ASSESSING AND PREDICTING HUMAN PERFORMANCE USING SIMULATOR DATA AND PROBABILISTIC METHODS

by

© Randy Joseph Billard

A Thesis submitted to the

School of Graduate Studies

in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

April 26, 2021

St. John's Newfoundland and Labrador

ABSTRACT

Lifeboat training is normally conducted in calm waters to minimize the risk to trainees and equipment. Practice in anything other than benign conditions is prohibited. Trainees are given little or no opportunity to practice in conditions that are probable in an emergency, including moderate sea states and reduced visibility. Coxswains are also expected to be able to deal with hazards and equipment faults although they are not exposed to these conditions in practice. Consequently, little is known about how trainees will perform in an actual emergency and the modeling of human performance in harsh environments has not been possible due to the scarcity of human performance data. With the advent of lifeboat simulator technology, it is now possible for trainees to practice in adverse weather conditions and to apply their skills in realistic emergency scenarios. Data can now be collected to assess how skills are acquired in training and how skills transfer to new operating scenarios. This data can be used to create models to investigate learning and to predict performance. The research in this proposal uses data collected from experimental studies performed with a simulator to study skill acquisition and retention, to predict human performance in emergencies, and to form models of competence that can be used to study this problem space. The thesis also provides insights on human performance and equipment limitation and uses numerical simulations to generate data of lifeboat launches into high sea states.

The thesis comprises of four research papers, presented as chapters. The first paper evaluates how the type of training received affects the performance of lifeboat operators based on their ability to complete tasks in an emergency scenario. In the second paper, Bayesian inference is used to produce models of human performance to investigate skills acquisition in new trainees transfer of skills to new scenarios. The third paper presents a method to create models of competence using

Bayesian Networks which are derived from expert prediction and experimental data. The final paper examines the performance of lifeboats in high sea states and the impact of coxswain timing on the launch performance, using data collected from numerical simulations.

The contribution of the research is 1) knowledge on the amount of practice needed to achieve and retain competence to launch an lifeboat, 2) an evaluation of how skills acquired in training transfer to new scenarios, 3) knowledge on how the type of training received affects performance in an emergency scenario, 4) insights on how much practice is needed to learn different lifeboat task types, 5) an increased knowledge of equipment performance limitations in weather conditions possible in an offshore emergency, and 6) methodologies to create probabilistic models of performance that can be used to study learning and adapt training. The study outcomes have relevance to training providers and presents methodologies that can be used to study other problem areas. The scope of work is performed in five studies using the outcomes of a human factors experiment and numerical simulations.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my supervisors Dr. Brian Veitch, Dr. Mashrura Musharraf, and Dr. Faisal Kahn for their support and guidance. I appreciate the direction and guidance provided by this talented team. I also appreciate the patience and persistence of this group to allow and encourage me to complete my studies while I continued my work career.

I extend thanks to the graduate students and researchers in the Faculty of Engineering and Applied Sciences at Memorial University with whom I have interacted over the past 7 years. A special thank you to Dr. Jennifer Smith who helped in conducting the experimental studies and provided valuable support in preparing and presenting my research.

I thank Antonio Simões Ré, Captain Anthony Patterson and again Dr. Brian Veitch for allowing me to build on the experimental research they started and for introducing me to this research team. The study of marine evacuation systems, lifeboats, and coxswain training is impactful research, and I am grateful to be able to contribute to this study area.

I would like to recognize the support provided by my colleagues at Virtual Marine who assisted in the research and who supported me while I completed my studies. A thank you to Robert Rees and Ryan Kelly who assisted in the numerical and experimental studies.

I thank the financial support provided by the Natural Sciences and Engineering Research Council of Canada - Husky Energy Industrial Research Chair in Safety at Sea, and the National Research Council Industrial Research Assistance Program.

Thank you to all the participants who volunteered for the study. This research would not have been possible without their interest and time.

I also give thanks to my family and friends for their support and encouragement. I extend a special thank you to my wife Tonya who provided support to me and our family while I completed my studies, often at the expense of our personal time.

Table of Contents

List of Tables

List of Figures

Nomenclature

AI	Artificial Intelligence
BI	Bayesian Inference
BN	Bayesian Network
CBT	Computer Based Training
CDF	Cumulative Distribution Function
CI	Credible Interval
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DNV-GL	Det Norske Veritas Germanischer Lloyd
ECD	Evidence Centred Design
FRC	Fast Response Craft
HPP	Human Performance Probability
H_s	Significant Wave Height
H_w	Wave Height
IMO	International Maritime Organization
ITS	Intelligent Tutoring Systems
LR	Life Raft
OIM	Offshore Installation Manager
PIW	Person in the Water
PLI	Pre-Launch Inspection
SB	Setback
SBA	Simulation Based Assessment
SD	Standard Deviation
SSM	Slow-Speed Maneuvering
SME	Subject Matter Expert
T	Period
T_p	Peak Period
VWT	Virtual Wave Tank

Note: additional variables are defined in Chapters 2-5 and are specific to the chapter.

1.0 CHAPTER 1: INTRODUCTION

1.1 Problem and Purpose Statement

Lifeboats are essential life-saving equipment for many types of vessels and offshore platforms, such as oil rigs. As the launch of a lifeboat is not a routine event, coxswains are required to practice regularly to maintain the requisite skills needed to launch a lifeboat in an emergency (International Maritime Organization, 2014). Lifeboat coxswains are expected to be able to launch and maneuver a lifeboat in environmental conditions that prevail in their location of operation. Coxswains are also expected to know operating procedures for inspecting and launching a lifeboat, and to be able to recognize and deal with waves, wind, reduced visibility, and hazards.

Although operators may experience challenging conditions in a real emergency, training is normally conducted in calm waters to minimize the risk to trainees and equipment. Trainees are given limited or no opportunity to practice skills in operational scenarios that represent offshore emergencies. For this reason, human performance in emergencies is difficult to predict due to the limited data that is available. Forecasts of coxswains' skill transfer to real-life operational scenarios have relied on experts' opinion. Industry studies have identified that coxswain skill has an impact on a successful lifeboat launch, yet benchmarking of lifeboat coxswain skill is difficult to assess based on the limitations in training (Robson, 2007). There is limited information on how much skills learned in lifeboat training transfer to new scenarios and adverse weather conditions. Human performance in harsh environments has not been possible to model due to the scarcity of data. The limitations of lifeboat launching equipment in high sea states has also not been fully explored. Field trials and experimental studies have investigated lifeboat performance in regular waves and wave heights as high as 10 m. The experimental studies used regular waves, which is a simplification of real conditions where wave shapes are irregular.

With the advent of lifeboat simulator technology, it is now possible for trainees to practice in weather conditions typical of their location of operation and to apply their skills in realistic emergency scenarios. Simulation provides the possibility to apply knowledge in highly contextualized environments that are representative of plausible emergencies. Data collected from a lifeboat simulator allows us to assess performance on tasks that were prohibitive to do in anything other than calm water training. In effect, new data are available to shape knowledge of human performance and investigate learning as participants practice tasks in simulator exercises. Environmental conditions used in the simulations can also be extended to irregular waves and higher wave heights to study the performance of lifeboats in extreme weather conditions.

The purpose of this research is to use simulators to investigate and predict human performance in weather conditions that could not be ethically investigated in field trials or experiments. Data are collected from experiments performed with simulators to investigate how skills are acquired in training and how these skills transfer to new scenarios that are representative of offshore emergencies. The data collected in the experiments are novel as practice could not previously be performed in the weather conditions that were used in testing. The experiments studied how the type of training received affects performance and identified tasks that require more training to achieve competence. The data were also used to create probabilistic models to predict coxswain performance in scenarios that included adverse weather and completion of multiple tasks and task types. The models also incorporated expert knowledge to improve the predictive accuracy.

An additional purpose of the research was to investigate equipment limitations in adverse weather and to study how human actions affected launch performance. Numerical simulations were used to evaluate the performance of lifeboats in extreme sea conditions that have not previously been tested. The simulations determined the weather conditions where the lifeboat could not successfully launch or clear from the launch platform due to high wave and wind forces. The numerical simulations also investigated the impact of timing of human actions on the lifeboat launch and evacuation from the launch platform.

The thesis investigates the following research objectives related to human performance and equipment limitations:

1. How much practice is needed for lifeboat coxswains to reach competence on launching tasks?
2. How does the type of training impact coxswains' ability to perform in a plausible emergency event?
3. What is the expected performance of new lifeboat coxswains as they apply skills learned in initial training to a new scenario?
4. How much practice is needed to acquire the procedural and psychomotor skills to launch and maneuver a lifeboat in plausible weather conditions?
5. Do specific tasks or task types require more initial training and practice to master?
6. Can we develop models to predict coxswain performance using experimental data and expert knowledge?
7. What is the impact of timing of human-performed actions on the probability of a successful lifeboat launch?

8. What are the limitations of human performance and equipment in high sea states?

The research aligns with current trends to use data and machine learning to investigate and improve domain knowledge. The thesis explores the use of Bayesian methods to form probabilistic models of human performance using data collected from a simulator experiment. In performing the studies, the thesis demonstrated how data collected from simulator studies are used to investigate the research questions by 1) using Bayesian inference to create cumulative distribution functions to quantify skill acquisition in trainees; and 2) creating a Bayesian Network model of competency using knowledge of task type and available performance measures. The thesis presents methodologies to generate numerical models that can integrate with artificial intelligence (AI) and machine learning algorithms. Data collected through simulator assessments can be used to model performance and gain a deeper understanding of how skills are acquired and to explore ways to improve training.

The study is relevant to training providers and researchers who aim to improve training outcomes using simulation-based assessments and numerical modeling. An outcome of the research is insight on how to apply the results, methodology, and models to study performance, improve expert assumptions, and extend to training applications where new data sets are being created. The models can be used to improve training programs, adapt training exercises to individual needs, and investigate human performance in other applications.

Chapter 1 describes the gaps in knowledge that are addressed in this thesis (section 1.2) and presents the research objectives and novelty (section 1.3). This chapter also discusses the organization of the thesis and how the research objectives were addressed in the subsequent chapters (section 1.4). This chapter also outlines how data were collected (Section 1.5) for each of the thesis chapters.

1.2 Statement of Knowledge and Gaps

The research investigates several theoretical frameworks and presents methodologies to use data collected from simulation studies to investigate learning in lifeboat coxswains. The literature review provides an overview of 1) related research with simulation, 2) background on skill acquisition and training techniques, 3) modeling of learning and competencies using probabilistic methods, and 4) lifeboat performance in high sea states. The section also identifies the knowledge gaps that are addressed in this thesis.

1.2.1 Related Research - Simulation Based Assessments

Simulators have been widely used to assess performance in operational conditions using scenario-based training exercises. Simulation-based assessment (SBA) has been used to measure cognitive and practical skills in adult learning and education (S. de Klerk, et al. 2015). Both high and low fidelity simulators have been used to investigate human performance in flight (McClernon et al. 2011) as well as medical (Stefanidis et al., 2007) and marine operations (Sellberg, 2017, Thistle et al. 2019). Lifeboat training data can now be collected to assess the amount of practice needed to acquire skills and to evaluate how skills learned in practice transfer to new scenarios. This thesis provides additional cases of how a simulator can be used to collect data and study human performance. The thesis uses data collected from SBAs to formulate probabilistic models to study performance of lifeboat operators as they learn and apply skills.

Simulators have been used to investigate human performance in lifeboat operations in ice (Power-MacDonald et al., 2011) and training in calm waters (Magee et al. 2016). The outcomes of these lifeboat studies provide background knowledge on the use of simulation to study specific problems, though the research does not include the study of skill retention, transfer of skills to new scenarios, or performance in adverse weather conditions. There are no existing studies to evaluate the amount of practice needed to acquire the skills needed to operate a lifeboat in weather conditions that are representative of an offshore emergency, as there has been no means to assess operator performance in conditions other than calm water. There are also no studies performed to evaluate how the skills acquired in training transfer to operational scenarios involving completion of multiple tasks, or in weather conditions that are typical of offshore operations. The research used human factors studies and numerical simulators to acquire data to investigate these knowledge gaps.

1.2.2 Skill Acquisition and Training

Lifeboat practice is normally conducted in benign weather conditions to minimize risk to trainees. Training conventionally included execution of a lifeboat drill that involves the launching of the lifeboat into the water, followed by the performance of simple maneuvering tasks (International Maritime Organization, 2014). The same scenario is often used in each practice event. As a result, trainees are given little opportunity to practice in different weather conditions, or to gain exposure to hazards or emergency situations. Training with the same scenario can increase comfort and decrease stress and cognitive difficulties in completing tasks (i.e. forgetting steps) for the scenario being practiced, but these benefits do not generalize to new scenarios (Baumann et al. 2011). Research has shown that gaining experience in scenarios that have similar cues and stressors as the operational environment helps trainees to improve decision making and develop mental models to improve performance (Klein, 2008). Studies have indicated that long term retention is dependent on the environment in which the actual performance will take place (Driskell et al., 1992). Similarity between the training environment and the operational environment can improve retrieval of information from memory (Arthur et al. 1998). Research has indicated variability in training encourages learners to focus on the structure of problems, providing beneficial results in training transfer (van Merriënboer et al., 2002). Some research has debated that variability in practice scenarios is not as important as the amount of practice performed, or the structure of the learning events (Van Rossum, 1990). For lifeboat training, there is little known about how skills transfer to operational scenarios that have not been practiced in training, or how learning occurs in operational scenarios that involve multiple tasks. This research thesis models learning and competence in lifeboat operators with data collected using practice events in plausible emergency scenarios.

Previous research has identified that different skill types require different amounts of practice to acquire and maintain competence. Complex tasks involving a variety of tasks, such as a lifeboat launch, are of interest as different skill types have different lengths of skill retention. Cognitive closed-loop tasks involving discrete responses and fixed sequences (e.g. pre-flight checks) are not as easily retained as continuous open-loop tasks involving tracking and problem solving (Arthur et al., 1998, Schendel, 1992, Wickens et al., 2013). Some tasks in a lifeboat launch are sequential and procedural closed-loop tasks requiring mental checks and recall of information. Other tasks are more physical and require application of motor skills to complete the tasks (e.g. opening a hook release, applying a throttle, steering). Retention of cognitive and physical skills is dependent on the type of practice performed and the amount of mental practice between assessment events (Arthur et al., 1998). Considering the type of tasks being performed, there is expected to be an establishment of both procedural memory and declarative memory as participants practice. For complex tasks involving both procedural and declarative components, it is suggested to train the procedural components first (Wickens et al., 2013). Previous research on lifeboat training has not fully identified the type of skills associated with completing lifeboat tasks. This thesis characterized lifeboat tasks based on the type of skills required to complete objectives and investigated the training needed for different task types. There has been little investigation on the difficulty of different lifeboat tasks or how much practice is needed to reach competence. The research in this thesis investigated how skills are acquired and retained for different task types.

The amount of training needed for lifeboat operators to reach competence has not been fully investigated. Research has indicated that multiple practice sessions are needed for lifeboat operators to maintain the skills needed to launch a lifeboat (Billard et al. 2018). Memory decay can occur between practice sessions, with additional practice reducing the amount of procedural and physical errors made. There is little known about how different techniques used for practice, such as simulators or live boats, affect learning and skill compared with existing training alternatives. The thesis researched the effect of using different training approaches on skill acquisition and performance on tasks in emergency scenarios.

1.2.3 Probabilistic Methods to Model Human Performance

Probabilistic methods have been used to quantify and study human performance and can combine both empirical data and assumptions to form models. This section discusses two approaches used to create probabilistic models: Bayesian inference and Bayesian Networks (BNs).

The research in this thesis demonstrates the suitability of Bayesian methods and performance data obtained from a simulator to measure learning and predict performance. Bayesian inference is a method of statistical inference which uses Bayes' theorem to update the probability of a hypothesis as more evidence or information becomes available. Studies have used Bayesian inference and data collected from simulators to investigate several problem domains where data are scarce (Groth et al., 2014, Musharraf et al. 2019). Similar to previous research, there is little data available on lifeboat coxswain performance. The thesis used Bayesian inference to develop models to study learning and evaluate task difficulty using a data set collected from a simulator study. The thesis researched how the models improved with new data and provide a method to improve models as new data is available.

