) and estimate a suitable frequency for retraining.

The second question is: should recurrency training be competence driven as opposed to time based? In this experiment, the observed differences in individuals' performance (both in learning and retaining skills) indicate that a fixed or standardized retraining interval may not be the best solution. Rather than identifying a standardized retention interval, further investigation should focus on evaluating the on-demand feature of simulation-based training (i.e. customize the frequency of training). Sui et al. (2016) suggest using metric driven scheduling to train and maintain skills. VE training has the flexibility to provide people with practice, assessment, and corrective feedback at customized scheduling to meet individual needs of each learner. Competence-driven VE training could reshape how recurrency training is provided for offshore egress. VE training could help transition recurrency programs from fixed-interval training (i.e. only meeting the needs of some individuals) to tailored training for individuals (i.e. adaptive training to meet each individual's learning needs) by providing custom frequency of practice intervals to maintain skills. This would provide the groundwork for a competence-driven training frequency based on participants' needs.

4.8. References

Anderson, J (1982) Acquisition of cognitive skill. Psychol Rev 89:369–406

Arthur W, Bennett W, Stanush P, McNelly T (1998) Factors that influence skill decay and retention: a quantitative review and analysis. Hum Perform 11:57–101

Atesok K, Satava R, Heest A, Hogan M, Pedowitz R, Fu F, Sitnikov I, Marsh J, Hurwitz S (2016) Retention of skills after simulation-based training in orthopaedic surgery. J Am Acad Orthop Surg 24:505–514

Canadian Association of Petroleum Producers (2015) Atlantic Canada offshore petroleum industry standard practice for the training and qualification of offshore personnel. CAPP. https://www.capp.ca/~/media/capp/customer-portal/documents/272259.pdf Accessed 16 March 2015

Cohen J (1988) Statistical power analysis for the behavioral science, 2nd edn. Lawrence Erlbaum Associates, New Jersey

Cook D, Brydges R, Zendejas B, Hamstra S, Hatala R (2013) Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. Acad Med 88:1178–1186

Corder G, Foreman D (2014) Nonparametric statistics: a step-by-step approach. Wiley, New Jersey

Darken R, Peterson B (2001) Spatial orientation, wayfinding, and representation. In: Stanney K (ed) Handbook of Virtual Environment Technology, 1st edn. Lawrence Erlbaum Associates, New Jersey, pp 493-518

Doody K (2018) Evaluation of performance in emergency response scenarios: a virtual environment skill retention study. Master's thesis, Memorial University of Newfoundland

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191

Fitts P (1964) Perceptual-motor skill learning. In: Melton A (ed) Categories of human learning. Academic Press, New York, pp 243-285

Gallagher A, Ritter E, Champion H, Higgins G, Fried M, Moses G, Smith C, Satava R (2005) Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in the surgical skills training. Ann Surg 241: 364–372

Hein C, Owen H, Plummer J (2010) A training program for novice paramedics provides initial laryngeal mask airway insertion skill and improves skill retention at 6 months. Simul Healthc 5:33–39

International Association of Oil and Gas Procedures (2010) Evacuation, escape and rescue OGP risk assessment data directory. http://www.ogp.org.uk/pubs/434-19.pdf Accessed 16 March 2015

International Association of Drilling Contractors (2009) Health, safety and environmental case guidelines for mobile offshore drilling units. http://www.iadc.org/iadc-hse-case-guidelines/ Accessed 16 March 2015

International Maritime Organization (2001) The standards of training, certification and watchkeeping code (STCW code) convention for seafarers. International Maritime Organization, London.

Kim J, Ritter F, Koubek R (2013) An integrated theory for improved skill acquisition and retention in the three stages of learning. Theor Issues Ergon Sci 14:22–37

Kobes M, Helsloot I, de Vries B, 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, Procedia Eng 3:37–51.

McGaghie W, Issenberg S, Barsuk J, Wayne D (2014) A critical review of simulation-based mastery learning with translational outcomes. Med Educ 48:375–385

Rasmussen J (1986) Information processing and human-machine interaction: an approach to cognitive engineering. Elsevier, New York

RStudio Team (2016) RStudio: integrated development for R. RStudio Inc. http://www.rstudio.com/ Accessed 26 June 2019

Sanli E, Carnahan H (2017) Long-term retention of skills in multi-day training contexts: a review of the literature. Int J Ind Ergonom 66:10–17

Seigel A, White S (1975) The development of spatial representations of large-scale environments. Adv Child Dev Behav 10:9–55

Smith J (2015) The effect of virtual environment training on participant competence and learning in offshore emergency egress scenarios. Master's thesis, Memorial University of Newfoundland

Smith J, Doody K, Veitch B (2018) Experimental investigation of the retention of emergency egress competence acquired in a virtual environment. Proc Int Conf Appl Hum Factors Ergonom 785:43–53

Smith J, Veitch B (2018) A better way to train personnel to be safe in emergencies. ASCE-ASME J Risk Uncertain Eng Sys Part B. https://doi.org/10.1115/1.4040660

Sui K, Best B, Kim J, Oleynikov D, Ritter F (2016) Adaptive virtual reality training to optimize military medical skills acquisition and retention. Mil Med 181:214–220

Taylor H, Brunye T, Taylor S (2008) Spatial mental representation: implications for navigation system design. Rev Hum Factors Ergonom 4:1–40

Transport Canada (2007) Practical skills training for marine engineers (TP 13720E). Transport Canada. http://www.tc.gc.ca/eng/marinesafety/mpsp-training-examination-certification-publications-125.htm Accessed 26 June 2019

McCallum M et al. (2000) A method for developing mariner assessment (CG-D-17-00). US Coast Guard Research and Development Center. https://apps.dtic.mil/dtic/tr/fulltext/u2/a384215.pdf Accessed 26 June 2019

VanLehn K (1996) Cognitive skill acquisition. Annu Rev Psychol 47:513–539

Waller D, Hunt E, Knapp D (1998) The transfer of spatial knowledge in virtual environment training. Presence- Teleop Virt 7:129–143

Wickens C, Holland J, Banbury S, Parasuraman R (2013) Memory and training. In: Campanella C (ed) Engineering psychology and human performance, 4th edn, Pearson Education Inc, New Jersey, pp 197-244

Witmer B, Sadowski W, Finkelstein N (2002) VE-based training strategies for acquiring survey knowledge. Presence- Teleop Virt 11:1–18

5. PREPARING FOR SKILL TRANSFER: A DECISION TREE TOOL FOR CURRICULUM DESIGN AND ASSESSMENT OF VIRTUAL OFFSHORE EMERGENCY EGRESS TRAINING

Jennifer Smith*, Mashrura Musharraf, Allison Blundon, and Brian Veitch

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

St. John's, NL, Canada, A1B 3X5

*Correspondence author: jennifersmith@mun.ca

5.1. Co-authorship Statement

A version of this manuscript has been submitted for publication in the International Journal of Training Research and is still under review. Author Jennifer Smith led the writing of this manuscript. Allison Blundon led the human reliability analysis experiment and shared the data. Dr. Mashrura Musharraf and Dr. Brian Veitch provided guidance, editorial changes, and recommendations for improvements to several drafts of this paper.

5.2. Abstract

To prepare personnel for offshore emergencies, safety training should focus on transferability. Virtual environment (VE) training is designed to support the transfer of acquired egress skills to novel offshore emergencies. Decision trees (DT) are useful tools to evaluate the transfer of training from both an individual learner and a systemic perspective. DTs use human performance data collected during VE training to model participants' behavioural patterns. An ideal decision strategy is used as a benchmark to

compare the correctness of the participants' data-informed behavioural patterns. Employing the diagnostic and predictive capabilities of DTs can indicate when a person is capable of responding to a wide variety of emergencies using inference. VE training efficacy is assessed by modeling participants' trees throughout the program as an indication of when the right amount of training has been achieved for each individual, or if further training is required. This study uses DTs as curriculum design and assessment tools to determine if the training curriculum adequately prepares participants to transfer their egress skills to new emergencies in the same virtual setting.

5.3. Introduction

Mandatory offshore safety training has predominantly been participatory-based and the resulting certification often represents nominal competence. To better prepare the offshore workforce for emergencies, operators and regulators need to pursue evidence-based safety training to guarantee workers achieve demonstrable competence. This gap in offshore training can be addressed by combining simulation technology with a well-designed, pedagogically informed, and data driven training program. Training simulators, such as virtual environments (VE), allow workers to rehearse safety protocols in various emergency conditions and provide formative feedback to help personnel learn and improve their performance. Simulation and VE technology collect performance data during the training that can provide valuable insight into the quality of learning that takes place and the efficacy of the training program as a whole. Training simulators are widely used for maritime education, however, few studies have considered the pedagogical aspects of simulation-based training (Sellberg et al., 2017). This paper draws on pedagogical frameworks and data-mining methodology to provide empirical and modeling evidence to inform the offshore and maritime industries on how to deliver training and how to assess trainee performance using VE technology.

To optimize VE training for offshore egress, regulators must determine how comprehensive, yet focused virtual offshore egress training should be to prepare personnel for the multitude of emergencies that could arise. Virtual practice exercises should match the conditions the workers are expected to experience in real situations. However, it is impracticable to rehearse for all possible situations. As a guiding pedagogical principle, VE technology should be designed to support the transferability of training specifically, the application, generalization, and maintenance of knowledge and skills learned in a training context to new situations (Blume et al., 2010). Therefore, VE training and supporting tools should assess when the trainee has achieved competence and is sufficiently equipped to apply their skills to new situations within the context of the training. Many factors influence the transfer of training, Grossman and Salas (2011) outlined three factors that are most relevant to training organizations:

(1) Trainee characteristics such as the learner's cognitive ability, self-efficacy, motivation, and their perceived utility of training;
(2) Training design such as behavioural modeling, error management, and realism of training environment; and
(3) The work environment such as transfer climate as well as the opportunity and support from management to allow workers to apply their training.

This paper focuses on factors that affect the training design, specifically the use of behavioural modeling to inform the design and assessment of VE training transfer. Decision trees (DT) are useful behavioural modeling tools to evaluate the efficacy of training transfer from both the individual and systemic perspectives. At an individual level, DTs can be used to inform how much training is needed to achieve competence for each learner. Musharraf et al. (2018) demonstrated the utility of DTs for identifying trainees' strategies and recommended ways to use DTs to assess individual learning as well as pedagogical approaches. The aim of Musharraf's et al. (2018) work was to understand how people made decisions in emergency egress in order to improve training and to create artificially intelligent agents with similar behaviours (Musharraf et al., 2018). DTs provide a visual representation of participants' decision-making strategies in the context of choosing the safest egress routes in emergencies. Decision tree modeling was selected for its visual simplicity and diagnostic capabilities when dealing with sparse data and categorical variables (Duffy, 2009; Musharraf et al., 2018). DTs employ supervised learning, which requires a repository of attributes of solved problems to draw inferences. From the individual learner perspective, the supervised learning mechanism of DTs help inform learning analytics by developing generalized decision rules from each participant's behavioural data collected during the scenarios in the VE training. At a systemic training level, the collection of DTs developed from the participants' data throughout the VE training help to identify emerging patterns of successful or unsuccessful behaviours. Musharraf et al. (2018) suggested that an effective training curriculum would result in the convergence of participant's DTs to strategies that lead to success and that systemic exceptions indicate gaps in the training approach itself.

This paper uses decision trees for their diagnostic and predictive capabilities to determine if a virtual offshore platform-training program has adequately prepared learners to transfer their egress skills to new emergencies in the same virtual setting. In particular, the research aims to answer the following questions:

(1) Can DTs help determine when participants are sufficiently prepared for training transfer at an individual and a systemic level?

(2) Can DTs show the development of credible heuristics throughout the VE training? and

(3) Can DTs predict participants' performance in new situations?

This study demonstrates the use of DTs as data-informed curriculum design and assessment tools. The scope of the research involves a longitudinal experiment using the AVERT simulator as a human behaviour laboratory. The three-phased experiment collected performance data during the skill acquisition (phase 1), retention and retraining (phase 2), and transfer of training to new emergencies (phase 3). The context of the study was to teach naïve participants the necessary egress skills to evacuate a virtual oil platform during an emergency. The experiment first taught basic egress skills, then assessed skill retention after a 6 to 9-month period, and subsequently tested the transfer of egress skills to new emergencies. The novel emergency scenarios differed from the training exercises by varying conditions such as the participants' proximity to hazards, their familiarity with the scenario's starting locations on the platform, and the availability of information (or lack thereof) about the situation. The AVERT simulator collected participants behavioural data in the training and testing scenarios throughout each phase of the experiment.