Bayesian methods can be used to develop competency models that use machine learning to improve training outcomes. As discussed by Millán et al. (2002), probability distributions can be incorporated in BNs to derive models of student competence to diagnose strengths and weaknesses in trainees. Machine learning and intelligent tutoring techniques can be applied using these models to improve student assessment and course design. BNs use a graphical structure to represent the relationship between several random variables. Research has studied the interaction between tasks using Bayesian Networks to derive models of student competence (Millán et al. 2002). These models can be used to study the relationship between training factors and to examine how practice on related tasks impacts performance. Training frameworks including Intelligent Tutoring Systems (ITS) (Millán et al., 2004) and Evidence Centered Design (ECD) use observable evidence and BN models to study skills acquisition and inherent competence (Mislevy et al., 2004). The formation of a student model using BNs offers additional means to apply probabilistic models to study relationships between variables that affect performance, including the type and amount of practice received. The probabilistic modeling of the BNs can be integrated with machine learning algorithms to build adaptive training applications to customize training material to an individual's strengths and weaknesses based on evidence gathered in training. The thesis used BNs to model competence and predict performance of trainees as they practiced tasks in simulator scenarios. The methodology can combine data sets and expert knowledge to create models. The thesis evaluated if expert knowledge can be used to improve the predictive accuracy of models that are created using small data sets. Tasks completed in simulator scenarios provided evidence that was used to evaluate the model accuracy. The thesis presents methodologies to generate numerical models that can integrate with artificial intelligence (AI) and machine learning algorithms. Data capture with a simulator provides a consistent and instantaneous means to capture performance measurements through computer tracking. Digital records are created as students practice with a simulator creating a database that can be accessed to form numerical models. The data can be used to model performance and gain a deeper understanding of how skills are acquired and to explore ways to improve training.

1.2.4 Lifeboat Performance in Harsh Environments

Studies performed by Simões Ré, et al. (2002) and Simões Ré and Veitch (2004) evaluated the capabilities of evacuation systems using scale model experiments. These studies assessed the launch performance of a lifeboat in a variety of weather conditions with varying wave height, wind, and wave steepness. A key outcome of these studies was the role of wave height on the ability to perform a successful lifeboat launch. The studies indicated that lifeboat setback, or backwards displacement of the vessel in a head sea, is a key measure of performance. High setback values could result in impact with the launching platform. The studies also determined that the position of launch on the wave affected the observed setback, with setback higher when the boat is released on the trough of the wave compared to the crest.

The studies resulted in several recommendations to improve the probability of a successful launch, including training to practice launching on a wave position that reduced lifeboat setback. The scale model experiments included launches in waves emulating sea states up to 10 m, although higher waves are possible in offshore oil and gas operations. The experiment also used a regular wave shape. The impact of human performance and the ability to perform a timely launch was not investigated.

The thesis builds on the outcomes of these experiments and uses simulations to study lifeboat performance in higher sea states and irregular seas. The research also evaluates the impact on human performance on the probability of successful lifeboat launch.

1.3 Research Objectives and Novelty

The objective of the research is to use data collected from simulation studies to shape knowledge of human and equipment performance in conditions that have not been previously studied. Specifically, the thesis is concerned with investigating performance in plausible emergencies involving the launch of a lifeboat. The research used data collected from simulation studies to explore learning in lifeboat operators and skills transfer to scenarios that previously were unable to be tested due to risk. The research also evaluated the performance of lifeboats in extreme weather conditions to determine the limits of launch equipment, and to evaluate how human actions impact launch performance. The use of simulation allowed for testing to be performed in new scenarios and created new data sets that were used to model skills acquisition and performance. As indicated in Figure 1-1, the thesis combines 1) data collected from simulation-based assessments, 2) human performance modeling techniques, 3) studies of skill acquisition and training, and 4) simulated numerical models of lifeboat to explore this problem area.

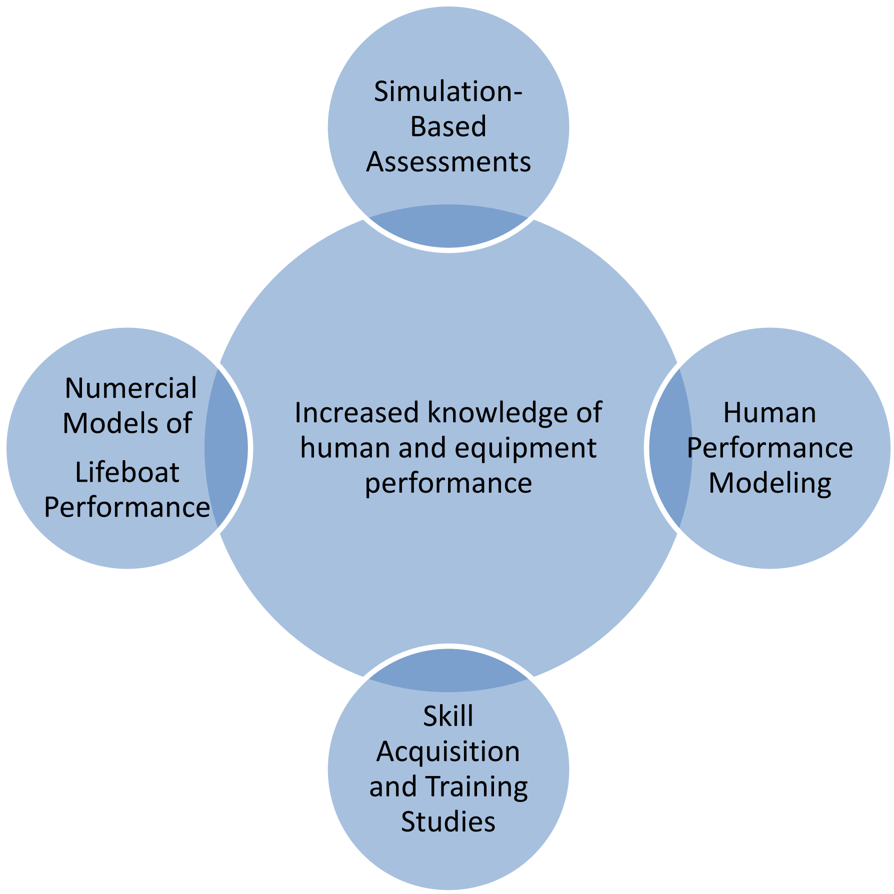

Figure 1-1: Research Overview

The thesis presents both data and methodologies to study a problem area that was previously not possible due scarcity of human performance data. This study presents outcomes specific to lifeboat training and launch equipment performance. The methodologies and approaches presented in this thesis can be applied to other problem areas where limited data are available. The thesis discusses how simulation can be used to collect novel data and how probabilistic methods can be used to model human performance and skills acquisition.

1.4 Organization of Research

The thesis includes four studies to investigate learning and model performance of lifeboat operators as they apply skills in operational scenarios. The PhD thesis is written in manuscript format and includes the following papers as chapters.

- Chapter 2 - Billard, R., Smith, J., Veitch B., (2019). *Assessing lifeboat coxswain training Alternatives using a simulator*. *The Journal of Navigation*. Published online by Cambridge University Press: 19 September 2019.
- Chapter 3 - Billard, R., Smith, J., Veitch B., (2020). *Using Bayesian Methods and Simulator Data to model lifeboat Coxswain performance.* WMU Journal of Maritime Affairs. June 2020 10.1007/s13437-020-00204-0.
- *Chapter* 4 - Billard, R., Smith, J., Masharraf, M., Veitch B., M. (2020). *Using Bayesian Networks to Model Competence of Lifeboat Coxswains.* Transnav International Journal of Marine Navigation and Safety of Sea Transportation. Journal Vol. 14., No. 3, September 2020.
- Chapter 5 – Billard, R., Rees, R., Veitch, B., Simões Ré, A. (2020). *Use of simulations to predict lifeboat survivability in extreme waves and the effectiveness of coxswain performed actions* (Unpublished Manuscript). Submitted to International Journal Maritime Engineering.

Chapter 2 evaluates how the type of training received affects the performance of lifeboat operators as demonstrated by their ability to complete tasks in an emergency scenario. This study investigated the performance of trainees who received different types of training over a year long period, representing the alternative training means available to lifeboat coxswains. These alternatives are: (1) using a lifeboat, (2) using Computer Based Training (CBT), and (3) using a lifeboat simulator. The three alternatives vary in the amount of hands-on practice provided in a training session, and the difficulty and variability of the scenarios that are used in training. The study investigated how these factors impacted skills acquisition and performance by making comparisons between separate groups of individuals trained in one of the three alternative ways.

Chapter 3 uses simulator data to examine lifeboat coxswain training and skill transfer as trainees practice lifeboat tasks for the first time. Bayesian inference is used to produce probabilistic models of human performance to study how skills transfer from initial training to a new practice scenario. The models are used to investigate the amount of practice needed to become competent and to compare the difficulty of different task types using evidence collected in a scenario with calm water conditions. An outcome is the creation of sets of cumulative distribution functions (CDFs) to quantify skill acquisition in a group of new trainees as they enter a training program designed to prepare coxswains for offshore emergencies involving a lifeboat.

Chapter 4 presents a methodology to evaluate the performance of lifeboat operators as they apply their skills in scenarios that are more difficult than scenarios used in initial training and practice. A BN is used to define a model of the competence of slow-speed maneuvering (SSM) based on tasks performed in adverse weather conditions. The model is derived from a combination of expert prediction and data collected from an experimental study. The methodology created a student model of SSM competence that can be used for the prediction of performance on tasks and the diagnostic study of causal relationships between model variables.

Chapter 5 investigated the performance of lifeboats in high seas states and the impact of timing of coxswain actions on launch performance. Numerical simulations were performed in extreme wave conditions that had not been explored, including irregular wave heights up to 14 m. Launch performance was evaluated based on the amount of setback and the ability to exit the lifeboat launch area in a timely manner. The study also investigated how the timing of human actions, including the application of throttle and release of hooks, impacted launch performance. Table 1-1 outlines the related research objectives and methodologies that are investigated in each chapter.

Table 1-1: Research Questions and Methodologies

Chapter	Research Question(Q) or Methodology (M)
2	• Q - How does the type of training impact coxswains' ability to perform in a plausible emergency event? • Q - Do currently utilized training alternatives provide enough practice to acquire the skills needed to perform a lifeboat launch in likely weather conditions?
3	• Q - What is the expected performance of new lifeboat coxswains as they apply skills learned in initial training to a new scenario? • Q - How much practice is needed for lifeboat coxswains to reach competence on launching tasks? • Q - Do specific tasks or task types require more initial training and practice to master? • M - Using Bayesian inference to create cumulative distribution functions (CDFs) to quantify skill acquisition in trainees
4	• Q - Can expert prediction and knowledge of task type be combined to model trainee competence? • M - Creating a Bayesian Network model of competency using knowledge of task type and available performance measures
5	• Q - What is the expected setback of a lifeboat in extreme regular waves and irregular waves? • Q - How is the time to clear the lifeboat from the launch structure affected by sea state? • How does delay in lifeboat throttle and hook release affect launch and evacuation of a lifeboat?

1.5 Study Resources – Simulator Experiment and Numerical Simulations

The first three studies (Chapters 2 to 4) use data collected from a human factors study performed with a live boat and simulator. The final study (Chapter 5) uses a numerical simulation to study performance in high seas. Further details on these studies are provided in the following sections.

1.5.1 Human Factors Experiment

The first three studies use outcomes of an experiment designed to evaluate lifeboat training programs. A test program was developed to emulate practice provided to new lifeboat coxswains. The test program was approved by the National Research Council of Canada Research Ethics Board. Initial training was provided at a shore-based facility using a live boat and presentation materials. Following initial training, over a one-year period participants received quarterly training in one of three ways: using live boats, computer-based training, or a simulator. Trainees then applied their skills in a simulator scenario that represented a plausible emergency requiring completion of lifeboat launch and on-water maneuvering tasks.

Data from the experiment is used to study skills retention and to compare different training alternatives (Chapter 2), to study skill acquisition and form models of human performance for initial training (Chapter 3) and to model competence in adverse weather conditions (Chapter 4). Each of the studies use specific data from the human factors experiment to provide insights on performance and present methodologies to study the problem space of lifeboat training. Additional details are provided in sections Chapters 2 to 4.

1.5.2 Test Equipment

Participants completed tasks in a simulator with a representative layout and equipment of the real lifeboat. The simulator is equipped with real lifeboat equipment (e.g. steering wheel, throttle, brake release, compass) allowing participants to operate the controls needed to launch the lifeboat in a simulation environment complete with visuals and sounds. The simulator is certified by Det Norske Veritas Germanischer Lloyd (DNV-GL) and Transport Canada as being capable of representing realistic situations needed for training. The simulated lifeboat motion, equipment, and layout were modeled to be the same as the real lifeboat. Components of the training were also performed with a real lifeboat. The type of lifeboat used in the study is currently used on offshore platforms in the North Atlantic. The lifeboat can carry up to 72 people and is approximately 9.4 m long, 3.5 m wide and 6 m high, with a draft of 2.9 m. Its empty weight is approximately 5806 kg and has a fully loaded weight of approximately 11,500 kg. Figure 1-2 shows the simulator and lifeboat used in the human factors experiment.

Figure 1-2: Lifeboat Simulator and Lifeboat used in Experimental Study

1.5.3 Numerical Simulation – Virtual Wave Tank

Chapter 5 uses a numerical simulation to study lifeboat performance in high sea states and the impact of human factors on launch success. The study measures the impact of timing on completion of key tasks, including releasing the lifeboat hook and applying throttle. Numerical simulations are performed using a simulator test environment developed by Virtual Marine. A virtual wave tank was adopted from Virtual Marine's simulator architecture designed to provide accurate models of small vessel water entry and maneuvering in waves. An image of the virtual wave tank is provided in Figure 1-3.

Environmental factors such as wave heights, periods, and wave shape (regular or irregular) can be changed in the virtual wave tank. Fast-time simulations can be performed in the virtual wave tank using a programmed driver to perform actions that would be conducted by a human in a lifeboat launch, including releasing the hooks and applying throttle. Timing can be controlled to delay the hook release and throttle application. These controls are used to examine the impact of timing of human actions on the successful launch of a lifeboat.

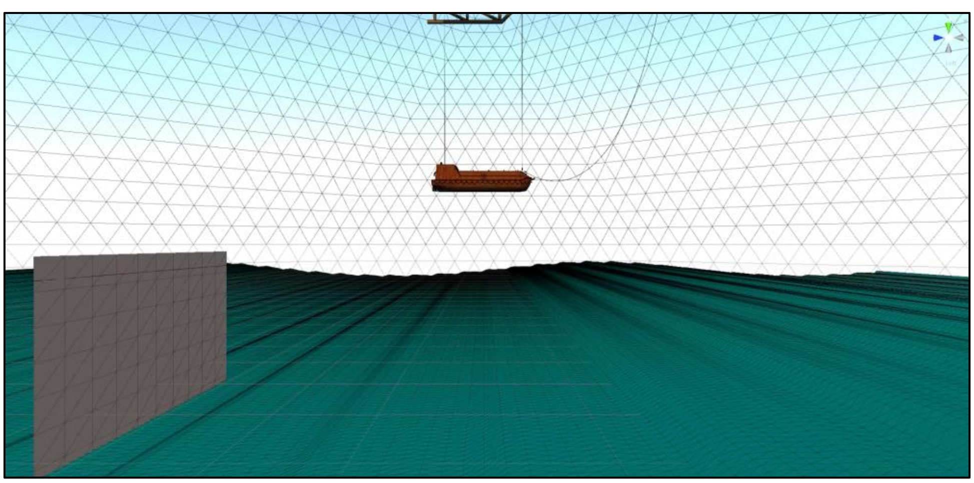

Figure 1-3: Virtual Wave Tank

1.6 References

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, **11(1)**, 57-101.

Baumann, M. R., Gohm, C. L., & Bonner, B. L. (2011). Phased training for high-reliability occupations live-fire exercises for civilian firefighters. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **53(5),** 548-557.

Billard R., Magee, L.E., Smith, J.J.E, Veitch, B (2018). Simulator Training for Offshore Oil and Gas Emergency Preparedness, International Training and Education Conference (ITEC) 2018.

de Klerk, S., Veldkamp, B.P., Eggen, T., (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. Computers & Education 85 (2015), 23-34.

Driskell, J.E., Willis, R.P., & Copper, C. (1992). *Effect of overlearning on retention. Journal of Applied Psychology*, **77(5),** 615-622.

Groth K., Smith, C., Swiler, L. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability and System Safety 128 (2014), 32-40

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). STCW including 2010 Manila Amendments, 2017 Edition.

Klein, G., (2008), Naturalistic decision making. *Human Factors: The Journal of Human Factors and Ergonomic Society*, **50(3)**, 456-460.

Magee, L.E., Smith, J.J.E., Billard, R., & Patterson, A. (2016). Simulator training for lifeboat maneuvers. Proceedings, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 16030.

Millán, E., Perez-De-La-Cruz, J.L., (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction 12: 281-330, Kluwer Academic Publishers, Netherlands

Millán , E., Loboda, T., Perez-de-la-Cruz, J.L. (2010). Bayesian networks for student model engineering. Computers and Education, 55, 1663-1683

Musharraf, M., Moyle, A., Khan, F., Veitch B. (2019) Using simulator data to facilitate human reliability analysis. Journal of Offshore Mechanics and Arctic Engineering. Vol. 141(2).

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(3), 207-218.

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design. CSE technical Report. Los Angeles: The National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://www.cse.ucla.edu/products/reports/r632.pdf.

Power-MacDonald, S.P., MacKinnon, S., Simões Ré, A., Power, J., & Baker, A. (2011). Effects of simulator training on novice operators' performance in simulated ice covered waters. National Research Council Canada, TR-2011-15.

Robson, J. K., (2007). Overview of TEMPSC performance Standards. Health and Safety Executive Research Report RR599. Norwich.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU Journal of Maritime Affairs, 16(2), 247-263.

Schendel, J., Hagman, J, (1982), On sustaining procedural skills over a prolonged retention interval. Journal of Applied Psychology, Vol 67 (5), 605-610.

Stefanidis, D., Korndorffer, J.R., Markley, S., Sierra, R., Heniford, B.T., & Scott, D.J. (2007). Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training? *Journal of the American College of Surgeons*, **205(2)**, 307-313.

Simões Ré, A., Veitch, B., and Pelley, D. 2002. Systematic investigation of lifeboat evacuation performance. Transactions, Society of Naval Architects and Marine Engineers, 110:341-3; and,

Simões Ré, A., Veitch, B., 2004, Evacuation Performance of Davit Launched Lifeboats, Proceedings of the 23rd OMAE 2004, Paper 51526.

Thistle, R., Veitch, B, (2019). An evidence-based method of training to targeted levels of performance. The Society of Naval Architects and Marine Engineers. Document SNAME-SMC-2019-030.

van Merriënboer, J. J. G., Clark, R.E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, **50(2)**, 39–64.

Van Rossum, J. H. (1990). Schmidt's schema theory: The empirical base of the variability of practice hypothesis: A critical analysis. Human Movement Science, 9(3), 387-435.

Wickens, C., Hollands, J., Banbury, S., Parasuraman, R. (2013). *Engineering Psychology and Human Performance, Fourth Edition,* Pearson.

2.0 CHAPTER 2: ASSESSING LIFEBOAT COXSWAIN TRAINING ALTERNATIVES USING A SIMULATOR

Randy Billard[1], Jennifer Smith[2], and Brian Veitch[3]

[1] Virtual Marine, [2]private consultant, [3]Memorial University of Newfoundland

2.1 Co-authorship Statement

This manuscript has been published in the Journal of Navigation (2019). Writing was led by Randy Billard, with results verified by author Jennifer Smith who assisted in conducting the experiment. Brian Veitch provided guidance in writing, presenting results, and revisions to the paper.

2.2 Abstract

Lifeboats are essential life-saving equipment for all types of vessels and offshore platforms. Lifeboat simulators have been created specifically for offshore personnel to practice in conditions that are normally too risky for live training. As simulation training is a relatively new alternative, there is a need to assess how training performed with a simulator compares to conventional training. A study was performed to evaluate how skills acquired with different training approaches transferred to an emergency scenario. Over a period of one year, participants received quarterly training in one of three ways: using live boats, computer-based training, or a simulator. Following training, participants were evaluated on their ability to launch and maneuver a lifeboat in a plausible emergency. The study suggests a benefit to performing training with realistic lifeboat controls and practicing using representative emergency scenarios. Insights are provided on how training can be modified to increase competence.

2.3 Introduction

Lifeboat operators are required to have the essential skills needed to launch a lifeboat in an emergency. As the launch of a lifeboat is not a routine event, coxswains are required to practice regularly to maintain the requisite skills. Lifeboat coxswains are expected to be able to launch and maneuver a lifeboat in environmental conditions that prevail in their location of operation. Coxswains are also expected to know operating procedures for inspecting and launching a lifeboat and be able to recognize and deal with waves, wind, reduced visibility, and hazards.

Although operators may experience challenging weather conditions in an emergency, training is normally conducted in calm waters to minimize risk to trainees and equipment. Lifeboat coxswains typically complete initial training at an onshore training facility and then perform regular recurrency training to maintain their skills. Recurrency training is normally conducted on the job every three months and has traditionally included execution of a lifeboat drill that involves the launching of the lifeboat into the water, followed by the performance of simple maneuvering tasks (International Maritime Organization, 2014). An alternative training means is to do onshore refresher training annually, or every two years, and to refresh skills quarterly through self-study (reading operations manuals or inspecting the launch equipment). A recent alternative, introduced in 2010, is to use immersive digital simulators for training. This alternative uses virtual cues and representative lifeboat equipment to perform training scenarios, instead of using a real lifeboat or onshore facilities.

Training alternatives vary in the amount of physical realism, scenario-based practice, and capacity to expose trainees to realistic conditions. With multiple training options available for lifeboat coxswains, there is an interest to assess the value and limitations of each alternative.

This study investigated the performance of trainees who received different types of training over a year long period, representing the alternative training means available to lifeboat coxswains. These alternatives are: (1) using a lifeboat, (2) using Computer Based Training (CBT), and (3) using a lifeboat simulator. The three alternatives vary in the amount of hands-on practice provided in a training session, and the difficulty and variability of the scenarios that are used in training. We investigated how these factors impacted skills acquisition and performance by making comparisons between separate groups of individuals trained in one of the three alternative ways.

The objectives of the research were to investigate the following:

1. How does the type of training impact performance in a plausible emergency event?
2. Do currently utilized training alternatives provide enough practice to acquire the skills needed to perform tasks in likely weather conditions?

The primary objective was to assess how the different training alternatives promoted skill development. Performance was assessed by comparing the ability to successfully complete all launch and maneuvering tasks in a simulated emergency exercise that included realistic weather conditions and a credible hazard. The performance in the scenario is an indicator of skills acquired during training and transferred to a plausible emergency event. A secondary objective was to observe individual task performance and investigate the common errors made by coxswains during assessment.

We investigated the launch of the lifeboat and on-water maneuvering separately. Launching the lifeboat is the primary duty of the coxswain as they evacuate from the oil and gas platform. Once in the water, coxswains may be required to participate in rescue exercises, rendezvous with other vessels, or rescue casualties. The ability to complete on-water tasks is dependent on the coxswain's ability to maneuver the boat in waves. Analyzing the performance on individual tasks allowed us to assess the difficulty of the task type and provided insights that may be used to direct training.

2.4 Background

Many factors can affect skill acquisition and retention. These include the similarity of the practice environment to the test environment, the variability in the training exercises, and the frequency and amount of training that is received.

Lifeboat practice is normally conducted in benign weather conditions to minimize risk to trainees. The same scenario is often used in each practice event. As a result, trainees are given little opportunity to practice in different weather conditions, or to gain exposure to hazards or emergency situations. Training with the same scenario increases comfort and decreases stress and cognitive difficulties in completing tasks (i.e. forgetting steps) for the scenario being practiced, but these benefits do not generalize to new scenarios (Baumann et al. 2011). Research has shown that gaining experience in scenarios that have similar cues and stressors as the operational environment helps trainees to improve decision making and develop mental models to improve performance (Klein, 2008). Driskell et al. (1992) also indicate that long term retention is dependent on the environment in which the actual performance will take place. Similarity between the training environment and the operational environment has been shown to improve retrieval of information from memory (Arthur et al. 1998). Billard et al. (2018) identified that changing the environmental conditions of a lifeboat launch impacted the ability to complete tasks that were previously mastered through training.

Variability in training exercise has been shown to encourage learners to focus on the structure of problems, providing improved training transfer to new exercises (van Merriënboer et al., 2002). Others have debated that variability in practice scenarios is not as important as the amount of practice given to master tasks, or the structure of the learning program (Van Rossum, 1990). Stepping up the difficulty of the task over time in training is a common practice that is well established in other training domains, such as flight training. Lim et al. (2009) report that several research studies have shown that variability of practice usually results in beneficial effects on transfer of training.

2.5 Methodology

A test program was developed to emulate practice provided to new lifeboat coxswains. Initial training was provided at a shore-based facility. This was followed by quarterly practice events. Participants received initial training (explained in section 3.1) and then received quarterly training in one of three ways. Following three quarterly practice events, an assessment exercise measured how skills acquired in the training program transferred to a plausible emergency event that required launch of a lifeboat in weather conditions typical of offshore operations. Figure 2-1 illustrates the elements of the study.

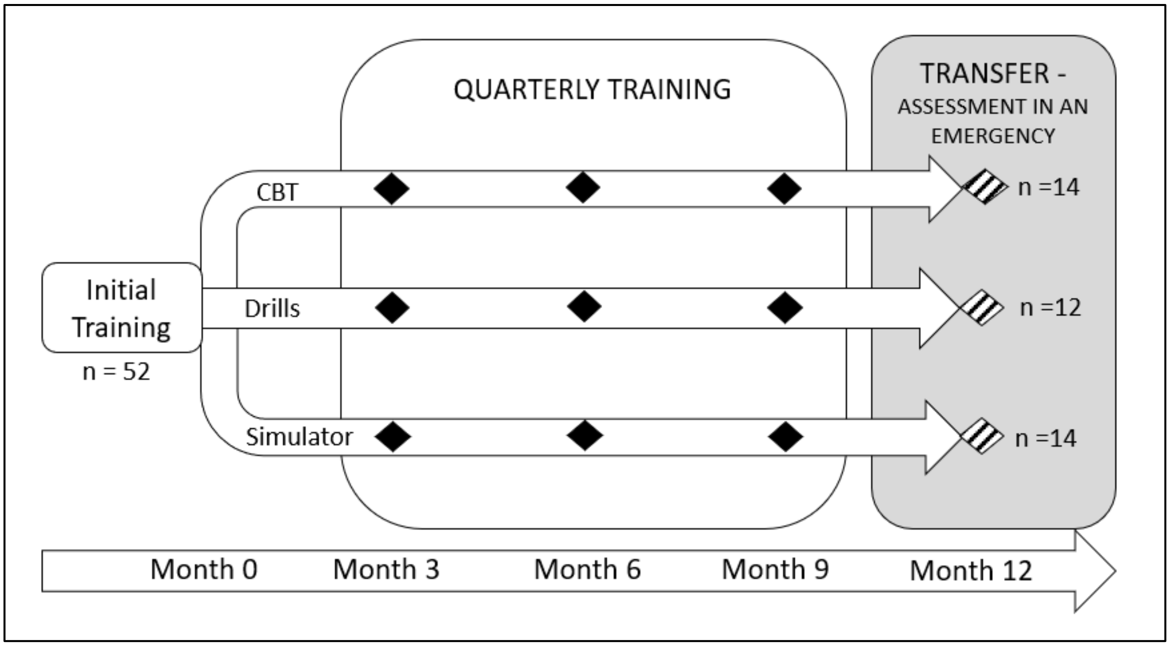

Figure 2-1: Study Timelines

In the study, we can examine launch and maneuvering tasks separately based on the skill types needed to complete tasks. The launch of a lifeboat is primarily a procedural cognitive task requiring the trainee to recall the steps required to perform a launch, the order of the steps, and the recognition of equipment faults. Maneuvering the lifeboat is primarily a psychomotor task and requires application of physical skills to control the lifeboat. Psychomotor tasks are usually less easily forgotten and require less frequent rehearsal to be remembered than cognitive tasks, which can require frequent practice and rehearsal (Stewart et al. 2008).

As noted by Arthur et al. (1998) the length of the retention interval, or non-practice period, is known to be a significant factor in skills retention. For the purpose of the study, we kept the retention interval consistent between the different training approaches. Training was provided at intervals that match industry practice of quarterly training (i.e. every 3 months). We compared differences in performance based on the type of training received.

Assessing performance in emergency conditions could only be done using a simulator due to the risks associated with live boat operations. Simulation has been widely used to assess human performance in flight (McClernon et al. 2011), as well as medical (Stefandis et al. 2007) and marine operations (Sellberg, 2017). Using a simulator that was representative of a real lifeboat provided a means to assess performance in an exercise that would otherwise be prohibitive due to risk.

Participants with no previous lifeboat experience or training were recruited for the study. Recruits were required to be unfamiliar with the lifeboat operation and launch procedure; they were not allowed to participate if they had previous lifeboat experience. Fifty-two volunteers between the ages of 18 and 65 were recruited. After initial assessment, two groups of 17 and one group of 18 participants were formed. Twelve participants dropped out of the study due to time commitment and scheduling conflicts, as the experiment was carried out over a year. Due to uneven attrition, two groups finished with 14 participants and the one group finished with 12 participants.

The training emulated practices used in industry, with controls added to make the training safe for the participants and to maintain consistency in training events.

2.5.1 Phase I – Initial Training and Grouping

Initial training of all participants consisted of a combination of classroom training from an instructor and familiarization exercises with a simulator. Participants were taught a sequence of actions needed to safely launch a lifeboat from davits, and to perform on-water maneuvering tasks using training designed to teach competencies identified in the Standards of Training, Certification and Watchkeeping for Seafarers (International Maritime Organization, 2010). The training materials covered basic operation of the lifeboat, coxswain duties, pre-launch inspection procedures, clear away procedures, navigation procedures and radio communication. The training course was conducted by an instructor who was experienced in small craft training and was proficient in operating the lifeboat used in the study.

Training also included a guided tour of a real lifeboat to familiarize the participants with the appearance and location of lifeboat equipment, thereby providing knowledge needed to inspect the lifeboat prior to launch. After training was completed, participants were given a fifteen-minute simulator exercise to become familiar with the lifeboat simulator.

Following initial training, participants completed a simulator assessment scenario designed to evaluate the fundamental skills required to operate a lifeboat, which included the launch and control of the lifeboat in calm weather conditions. To ensure a baseline competence was achieved, participants repeated the scenario until they were able to complete all tasks. The assessment scenarios was used to score personnel on their ability to complete launch and maneuvering tasks, with the number of trails to criterion used to rank performance. The rankings were used to balance the groups as evenly as practicable.

2.5.2 Phase II – Quarterly Training

Participants practiced quarterly three times using the training approach assigned to their group. Table 2-1 summarizes the type of training received by each group and the tools used for training. As outlined in this table, the training received by each group was configured to match the conditions used in practice.

Table 2-1: Training received by Group Designation

Group	Representative Training Practice	Launch Tasks	Maneuvering Tasks	Scenario Parameters	Faults and Hazards
Group 1 - Drills	**Live offshore quarterly drills from an offshore platform**	Practiced in simulator with real lifeboat equipment	Practiced on-water using real boat	Same scenario each training session, limited to calm waters,	None
Group 2 - CBT	**Annual refresher training with skills maintained quarterly through self-study**	Desktop CBT based on operating manuals	Desktop CBT based on operating manuals	N/A – no scenario practice used	Covered in CBT
Group 3 - Simulator	**Simulation-based training programs in use in Oil and Gas Training**	Practiced in simulator with real lifeboat equipment	Practiced in simulator with real lifeboat equipment	Progressive with each training session, calm to moderate sea state	Introduced as scenarios progressed

Drills training, assigned to Group 1, consisted of practice emulating live offshore quarterly drills, which typically include the launch of a lifeboat in calm water and simple maneuvering exercises in the water. To minimize risk to trainees, the launch task was performed using a simulator. Maneuvering was performed using a real lifeboat in calm weather conditions. To emulate industry practice, the launching conditions for each practice drill was calm water with no equipment faults or hazards.

Group 2 participants were representative of trainees who perform annual refresher training onshore and who do not perform regular practice drills or scenarios following initial training. For this type of trainee, skills are maintained through self-study of launch procedures and operating manuals.

A Computer Based Training (CBT) program was developed to provide participants with relevant training materials delivered via a desktop computer. Course materials were derived from operation manuals of the davit and lifeboat, and included materials to familiarize trainees with the layout and operation of lifeboat equipment and the launching procedure.

Group 3 represented users of lifeboat simulators. The members of this group were provided with quarterly training in scenarios of escalating difficulty. Launching and maneuvering were practiced in the simulator. The difficulty of the scenarios increased every quarter, with initial scenarios starting with calm water (low difficulty) and progressing to launches in moderate seas in the final quarterly training period. Scenarios were developed by a subject matter expert to increase the difficulty of tasks over time and provide exposure to scenarios that are not practiced in the real boat due to associated risks.

For the Drills and Simulator groups, participants practiced in each quarterly training session until they achieved the baseline competence, meaning they were able to complete all launch and maneuvering tasks successfully in the practice scenarios. For the CBT group, participants had to pass a multiple-choice test based on the training materials to demonstrate that they knew the course materials.

2.5.3 Phase III – Transfer to an Emergency Scenario

After completion of three quarterly training sessions, and following an additional 3 months without practice, participants performed a scenario in weather conditions that were representative of common operating conditions in the North Atlantic (C-Core, 2015). The parameters of the scenario were set to night time with clear visibility, 13 knot winds, and a 3-meter wave height. To eliminate risks to the participants, all testing made use of the simulator as the test environment for both launch and maneuvering tasks.

All participants were given the same scenario and scenario briefing. The briefing indicated that an explosion had been heard on the platform, followed by a fire alarm. The Offshore Installation Manager (OIM) had ordered an evacuation from the platform and the duty was to launch the lifeboat and assist in a search and rescue exercise once in the water. The evaluation scenario was more difficult than the hardest scenario that was provided in any of the training exercises, including those given to the simulator group. Figure 2-2 provides an overview of the emergency scenario. This image was provided to trainees in the briefing.