Portions of the participants’ performance data were used to iteratively create DTs for each phase of the experiment. Multiple DTs were developed for each participant to determine if the participant’s strategies changed from the skill acquisition, retention, and transfer phases. This process was used to see how people develop and modify their route strategies or heuristics throughout training. An ideal decision strategy for responding to offshore egress emergencies was defined based on the results from Smith et al. (2017). The ideal DT depicted in Figure 5.1 was created to guide the development of the curriculum and used as a benchmark to assess the correctness of the participants’ DT behavioural patterns. As shown in Figure 5.1, the expected strategy was for the participant was to listen to the public address (PA) announcement for the safest route and choose a route accordingly. If the PA indicated that the primary route was the safest option, then the participant would follow the primary route. Alternatively, if the PA indicated that the secondary route was the safest option then the participant would follow the secondary route. When no route direction was provided in the PA, the participant was expected to follow their preferred route. The training curriculum was designed to encourage participants to develop the intended decision strategy.

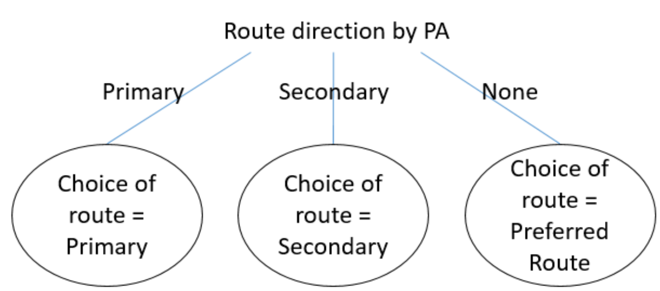


Figure 5.1: Curriculum Design – Ideal decision tree for emergency egress

The ideal decision strategy was used as a benchmark to compare the participants' behavioural patterns identified by their DTs to check their competence. Figure 5.2 summarizes the DT comparison process and the expected outcomes from the comparison. This comparison determined whether the amount of training the participants received was sufficient to prepare them for emergencies. The compilation of DTs developed throughout the VE training program were used to identify emerging group patterns and to evaluate the efficacy of the training curriculum. The final DTs at the end of the VE training were used to predict the transfer of skills to new emergency scenarios (e.g. how people will perform in new situations).

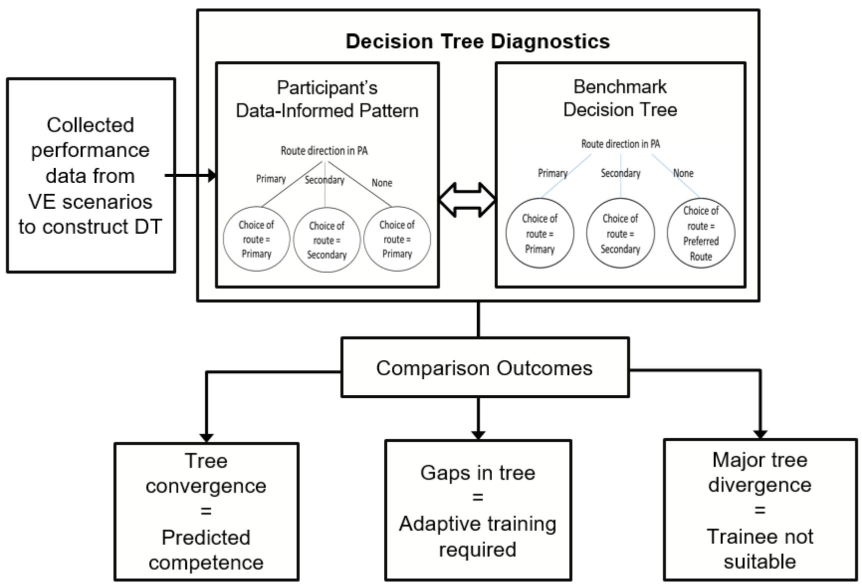


Figure 5.2: Process for comparing participants' route strategies with ideal DT

Section 2 of the paper describes the longitudinal study and the DT development. Section 3 describes the DT curriculum design and assessment methodology. Section 4 presents and discusses the results. Section 5 summarizes and concludes the findings.

5.4. Background

5.4.1. Longitudinal Pedagogical Experiment

The longitudinal experiment was conducted in 3 phases to investigate the skill acquisition, retention, and transfer of egress skills taught using a virtual offshore platform. The experiment first taught participants basic egress skills in phase 1, then assessed their skill retention after a 6 to 9-month period in phase 2, and subsequently tested their transfer of egress skills to novel emergencies in phase 3. Figure 5.3 depicts the three phases of the experiment.

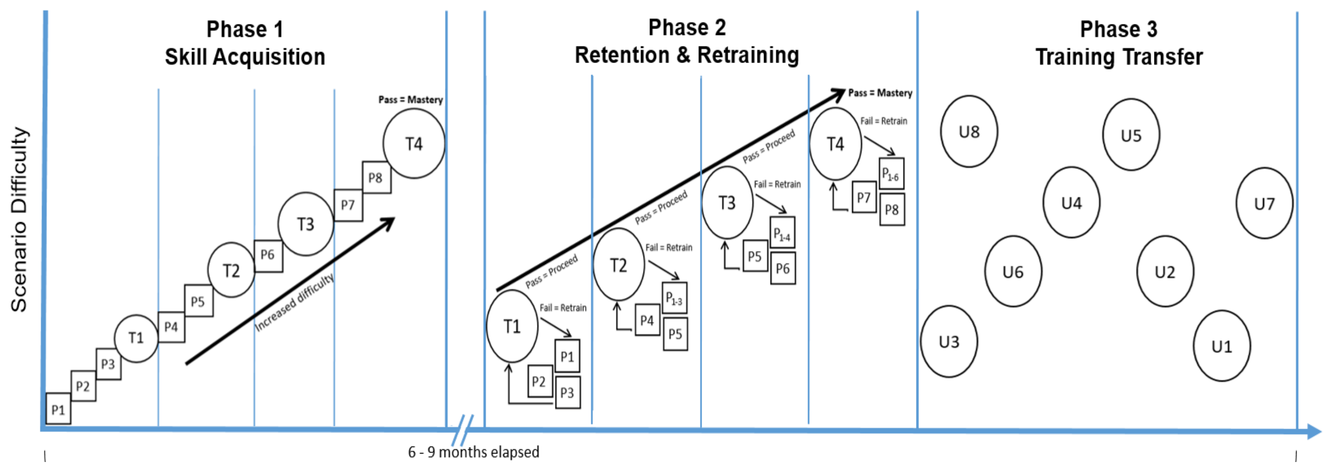


Figure 5.3: Longitudinal Experiment Timeline

The first phase of the experiment used a simulation-based mastery learning (SBML) pedagogical approach to teach naïve participants the necessary skills to evacuate safely (McGaghie et al., 2014; Smith & Veitch, 2018). The core learning objectives included familiarity with the platform layout, emergency alarms, egress routes, safety protocols, and mustering procedures. Participants were required to repeat the practice and testing scenarios until they demonstrated competence in all performance criteria. The practice and test scenarios are denoted in Figure 5.3 as P1-8 and T1-4, respectively. The scenarios involved participants practicing their egress routes in varying conditions from muster drills (T1 & T2) to emergency evacuation (T3 & T4). Table 5.1 provides a detailed description of the four test scenarios. Full details of the skill acquisition phase, including the training curriculum and performance metrics are described in Smith & Veitch (2018).

In the second phase of the experiment, participants returned after the retention interval and were tested on their ability to respond to the same egress scenarios (T1-4). Participants who had trouble remembering the egress procedures were provided retraining on deficient skills (e.g. practice scenarios P1-8). Participants were required to complete retraining for any deficiencies before moving on to the third phase. Full details of the retention phase are described in Smith, Doody & Veitch (2019).

The third phase of the experiment involved testing participants on emergency scenarios that differed from the training exercises to which they were exposed in the acquisition and retention phases. As depicted in Figure 5.3, after completing phase 2 of the experiment, participants performed the transfer scenarios in a random order. The emergency scenarios in phase 3 differed from the training exercises in the earlier phases by varying conditions such as the participants' proximity to hazards, their familiarity of the scenario's starting locations on the platform, and the availability of information (or lack thereof) about the situation. Phase 2 and 3 are explained in more detail in Section 5.5.

Table 5.1: Description of the test scenarios (after Smith, Doody, & Veitch, 2019)

Test Scenario	Scenario Description
T1 Wayfinding Drill	This scenario assessed the participants' spatial knowledge of the platform. Participants were asked to meet their supervisor at their assigned lifeboat station by following their primary or secondary egress routes.
T2 Muster Drill	This scenario assessed the participants' understanding of alarms and muster procedures. Participants were tasked with responding to a muster drill (General Platform Alarm). During this alarm, all personnel were required to collect their safety equipment and muster at their primary muster station.
T3 Blocked Route	This scenario assessed the participants' ability to deal with obstructions to their planned egress route. Participants were required to respond to the alarm, listen to the announcement, and follow the muster procedures. The PA announcements provided information to help the participants select the most effective route.
T4 Emergency	This scenario assessed the participants' ability to avoid hazards and follow the safest available route to their lifeboat station. Participants were tasked with responding to an emergency involving a General Platform Alarm (GPA) due to fire in the galley. The fire compromised the muster station with smoke and the situation escalated to a Prepare to Abandon Platform Alarm (PAPA). Initially all personnel were required to go to the muster station, but were forced to re-route to the lifeboat station because of the compromised muster station.

5.4.2. Decision Tree Development for Phase 1:

Decision tree models depicted participants' decision-making in selecting the safest egress route during an emergency. The participants' performance data recorded during the practice scenarios in the VE training (phase 1 of the experiment) was separated into training and testing datasets used to develop and verify generalized decision rules. The training dataset, which consisted of a knowledge base (KB) was used to 'train' the decision tree algorithm. The testing dataset was put aside to check the accuracy of the trees to predict the participants' route choice in future scenarios.

An individual KB was developed for each participant by tabulating the participant's successful performance in the practice scenarios and storing it in a two-dimensional matrix. Each row in the KB contained the different programmed attribute values and the corresponding route choice taken by the participants in the scenarios. Table 5.2 provides a list of attributes and the possible values. Table 5.3 provides an example of the KB for a sample participant after finishing all of the training in phase 1. Each row in the KB contains the different attribute values for the scenario and the corresponding route choice. The KB was a single data file for each participant.

Table 5.2: Description of scenario attributes (Smith et al., 2017)

Attributes	Possible Values
End Location	Muster, Lifeboat
Alarm type	None, GPA, PAPA
Route directed by PA	None, 1st, 2nd
Hazard presence	No, Yes
Obstructed route	None, 1st, 2nd
Previous route selected	1st, 2nd

Musharraf et al.'s (2018) decision tree algorithm was applied to the KBs to identify participants' egress route selection strategies. During the decision tree induction, the algorithm (employing the ID3 engine) iteratively classified data using the attribute that has the highest information gain. This process was repeated until no attributes were left for classification, or the dataset was empty, or data in each group belonged to the same class and no further classification was needed. Methodology for DT development is described in detail in Musharraf et al. (2018) and Smith et al. (2017).

Table 5.3: Sample participant knowledge base (KB) developed from data in phase 1

Scenario	Attributes						Route choice
	End Location	Alarm	Route by PA	Hazard	Obstructed Route	Previous Route	
P1	Muster	None	1st	No	None	N/A	Primary
P3 (F1)	Lifeboat	None	1st	No	None	1st	Primary
P3 (F2)	Muster	None	2nd	No	None	1st	Secondary
T1	Lifeboat	None	None	No	None	2nd	Secondary
P4	Muster	GPA	1st	No	None	2nd	Primary
P5	Muster	GPA	None	No	None	1st	Primary
T2	Muster	GPA	1st	No	None	1st	Primary
P6	Lifeboat	PAPA	1st	No	2nd	1st	Primary
T3 (F1)	Muster	GPA	None	No	None	1st	Primary
T3 (F2)	Muster	GPA	2nd	No	1st	1st	Secondary
P7	Lifeboat	PAPA	2nd	Yes	1st	2nd	Secondary
P8 (F1)	Muster	GPA	1st	Yes	2nd	2nd	Primary
P8 (F2)	Lifeboat	PAPA	1st	Yes	2nd	2nd	Primary
T4 (F1)	Muster	GPA	None	Yes	None	1st	Primary
T4 (F2)	Lifeboat	GPA	2nd	Yes	1st	1st	Secondary
T4 (F3)	Lifeboat	PAPA	2nd	Yes	1st	1st	Secondary

The DTs provided a depiction of the participants' understanding of emergency egress. Figure 5.4 shows the strategy of the sample participant. This participant's strategy was to listen to the public address (PA) announcement for information regarding the safest route direction and choose a route accordingly. When no route direction was provided in the PA, the participant followed the primary or secondary route based on their intended end location. If they were required to go to the muster station, then they would take the primary route, and if they were required to go to the lifeboat station, then they would take their secondary route.