Participants were required to perform a pre-launch inspection (PLI) of the lifeboat and then launch the lifeboat and clear from the oil and gas platform. As a fire was active, participants had to turn on the air and sprinkler system to minimize the chance of harm. Once in the water, the participant had to participate in a rescue exercise that included locating and recovering persons in the water (PIWs) and transferring personnel to a fast response craft (FRC). The weather increased the difficulty of some launch tasks practiced in training, including releasing the lifeboat when in the water. It also made on-water tasks more difficult.

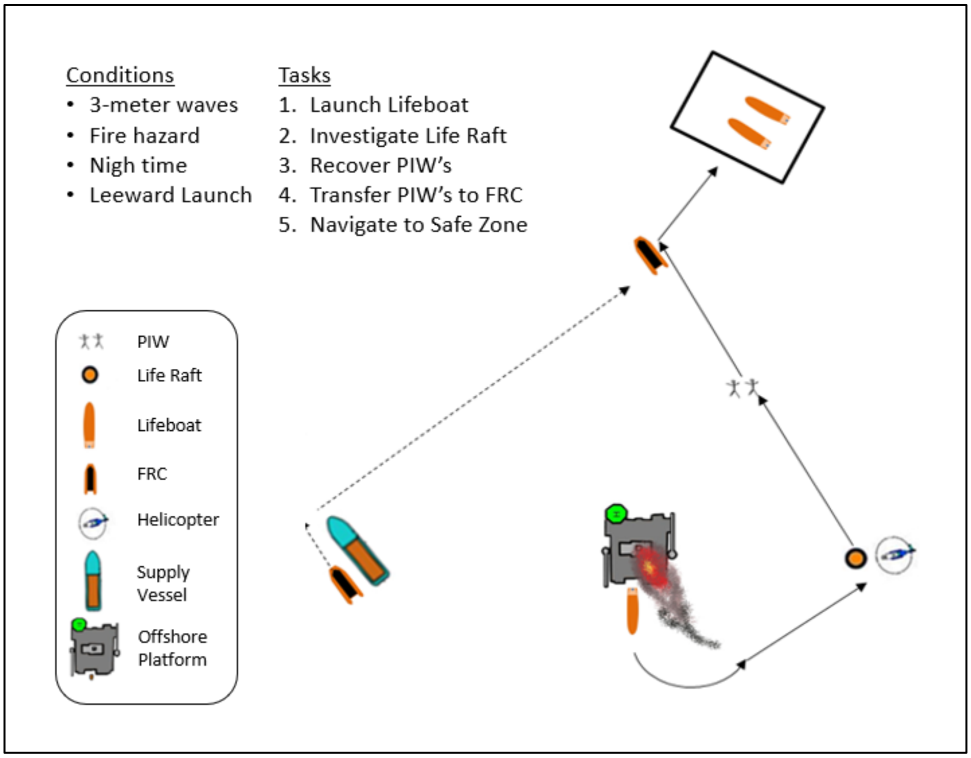

Figure 2-2: Emergency Scenario

2.5.4 Performance Measurements

A scoring rubric was established to measure performance for launching and maneuvering tasks in the practice sessions and the emergency test scenario. The criteria for task completion were established by a subject matter expert to reflect a standard of proficiency as identified in recognized training standards, including the Standards of Training, Certification and Watchkeeping for Seafarers (International Maritime Organization, 2010). This standard is commonly used to model lifeboat training courses. Table 2-2 provides a list of tasks and objectives used to measure performance. Launch actions are expected to be performed in sequence, except for operating the sprinkler system, which can occur any time before entering the water. The order, type, and number of on-water tasks are dependent on the scenario being practiced.

For the Drills Group (Group1), quarterly training sessions allowed for practice of 9 of 10 of the launch tasks identified in Table 2-2, as there was no training scenario that required the operation of the sprinkler and air system. Training on the operation of the air and sprinkler system was provided to all group members during the initial training session. On-water tasks comprised of maneuvering a short distance followed by a single task of approaching and stopping next to a vessel. The Simulator Group (Group 2) practiced navigating by compass and PIW pickup in each of the quarterly practice sessions, and practiced stopping next to a vessel in one practice scenario. The Simulator Group did not practice in environmental conditions as difficult as the emergency test scenario. No launching or maneuvering practice was provided to the CBT Group (Group 3). The CBT Group reviewed launch procedures and the equipment manuals. This allowed for mental rehearsal of the launch tasks but did not allow a practice launch in a scenario.

In the emergency test scenario, a total of 10 launch tasks were required to be completed. The on-water tasks included two navigate-by-compass tasks and two PIW pickups, resulting in 10 total on-water tasks. All launch tasks were completed in the simulator, except for the PLI, which was conducted prior to starting launch. A PLI normally involves visual examination of the lifeboat's exterior and interior. In the study, pictures were used instead and the trainee had to determine if a given picture represented a correct or incorrect state of equipment. Identification of correct state of equipment was needed to allow for a safe launch (i.e. removal of maintenance pendants, brake cable present).

Table 2-2: Task Objectives

	Task Name	Task Objective
Launching Tasks	**Pre-Launch Inspection (PLI) – Critical Errors**	Perform visual inspection of lifeboat in preparation for launch and ensure no equipment is stopping vessel launch.
	Permission to Launch	Obtain permission to launch from OIM.
	Inform Crew Prior Launch	Inform crew prior to launch – "Launching, Launching."
	Lower w/o stopping	Pull brake release, lower lifeboat without stopping by keeping tension on release.
	Air and Sprinkler	Order the use of air system and sprinkler after being informed of gas, smoke or fire.
	Engine Started	Ensure engine is started before lowering/splashdown using engine turn key.
	# of re-entries	Ensure lifeboat completely enters water and is fully buoyant before releasing hooks by looking at hydrostatic indicator on hook release or visual cue.
	Splashdown zone	Promptly release hooks using hook handle release and apply throttle
	Contact with platform	Maneuver vessel and do not make contact with platform after release of hooks.
	Clear Away Zone	Safely leave clear away zone by moving away from platform quickly and avoiding hazards.
On-water Tasks	**Navigate by compass**	Maintain a compass heading with minimal veer from target heading and control heading.
	Approach a Mark	Approach a static object accounting for wind and wave direction. Use a speed to allow stopping.
	Stop at a Mark	Stop close to landmark (2-3 boat lengths) and maintain position.
	Approach a Person in the Water (PIW)	Approach a drifting PIW accounting for wind and waves to minimize chance of contact. Use a speed to allow stopping.
	Recover a PIW	Stop close enough to PIW to allow pickup and maintain position.
	Navigate to a landmark	Maintain a heading in line with a target landmark and control heading and veer.
	Approach a vessel	Approach a static object accounting for wind and wave direction. Use a speed to allow stopping.
	Come alongside a vessel	Stop next to vessel close enough and at an angle to allow personnel transfer and maintain position.

For each of the assessments and training sessions performed using a simulator or live boat, an instructor evaluated the participants based on the rubric. For practice scenarios requiring voice command (e.g. instructing someone to turn on the sprinkler, requesting permission to launch) the instructor role-played as a crew member and OIM as circumstances warranted. If the student requested assistance from the instructor to complete any task, the task was considered to be incomplete. For the CBT group, completing the quiz at the end of the review of materials was an indicator the trainee knew the materials that had been presented.

2.5.5 Simulator and Lifeboat

The type of lifeboat used in the study is currently used on offshore platforms in the North Atlantic. The lifeboat can carry up to 72 people and is approximately 9.4 m long, 3.5 m wide and 6 m high, with a draft of 2.9 m. Its empty weight is approximately 5806 kg and has a fully loaded weight of approximately 11,500 kg. For the study, the lifeboat was empty except for the participants and an operator who monitored student performance and ensured safe operation of the vessel. All testing performed with the lifeboat was conducted in a sheltered harbor. Figure 2-3 show the simulator and the lifeboat used in the study.

Participants completed tasks in a simulator with a representative layout and equipment of the real lifeboat. The simulator is equipped with real lifeboat equipment (e.g. steering wheel, throttle, brake release, compass) allowing participants to operate the controls needed to launch the lifeboat in a simulation environment complete with visuals and sounds. The simulator is certified by Det Norske Veritas Germanischer Lloyd (DNV-GL) and Transport Canada as being capable of representing realistic situations needed for training. The simulated lifeboat motion, equipment, and layout were modeled to be the same as the real lifeboat. This simulator has been used in previous studies to measure skill transfer (Magee et al., 2016) and skill retention (Billard et al., 2018).

Figure 2-3: VMT Lifeboat Simulator Interior and Lifeboat

2.6 Results

Our principal measure of performance was the ability of participants to complete all tasks successfully on their first attempt in the emergency test scenario. The results of first attempts provides an indicator of skills retained from practice and an indicator of skills transferred from practice to a new plausible event. In a real emergency scenario, a successful launch of the lifeboat would require all steps to be completed correctly based on the established performance criteria. Failure to complete any of the tasks could result in harm to the crew or damage to the lifeboat. To make comparisons between the type of training received, we analyzed the performance on tasks on the first attempt at the test scenario launching and maneuvering tasks.

The frequency of errors made during launching and maneuvering was also investigated. An analysis of individual tasks provides a more granular measure of performance and indicates which tasks resulted in the most frequent errors for each of the sub-groups. Launching and maneuvering tasks were compared separately as they require different types of skills.

The final group sizes create some statistical uncertainty in comparison of performance between groups. Statistical comparisons assume small sample sizes and use appropriate significance levels for acceptance of differences.

2.6.1 Success on First Attempt at an Emergency Launch and Maneuvering Tasks

The results shown in Table 2-3 summarize the performance of each group on their first attempt to launch the lifeboat in the emergency test scenario. The table shows that six members of the Simulator Group and two members of the Drills Group were successful on their first attempt at the launch task, and that no one in the CBT Group was successful.

Table 2-3: Frequency of Success on First Attempt at the Launch task

	Group		
	Simulator	Drills	CBT
Success	6	2	0
Failure	8	10	14
	14	12	14

As indicated in Figure 2-4, this corresponds to 17% of the Drills Group and 43% of the Simulator Group being able to complete all tasks successfully on first attempt. Overall, 20% of all participants were able to successfully launch the lifeboat on first attempt.

Pair-wise comparisons of the groups were conducted using Fisher Exact Probability Tests (one tailed) with $p \leq 0.10$ as the critical value for rejecting the hypothesis that the performance of the groups is the same. The Simulator Group did not have reliably more successes than the Drills

Group, $p > 0.10$, but did have reliably more successes than the CBT Group, $p \leq 0.10$. There is no reliable difference in the CBT Group and the Drills Group, $p > 0.10$.

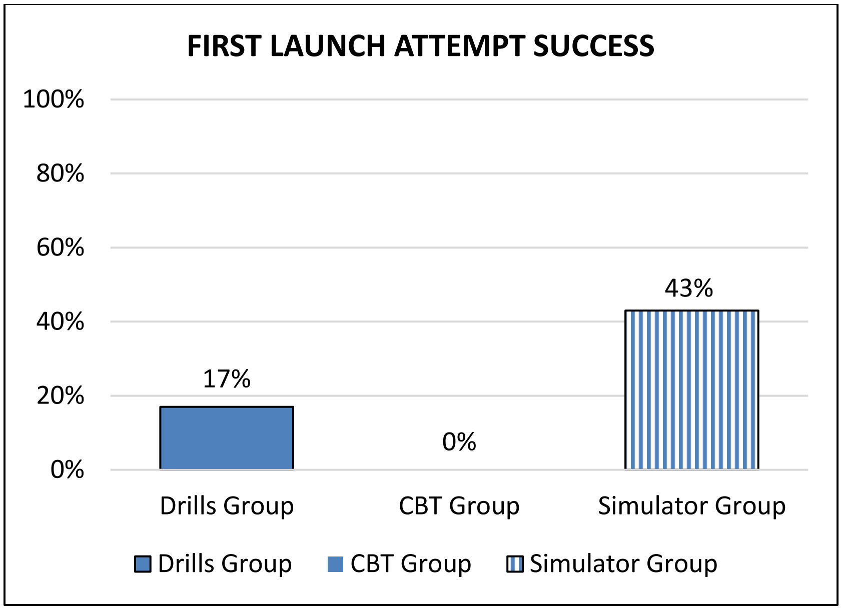

Figure 2-4: Percentage of Success on First Launch Attempt

No participants in either group could successfully complete all maneuvering tasks on their first attempt in the emergency test scenario. The inability to achieve the minimum standard of performance for one or more subtasks accounts for the low success rates for the maneuvering task on first attempt, since success depended upon doing all component subtasks. An analysis of frequency of errors for each subtask is made in the next section to compare the performance of the groups on maneuvering tasks.

2.6.2 Frequency of Errors Made on Launching and Maneuvering Tasks

Figure 2-5 shows the frequency of errors made by each of the groups during the launch and maneuvering tasks. A lower number of errors is an indicator of higher performance.

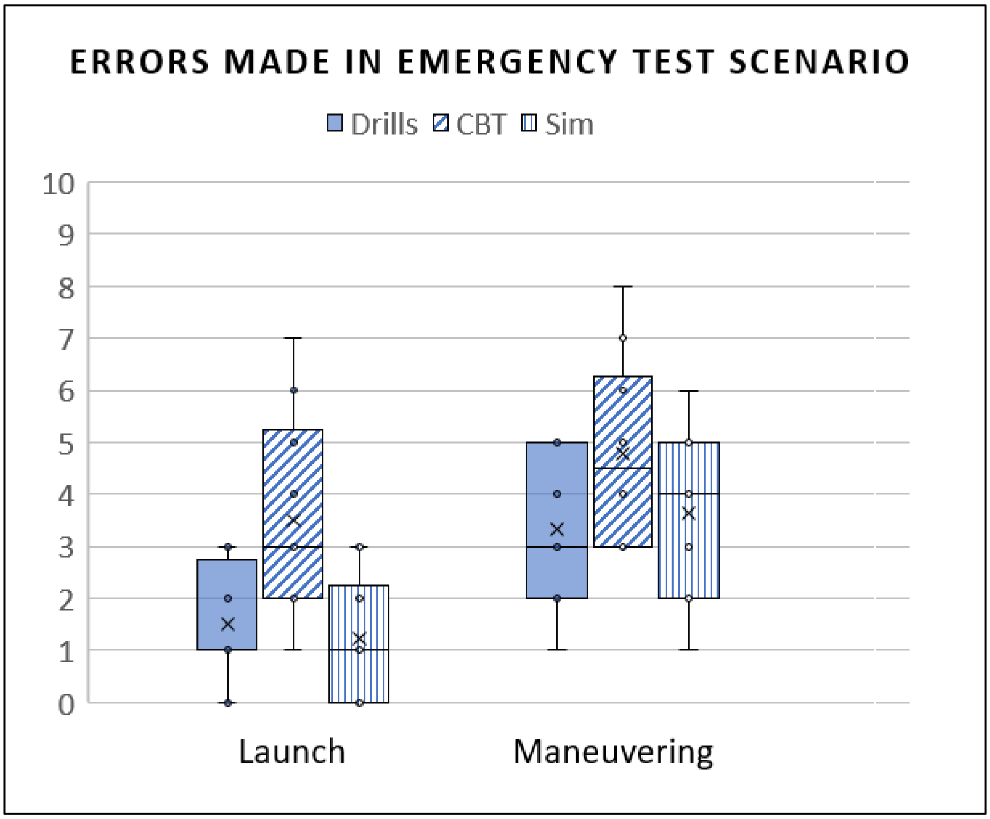

Figure 2-5: Frequency of Errors made during Launching and Maneuvering Tasks

The Simulator Group averaged the lowest amount of errors during launching, with 1.2 errors made on first attempt of tasks (median 1, mode 0). The Drills Group made an average of 1.5 errors on first attempt (median 1, mode 1), and the CBT Group made an average of 3.5 errors (median 3, mode 3). A Kruskal-Wallis test of differences among the medians indicated significant differences between groups ($H = 12.13$, $p < 0.05$). A one-way ANOVA (95% confidence) on ranked data of the three groups indicated a significant difference between the group means. Tukey Comparison Tests (95% confidence) determined the CBT Group mean is significantly different than the mean of the Drills and Simulator Groups. No significant difference between the means of Drills and Simulator Groups was determined from this analysis.

Overall, the participants made more errors in maneuvering tasks than launching tasks. The CBT Group had the highest number of errors in maneuvering on first attempt with an average of 4.8 errors (Median 4.5, mode 3). The Drills Group averaged the lowest number of errors with a mean of 3.33 (median 3, mode 5). The Simulator Group averaged 3.6 errors (median 4, mode 5) on first attempt on maneuvering tasks. Kruskal-Wallis tests did not indicate a significant difference between medians of the groups and a one-way ANOVA on ranked data did not indicate a significant difference in sample means. This outcome suggests additional testing is needed to discern if there is a difference in group performance on maneuvering tasks.

2.6.3 Analysis of Individual Tasks

Analysis of individual job tasks provides further insights into the performance of each group. Figure 2-6 shows the percentage of successful task completions for each of the groups and the overall average of successful task completions for all participants.

The CBT Group performed worse than the overall group average on 8 of the 10 launch tasks. The Drills and Simulator Groups scored as good as or better than the overall group average on 9 of 10 of the launch tasks. This observation further indicates the CBT Group did not perform as well as the Drills or Simulator Groups based on the comparison of performance on individual launch tasks.

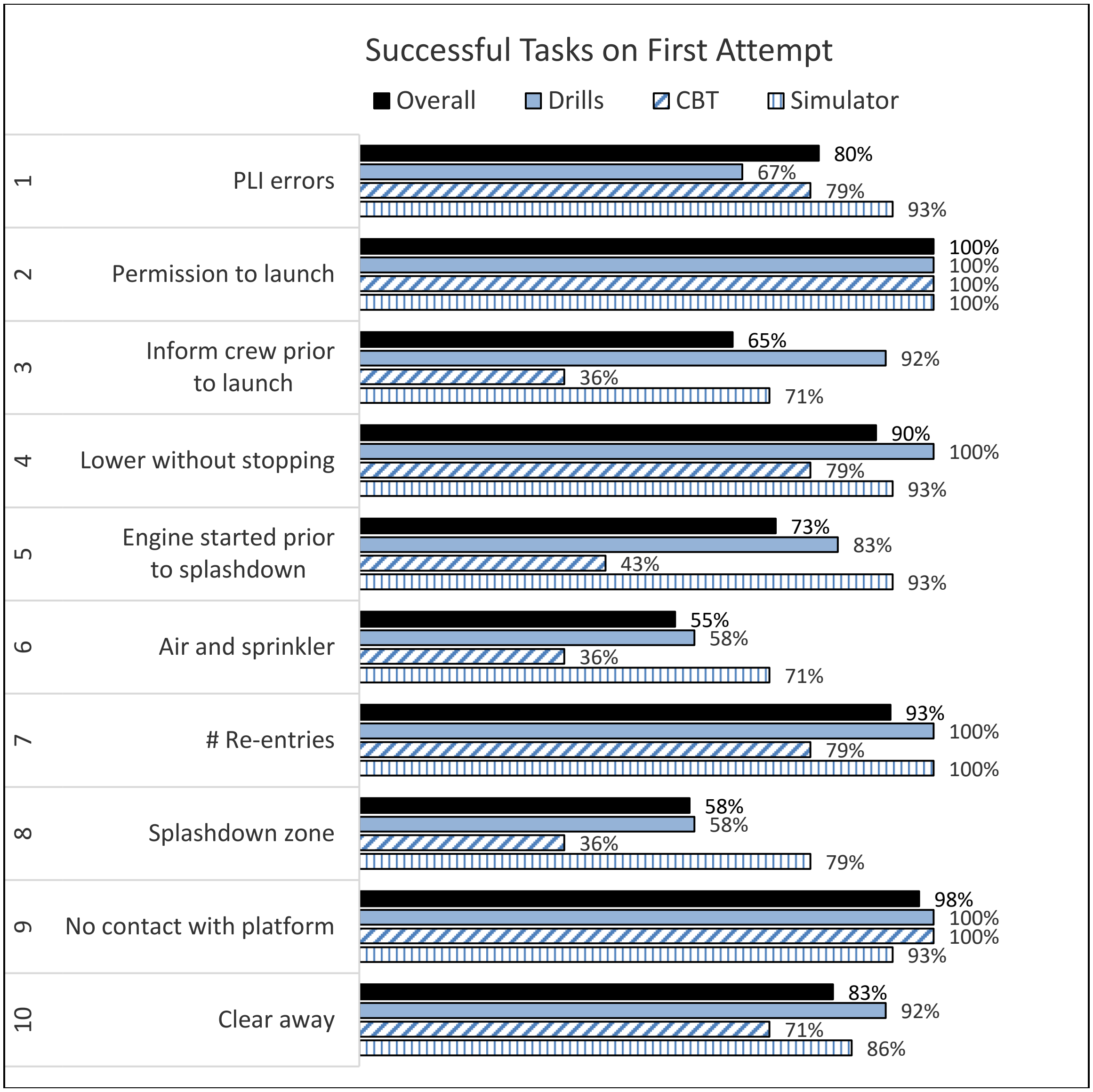

Figure 2-6: Successful Task Completions - Launching

The analysis of individual tasks also allows us to discern the tasks that had the highest and lowest success rates, thereby providing an indicator of transfer of skills to the emergency test scenario and the difficulty of the tasks. The completion of the *Air and Sprinkler* and *Splashdown zone* tasks (see Table 2-2) scored lowest for all groups with less than 58% overall success. The Simulator Group was the only group that had previous practice in identifying a fire hazard and launching into sea states, and showed the highest success rate for these tasks. All other tasks showed an overall success rate of 65% or higher.

Figure 2-7 shows the percentage of successful maneuvering task completions.

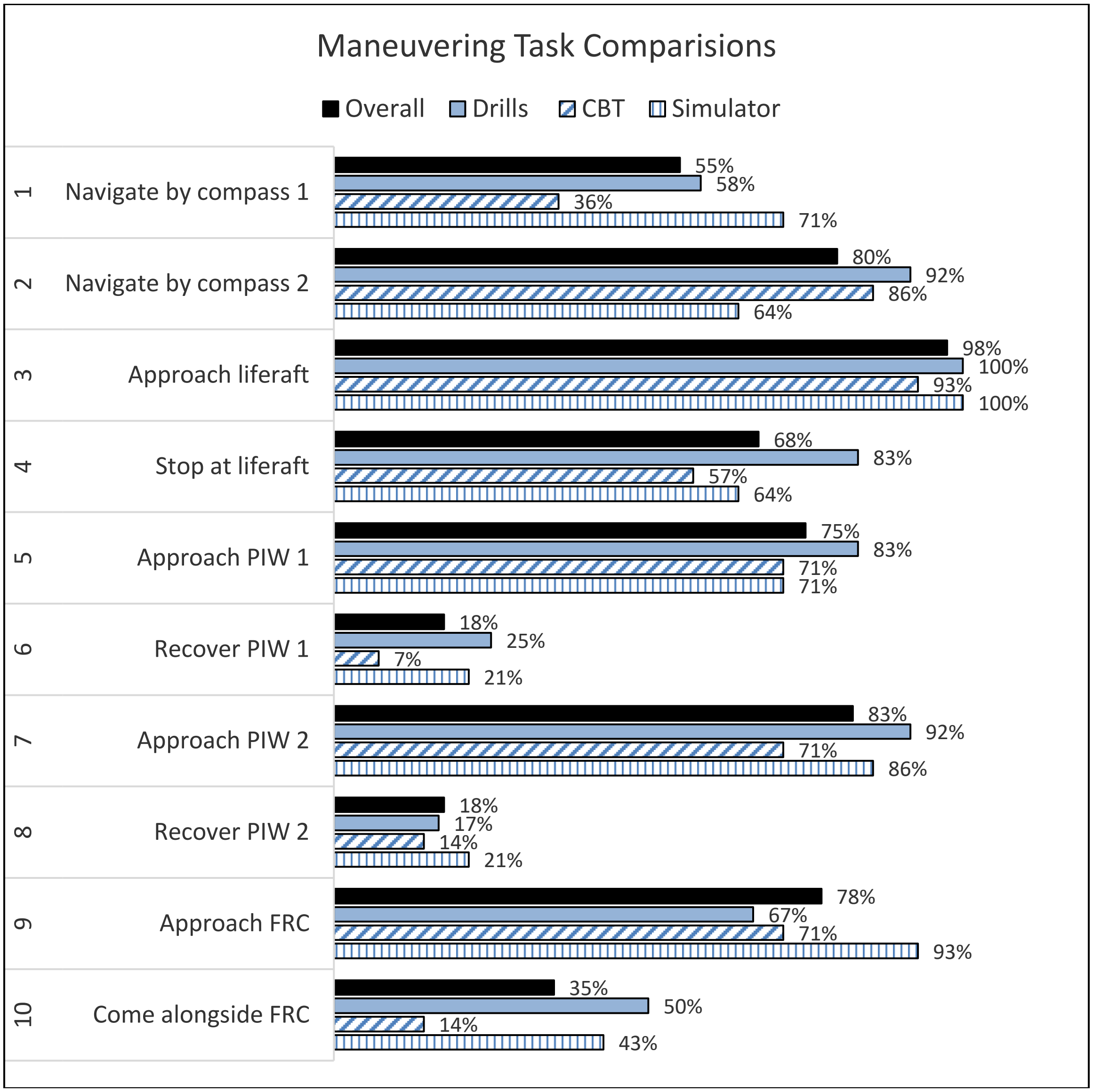

Figure 2-7: Successful Task Completion – Maneuvering

The CBT Group scored lower than the overall group average on 9 of 10 of the maneuvering tasks. The Simulator Group scored higher than the overall group average on 7 of 10 tasks and the Drills

Group scored higher than the overall group average on 8 of 10 tasks. This shows that the CBT Group did not perform as well as the Drills and Simulator Groups on maneuvering tasks. Tasks related to stopping the vessel showed the lowest success rate for all groups. The task *Recover a PIW* was identified as the most difficult, with less than 20% of all participants able to the complete the task on first attempt. *Coming alongside a vessel* had an overall success rate of 35% on first attempt.

2.7 Discussion of Results

The results of the study provide evidence regarding skill acquisition and transfer in relation to the type of training. There is a benefit to performing training using realistic scenarios and real lifeboat equipment. The participants in the Drills Group and the Simulator Group practiced using lifeboat equipment during launching and maneuvering training and were able to practice tasks in live scenarios. The CBT Group refreshed their knowledge of procedures and skills learned during initial training through mental practice. The Drills and Simulator Groups outperformed the CBT Group in launch tasks in all comparisons.

On the overall task of successfully launching a lifeboat on first attempt, the Drills and Simulation Group outperformed the CBT Group. Combining the Drills and Simulation Groups, 31% of participants who received hands-on training were able to successfully launch the lifeboat on first attempt, compared to 0% of CBT Group participants who did not get hands-on training. On a task by task analysis, we observed that specific tasks that could not be rehearsed in the CBT training resulted in inferior performance. These include procedural tasks during launch, including turning on the lifeboat engine, informing the crew prior to launch, and timing the release of the lifeboat.

While these tasks were identified in the quarterly training material, the CBT Group did not rehearse in a scenario, or have the opportunity to practice using real equipment. The CBT Group also performed worse than the Simulator Group and Drills Group on psychomotor maneuvering tasks, including the more difficult slow maneuvers.

Practicing in progressive scenarios with exposure to weather conditions and hazards provides an incremental improvement in performance during launch tasks. This is evident in the successful completion of all tasks during first launch attempt. Although no significant difference between groups was established based on success on first launch, the analysis of individual tasks indicates the Simulator Group was able to recognize hazards and deal with weather conditions better than the Drills Group, who practiced in calm water conditions without hazards. These outcomes provide further evidence that providing similarity between the test environment and the learning environment enhances the retrieval of information (Arthur, 1998, Wickens, 2003) and skill retention (Driskell et al., 1992).

No one was able to successfully complete all maneuvering tasks on first attempt. This result indicates that no group received enough practice to successfully perform these skill-based tasks. The study suggests that specific tasks, such as stopping a vessel in waves, are difficult and require more practice to master. The amount of practice received on maneuvering tasks in calm water or lower sea states did not provide opportunity to achieve competence for tasks in moderate waves. The poor performance of a lifeboat at low speeds could also be a factor contributing to the low success rate.

An analysis of all groups indicates lifeboat launch training is a task that is susceptible to skill fade. For the study, the interval of training was kept constant to match industry practice and to make comparisons based on the type of training received. While some failures were due to hazards that were not practiced in training, the analysis suggests that errors were made in tasks that were practiced to competence during quarterly training sessions including performing pre-launch inspections, starting the lifeboat, and issuing commands to crew members.

2.8 Conclusions and Recommendations

The study provides insights on the difficulty of performing tasks in hazard scenarios and environmental conditions that are plausible in real-life operations. Although the participants received multiple practice sessions, overall there was low success rate on launching and no participants were able to do all the maneuvering tasks. This outcome indicates a need to investigate further the acquisition and loss of maneuvering skills over time, and the transition of skills from benign conditions to challenging conditions. These results may be of interest to stakeholders in oil and gas, shipping, defense, security, and cruise ship industries who are required to maintain the competence of personnel in emergency evacuation tasks or launching procedures.

The training regime used in the experiment was designed to emulate training practices currently utilized in industry. The results of the study suggest that some skills may not be mastered using the type, amount, and frequency of training given in practice. The study of performance on the cognitive tasks associated with launching a lifeboat indicate some skill fade, which has been found in similar research (Stewart et al., 2008). The low performance on maneuvering tasks suggests these skills were not mastered in the training provided, as we would expect higher performance on these psychomotor tasks if competency was achieved. The study indicates that current practices may not provide enough training time to acquire the needed skills in a year-long training program. Subsequent training beyond the one year would provide more training events and may increase training performance.

The study did not measure physiological information on trainee stress, though the context of the emergency exercise and increased difficulty was expected to create more mental demand and stress. Allowing trainees to practice in stressful environments has been shown to improve performance in stressful operations (Mclernon, 2011) and can reduce cognitive difficulties in high-reliability occupations such as firefighting (Baumann et al. 2011). Research has shown that practice in scenarios with representative events and difficulty helps development of mental models to improve decision making (Klein, 2008). Similar benefits are expected for lifeboat operators if they receive training in difficult and complex scenarios.

Maritime education and training instructors can apply the results to improving training practices and outcomes. We see a benefit to performing training using real lifeboat equipment and practicing in scenarios of progressive difficulty, collectively creating a training environment that is representative of a real emergency. Training in scenarios with stressors and hazard that are possible in a real emergency is expected to increase trainee confidence and performance in an actual emergency event.

Psychological principles of learning, such as overtraining, could be employed to improve retention between practice sessions and to improve performance in new events. They could also be employed to promote training adaptability and to provide training that addresses recognized weakness in skills. More frequent training events, shortened intervals between training, and adaptive training can improve skill acquisition and limit skill fade. Future studies will examine the impact of these factors on achieving and maintaining competence in lifeboat coxswains.

2.9 Acknowledgements

We thank Petroleum Research Newfoundland and Labrador and the Industrial Research Assistance Program of the National Research Council Canada who sponsored the study.

2.10 References

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, **11(1)**, 57-101.

Billard, R., Smith, J., Magee, L., Veitch, B., (2018). Simulator training for offshore oil and gas emergency preparedness. *ITEC Proceedings*, Stuttgart.

Baumann, M. R., Gohm, C. L., & Bonner, B. L. (2011). Phased training for high-reliability occupations live-fire exercises for civilian firefighters. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, **53(5),** 548-557.

C-Core (2015). *Metocean climate study offshore Newfoundland and Labrador, Nalcor Energy Report*. http://exploration.nalcorenergy.com/wp-content/uploads/2016/09/Nalcor-Metocean-Study-Final-Report-Volume-2-27-May-2015.pdf

Driskell, J.E., Willis, R.P., & Copper, C. (1992). *Effect of overlearning on retention. Journal of Applied Psychology*, **77(5),** 615-622.

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). *STCW including 2010 Manila Amendments, 2017 Edition*.

International Maritime Organization. (2014), International Convention for the Safety of Life at Sea (SOLAS), *Consolidated Edition*. London: International Maritime Organization.

Klein, G., (2008), Naturalistic decision making. *Human Factors: The Journal of Human Factors and Ergonomic Society*, **50(3)**, 456-460.

Lim, J., Reiser, R., Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a cognitive skill. Educational Technology Research and Development, **57,** 61-77

Magee, L.E., Smith, J.J.E., Billard, R., & Patterson, A. (2016). Simulator training for lifeboat maneuvers. *Proceedings of the Inteservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*. Paper number 16030.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. *WMU Journal of Maritime Affairs*, **16(2)**, 247-263.

Stefanidis, D., Korndorffer, J.R., Markley, S., Sierra, R., Heniford, B.T., & Scott, D.J. (2007). Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training? *Journal of the American College of Surgeons*, **205(2)**, 307-313.

Stewart, J., Johnson, D., Howse, W. (2008). Fidelity requirements for army aviation training devices, Army Research Institute for the Behavioral and Social Sciences, Research. *Report 1887, U.S. Army Research Institute*.

van Merriënboer, J. J. G., Clark, R.E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development*, **50(2)**, 39–64.

Van Rossum, J. H. (1990). Schmidt's schema theory: The empirical base of the variability of practice hypothesis: A critical analysis. *Human Movement Science*, **9(3)**, 387-435.