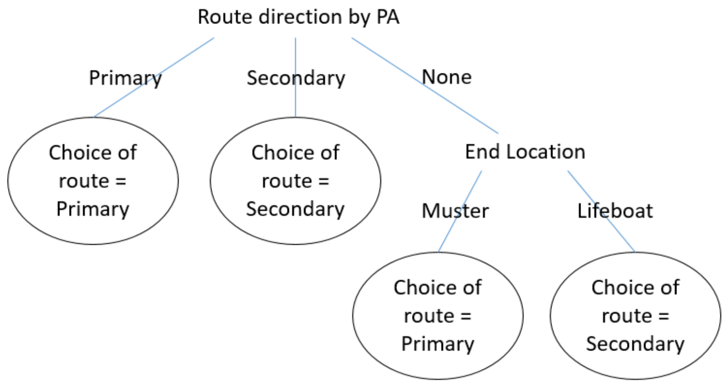


Figure 5.4: Sample participant DT developed from the KB shown in Table 5.3

5.5. Methods

This paper focuses on the retraining and transfer results from phase 2 and 3 of the longitudinal experiment. The goal was to use data from phase 2 of the experiment to iteratively model participants' decision strategies in the form of DTs and use these DTs to predict training transfer in phase 3. This section describes the details of phase 2 and 3 of the experiment including the participants, AVERT simulator, experimental design, updating the decision trees, and application of the decision trees.

5.5.1. Participants

Participants from the skill acquisition phase (n = 55) were invited to return after a period of six months to participate in the retention assessment and transfer phases of the experiment. Thirty-eight participants completed the retention and transfer phases. Twenty-eight participants were male and ten participants were female. Participants ranged in age from 19 to 54 years (M = 28 years, SD ±8.7 years).

5.5.2. AVERT Simulator

All phases of the experiment used the same emergency preparedness training simulator called AVERT. AVERT is a desktop virtual environment that allows participants to interact with the virtual offshore platform using a gamepad controller (Xbox). The virtual environment depicts a realistic representation of an offshore Floating Production Storage and Offloading (FPSO) vessel (Smith et al., 2019). Participants can move onboard the FPSO by controlling a first-person perspective avatar of an offshore worker. AVERT was designed as a human behaviour virtual laboratory to train people in basic offshore emergency duties such as how to navigate the platform and muster at their designated stations during an emergency.

5.5.3. Experimental Design

5.5.3.1. Phase 2: Retention and Retraining

The objective of phase 2 of the experiment was to investigate the long-term retention of emergency egress competence obtained using a virtual offshore platform and determine the efficacy of the VE-based retraining on maintaining egress skills. In phase 2, the participants from phase 1 returned and were tested on the same test scenarios from phase 1 to assess their retention of egress skills. If participants were successful at demonstrating competence in the test scenario, they advanced to the next test scenario. If participants forgot skills or made errors in performing the task in the test scenarios, then they were required to complete retraining scenarios specific to the errors they made. Participants repeated the test scenarios and practiced the retraining scenarios until they reached competence in all tasks. The retraining dataset served an additional purpose to iteratively model participants' heuristics and identify emerging group behaviours to inform the efficacy of the training curriculum.

5.5.3.2. Phase 3: Transfer Scenarios

Phase 3 of the experiment was designed to collect quantitative human performance data for human reliability analysis (HRA) in emergencies in a virtual environment using Bayesian network (BN) modeling (Musharraf et al., 2019). As an alternative to expert judgement VEs can be used to collect data for human reliability assessment (HRA), a technique used to predict how people would respond to situations (Musharraf et al., 2014). The HRA experiment used the same participants and AVERT simulator as phases 1 and 2 to investigate participants' sensitivity to different scenario attributes. Consequently, the HRA dataset served an additional purpose to inform the transferability of the participants' training to new emergencies. This HRA dataset was repurposed to perform the decision tree analysis retroactively. A two-state, three-factor experiment in AVERT was designed to investigate the interdependencies amongst the scenario attributes and the participants' task performance in the emergencies. Three attributes were varied, including: (1) the amount of information (or lack thereof) provided over the PA announcements, (2) the participants' proximity to hazards during the emergency, and (3) the participants' familiarity to the starting locations on the platform. Each of the attributes were assigned a high and low level. Eight scenarios were developed to incorporate all the high and low level combinations of the attributes (as shown in Table 5.4).

Table 5.4: Phase 3 Scenarios for AVERT (after Musharraf et al., 2019)

#	Set	Quality of Communication		Proximity to Hazards		Situation Familiarity	
U1	1	+	Clear PA with all relevant info	+	Hazard not accessible	+	Starts in cabin
U2		-	Poor PA that lacked info	+	Hazard not accessible	+	Starts in cabin
U3		+	Clear PA with all relevant info	-	Hazard blocking route	+	Starts in cabin
U4		-	Poor PA that lacked info	-	Hazard blocking route	+	Starts in cabin
U5	2	+	Clear PA with all relevant info	+	Hazard not accessible	-	Starts in Bridge
U6		-	Poor PA that lacked info	+	Hazard not accessible	-	Starts in Bridge
U7		+	Clear PA with all relevant info	-	Hazard blocking route	-	Starts in Bridge
U8		-	Poor PA that lacked info	-	Hazard blocking route	-	Starts in Bridge

Varying the high and low values of the attributes affected the scenario difficulty. The scenario difficulty is related to the transfer proximity as defined by Barnett and Ceci (2002). Training transfer is considered near transfer when the test setting is very similar to the knowledge covered in the training and far transfer occurs when the test setting is very different from the training (Ford et al., 2018). For example, in Table 5.4, the scenarios are listed in increasing order of difficulty by design, (scenario U1 being the least difficult or a near transfer setting and scenario U8 being the most difficult or a far transfer setting from the training). The first set of four scenarios (U1-4) had the same starting location that participants had trained for in the skill acquisition phase. The second set of four scenarios (U5-8) had a new starting location, different from where participants had practiced, but a location they had seen before.

Scenarios U1 and U2 closely resembled the conditions of the skill acquisition and retention test scenarios. The scenarios gradually increased in difficulty by removing information from the PA announcement and adding hazards to block viable routes. For example, scenario U3 assessed the participants' egress skills associated with listening to the PA and avoiding obstructed routes. Scenarios U7 and U8 were the most different from the conditions of the training and retention scenarios. For example, scenario U8 assessed the participants' ability to respond to a situation where they started in an unfamiliar location, were provided little information from the PA, and had to manage hazards blocking their egress path. Table 5.5 shows the context for each of the transfer test scenarios.

Table 5.5: Description of the transfer scenarios (after Blundon, 2019)

Test Scenario	Scenario Description
U1	Starting in their cabin, participants had to respond to a GPA alarm followed by a PA announcement about reports of smoke in the turret. All personnel were required to go to their muster station by following either the primary or secondary egress routes.
U2	Starting in their cabin, participants had to respond to a GPA alarm followed by a PA announcement with no information about the situation. Participants were unaware of an explosion and fire in the process module. All personnel were required to go to their muster station by following their primary or secondary egress routes.
U3	Starting in their cabin, participants had to respond to a GPA alarm followed by a PA announcement about a gas leak in the external stairway of the accommodation block near B-deck. All personnel were required to go to their muster station by following their primary route.
U4	Starting in their cabin, participants had to respond to a GPA alarm followed by a PA announcement with no information about the situation. All personnel were required to go to their muster station. Participants were unaware of a fire in the stairwell on C-deck that blocked the primary egress route. Participants had to re-route from the obstructed route and take their secondary route to the muster station.
U5	Starting in the bridge of the vessel, the GPA alarm sounded followed by a PA announcement about a gas leak in the wellhead bay. Participants were required to go to their muster station by following their primary or secondary egress route. Once at

	the muster station, the situation escalated to a PAPA alarm. Participants were required to muster at their lifeboat stations and don their immersion suit.
U6	Starting in the bridge of the vessel, the GPA alarm sounded followed by a PA announcement with no information about the situation. Participants were required to go to their muster station by following their primary or secondary egress route. Participants were unaware of an explosion in the engine room. Once at the muster station, the situation escalated to a PAPA alarm. Participants were required to muster at their lifeboat stations and don their immersion suit.
U7	Starting in the bridge of the vessel, the GPA alarm sounded followed by a PA announcement about smoke in office space on D-Deck and that the internal stairs were filled with smoke. Participants were required to go to their muster station by following their secondary egress route. Once at the muster station, the situation escalated to a PAPA alarm. Participants were required to muster at their lifeboat stations and don their immersion suit.
U8	Starting in the bridge of the vessel, the GPA alarm sounded followed by a PA announcement with no information about the situation. All personnel were required to go to their muster station. Participants were unaware of smoke and fire on external stairway of C-deck that activated the deluge system and blocked the secondary egress route. Participants were required to re-route from the obstructed route and take their primary route to the muster station. Once at the muster station, the situation escalated to a PAPA alarm. Participants were required to muster at their lifeboat stations and don their immersion suit.

5.5.4. Decision Tree Modeling for Phases 2 & 3

A 3-step process was used to develop and validate the DTs that involved: (1) updating the existing KB with retraining data, (2) using the decision tree algorithm to form new DTs, and (3) testing the classification accuracy of the DTs against the retention test scenarios. Each step is described.

5.5.4.1. Step 1: Iteratively Update the KB with Retraining Data

Participants completed the same test scenario from phase 1 in phase 2 after the retention interval. Thus, the KB was gradually updated with new information as participants completed the retention and retraining. Throughout phase 2, it was assumed that participants remembered what they had learned 6 months prior. If participants showed evidence of forgetting by making mistakes in a test scenario, they were retrained and their KB was updated accordingly. Only the participants' successful attempts at the scenarios were stored in the new KB. The phase 2 entries replaced the participant's past attempts at the scenarios from phase 1 (as depicted in Figure 5.5). Other records in the KB remained unchanged as they were assumed to be remembered by the participant.

Phase 1 Knowledge Base

	Attributes						Action
Scenario	End Location	Alarm	Route by PA	Hazard	Obstructed Route	Previous Route	Egress route Choice
T1	Lifeboat	None	None	No	None	2nd	Secondary
P4	Muster	GPA	1st	No	None	2nd	Primary
P5	Muster	GPA	None	No	None	1st	Primary
T2	Muster	GPA	1st	No	None	1st	Primary
P6	Lifeboat	PAPA	1st	No	2nd	1st	Primary
T3	Muster	GPA	2nd	Yes	1st	1st	Primary
P7	Lifeboat	PAPA	2nd	Yes	1st	1st	Secondary
P8	Muster	GPA	1st	Yes	2nd	2nd	Primary
T4	Lifeboat	PAPA	2nd	Yes	1st	1st	Secondary

Replaced old entries with new outcomes from retention tests and retraining.