Wickens, C., Hollands, J., Banbury, S., Parasuraman, R. (2013). *Engineering Psychology and Human Performance, Fourth Edition*, Pearson.

3.0 CHAPTER 3: USING BAYESIAN METHODS AND SIMULATOR DATA TO MODEL LIFEBOAT COXSWAIN PERFORMANCE

Randy Billard[1], Mashrura Musharraf[2], Jennifer Smith[3], and Brian Veitch[2]

[1]Virtual Marine, [2]Memorial University of Newfoundland, [3]private consultant

3.1 Co-authorship Statement

This manuscript has been published in the World Maritime University (WMU) Journal of Maritime Affairs (2020). Writing was led by Randy Billard, with assistance on modeling and results interpretation provided by Mashrura Musharraf and Jennifer Smith. Brian Veitch and Mashrura Musharraf provided guidance in writing and provided revisions to the paper.

3.2 Abstract

Lifeboat training is normally performed in controlled conditions to minimize the risk to trainees and equipment. Participants are given limited or no opportunity to practice skills in operational scenarios that represent offshore emergencies. For this reason, human performance in plausible emergencies is difficult to predict due to the limited data that is available. Simulation provides a means to collect novel data on human performance and learning in situations that are otherwise prohibitive due to risk. In this study, we use simulator data to shape knowledge of the problem space of lifeboat coxswain training and skill transfer. We use Bayesian inference to produce human performance probabilities (HPPs) to model the performance of lifeboat coxswains as they practice lifeboat tasks for the first time. Data collected in an experiment are used (1) to generate

probability distributions to predict the amount of practice needed for new coxswains to achieve competence on lifeboat launching and maneuvering tasks, (2) to study how skills learned in training transfer to a new scenario and, (3) to make comparisons between task difficulty. The methodology can be applied to other problems to assess training effectiveness and improve instructional design. Models can be continuously strengthened with additional data to improve predictive accuracy. Probability distributions can be used to assess competence in new scenarios and to diagnose strengths and weaknesses using machine learning.

3.3 Introduction

Lifeboat operators are required to have the skills to launch and operate a lifeboat in environmental conditions that prevail in their location of operation. Although operators may experience challenging conditions in a real emergency, initial training is normally conducted in calm waters to minimize the risk to trainees and equipment. Training typically includes a combination of lectures, demonstrations, and group practice sessions. Competence is normally assessed based on the trainee's ability to demonstrate completion of tasks trained individually in course curriculum. Practice drills following training are performed in calm water and involve launching the lifeboat and simple maneuvering tasks to re-familiarize trainees with the operation of the lifeboat (International Maritime Organization, 2014). It is assumed that skills acquired in training can be transferred to more difficult scenarios, such as emergencies involving a lifeboat in adverse weather.

Industry studies have identified that benchmarking of lifeboat coxswain skill is difficult to assess based on the limitations in training, yet coxswain skill has an impact on a successful lifeboat launch

(Robson, 2007). There is little information available on the amount of practice needed to acquire and master specific skills, and how the skills learned in training transfer to operational scenarios. Little is known about coxswain performance in sea states other than calm water, or in scenarios where coxswains must complete a combination of launch and maneuvering tasks as they would in a real emergency. Forecasts of coxswains' skill transfer to real-life operational scenarios has relied on experts' opinion.

With the advent of simulator technology, it is now possible for trainees to practice in weather conditions typical of their location of operation and to apply their skills in realistic emergency scenarios. Data collected from the simulator provide an opportunity to measure competence in lifeboat coxswains as they demonstrate skills in scenarios that previously could not be performed including weather conditions that are considered too dangerous for live training. Using simulation-based assessment (SBA) to measure cognitive and practical skills is increasing in learning and education (S. de Klerk, et al. 2015). Simulators have been used to investigate human performance in marine operations (Sellberg, 2017, Power-MacDonald et al., 2011, Magee et al. 2016, Thistle et al. 2019), and specifically to study skill acquisition and transfer for lifeboat coxswain training for credible emergencies (Billard et al. 2019). The paper discusses a method that uses data collected from SBAs to formulate probabilistic models to study human performance.

We define human performance probability (HPP) as the probability that a trainee will successfully complete a task in a given scenario. Studies have used Bayesian inference and data collected from simulators to estimate performance probabilities for several problem domains to deal with scarcity of data (Groth et al., 2014, Musharraf et al. 2019). The Bayesian inference method is used in this paper to develop a probabilistic model that updates the prior beliefs about HPPs of lifeboat coxswains using data collected from a simulator experiment. The model provides a quantitative look at the problem space of initial skills acquisition in lifeboat coxswains.

The HPP cumulative distribution functions (CDFs) are used to investigate the following research questions:

1. What is the expected performance of new lifeboat coxswains as they apply skills learned in initial training to a new scenario?
2. How much practice is needed to acquire the procedural and psychomotor skills to launch and maneuver a lifeboat in plausible weather conditions?
3. Do specific tasks or task types require more initial training and practice to master?

An outcome is the creation of sets of CDFs to quantify skill acquisition in a group of new trainees as they enter a training program designed to prepare coxswains for offshore emergencies involving a lifeboat. We compare HPP CDFs to evaluate the relative difficulty of tasks and investigate the amount of practice needed to acquire skills in initial training. We compare the statistical measures of the CDFs to evaluate the strength of the modeling approach.

Some lifeboat tasks can be categorized as either procedural or motor-skill based, while others require a combination of both physical and cognitive skills. Studies have shown there are differences in the type and amount of practice needed to acquire and retain skills for each task type (Sauer et al., 2000). We investigate performance on individual task types to study the difference in skill acquisition for tasks involving combinations of procedural skills and physical motor skills.

The paper presents an analysis that has relevance to lifeboat training providers and presents a methodology that can be used to study other problem areas. The study outcomes can be used by trainers and industry stakeholders to assess competence in new trainees using the HPPs as a benchmark of performance. The methodology can be applied to additional problems including those where simulation is extending training to conditions to challenging conditions that could previously not be practiced. As additional data is collected on user performance, Bayesian methods can be applied to improve the predictive accuracy of probability-based models. The probability distributions created in this study can be used to diagnose competence and adapt training curriculum to individual needs. The outcomes can be used to evaluate risk, improve training programs, performance, and investigate ways to accelerate time to competence.

3.4 Methodology

The study uses Bayesian inference (BI) to generate HPPs for a group of new coxswains as they apply their skills in an operational scenario involving completion of multiple tasks. Similar to other studies (Groth et al. 2014, Musharraf et al. 2018), we use experimental data sets from a simulator experiment to update prior beliefs about the HPPs and to create a posterior distribution that is informed by new data.

Experimental data was collected for participants who received initial training on the operation of a lifeboat and then applied their skills in a new scenario involving completion of lifeboat launch and maneuvering tasks. Seven different tasks had to be completed in the scenario. The scenario was repeated multiple times until participants completed all tasks successfully allowing for additional learning through practice.

Probability distributions for each task were created for each scenario attempt. Sets of HPPs were created for each task using the experimental group's performance outcome for each attempt at task. The sets of HPPs were used to evaluate the learning transfer from initial training to a new scenario, and to investigate skills acquisition as tasks were repeated. Tasks were grouped based on the type of skill required to complete tasks to make comparisons between task types. We also created a distribution model for overall competence, based on successful completion of all tasks in the scenario.

3.5 Bayesian Inference Study Approach

The BI process uses Bayes Theorem (Equation 1) to update a hypothesis, or belief, *H*, based on new data. *D*.

$$\boldsymbol{P(H|D) = P(H)\frac{P(D|H)}{P(D)}} \qquad (1)$$

Bayes theorem is applied to obtain the posterior distribution $P(H|D)$ given new data is provided. $P(\mathrm{D})$ is the marginal probability of the data. This is normalized over all specific hypotheses and remains constant, independent of H. We collected data in a format that can be used to update prior distribution $P(H)$, using a likelihood model to translate observable data into probabilistic information. The likelihood model expresses the probability of the data given the truth of the hypothesis H, and is defined as $P(D|H)$. We use likelihood functions and prior distributions that are conjugate pairs, allowing for a closed form of integration to calculate the continuous distribution of $P(\mathrm{D})$ as discussed in similar studies using BI (Groth et al. 2014, Musharraf et al. 2019).

The BI process used in the study involved the following steps:

1. Define the likelihood function for the performance distribution: We assume a binomial distribution as the probabilistic distribution to identify the number of successful attempts on completion of tasks, assuming independence among training events in group. The binomial distribution describes uncertainty on the number of successful attempts (*x*) in a number of cases *(n)*, assuming the probability of success. In our case, *x* is the number of successful demonstrations of completion of a task for a specific attempt and *n* is the total number of participants tested in a scenario. The distribution is as follows, with unknown parameter *p* (i.e. the HPP for event *X* is *p)*,

$$\Pr(X = x) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{2}$$

2. Identify sources of information consistent with the distribution model: For data obtained in the experimental study, we assume tasks of all types are either demonstrated successfully, or else they are a failure, using an established rubric to measure successful completion. The study uses groups of participants of total number *n* for each attempt measured, and *x* defines the number of successful task completions in the group.
3. Specify the initial prior distribution: The prior distribution of the HPP is defined as p_o. The Beta distribution is a conjugate pair of the binomial distribution and is commonly used as the prior distribution to define *p.* The probability density function (PDF) for the Beta distribution is as follows:

$$PDF = f(p\colon \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \tag{3}$$

$0 \leq p \leq 1,$ standard Beta distribution.

where $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \Gamma\ is\ the\ Gamma\ Function$

The CDF is therefore defined as the regularized incomplete beta function $I_x(\alpha, \beta)$, and is defined as follows:

$$\boldsymbol{CDF = P(X \leq t) = \int_0^t \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dt} \tag{4}$$

Parameters of the beta distribution (α and β) are initially estimated to form a prior distribution model. As discussed in Section 2.4, the participants in the study are provided with a basic lifeboat training and familiarization, and we assume trainees acquired some skill in operation of the vessel and basic maneuvering. Little also is known about the background skill of the participants that could be transferred to the operation of the lifeboat. For this reason, we assume a Jeffreys prior, Beta (0.5,0.5) for the first attempt at task, assuming an equal chance of successful task completion on first attempt for all tasks in a new scenario.

4. Perform Bayesian updating to create posterior distribution: The prior distribution (*Beta*($\alpha_{prior}, \beta_{prior}$)) and the likelihood function *(Binomial (n,p))* are conjugates and with Bayesian updating, the posterior distribution is also a Beta distribution. Posterior values can be calculated through the updating of the Beta distribution parameters (α and β) using available data. For each set of attempts on tasks, the simulator data records x successes in n cases, providing new evidence on the probability model. Values of updated α and β are found using the following equations:

$$\alpha_{post} = \alpha_{prior} + x \quad (5)$$

$$\beta_{post} = \beta_{prior} + n - x \quad (6)$$

5. Perform BI for additional attempts: The study collected group performance data (successful completions) for each attempt (A) made in the simulator assessment scenario. The data collected on the first attempt is used to update the prior belief about the HPP (Jeffrey's prior) and generate a posterior Beta distribution based on observed outcomes. The refined distribution serves as an indicator of group probability of success on the first attempt at tasks. The posterior is assumed to be a more accurate indicator of probability compared to a Jeffreys prior. The posterior from the first attempt A_1 is used as the prior to the next attempt A_2. Data collected on A_2 is then used to perform the updating using BI. The same methodology is performed to form HPP distributions for 3 attempts at task. Table 3-1 provides a breakdown of the parameters used for each HPP CDF.

Table 3-1: Distribution Parameters for Each Task Attempt:

	Task Attempt (A)		
Parameters	**A = 1**	**A = 2**	**A = 3**
Total participants in Attempt	n_1	n_2	n_3
Number of successes in Attempt	x_1	x_2	x_3
$\boldsymbol{\alpha_{prior}}$	0.5	α_{post1}	α_{post2}
$\boldsymbol{\beta_{prior}}$	0.5	β_{post1}	β_{post2}
$\boldsymbol{\alpha_{postA}}$	$\alpha_{prior} + x_1$	$\alpha_{prior2} + x_2$	$\alpha_{prior3} + x_3$
$\boldsymbol{\beta_{postA}}$	$\beta_{prior} + n_1 - x_1$	$\beta_{prior2} + n_2 - x_2$	$\beta_{prior3} + n_3 - x_3$

3.5.1 Assessing Transfer and Learning using Human Performance Probability Distributions

The outcome of this process is sets of HPP distributions. Data collected on completion of tasks in the new scenario provides a measure of skills transfer from initial training to a new scenario. Subsequent attempts provide data on how additional practice impacts the probability of completion of tasks. With each attempt, learning is expected to occur. Data from each attempt is used to refine the probability distribution for the task attempt considering the number of successfully completions observed in the group of coxswains. The HPP distributions are presented as CDFs to provide a visual comparison of performance for different task types and give insights on the group performance after multiple attempts. As shown in Figure 3-1, for each task type a set of three HPP CDFs are generated, one for each of the attempts made for a specific task.

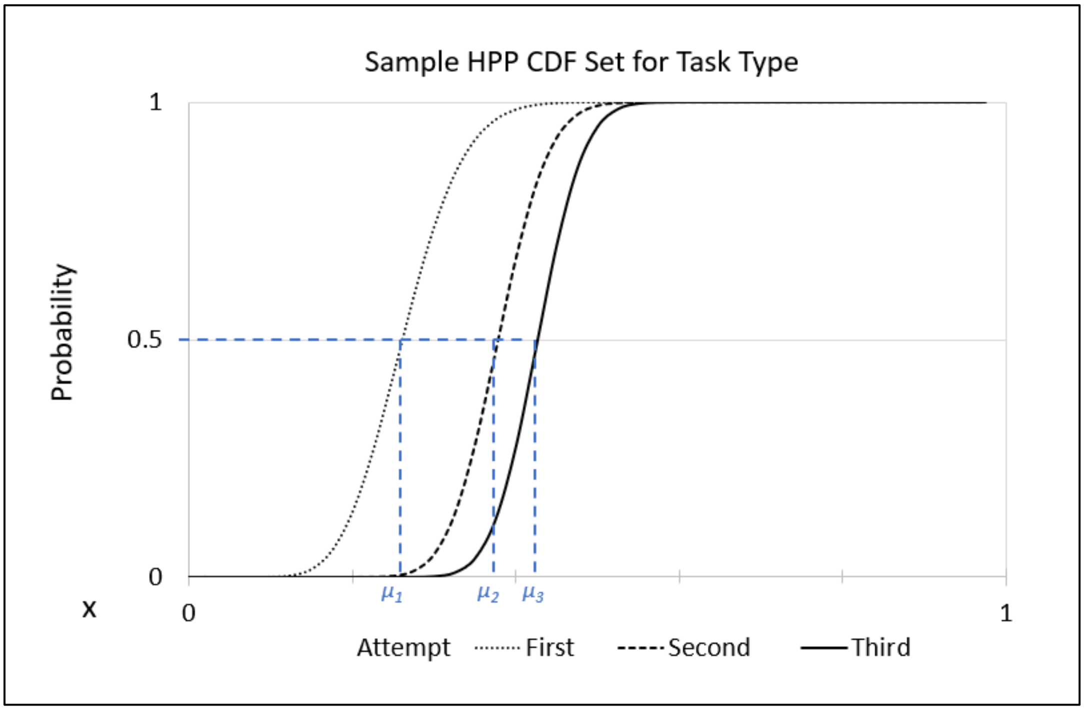

Figure 3-1: HPP Distribution Set

3.5.2 *Making Comparisons and Assessing Strength of BI Approach for HPP Modeling*

We make comparisons between the task types and attempts at tasks using the mean of the beta distributions. The mean of the distribution changes with each task attempt and is calculated based on the α and β specific to the task type and attempt performed. The mean of the distribution, E, is defined as follows:

$$E[X] = \mu_A = \frac{\alpha_A}{\alpha_A + \beta_A} \tag{7}$$

Where (α_A and β_A) are the beta parameters for each attempt (A). The mean of the distribution is $\mu_A = 0.5$ on the CDF, as shown in Figure 3-1.

We evaluate the strength of the modeling approach used in the study through comparison of the standard deviation and credible intervals of the HPP CDF's calculated parameters. A reduction in these parameters suggests a reduction in uncertainty.

The standard deviation (SD) of the beta distribution is calculated as follows using the calculated beta distribution parameters:

$$SD_A[X] = \sqrt{\frac{\alpha_A \beta_A}{(\alpha_A + \beta_A)^2(\alpha_A + \beta_A + 1)}} \tag{8}$$

A Credible Interval (CI) is commonly used in Bayesian statistics to summarize the uncertainty related to calculated parameters (Makowski et al. 2019). The CI provides a range containing a percentage of probable values of the modeled posterior distribution from the Bayesian inference. The shorter the CI, the lower the uncertainty. A 95% credible interval is used, which assumes a central portion of the posterior distribution (upper and lower 2.5% removed) and considers the percentile from the lower to upper cumulative distribution function of the beta distribution.

3.5.3 Experimental Data from Lifeboat Simulator Study

Experimental data was collected from a group of participants with no prior lifeboat experience. The experiment consisted of the following:

1. Initial training of naïve participants to outline the key equipment and operation of the lifeboat, followed by familiarization with the simulator that would be used in the study.
2. Completion of an assessment scenario in a simulator in a more difficult environment than used in initial training. The scenario involved completion of multiple task types learned in initial training. Participants practiced the scenario until they demonstrated competence on all tasks.

54 participants were recruited for the study. Participants were between the ages of 18 and 65. Initial training of all participants consisted of a combination of classroom training from an instructor and familiarization exercises with a simulator. The training curriculum covered competencies identified in the Standards of Training, Certification and Watchkeeping for Seafarers (International Maritime Organization, 2010) and emulated a training course provided to coxswains in industry. The training covered basic operation of the lifeboat, coxswain duties, pre-launch inspection procedures, clear away procedures, navigation procedures, radio communications, and layout of equipment. After training was completed, participants were given a fifteen-minute simulator exercise to become familiar with the lifeboat simulator and boat handling characteristics.

All training on the simulator was performed in calm water, consistent with industry practice, and operation of the vessel in higher wind and waves was only covered in classroom materials.

Approximately 5 days following initial instruction, participants completed a simulator assessment scenario designed to evaluate the fundamental skills required to operate a lifeboat, which included the launch and control of the lifeboat in weather conditions that required the operator to consider the impact of light waves and wind when maneuvering the boat. The scenario was designed by Subject Matter Experts (SMEs) and scored personnel on their ability to complete launch, navigation, and slow-speed maneuvering tasks in an exercise that required completion of these types of tasks. The parameters of scenario were set to daytime with clear visibility, 9 knot winds, and a Beaufort scale of 3 with large wavelets. Wind and waves did not have a high impact on vessel performance but could affect boat during slow-speed maneuvers. Users also had to consider the direction of wind when determining the correct direction to approach the target. The scenario diagram is provided in Figure 3-2.

The scenario consisted of two launch tasks and five on-water tasks. Participants were not permitted to move to on-water tasks until they successfully completed all launch tasks. The scenario was stopped immediately during the launch and clear away if participants made critical errors that could result in damage to the vessel or harm to the crew such that the scenario could not progress (i.e., forgetting to ensure launch equipment was operational, making contact with platform). In these cases, the launch was restarted. Tasks 1 and 2 in Figure 3-2 denote the launch and clear away tasks. When a successful launch (Task 1 and 2) was achieved, the participant immediately proceeded to the maneuvering course (Scenario Event ST, Figure 3-2) and started at this point for the following attempts. Multiple attempts at the on-water tasks (Tasks 3-7) were provided as needed until all tasks were completed successfully. Additional details on the experiment and simulator used in the study are discussed in Billard et al. (2019).

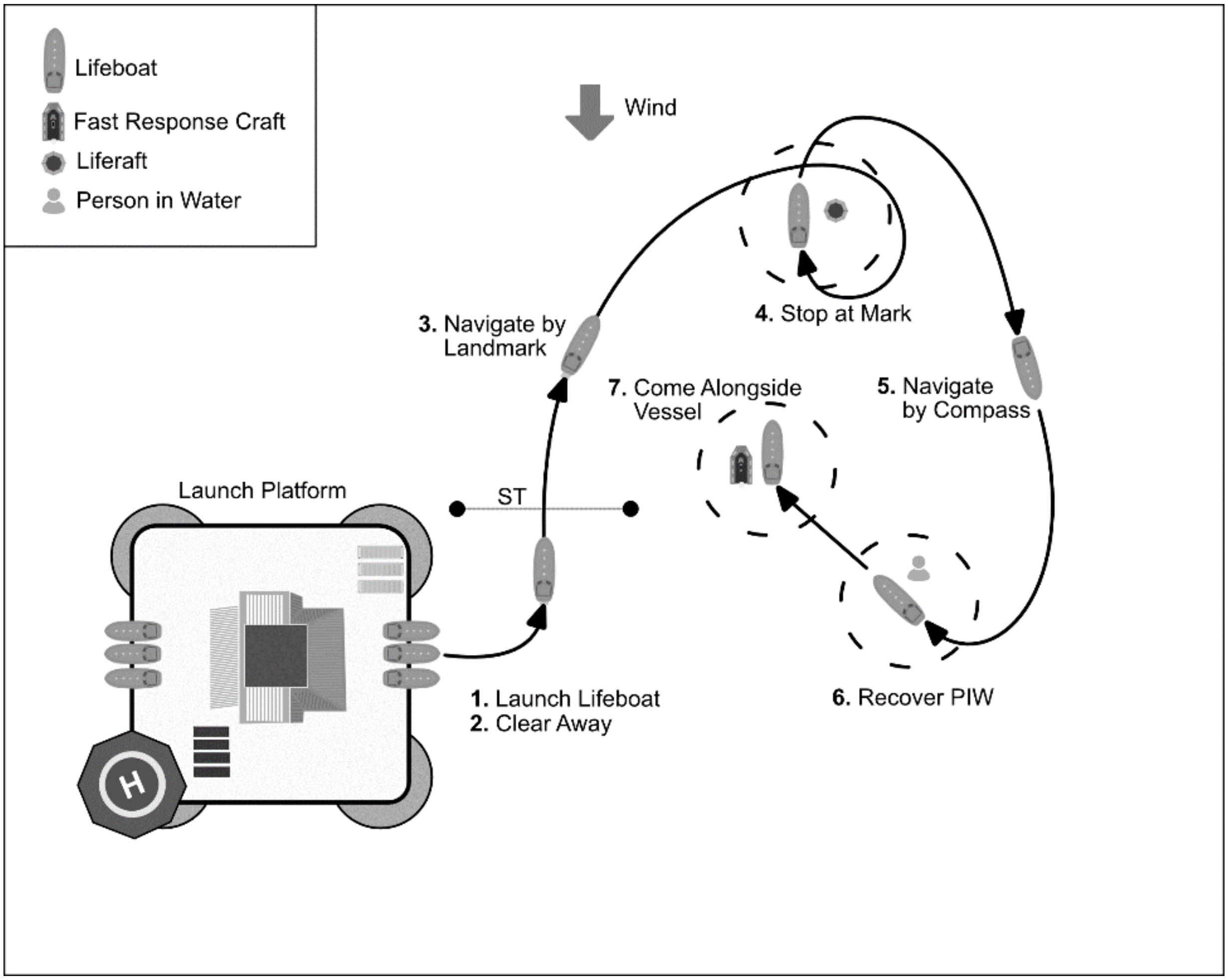

Figure 3-2: Assessment Scenario

3.5.4 Performance Measurements

A scoring rubric was established to provide a consistent measure performance for launching and maneuvering tasks in the assessment scenario. The criteria for task completion reflect a standard of proficiency as identified in recognized training standards (International Maritime Organization, 2010). Scoring measures were determined by SMEs with experience in delivering lifeboat training. Tasks were categorized based on the type of activity that had to be performed in the scenario and the type of skill being assessed. Table 3-2 provides a list of tasks and objectives used to measure successful completion and identifies the categories used for comparison of task types. Each task category contained subtasks. An attempt at task was considered successful only if all subtasks were completed successfully. Additional information on the task type and rubric is provided in Billard et al. 2018.

Table 3-2: Task Categories

Task Category	Description of Task	Scenario Task
Procedural launching tasks	A combination of procedural tasks, performed in the correct order, to prepare lifeboat for launch and lower lifeboat to the water surface. Tasks include inspection of equipment prior to launch, communication with Offshore Installation Manger and crew members to ensure situation is safe for launch, preparing boat for launch, and lowering to water surface.	1
Clear away tasks	Physical task of ensuring lifeboat is buoyant, releasing lifeboat from davit, proceeding to a safe zone as quickly as possible, and avoiding contact with platform.	2
Navigation tasks	Maintaining control while steering a lifeboat to a desired location using visual cues to steer to a landmark or using a navigation aid (i.e. compass) to steer on a heading.	3, 4
Slow-speed maneuvering tasks	Skill-based task of maneuvering a lifeboat close to an object and maintaining the position of the lifeboat at low speeds while applying strategy to deal with environmental forces such as winds and waves. Types of tasks include stopping next to a mark (i.e. a life raft), stopping next to a PIW for recovery, and stopping next to another vessel (a Fast Response Craft) to transfer personnel.	5,6,7

3.6 Results

The results of the methodology are sets of HPP CDFs for individual task types that can be analyzed to evaluate learning in the experimental participants as they practice tasks. We make comparisons within and between sets of CDFs to evaluate transfer and skill acquisition and to assess task difficulty.

The calculated means of the beta distribution CDFs, as derived from the BI methodology, are used to make numerical comparisons between HPP CDFs. Comparisons are made between different task types to evaluate relative task difficulty, and within set of tasks HPP CDFs to assess how practice on the same task impacted performance. A visual comparison of the CDFs can be used to observe differences in group performance. As probability of success increases, as indicated by an increase in the CDF mean, the curve will move to the right. More movement indicates a bigger step in group performance and shows the impact of additional practice on skill acquisition.

The HPP CDFs are grouped based on the task categories identified in section 2.4.1. A discussion is provided for each set of tasks to illustrate how the HPP CDFs are used to investigate the research objectives identified in section 1.0. The impact of practice is examined for each task type and insights on the factors impacting task difficulty are provided. A comparison of all tasks is provided to illustrate the relative difficulty of the task types. A model of overall coxswain competence assuming successful completion of all tasks in the simulator scenario is also presented to relate the study outcomes to participant preparedness for a plausible emergency event. The strength of the modeling approach is also discussed through comparison of standard deviation and credible intervals for the sets of distributions. The results are presented in detail below.

3.6.1 Summary of Group Data by Attempt

Participants attempted tasks until all tasks were completed successfully in the same scenario. As discussed in section 2.3, *procedural launch* and *clear away* tasks were repeated if critical errors were made, after which participants were able to start maneuvering and navigation tasks. As result of this setup, there were a different number of participant cases (*n*) for attempts at task. The number of successful attempts for each case, *x*, was the number of times the task was successfully completed by participants in the group. Table 3-3 provides a breakdown of the experimental outcomes.

Table 3-3: Successful Task Completions by Attempt

	First Attempt (A1)		Second Attempt (A2)		Third Attempt (A3)	
Case	n_1	x_1	n_2	x_2	n_3	x_3
Procedural Launch	54	14	44	23	24	14
Clear Away	54	26	41	32	13	10
Navigate by Landmark	54	48	51	41	46	43
Navigate by compass	54	33	51	39	46	35
Stop at Landmark	54	23	51	31	46	31
PIW Pickup	54	13	51	13	46	20
Come Alongside a vessel	54	25	51	28	46	34

The posterior Beta distribution parameters and resultant HPP CDF's were derived from the experimental data using the approach outlined in section 2. The CDFs are shown in Figures 3-3-3-5.

3.6.2 Launching Tasks

The distributions indicate the initial success rate of the *procedural launch* tasks and *clear away* tasks to be 26% and 48% respectively. This result indicates initial training did not provide enough practice to for trainees to acquire the skills needed to perform these tasks in a new scenario. For all the attempts, the group performance was higher in the *clear away* task than for the *procedural launch* tasks. The mean HPP for the *procedural launch* tasks increased to 42% and the mean HPP for the *clear away* tasks increased to 63% after three attempts. The *procedural launch* tasks were more difficult to complete than the skill-based launch task. A distribution mean of less than 50% on the procedural tasks suggests that initial practice and three practice attempts did not provide enough training for the majority of the group to acquire the skills needed to complete the tasks. An explanation for the low success rate is the number of procedural items that had to be performed and the requirement to complete the tasks in a specific order resulted in a high task difficulty.

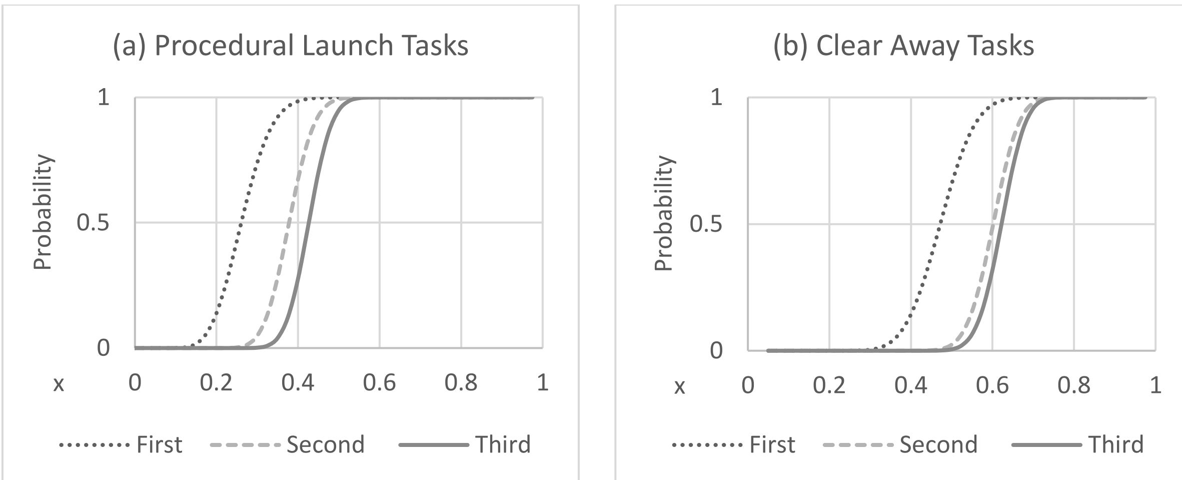

Figure 3-3: Launch Task HPP CDFs

3.7 Navigation Tasks

Analysis of the navigation tasks suggests this task type was easier to complete compared to the launch tasks. As shown in Figure 3-4a, the mean of the distribution for the task of *navigating by landmark* was above 80% on for all attempts and did not change substantially with additional practice. This result indicates initial training provided sufficient practice to achieve competence on the task. With reference to Figure 3-4b, a comparison of the distributions suggests maintaining a constant heading when *navigating by compass* is more difficult than *navigating by a landmark*. Task difficulty is increased when using a compass. The user must maintain control of a boat while monitoring a compass heading that has erratic behavior, which takes more skill to maintain a heading than navigating to a landmark that is viewed at a distance from the lifeboat. The operation requires driving skill to correct changing values shown on the navigation equipment while maintaining control of the vessel. The results show an increase in performance with practice. The mean of the distribution for *navigating by compass* increased from 61% to 71% after 3 attempts, showing additional practice improved performance.