Repeated Scenario: T3R	
Attribute	***Value***
End location	Muster
Alarm Type	GPA
Route direction in PA	2nd
Presence of hazard	Yes
Obstructed route	1st
Route taken in prev. scenario	1st
Action: Choose egress route = Secondary	

New Phase 2 Knowledge Base

	Attributes						Action
Scenario	End Location	Alarm	Route by PA	Hazard	Obstructed Route	Previous Route	Egress route Choice
T1R	Lifeboat	None	None	No	None	n/a	Primary
P4	Muster	GPA	1st	No	None	2nd	Primary
P5	Muster	GPA	None	No	None	1st	Primary
T2R	Muster	GPA	1st	No	None	1st	Primary
P6	Lifeboat	PAPA	1st	No	2nd	1st	Primary
T3R	Muster	GPA	2nd	Yes	1st	1st	Secondary
P7	Lifeboat	PAPA	2nd	Yes	1st	1st	Secondary
P8	Muster	GPA	1st	Yes	2nd	2nd	Primary
T4R	Lifeboat	PAPA	2nd	Yes	1st	1st	Secondary

Figure 5.5: Process for updating the KB with retraining data

5.5.4.2. Step 2: Form New Decision Trees Using the DT Algorithm

Similar to the methods used in phase 1, the performance data from the VE scenarios were separated into training and testing datasets. The KB was divided into two training segments; after the initial acquisition and after the refresher training (respectively denoted 'End of

ACQ' and 'End of RET'). Each segment captured the participant's prior performance in scenarios they completed leading up to the next test. The retention test scenarios (phase 2) were set aside to check the predictive validity of the current DT. Once the performance from the test scenario was used to check the validity of the DTs, the data was used to update the KB. The segments of the KB were gradually updated with information from the participant's performance in the test and retraining scenarios. The decision tree induction process classified the content in the KB to create generalized decision rules. The output was a DT that could predict the participant's choice of route (i.e. primary or secondary) based on the value of the programmed attributes of a new emergency scenario. The decision tree algorithm produced multiple DTs that described the participant's behavioural pattern prior to the test scenarios (T1, T2, T3, T4 in phase 2 and the suite of transfer scenarios in phase 3). The extent of the updating depended on the amount of retraining the participants received. The resulting DTs were used to visualize how participants formed emergency egress decision rules based on the content in the KB.

5.5.4.3. Step 3: Test Classification Accuracy

Once the DTs were generated based on the retention and retraining dataset, the testing dataset was used to check the prediction accuracy of the decision trees. The DTs were compared against the participants' route choices in the four retention test scenarios. The prediction accuracy was calculated as a percentage of correctly matched trees for each test scenario.

5.5.5. Application of Decision Trees

Three applications of DTs were used to assess the training curriculum: (1) comparing the participants' DTs to the ideal DT as an indicator of whether learning occurred, (2) tabulating the different DTs formed throughout retraining as an indicator of whether the curriculum adequately prepared participants for transfer, and (3) using the DTs at the end of the retraining (phase 2) to predict performance in the transfer scenarios (phase 3).

5.5.5.1. Comparing Participants DTs with Ideal DT

The ideal design strategy (depicted in Figure 5.1) was used as benchmark to compare with each participants' behavioural patterns. The participants' DTs at the end of the acquisition phase and at the end of the retraining dataset were compared to the intended DT. Figure 5.6 shows an example of the possible outcomes of the comparison. This comparison was performed to check that the participants' DTs met the curriculum criteria.

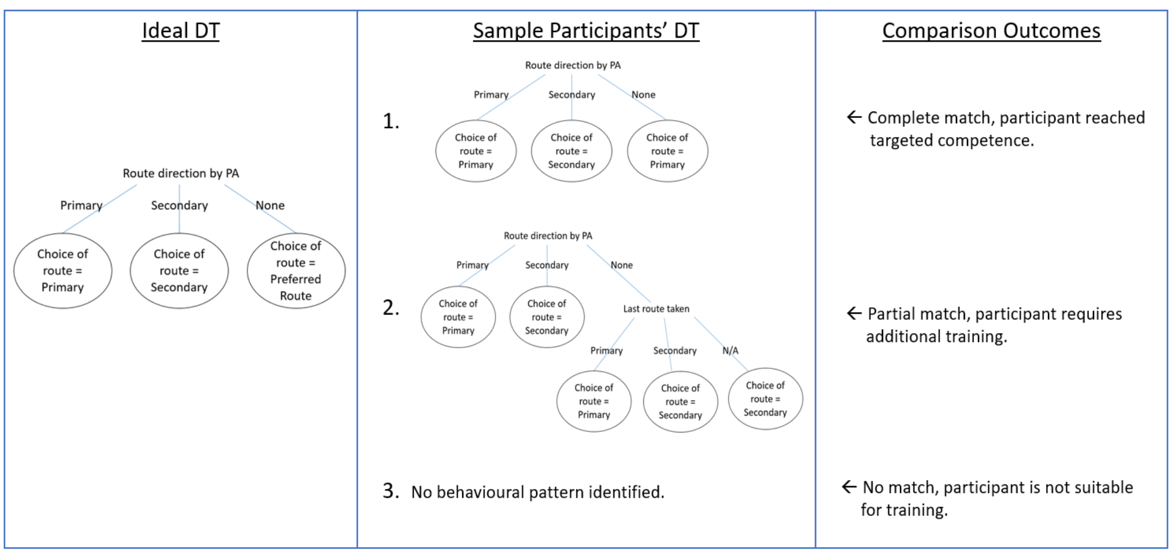


Figure 5.6: Comparing ideal DT to participants DT at end of phase 2

As shown in Figure 5.6, if a participant's DT matched the ideal DT, this was a predictor that the participant had reached the desired competence and was ready for training transfer scenarios. If a participant's DT did not match the ideal DT, this was a predictor that the participant was not prepared for the new transfer scenario. The incomplete or incorrect trees were an indicator that the participant required additional adaptive training before they would be successful in the transfer scenarios. If no behavioural pattern could be determined from the participant's performance data, it was likely that the training was not suitable for this person.

5.5.5.2. Tabulating DTs to Evaluate Training Curriculum

As participants completed the retraining, data in their KB was updated. This iterative updating resulted in the formation of different DTs throughout the retraining. The DTs were tabulated to identify emerging group patterns to inform the training curriculum efficacy.

Several heuristics or strategies emerged from the participant performance data. Figure 5.7 shows possible heuristics developed by participants during the retraining (Smith et al., 2017). These heuristics represent decision rules that were directed by the end location, the alarm type, and whether or not the travelled route is obstructed. These heuristics were not explicitly taught in the training but are acceptable decision rules in many situations.

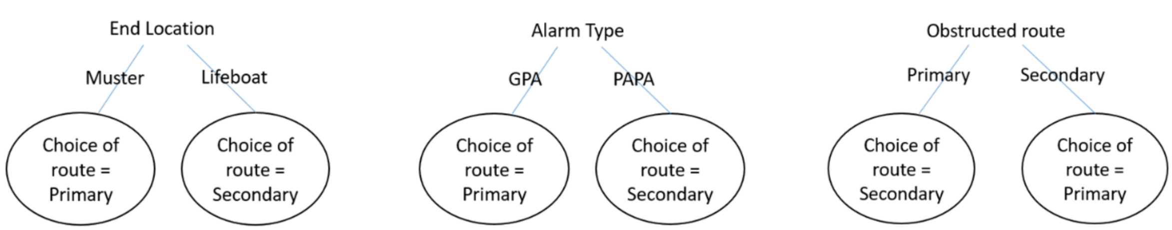


Figure 5.7: Curriculum Design - Accepted heuristics for emergency egress

The percentage of participants whose DT converged to the ideal DT or an accepted heuristics (shown in Figure 5.7) were tabulated. This grouping of DTs was used to determine if the training program adequately prepared participants to apply their egress skills to new emergencies. If the participants' DTs converged to the ideal DT or accepted heuristics, this was an indicator that the training curriculum was successful in preparing participants for the new transfer scenario. If the participants' DTs diverged from the ideal DT or developed unacceptable decision rules, this was an indicator that there were gaps in the training curriculum.

5.5.5.3. Use of DTs to Predict Skill Transfer to New Scenarios

The resulting DT, formed from the full retraining dataset, represented the participants understanding of egress procedures. Therefore, the DTs can also be used to predict the participants' choice of route for a given transfer scenario in phase 3. The final tree from the retraining dataset shows the decision rules a participant is expected to follow based on their previous performance in similar scenarios. To predict how the participants would perform in the eight phase 3 transfer scenarios, the modelled DTs from 'End of RET' were used. The predicted performance was compared against the participants' actual route choices in the transfer scenarios, which was the basis of a prediction accuracy calculation. The prediction accuracy was the calculated for the eight scenarios.

5.6. Results & Discussion

A combination of DT modeling and empirical results were used to investigate the transferability of the training. To inform training transfer from the perspective of the individual learner and systemically the training curriculum, three metrics were used: (1) the DTs diagnostic assessment of participants' performance and the training curriculum efficacy; (2) the overall performance scores in the retention and transfer scenarios; and (3) the DTs prediction of participants' performance in the transfer scenarios.

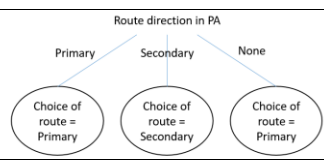
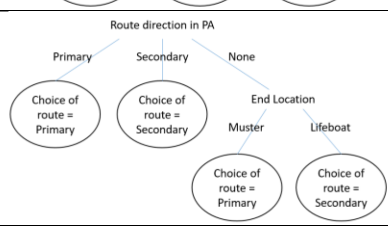
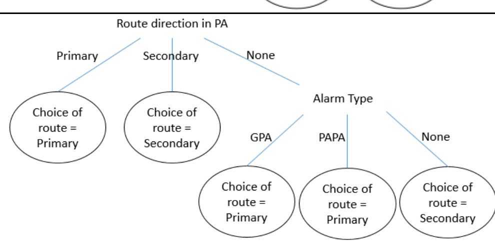
5.6.1. Learning Analytics and Curriculum Assessment with DTs

At an individual level, comparing participants' DTs with the intended decision rules is a useful indicator of how much each participant learned from the retraining in phase 2. At a systemic level, grouping emergent DT patterns is also useful predictor of how well the training curriculum equipped participants to apply their egress skills to new situations in phase 3 of the experiment. This section describes the iterative transformation of the participants' DTs as they completed the retraining, provides a comparison of the participants' DTs with the curriculum designed ideal tree and discusses the implications of emergent DT patterns on the efficacy of the training curriculum.

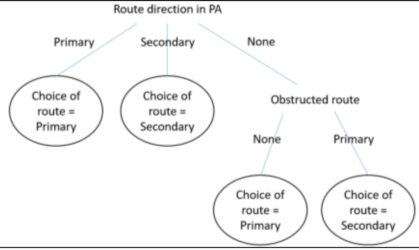
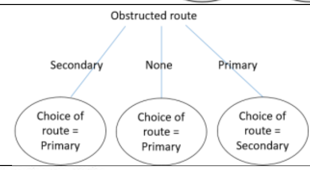
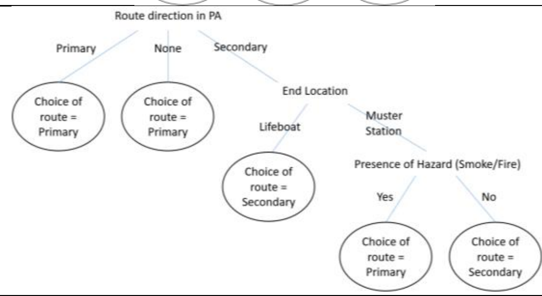
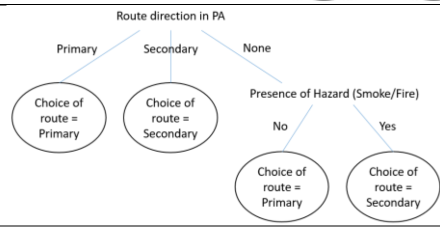
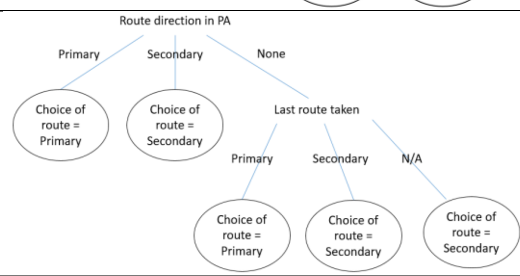
As each participant completed the retraining scenarios in phase 2, their corresponding KB was updated with their performance. Applying the decision tree algorithm to the iteratively updated KB produced either the same or new decision rules. For 55% of participants, the updated data produced the same DT throughout the updating in phase 2. For 42% of participants, the modelled DTs changed between each of the retention tests in phase 2. For some participants, the KB did not have a recognizable pattern and the decision rules were not drawn.

DT patterns that emerged from the training were tabulated. Eight different types of DTs were modelled using the DT updating process throughout the retraining in phase 2. Table 5.6 shows the percentage of participants who formed each type of DT at the beginning and the end of the retraining in phase 2. The column ‘End of ACQ’ depicts the percentage of participants who formed each DT developed from data collected at the end of the acquisition phase. The column ‘End of RET’ depicts the percentage of participants who formed each DT developed from data collected at the end of the retraining phase.

Table 5.6: Decision trees after acquisition (ACQ) and retraining (RET) phases

<table>
<tr><th rowspan="2">Type</th><th rowspan="2">Decision Rules</th><th colspan="2">% Participants</th><th rowspan="2">Compared to Ideal DT</th></tr>
<tr><th>End ACQ</th><th>End RET</th></tr>
<tr><td>1</td><td>
</td><td>60%</td><td>68%</td><td>Match</td></tr>
<tr><td>2</td><td>
</td><td>16%</td><td>10%</td><td>Match</td></tr>
<tr><td>3</td><td>

*One participant had a similar tree but reversed rules for PAPA and None</td><td>10%</td><td>13%</td><td>Match</td></tr>
</table>

4		3%	0%	Match
5		3%	0%	Incomplete
6		3%	0%	Incorrect
7		0%	3%	Incorrect
8		0%	3%	Incorrect
	*No behavioural pattern or strategy identified	5%	3%	Trainees not suitable

The ideal DT, depicted in Figure 5.1, represents what was envisioned as the correct decision rules needed to respond to a multitude of emergencies. The ideal DT was used as a benchmark for comparison to determine if the participants' behavioural patterns, identified by their DTs, converged to the intended strategies. This comparison was performed to determine if the participants achieved the target competence. The ideal DT was compared to participants' modelled DTs at two stages: after the initial acquisition (End of ACQ) and at the end of the refresher training (End of RET). Four types of DTs were identified as successful matches: types 1, 2, 3, and 4 (depicted in Table 5.6, column 'Match'). Four other DTs did not meet the requirements for safe evacuation: types 5, 6, 7, and 8 (depicted in Table 5.6, column 'Incomplete' or 'Incorrect').

Grouping emergent DT patterns (as shown in Table 5.6) provided an indication of whether the training curriculum adequately prepared participants to apply their egress skills to new emergencies in the virtual offshore platform. At the end of skill acquisition, 89% of participants' behaviour represented the ideal DT or an accepted variation of the tree, and the remaining 11% of participants (four participants) did not form a tree to match the intended decision strategy. At the end of retraining, 91% of participants' behaviour represented the ideal DT or accepted variation of the tree and the remaining 9% of participants (three participants) formed incorrect trees. The following subsections discuss the implications of the emergent DT patterns on measuring participants' competence and assessing the efficacy of the training curriculum.

5.6.1.1. Description of Successful Decision Trees

The training curriculum taught participants three main tasks for responding to emergencies: (1) to listen to the alarm type as it indicated the severity of the situation, and indirectly, the muster location; (2) to listen for the PA announcement as it sometimes provided information on the safest egress route; and (3) to re-route if their egress route was obstructed by hazards. From the perspective of the individual learner, successful decision rules included the DTs that all started with information for the PA announcement (e.g. types

1, 2, 3, and 4). For these types of DTs, the participant's route selection was decided based on their understanding of the PA announcement. For example, the type 1 decision rules specified that if the PA directed them to the safe route, then the participant followed that route. If the PA did not provide information about the safest route, then the participant's choice defaulted to their primary egress route. The majority of participants' DTs relied on information from the PA in choosing the egress route (type 1), representing 68% of participants at the end of retraining.

From the systemic perspective of the training curriculum, three DTs developed contingency branches for situations where the PA provided little route information. In such cases, the decision rules included using attributes such as the end location (type 2), alarm type (type 3), and route obstructions (type 4) in making the route choice. These three types of DTs represent heuristics that participants formed throughout the training. Although they were not explicitly taught, these DTs are acceptable route selection strategies for most situations. From a systemic training perspective, these emerging behavioural patterns show how the training resulted in learning strategies that surpassed the intended learning objectives.

The decision rules for type 2, 3 and 4 emerged from the performance data and are good examples of decision strategies that participants developed on their own in the training program. For type 2 decision rules, in circumstances when the PA did not provide situational information, the participant's choice was based on the intended end location. If the participant was required to go to the muster station, they would take the primary route, and if the participant was required to go to the lifeboat station, they would take their secondary route. For type 3 decision rules, the participant's choice was based on the presence of an alarm. If the General Platform Alarm (GPA) or Prepare to Abandon Platform Alarm (PAPA) sounded, the participant would take the primary route, and if there was no alarm, they would take their secondary route. Some participants' DTs relied on the end location (type 2) and alarm type (type 3), representing 10% and 13% of participants respectively at the end of retraining. For type 4 decision rules, in the absence of a PA announcement the participant's choice was based on whether the route was obstructed by a hazard. If the primary egress route was obstructed, the participant would re-route to the secondary route, and if the secondary egress route was obstructed they would re-route the primary route. Only 3% of participants at end of the acquisition phase modelled decision rules that focused on the obstructed route.

5.6.1.2. Description of Incorrect Trees

Four other trees developed from the participants' data represented decision rules that were either incomplete, too specific, or incorrect (types 5, 6, 7, and 8). From the perspective of the individual learner, these are examples in which the participants may not have received sufficient training to develop the ideal DT, or they developed their own heuristics that are not safe strategies for emergency egress. For these types of DTs, there was too much emphasis placed on incorrect attributes. For example, the type 5 decision rules were incomplete and focused solely on whether or not the egress routes were obstructed. Type 6 decision rules had too many specific branches due to overfitting the data (i.e. representing a specific rule for each scenario attribute). For type 7 decision rules, the participant's choice was based on the presence of hazards. If the scenario had a hazard, the participant would take the primary route, and if there was no hazard, they would take their secondary route.

This rule suggests that the primary route is the only safe route if there are hazards in the scenario and incorrectly presumes that there will never be hazards in the accommodation block. For type 8 decision rules, the participant's choice was based on the last route taken in the previous scenario. This participant would indiscriminately alternate between egress routes regardless of the scenario attributes. If the participant had taken the primary route in the previous scenario, then the participant would alternate and take the secondary route in the next scenario and vice versa. Rules based on previous route choices provided no advantage in responding to emergencies.

These unsatisfactory trees identify weaknesses in the individual trainee as well as training curriculum. From the systemic perspective of the training curriculum, these weaknesses can be addressed by providing individuals with more practice exercises to target the errors they are making and reinforce the decision rules of what to do in these circumstances (e.g. offering experience to build the correct contingency branches). In the case of building decision rules based on the presence of hazards (type 7), this is an indicator that the participant has not received enough practice scenarios to know what to do if a hazard occurs on their secondary route. Additional training scenarios should be provided to modify the branches on this participant's DT. Similarly, in the case of type 8 decision rules, this is evidence that the participant might not be challenged by the exercise and is simply gaming the experiment, by taking the opposite route for each scenario. This participant should receive training scenarios to challenge them and reinforce the importance of the training.

In a few situations, the VE training was not compatible with the learner. Some participants' performance data did not have a recognizable pattern and the decision rules were not drawn. If no behavioural pattern could be determined from the participant's performance data, it was likely that the trainee was not suitable for this type of VE training.

5.6.2. Empirical Results: Performance in Transfer Scenarios

The performance scores provided another indicator of how well participants were able to apply their egress skills to new situations in phase 3 of the experiment. Table 5.7 shows the descriptive statistics for participants' performance in the four retention and eight transfer scenarios. Figures 5.8 and 5.9 provide a visual representation of the retention and transfer data in Table 5.6 using boxplots.

Table 5.7: Performance scores for all the phase 3 test scenarios

Test	Phase	n	Performance Scores (%)				
			Mean	St. Dev	Median	Min	Max
T1	2	35	72.3	26.9	81.0	9.0	100
T2		35	83.7	18.0	91.0	35.0	100
T3		36	96.8	8.0	100	63.0	100
T4		34	95.7	13.4	100	29.0	100
U1	3	38	98.5	5.1	100	78.3	100
U2		38	99.1	4.1	100	78.3	100
U3		38	95.8	14.3	100	28.6	100
U4		38	80.8	4.8	82.1	66.1	100
U5		37	93.4	11.7	100	59.5	100
U6		36	96.2	7.9	100	66.7	100
U7		36	92.4	10.8	100	75.0	100
U8		37	77.0	19.7	80.8	19.2	100

As shown in Figure 5.8, the elapsed time (of 6 to 9-months) between phase 1 and phase 2 affected the performance of participants in the first two retention scenarios (T1 & T2). The participants returned to competence in the final two retention scenarios (T3 & T4).

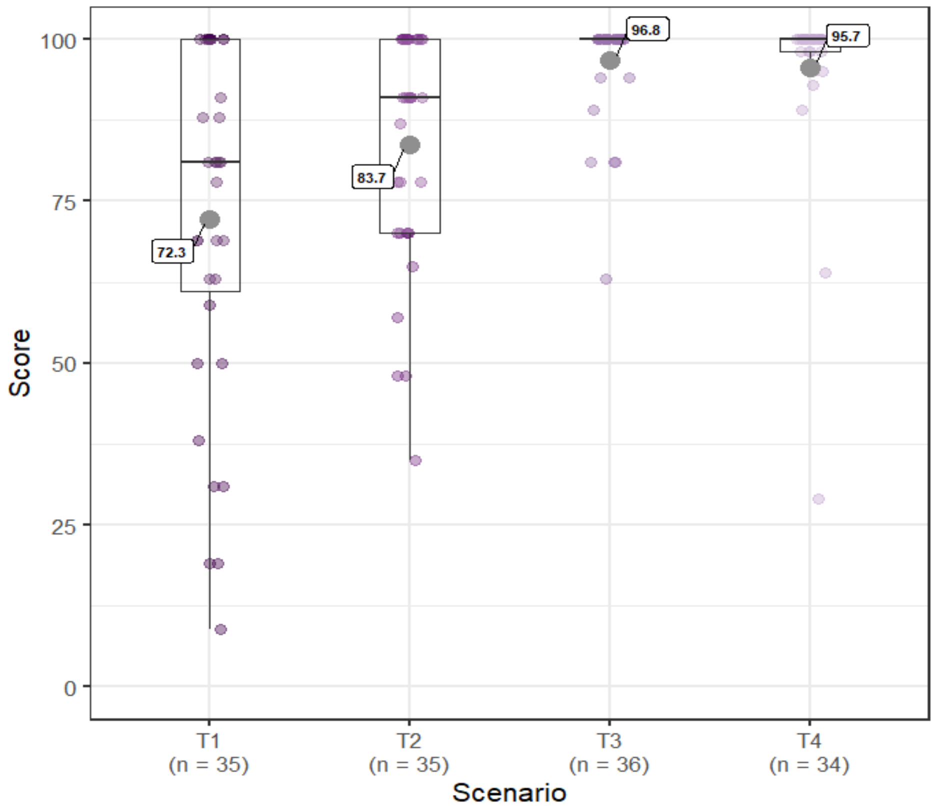


Figure 5.8: Boxplot of performance scores for the 1st attempt of retention scenarios

As shown in Figure 5.9, the retraining prepared participants to demonstrate skill transfer for the majority of the phase 3 scenarios. The overall performance showed that participants successfully applied their egress skills in the transfer scenarios that were similar to the training conditions (e.g. near transfer for test scenarios U1, U2, and U3). The conditions of these scenarios involved egress from the cabin. If there was a hazard in the scenario, its location was often identified in the PA announcement. The participants' performance dropped in scenario U4. This scenario also involved egress from the cabin and a hazard blocking the primary egress route, but in this case, there was no information from the PA announcement about the hazard. The 97% of participants came across the fire in the main stairwell, blocking their primary egress route, and were forced to re-route to the secondary egress route to their muster station.

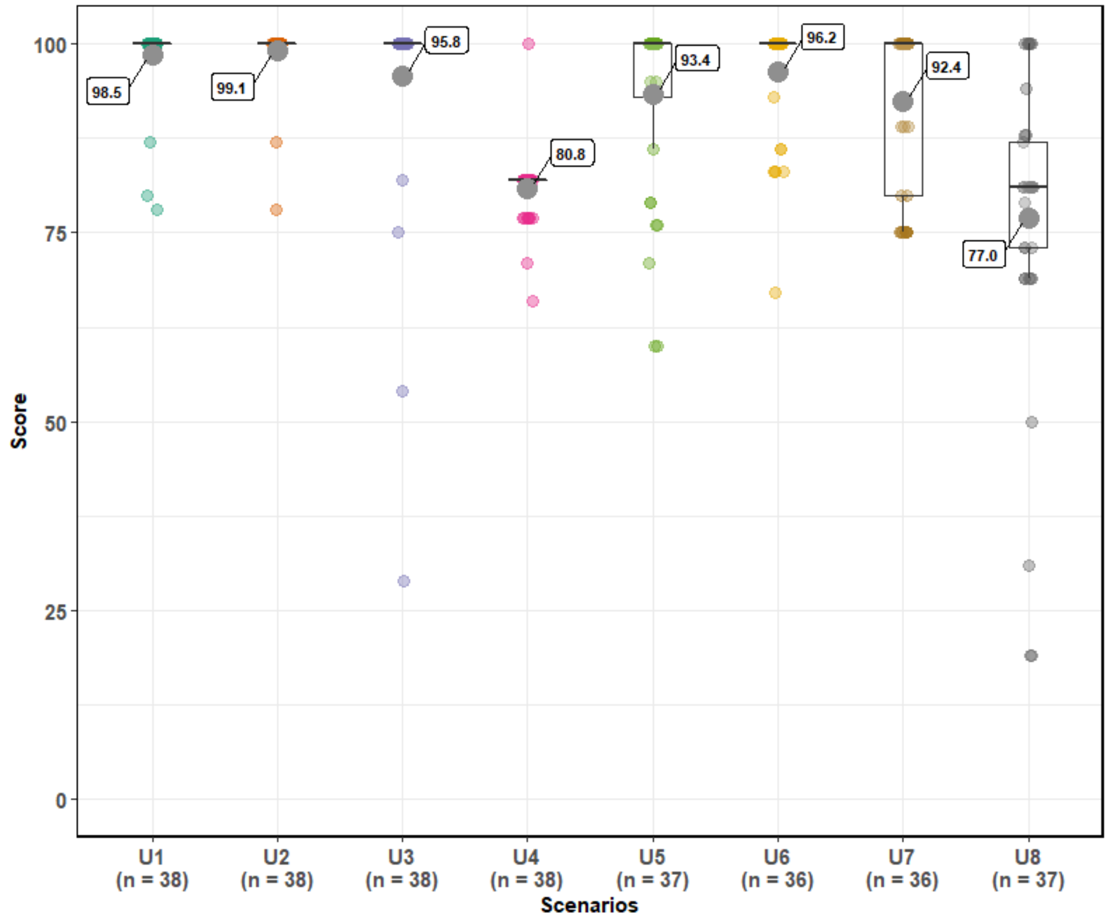


Figure 5.9: Boxplot of performance scores for the transfer scenarios

As the conditions in the transfer scenarios moved beyond the scope of the training, the overall performance of the participants decreased (e.g. for far transfer test scenarios U5, U6, U7, and U8). The conditions in these scenarios involved egress from a less familiar starting location, the bridge. The high and low values of the scenario attributes for hazards and the PA were also varied (e.g. scenario U6 provided little information about the emergency over the PA and scenario U7 had a major egress route obstructed by hazardous conditions). In particular, participants' performance dropped in scenario U8. This scenario deviated the most from the training conditions. U8 assessed the participants' ability to respond to a situation starting in the bridge, with smoke and fire hazards blocking their egress path and little information from the PA. Seventy-nine percent of participants came across the fire and the deluge system in the external stairway, blocking their secondary egress route, and were forced to re-route to the primary route to reach their muster station.

5.6.3. DTs to Predict Training Transfer

The DTs from phase 2 (depicted in Table 5.6, column 'End of ACQ') were used to predict the participants' route choices in the retention scenarios. The final DTs from the retraining (depicted in Table 5.6, column 'End of RET') were used to predict participants' route choices in the transfer scenarios. The prediction accuracies were the calculated for the retention and transfer scenarios by comparing the modelled DTs to the participants' corresponding route choices in the 12 scenarios. Table 5.8 shows the percentage of DTs that predicted the participants' successful performance in phases 2 and 3.

Table 5.8: Percentage of DTs that predicted participants' successful performance

Decision Tree Type	% DTs predicted successful performance											
	Phase 2				Phase 3							
	Retention				Near transfer				Far transfer			
	T1	T2	T3	T4	U1	U2	U3	U4	U5	U6	U7	U8
All Types	22	27	79	77	66	85	82	87	13	16	53	16
Type 1	11	18	55	61	47	66	58	63	5	13	40	8
Type 2	3	3	13	8	8	8	11	11	0	0	5	3
Type 3	5	3	8	5	11	11	13	13	8	3	8	5
Type 4	3	3	3	3	-	-	-	-	-	-	-	-

5.6.3.1. Classification Accuracy of Decision Trees

The DTs at the initial stage of the retraining phase (from ‘End of ACQ’) produced low classification accuracy for the first two test scenarios (T1, T2). This poor accuracy is not a reflection of the prediction capabilities of DTs, but is a direct result of participants’ weak performance in the retention scenarios (e.g. due to forgetting their egress routes after the long retention interval). As an example, in scenario T1 the decision trees were compared to the participants’ route choice after 6 to 9-months. Thirty-two percent of participants (12 people) passed scenario T1 and the DTs predicted the route choice of 55% of participants (21 people). However, only 22% of participants (8 people) passed the scenario and had their DT predict their performance in T1. This low percentage is likely due to participants forgetting their egress route. Eighteen percent of participants (7 people) DT correctly predicted their route choice but they were unable to continue to follow their route because they forgot the full route. The DTs do not have a mechanism to model forgetting, so this result affected the classification accuracy of the trees in the retention phase. This is not a limitation of DTs, but an example of how DTs can help diagnose the strengths and weaknesses of individuals completing the VE training.

After participants received sufficient retraining, their performance improved in the last two retention scenarios (T3, T4). Similarly, the participants’ corresponding KB was updated with the full dataset from the retraining and the classification accuracy of the revised DTs improved. For example, in scenario T4 the DTs were compared to the participants’ route choice after receiving retraining. Eighty-nine percent of participants (34 people) passed the scenario T4 and the DTs predicted the route choice of 87% of participants (33 people). However, only 77% of participants (29 people) passed the scenario and had their DT predict their performance in T4.

5.6.3.2. Decision Trees to Predict Transfer

The DTs at the end of the retraining (depicted in Table 5.6, column ‘End of RET’) had mixed success in predicting the route selection choices of participants in the phase 3 scenarios. The prediction accuracy of the DTs for phase 3 was sensitive to the degree of transfer (i.e. near and far transfer). Near transfer scenarios tested participants’ application of egress skills in a transfer setting that was very similar to the training (e.g. the participants’ familiarity to the scenarios start location). All the training (phase 1) and retraining (phase 2) scenarios were designed to teach participants egress route from their cabin. The first set of four scenarios (U1-4) in phase 3 tested near transfer and started in the cabin, the same location as the training scenarios. Far transfer scenarios tested participants’ generalization of egress skills in a transfer setting that was very different from the training. The second set of four scenarios (U5-8) in phase 3 tested far transfer and started in the bridge. The DT prediction accuracy was higher for the scenarios that started in the cabin (66-87%) than the scenarios that started in the bridge (13-53%). As a result, the DTs could not predict successful performance in all the transfer scenarios.

The poor prediction accuracy of the second set of transfer scenarios is not a reflection of the assessment quality of the DT modeling approach, but an indicator of the gaps in the training curriculum. This indicator shows that the VE training did not adequately prepare participants to respond to the emergency scenarios in the transfer phase because some of the training transfer scenarios (U5-8) were beyond the scope of the training. DTs use supervised learning, which involves a database of solved examples to infer choices or actions in future cases. Supervised learning is applicable when the future cases are within the same parameters of the trained model. Therefore, if the scope of the transfer scenarios reach beyond the training context, the DTs will not be able to predict actions for unforeseen or ill-defined future states. This can be considered a similar mechanism to how trainees learn, apply, and generalize knowledge. Trainees learn information in training context and they are expected to apply their skills to real-world situations. It is unreasonable to expect a trainee to transfer skills to situations well beyond what they have been trained for. Thus, if the DT depiction of the participants' training experience cannot adequately predict their performance in the transfer scenarios, it is possible that the reason for the poor prediction is that the training did not adequately prepare the participants for the new emergency scenarios. From a systemic perspective, the DT prediction analysis provides a valuable diagnostic lens to assess the efficacy of a training program.

5.6.4. Recommendations to Improve Transferability

The diagnostic DT comparison, tabulated emergent DTs, and prediction methods in this paper identified gaps in the design of the VE training curriculum. This process helped identify the limitations of the benchmark DT and the size of the participants KB, both of which inform the design and assessment of the curriculum.

The ideal DT must be improved to address the participants' deteriorating performance and the reduced DT prediction accuracy for complex transfer scenarios. A way forward is to modify the ideal DT to improve the learners' preparedness for both near and far transfer scenarios. Figure 5.10 provides an example of an improved DT with additional branches to address the situations participants faced in the far transfer scenarios (e.g. U5-8, when the PA does not provide information about the situation and the participant is required to make a decision based on obstructed egress routes).

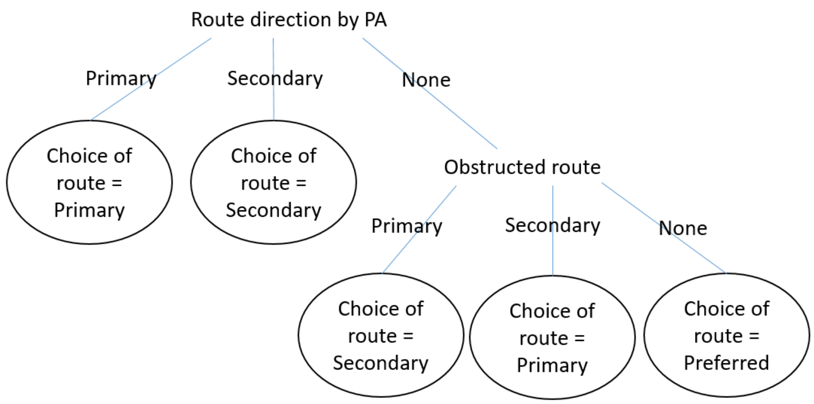


Figure 5.10: Revised decision tree for emergency egress

To achieve the revised DT, participants should be provided with more training scenarios to target the intended decision strategies and thereby increase the number of entries in their corresponding KB. Adding more practice scenarios (e.g. situations without a PA announcement and obstructed routes) will improve the participants' repository of solved examples, which is necessary to develop the contingency branches to match the revised DT (in Figure 5.10). However, practice scenarios should target the needs of each individual leaner. Combining DT modeling with adaptive training would provide a customized tool to ensure each learner achieves demonstrable competence by developing a DT to match the intended decision strategies (e.g. developing the contingency branches needed to help learners respond to many plausible emergencies).

Another gap identified by this process was that the existing DT target captured the protocols as defined by regulations. However, this DT was found to be not comprehensive enough to predict performance in the complex transfer scenarios. The transfer scenarios in this experiment were plausible scenarios that challenged the scope of the VE training by testing situations that go beyond the competencies dictated by industry regulations in order to measure human reliability in emergency conditions. This is important to highlight because emergencies are characterized by complexity, time pressure, and uncertainty. Offshore safety training must be robust enough to apply to a wide range of circumstances and therefore VE training should be designed to encourage training transferability.

5.7. Conclusion

This research used decision trees as data-informed curriculum design and assessment tools to evaluate the transferability of virtual offshore training. Visualizing participants' decision strategies can help instructional designers determine if participants are adequately prepared for new training transfer situations. The results of this work showed that DTs can model participants' decision making strategies throughout the acquisition, retention and retraining, and transfer phases of the experiment. The majority of participants demonstrated training transfer in test scenarios that were similar to the training conditions (near transfer test scenarios U1-4). Similarly, the majority of participants' DTs converged on intended strategies. From the individual learner perspective, this convergence is an indicator that the participants achieved the intended demonstrable competence. DT analysis provided evidence that participants accrued sufficient training to transfer their skills (e.g. promote the application and generalization of skills).

However, not all participants developed the ideal DT which shows the limitations of the trainee and the training. For some participants, their performance data developed credible heuristics (e.g. specific decision rules that were not explicitly taught). This finding highlights the diagnostic utility of DTs as they show how participants develop strategies that do not always match the intended training material. From a systemic perspective, DTs are useful for identifying the strengths and weaknesses of a training curriculum. In this case, some participants required more opportunities to practice scenarios with varied attributes to ensure they develop the targeted route selection strategies. This retroactive analysis could be performed in real time as a basis of adaptive training. VE training can be further optimized by combining DT methodology with adaptive training mechanisms to provide participants with customized scenarios to meet their specific learning need and to prepare them for training transfer.

Decision trees also have the potential to predict when a person is prepared or not prepared for more advanced situations. Overall, the participants' data-informed DTs were successful in predicting the participants' performance in new scenarios that represented near training transfer (U1-4). As a group, the performance decreased as scenario conditions moved beyond the scope of the initial training (far transfer test scenarios U5-8). Similarly, the decision rules developed throughout the training were not good at predicting participants' route selection strategies in situations that went beyond the training context. The poor DT prediction of participant performance in the transfer scenarios is an indicator of gaps in the VE training. The far transfer scenarios were plausible scenarios that challenged the scope of the VE training by testing situations that go beyond the competencies prescribed by regulations. This finding suggests that DT can push the boundaries of existing training by highlighting its shortcomings and challenging instructional designers to develop training that prepares people for emergencies, rather than for the somewhat nominal requirements of regulations.

5.8. Compliance with Ethical Standards

Memorial University's interdisciplinary committee on ethics in human research approved the experimental protocol. Following the approved research protocol, informed consent was obtained from all participants prior to completing the longitudinal study.

5.9. Acknowledgements

The authors acknowledge with gratitude the support of the NSERC-Husky Energy Industrial Research Chair.

5.10. References

Barnett, S. & Ceci, S. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.

Blume, B., Ford, J., Baldwin, T., & Huang, J. (2010). Transfer of training: a meta-analytic review. *Journal of Management*, 36(4), 1065–1105.

Blundon, A. (2019). *Human reliability assessment for offshore emergency situations using a virtual environment* (Master's thesis). Memorial University of Newfoundland, St. John's NL, Canada.

Duffy, V. (2009). *Handbook of digital human modeling: research for applied ergonomics and human factors engineering* (pp 37–44). Boca Raton, FL: CRC Press Taylor & Francis Group.

Ford, J., Baldwin, T., & Prasad, J. (2018) Transfer of training: The known and the unknown. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 201–225.

Grossman, R., & Salas, E. (2011). The transfer of training: what really matters. *International Journal of Training and Development*, 15(2), 103–120.

McGaghie, W., Issenberg, S., Barsuk, J., Wayne, D. (2014). A critical review of simulation-based mastery learning with translational outcomes. *Medical Education*. 48, 375–385

Musharraf, M., Bradbury-Squires, D., Khan, F., Veitch, B., Mackinnon, S., and Imtiaz, S. (2014). A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. *Reliability Engineering & System Safety*. 132, 1–8. https://doi.org/10.1016/j.ress.2014.06.016

Musharraf, M., Moyle, A., Khan, F., & Veitch, B. (2019). Using simulator data to facilitate human reliability analysis. *Journal of Offshore Mechanics and Arctic Engineering*. 141(2), 021607.
https://doi.org/10.1115/1.4042538

Musharraf, M., Smith, J., Khan, F., & Veitch, B. (2018). Identifying route selection strategies in offshore emergency situations using decision trees. *Reliability Engineering & System Safety*. https://doi.org/10.1016/j.ress.2018.06.007

Sellberg, C., Lindmark, O., & Rystedt, H. (2017). Learning to navigate: the centrality of instructions and assessments for developing students' professional competencies in simulation-based training. *WMU Journal of Maritime Affairs*. 17, 249–265.

Smith, J., & Veitch, B. (2018). A better way to train personnel to be safe in emergencies. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems Part B*, 5(1), 011003. https://doi.org/10.1115/1.4040660

Smith, J., Doody, K., & Veitch, B. (2019). Being prepared for emergencies: a virtual environment experiment on the retention and maintenance of egress skills. *WMU Journal of Maritime Affairs.* https://doi.org/10.1007/s13437-019-00174-y

Smith, J., Musharraf, M., & Veitch, B. (2017). Data informed cognitive modelling of offshore emergency egress behaviour. *Proceedings of the 2017 International Conference on Cognitive Modeling*, Coventry, UK.
https://iccm-conference.neocities.org/2017/ICCMprogram_files/paper_17.pdf

6. CONCLUSIONS

6.1. Conclusions

VE training will probably not replace conventional offshore safety training used in the real-world operating environment. However, the results of this research demonstrated that pedagogically designed and data-informed VE training technology can enhance conventional training. This work provided evidence for operators and regulators in offshore and maritime domains to support the adoption of VE training to improve workers' overall competence and compliance during emergencies. More specifically, VE training was used in this research for dual purposes: to provide trainees with artificial experience to fill in competency gaps, and as a human behaviour laboratory to assess the implications of new training interventions prior to implementing them offshore.

To evaluate VE training, a longitudinal experimental program was used to provide empirical and modeling evidence as a means to address pedagogical gaps in conventional training. Four main research questions within VE training were the foci of this work:

(1) how to design and deliver VE training to address individual variability in learning skills and assure demonstrable competence in the workforce,

(2) how to assess data modeling tools to improve the assessment of different VE training interventions,

(3) how to evaluate VE training to address skill retention and better inform the mandated recurrency training schedule, and

(4) how to improve VE training to foster skill transfer and prepare the workforce for a wide variety of emergencies.

Two perspectives were used to structure the evaluation of the VE training efficacy: 1) pedagogical theory of simulation-based mastery learning, and 2) data-mining methodology of decision tree modeling. The investigation first evaluated the design and delivery of VE training using the simulation-based mastery learning pedagogical approach by collecting trainees' performance data over the course of the longitudinal study at three critical learning phases: skill acquisition, skill retention, and skill transfer. The investigation then examined methods from the data-mining domain to support the assessment and optimization of VE training. Decision tree modeling was used to identify participants' data patterns (as a proxy to understanding their learning strategies) and to inform the efficacy of VE training on a systemic level. This analysis uncovered participants' heuristics that are otherwise not easy to identify using conventional assessment methods. By using decision trees to mine the performance data, the efficacy of VE training could be assessed more thoroughly at three critical learning phases. Results of the analysis were reported using a combination of performance metrics and decision tree models of participants' data patterns. The significance of the main empirical and modeling findings are summarized below.

6.1.1. Empirical Evidence

Empirical evidence from this work offered a conventional approach to evaluate the efficacy of VE training. Overall, the SBML framework applied to the VE training was effective during the skill acquisition and retraining phases; however, some limitations were observed at the skill retention and transfer phases of the experimental program.

At the skill acquisition phase, the SBML framework reinforced learning, addressed performance variability, and assured competence. This addresses the first pedagogical gap in offshore safety training that *conventional training does not account for individual variability in learning styles and paces*. Due to regulatory, logistical, and financial constraints, conventional safety training is often provided using fixed-time instruction and lacks the structure and resources to assure competence is acquired and maintained. However, a workforce with unknown or variable competence is a safety concern. Conversely, the VE training recorded participants' performance and monitored their progress through the training. As a direct result, the SBML guided VE training assured that all participants' achieved competence upon the completion of the training. Overall, evidence from this experiment demonstrated the utility of VE training in providing structure, standardization, and accountability to offshore egress training. Employing VE training in industry has the potential to improve the overall safety of offshore operations.

At the skill retention phase, the performance results indicated that emergency egress skills taught using the VE training were susceptible to skill decay over a period of 6 to 9 months. The decay in performance was largely due to a reduction in procedural compliance and the degradation of spatial knowledge (e.g. remembering vital egress routes). Therefore, a shorter retention interval is necessary to maintain egress skills. This evidence should encourage operators and regulators to address the second pedagogical gap in offshore safety training that *conventional training is forgotten before the mandated recurrency training is scheduled*. The findings from the second phase of the experiment challenge the convention that recurrency training should be administered at a standardized or fixed-interval regardless of individual's learning and retention tendencies. This research demonstrated the potential of VE training as an alternative on-demand training solution (i.e. a customized retraining frequency).

To address the degradation of participants' egress skills in the experiment, the mastery-learning theory (Bloom, 1971; Gusky, 2007) was used to develop an adaptive retraining matrix. The adaptive matrix assigned participants to VE training exercises based on the specific errors they made (Doody, 2018). This approach accommodated the different learning paths and paces of participants. As a direct result, the retraining was successful in bringing all participants back to demonstrable competence, and did so quickly. These results show that the VE training can help maintain egress skills for workers who have been away from the platform for an extended period. Employing VE training in industry has the potential to change how recurrency training is provided. VE technology can facilitate a shift in safety operations from fixed-interval training to a competence-driven training at an on-demand frequency that is based on each individual workers' needs.

At the skill transfer phase, the VE training was able to provide participants with sufficient artificial experience to apply their knowledge and skills to multiple emergencies. This phase of the experiment supports the conclusion that VE training addresses the third pedagogical gap in conventional offshore safety training that *the training does not measure learning outcomes and as a result does not inform the transfer of training*. From a training transfer perspective, the efficacy of VE training could be measured by employing capabilities of VE training to record and track participants' performance throughout the training. The performance results in the transfer scenarios revealed that participants were able to apply their egress skills to new situations in the same virtual setting. However, there were limitations to the extent participants could transfer their training to novel emergencies.

The overall performance results showed that participants successfully applied their egress skills in the transfer scenarios that were similar to the training conditions, but that the participants' performance decreased as the conditions in the emergency scenarios moved beyond the scope of the training, as expected. This represents a gap in the training curriculum scope that trainees should not be expected to fill on their own. These results provide evidence to highlight the importance of addressing the fourth pedagogical gap in mandatory offshore safety training that *conventional training is not representative of the conditions in real emergencies.* The finding that participant performance decreased as the complexity of the test scenarios increased and deviated from the training context is not surprising. It is unrealistic to expect people to apply skills to situations that differ drastically from the training. Therefore, it is naive to assume that muster drills in benign conditions adequately prepare offshore workers for the complexities and time-sensitive procedures required of real emergencies offshore. Although this experimental program revealed limitations in how the VE training was delivered from a training transfer perspective, overall the VE technology and experimental design employed in this research were effective at identifying deficiencies in the assumptions of conventional offshore training and offering recommendations on how to use VE training to improve offshore safety. Employing VE training in industry has the potential to provide transferable experience with plausible emergency conditions that cannot be practiced safely in offshore evacuation exercises.

6.1.2. Modeling Evidence

This research applied a data-mining technique to expand the capabilities of VE training. Building on the work of Musharraf et al. (2018), who demonstrated the utility of DT modeling to assess training at an individual level, DT modeling was used to evaluate VE training at a systemic level. The modeling evidence from this work offered a more comprehensive assessment of the VE training at the three learning phases. Overall, the DT modeling was effective at diagnosing the strengths and weaknesses of the VE training at the skill acquisition phase and was able to provide a means to diagnose and predict when individuals received sufficient training to use their skills efficiently in multiple emergencies at the skill retention and transfer phases. However, some limitations were observed at the retention and transfer phases of the experimental program.

At the skill acquisition phase, the DT modeling results identified systemic strengths and weaknesses in the delivery of VE training. This was identified by comparing the DTs from two groups training using different methods: LBT and SBML training. The DTs were able to diagnose when participants received sufficient training to be competent and were able to predict participants' behaviours in the test scenarios (e.g. egress route choices). The results of the DT analysis also showed that DTs were useful tools for the design and assessment of VE training because they offered a visual representation of an individual's heuristics that was easy to interpret. Overall, decision trees were shown to improve the design and delivery of VE training by comparing the changes in trainees' decision-making patterns, in response to different training interventions, to the intended learning objectives. This is a key finding because understanding how trainees develop learning strategies or heuristics is important for addressing individual variability and assuring demonstrable competence in the workforce.

At the skill retention and transfer phases, the DT modeling results again identified systemic strengths and deficiencies in the SBML approach applied to VE training. The DTs were able to diagnose when the SBML-trained participants received sufficient retraining to be competent (e.g. when DTs converged on the intended learning objectives and represented safe behaviours). However, the DTs were limited in predicting when participants were ready to transfer skills to new situations (e.g. test scenarios different from the training context). The poor prediction accuracy of the DTs in forecasting training transfer was an indicator of gaps in the VE training. The transfer scenarios were plausible scenarios that challenged the scope of the VE training by testing situations that go beyond the competencies prescribed by regulations. Additionally, the DT algorithm uses inference to predict possible choices to future events, thus some of the training transfer test scenarios were misaligned or outside the scope of the training (e.g. the trainees were not sufficiently trained for test scenarios that were too advanced). This disconnect between the DTs and the transfer scenarios can be addressed by making modifications to the design of the VE training (specifically the practice scenarios). Overall, the DT analysis demonstrated that these data-mining tools play an important role in learning analytics, but have some limitations (e.g. can only infer and not predict the transfer of skills beyond the context of the training).

DT modeling can improve the assessment of VE training by identifying when the prescribed VE training does not match the difficulty level of the test scenarios. This is especially important when the goal of VE training is to prepare the workforce for a wide variety of emergencies (as conventional safety training does not represent the conditions of real emergencies). Incorporating DT modeling into VE training can push the boundaries of existing training by highlighting its shortcomings and by challenging instructional designers to develop training that prepares people for emergencies, rather than for the somewhat nominal requirements of regulations. These DT tools have potential to improve future learning applications, specifically to support adaptive training programs. Pairing DTs' diagnostic and predictive tools with VE training offers the flexibility to provide people with practice, assessment, and corrective feedback on-demand and at a customized schedule that meets the needs of each learner. This work provides evidence to operators and regulators in offshore and maritime domains to support the adoption of VE training to improve workers' overall competence and compliance during emergencies.

6.2. Technical Challenges and Limitations

The following describes the technical challenges and limitations that arose from this research. Explanations for the design choices are provided and recommendations are made for future studies.

1. This study used a convenience sample in that the recruited participants did not directly represent the target population (Ritter et al., 2013). Most of the participants in the experiments were undergraduate and graduate students. All participants had no prior offshore experience and no exposure to the AVERT simulator prior to the study. The decision to limit recruitment to naïve subjects was made to control for experience (e.g. remove past spatial and procedural experience from the study). As

such, people who worked in the offshore oil and gas industry were excluded because their prior offshore experience would confound the measurements of learning egress skills in the VE training (e.g. it would be difficult to discern if the participants learned the egress skills from VE training or applied their prior knowledge when performing in the scenarios). Controlling for experience does have its trade-offs. The data collected from naïve participants (convenience sample) limits the generalization of the results, as it may not provide the same conclusions as testing with the offshore workforce population. If the experiment were repeated with experienced offshore workers the outcomes could be different.

2. Designing and managing the logistics of a longitudinal experiment presented participant attrition challenges. Efforts were made to recruit a sufficient number of participants at the initial phase of the experiment to accommodate the anticipated attrition so that the statistical power was maintained at the retention phase of the experiment. The target sample size for the retention phase of the study was 40 participants. Sixty participants were recruited for the first phase of the study with an expectation of 25% attrition for the longitudinal portion of the study (e.g. loss of 15 participants). Fifty-five participants completed the skill acquisition phase. Five participants withdrew at the onset, due to simulator sickness and difficulty with the controller. Seventeen participants opted out of the longitudinal study during the 6 to 9-month retention interval. The remaining 38 participants completed the retention phase. Two were identified as outliers (completed the retention assessment at 4 and 10 months) and were removed from the retention analysis.

3. In the peer-review of this research, it was identified that combining the retention assessment and retraining into one phase was a limitation in the experimental design. This experiment used an adaptive retraining matrix to retrain participants directly after they failed the test scenarios. Merging the methods made testing the retention of complex skills difficult because participants were brought back to competence in foundational skills before being tested on more advanced emergency scenarios. As a result, this design did not allow for conclusive answers as to why participants performed well in the more complex test scenarios. The reviewer suggested that a better design would be to separate the retention assessment from the retraining (as opposed to weaving them together). This experiment may have benefitted from a separate baseline group to test retention (e.g. assessing participants on all 4-test scenarios before the retraining them). However, a substantially larger sample size would have been necessary at the initial skill acquisition phase to ensure the returning sample for the retention phase was large enough to accommodate a separate group to test baseline retention. Recruiting a larger sample size for a longitudinal study was not feasible due to logistical and attrition implications. The decision to combine the retention assessment and retraining into one phase was made in an effort to strike a balance between experimental control, ecological validity, and practicality of the training delivery.

4. This experiment used a proxy measure of training transfer by comparing how training in a VE in one context helped participants apply their newly acquired skills to a novel or unforeseen context in the same virtual setting (Wickens et al., 2012).

Traditional training transfer studies measure skill transfer by first training skills, for example using a VE or simulator, and then evaluating the skills in the real environment. This was a limitation to the experimental design due to ethical, safety, and logistical constraints. It was not feasible or safe to assess trained participants on how their skills transferred to the real environment (on an offshore platform). Therefore, the study was designed around measuring training transfer using novel situations in the same virtual setting. This was accomplished by repurposing the scenarios and data collected originally for human reliability analysis (Blundon, 2019; Musharraf et al., 2019).

5. Decision trees are one of many data-mining methods that are useful for decision modelling. There are other data-mining approaches that could be used to make VE training more adaptive, such as Bayesian network (BN), artificial neural networks (ANN), and support vector machines (SVM). In some cases, these methods have better diagnostic capabilities and higher prediction accuracy, however, they are harder to interpret and communicate to non-domain experts such as instructional designers. One such example is Bayesian Networks (BN), a statistical learning method that models the probability relationship based on performance data (i.e. the likelihood of outcomes). Musharraf et al. (2017) and Blundon (2019) used BN to investigate human reliability in emergency egress tasks using a virtual environment. Decision trees were well suited for the application of evaluating training efficacy because of their relative ease of construction, visual simplicity, high interpretability, and transparency compared to the other data-mining methods (Liu, 2009; Romero

and Ventura, 2010). Decision trees are especially suited to assist instructional designers in the interpretation of the VE performance data (e.g. understand the participants' behavioural patterns or strategies) and to make design changes to the training curriculum (e.g. revise the content, structure, and delivery of the training accordingly).

6. This experiment used decision trees, which are a logic-based supervised learning approach to modeling decision-making. This approach resulted in a simplified representation of decision-making that may not accurately reflect how people actually learn or make complex decisions. Klein (2008) suggests that people tend to develop heuristic approaches to decision-making (e.g. simplified intuitive responses or rules) instead of optimal judgement or systematic strategies. Insights into complex decision-making in emergencies may not be easily modeled using performance data and decision tree modeling. Although DTs do not replicate how people actually form decisions, they do provide perspective on patterns in the data that can be used as a proxy for participants' strategies. DTs are useful decision-making representation tools to help understand what information people might be attending to when making decisions in emergencies. That is valuable in understanding when participants have reached competence or when they might require more training.

7. For the decision tree modeling, a *split training set* approach was used to calculate the classification or prediction accuracy of the decision trees. The *split training set* approach divides the dataset, uses 2/3 of the dataset for training, and holds the other

1/3 of the dataset to test the classifier’s performance. In the peer-review process reviewers expressed limitations with the *split training set* method and suggested that cross-validation would be a more suitable method for evaluating the DT classification accuracy. The cross-validation approach involves dividing the data set into mutually exclusive, equal sized, training subsets for which the DT algorithm is trained and then the resulting DT model is tested on all of the subsets (Kotsiantis, 2007; Han et al. 2011).

6.3. Future Work

This research has opened two important lines of inquiry to improve offshore emergency training. Future work should address: 1) implementation of decision trees into VE training to support adaptive training, 2) investigation of the implications of VE training on more complex decision-making tasks.

1. Adaptive offshore egress training can be achieved by implementing decision tree modeling into the VE technology. This integration would provide real-time performance assessment and customized training exercises for each trainee (i.e. to better meet the individual’s learning style and pace). As a result, safety training could be provided on demand as opposed to crews waiting for the next recurrency interval. The very idea of this could change how training is provided to the offshore industry. Decision tree informed adaptive VE training has the potential to shift recurrency training from a standard frequency to an individualized maintenance schedule that reflects each person’s tendencies to remember or forget training. Since

this research has demonstrated the utility of applying DTs to VE training, future work should focus on embedding decision tree analysis into an adaptive VE training model.

2. VE training has the potential to teach complex decision-making in multiple emergency contexts. This research demonstrated the utility of VE training for teaching decision-making skills that were mostly procedural based (e.g. easy to follow safety protocols and selecting the safest egress route for the emergency conditions). The targeted VE training brought trainees competence in these simple decision-making tasks.

   The next step is to investigate VE training for more complex tasks that require a higher order decision-making. For example, members of the emergency response team offshore have larger responsibilities during emergencies and require a higher level of proficiency and expertise, as described by Dreyfus (1997) and Griswold-Theodorson et al. (2015) as a five-staged training framework. Training emergency response teams is not as straightforward as teaching general personnel basic emergency egress skills because emergency situations are always different and preparing people for a variety of possible scenarios is difficult (if not impossible) by just providing practice.

   To provide a more holistic training experience with VE training, research should focus on teaching transferable emergency response skills, such as 1) how to recognize important cues, and 2) how decision-making strategies may change in various emergency contexts. Teaching to competence is not enough for complex situations as this form of training only establishes the rules and may not provide enough exposure (or experiential learning) for people to practice testing the rules in various situations in order to see what works, what does not, and what the consequences of the actions are in different situations.

Recognition-primed decision making (RPDM) is a framework that describes how expert operators collect and store patterns during complex situations (e.g. storing information like the causation, expectancies, goals, and reactions to the situations) and how experts revert to this information to help make decisions in new situations (Klein, 2008). RPDM is applicable to complex offshore emergency decision-making situations and these training methods are needed to target proficiency (i.e. recognizing the situation) and expertise (i.e. knowing what strategies are suited to the situation). To close performance gaps with emergency response teams, researchers should investigate the use of VE technology and RPDM theory for training complex decision-making.

6.4. References

Bloom, B., S (1971). Mastery Learning. In J. H. Block (Ed.), Mastery Learning: Theory and practice (pp.47-63). New York: Holt, Rinehart & Winston.

Blundon, A. (2019). *Human reliability assessment for offshore emergency situations using a virtual environment* (Master's thesis). Memorial University of Newfoundland, St. John's NL, Canada.

Doody, K. (2018). *Evaluation of performance in emergency response scenarios: a virtual environment skill retention study* (Master's thesis). Memorial University of Newfoundland, St. John's NL, Canada.

Dreyfus, S., & Dreyfus, H. (1980). A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition. Fort Belvoir, Va: Defense Technical Information Center; 1980.

Liu, Y. (2009). Data mining and its applications in digital human modeling. In V. Duffy (Eds), *Handbook of digital human modeling: research for applied ergonomics and human factors engineering* (pp 37-1 – 37-44). Boca Raton, FL: CRC Press Taylor & Francis Group.

Griswold-Theodorson, S., Ponnuru, S., Dong, C., Szyld, D., Reed, T., and McGaghie, W. (2015). Beyond the Simulation Laboratory: A Realist Synthesis Review of Clinical Outcomes of Simulation-Based Mastery Learning, Academic Medicine, 90(11), pp. 1553-1560.

Gusky, T., 2007. Closing achievement gaps: revisiting Benjamin S. Bloom's "Learning for Mastery". *Journal of Advanced Academics*, 19, 8-31.

Han, J., Kamber, M., & Pei, J., 2011. Data mining: concepts and techniques. Elsevier.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50(3), 456-460.

Kotsiantis, S. (2007). Supervised machine learning: a review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies*. Edited by I. Maglogiannis, et al., IOS Press, 2007.

Musharraf, M., Smith, J., Khan, F., & Veitch, B. (2018). Identifying route selection strategies in offshore emergency situations using decision trees. *Reliability Engineering & System Safety*.

Ritter, F., Kim, J., Morgan, J., & Carlson, R. (2013). *Running behavioral studies with human participants: A practical guide*. SAGE Publications.

Romero, C., and Ventura, S. (2010). Educational data mining: a review of the state-of-the-art. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews.* 40(6), 601-618.

Wickens, C., Holland, J., Banbury, S., & Parasuraman, R. (2013). Chapter 7: Memory and Training, *Engineering Psychology and Human Performance*, Fourth Edition, Pearson Education Inc. p234.