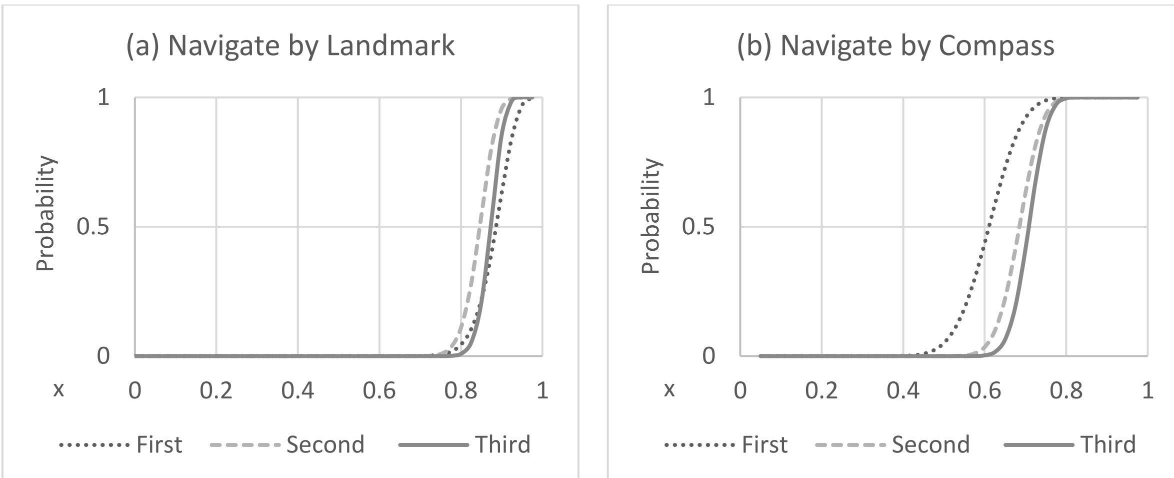

Figure 3-4: Navigation Task HPP CDFs

3.7.1 Slow Speed Maneuvering Tasks

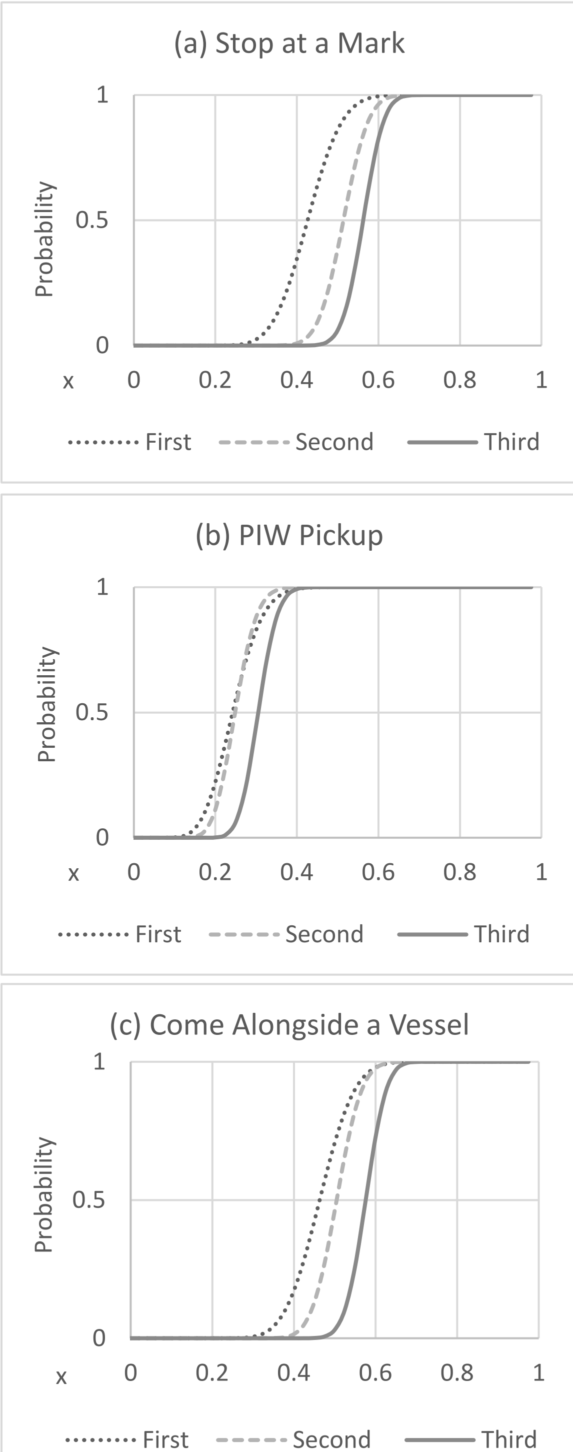

Figure 3-5: Slow Speed Maneuvering Tasks HPP CDFs

The HPP's for all the slow speed maneuvering tasks show an initial success rate of lower than 50%, indicating initial training did not enable most participants to acquire the skills needed to perform this task type. As shown in Figure 3-5a and 3-5c, the overall probability of success increased for the tasks of *stop at a mark* and *come alongside a vessel*, and most group participants were able to complete these tasks after three attempts (mean probability of success increased to greater than 50%). As shown in Figure 5b the task of *PIW pickup* had a group success rate of 25% on first attempt, increasing to 31% on the third attempt, suggesting that this task was the most difficult of the slow-speed maneuvering tasks that were practiced.

The PIW task was most difficult to perform due the challenge of losing sight of the small target. The PIW cannot be seen from the coxswain's point of view when close to the lifeboat due to the location of the chair and windows which are located high in the lifeboat. The participant had to rely on the recommended practice of communicating with an assistant in the boat who provided information on distance to the target using voice relays. For the tasks of *stopping at a mark*, and *coming alongside a vessel,* the target vessels could be more easily seen from the coxswain's viewpoint and less communication was needed to determine distance. Light contact was also permitted for the vessels, but not for the PIW as contact would result in injury to the person.

Each of the slow-speed maneuvering tasks had a common objective of approaching the target from the right direction and stopping the vessel. Given the similarity of the three slow-speed maneuvering tasks and performance measures, it is likely that practice on subsequent tasks improved performance on following tasks. Each scenario attempt included three slow-speed maneuvering tasks. The difficulty of the *PIW pickup* task is further emphasized, as a low probability of success was evident even with additional practice.

3.7.2 Comparison of Task Types

Figure 3-6 shows the HPP CDF means for each of the task types and attempts. This outcome indicates the navigation tasks are the easiest to complete given the practice provided, and the slow-speed maneuvering tasks appear to be the most difficult. The tasks *launching the lifeboat* and *PIW pickup* both had a probability of success of less than 30% on first attempt and did not increase to above 50% after three practice attempts. This result indicates that additional initial training or supplemental practice is needed for most of the group to complete these tasks. All other tasks resulted in a group performance of greater than 50% after three attempts, though the group only achieved a probability of greater than 70% after three attempts on two of the seven tasks. Most tasks show an increase in probability of success with additional attempts, suggesting that students were able to increase skills with practice.

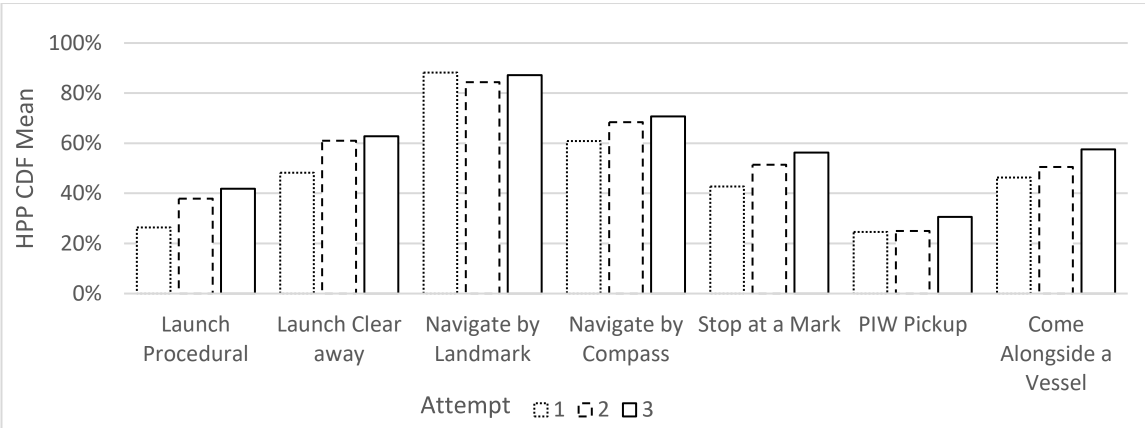

Figure 3-6: Task Distribution Means

3.7.3 Assessment of Overall Competency

We can use the methodology to investigate overall coxswain competency considering the ability to successfully complete all tasks in the scenario in one attempt. The overall probability of success is influenced by individual task probabilities, with incompletion of any tasks resulting in an unsuccessful completion of the scenario. As shown in Figure 3-7, the HPP for the first attempt was 3% and increased to 12% after three attempts.

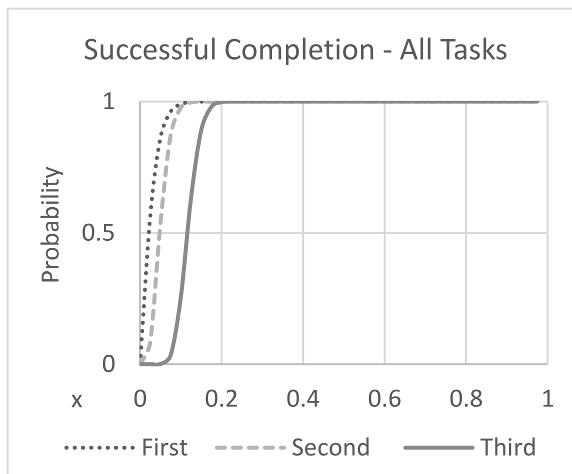

Figure 3-7: Overall Competency HPP CDFs

Performance increased with each attempt, though the low overall probability of success of the group after three attempts suggests that the initial training and additional practice was not enough to achieve competence in completing an exercise involving multiple tasks.

3.7.4 Uncertainty Measures

As a measure of uncertainty, SD and CI were investigated for each task attempt. As shown in Figure 3-8, the SD of the distributions decreased with each attempt (A) for all tasks, indicating a reduction in uncertainty in the HPP CDFs for each additional attempt. As shown in Figure 3-8, the CI for the HPP CDFs also decreased with each attempt, for all tasks. The uncertainty in the CDFs reduced with each attempt, suggesting strengthening of the model as more data was available and improvement of the assessment of HPPs in each iteration.

Choosing Jeffrey's prior as the prior distribution allowed the model to be structured enough to learn, but weak enough to learn from a limited amount of data. In case of a more dominating prior, much more data might be required to counterbalance the effect of the prior and observe a learning effect demonstrated by a reduction of uncertainty.

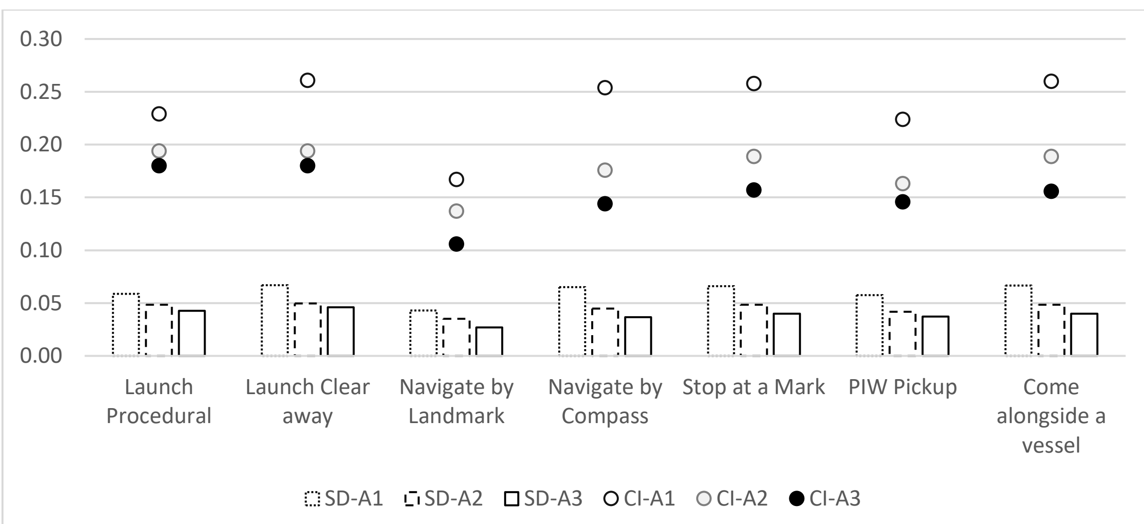

Figure 3-8: HPP CDF Standard Deviation and Credible Interval

3.8 Discussion

The goal of this study was to use data obtained from a simulator experiment to create probability distributions that can be used to study the skill acquisition of new trainees as they enter a training program designed to prepare coxswains for offshore emergencies involving a lifeboat. Using BI to form probability distributions allowed us to use available data to measure and compare performance of the group as they acquired and applied skills. We were able to study skills acquisition and transfer through the breakdown of competency into specific tasks that could be analyzed independently or collectively to compare task type. The methodology of using HPP CDF parameters calculated from previous attempts as estimates for future attempts, combined with new available data, results in a reduction in uncertainty of the modeled HPP parameters used for performance comparisons. The study indicated the models were strengthened as new data was used in sets of distributions.

The HPPs generated provide an indicator of the amount of practice needed to achieve competence on newly trained tasks, as assessed by the probability of successful completion from a group of individuals who had not performed the tasks prior to the study. Revisiting the research questions identified in section 1.0, we were able to use the HPP CDFs to assess the performance of the participants as they acquired skills, and to gain insights on the effectiveness of the training and practice received. We were also able to make comparisons between tasks to evaluate the relative task difficulties. A summary of the research outcomes is as follows:

1. The initial training program did not result in a high group performance on their first attempt at tasks in a plausible emergency scenario. The modeled probability of successful

completion on most tasks is less than 50%, indicating most participants would not be able to complete the launch and maneuvering tasks on their first attempt. Overall, the probability that the coxswains would be able complete all scenario tasks in order, as would be required in a real lifeboat evacuation, was very low.

2. Additional practice resulted in an increase in probability of successful completion for most tasks. After three practice attempts, the modeled probability of successful completion was greater than 50% for the tasks of *clear-away*, *stopping at a landmark*, and *coming alongside a vessel*. The results indicate improved performance with additional practice attempts. Learning was still occurring after three practice attempts. The HPP CDFs indicated that three practice attempts did not result in a model mean that was greater than 50% for the tasks of *launching the lifeboat* and *PIW pickup*, indicating these tasks require more practice to master compared to other tasks.
3. The tasks of *navigating by compass* and *navigating by landmark* appear to be the least difficult to complete, with a modeled probability of success above 50% for first attempt at tasks trials and increasing with additional practice. The procedural task of *launching the lifeboat* and performing a *PIW pickup* appear to be the most difficult to perform, based on the modeled probability of successful completion on first attempt and after additional practice attempts.

We can use the study outcomes to provide insights on the amount of training and practice needed to acquire skills to launch and maneuver a lifeboat, and to suggest ways to improve performance. The results suggest additional training and practice were needed for trainees to build the mental models needed to perform the procedural tasks required to launch a lifeboat, as found in similar research (Stewart et al., 2008). The low performance on the driving tasks of maneuvering the boat suggests the skills were not mastered in the initial training and subsequent practice. Studies performed to evaluate coxswain training programs (Billard et al., 2019) also found that low speed maneuvering tasks require a high amount of training and are difficult to perform in new scenarios, including those with more severe weather. The difference in the practice provided in initial training environment and the scenario used in assessment scenarios also needs to be considered. The scenario used for assessment was different than scenarios used in initial training. Research has shown that practice in scenarios with representative events and difficulty helps development of mental models to improve performance (Klein, 2008). Similar benefits are expected for lifeboat operators if they receive training in scenarios that represent real emergencies.

Consideration must be given to the size of the data set that was used to estimate the performance probabilities. For attempts with a lower number of group participants, the Beta distribution means may not change significantly due to the lower number of cases and the approach, which relied on the use of the prior distribution parameters. Additional data collection can be performed to improve the accuracy of the distributions.

3.9 Conclusions

To summarize, the methodology of using BI to create HPPs to make comparisons based on group performance enabled us to study learning in new lifeboat operators. We were able to discern the tasks that required more practice and the relative difficulty of tasks using the probability distributions created using the experimental data.

HPP CDFs can be developed for other problems to measure the effectiveness of training programs with targeted measures of competence and performance. The methodology can be applied to study performance of novice and expert trainees as they apply their skills in new scenarios. Training providers can use HPPs to set performance targets for evaluation of trainee competence at various stages of training as skills are learned. This may be of value to instructors who wish to determine if a group or individual's performance is better or worse than expected, and training and instructional approaches can be adjusted accordingly.

Specific to offshore emergency training, data collected from training and experimental studies can be used to improve courses designed for new lifeboat coxswains As new data is collected on performance of new coxswains, HPPs can be updated to improve the predictive and diagnostic accuracy of the probability distributions. For cases where little information is known on performance, such as launching a lifeboat in severe sea states, data collected on expert and novice trainees can be used to gain insight on human performance as coxswains apply skills in more challenging conditions. As training is extended to more severe weather, limitations of the human performance and equipment will also become apparent. The methodology used in this paper can be used to research these areas.

The study demonstrates the suitability of Bayesian Methods and performance data obtained from a simulator to investigate new problem areas. The methodology can be extended to gain a deeper understanding of how skills are acquired and to explore ways to improve training. The distributions created using the BI methodology can be incorporated into models to study learning and adapt training to individual trainees. Bayesian methods can be used to develop competency models that utilize machine learning to improve training outcomes. As discussed by Millán et al. 2002, probability distributions can be incorporated into Bayesian Networks to derive models of student competence to diagnose strengths and weaknesses in trainees. Machine learning and intelligent tutoring techniques can be applied and improve student assessment and course design.

3.10 Acknowledgements

We thank Petroleum Research Newfoundland and Labrador and the Industrial Research Assistance Program of the National Research Council Canada who sponsored the study. The authors acknowledge with gratitude the support of the NSERC/Husky Energy Industrial Research Chair in Safety at Sea.

3.11 References

Billard, R., Smith, J.J.E. (2018). Using simulation to assess performance in emergency lifeboat launches. Proceedings, e Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 19179.

Billard, R., Smith, J., Veitch B., (2019) Assessing lifeboat coxswain training Alternatives using a simulator. The Journal of Navigation, Published online by Cambridge University Press: 19 September 2019.

Groth K., Smith, C., Swiler, L. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability and System Safety 128 (2014), 32-40

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). STCW including 2010 Manila Amendments, 2017 Edition.

International Maritime Organization. (2014). International Convention for the Safety of Life at Sea (SOLAS), Consolidated Edition. London: International Maritime Organization.

Klein, G., (2008), Naturalistic decision making. Human Factors: The Journal of Human Factors and Ergonomic Society, 50(3), 456-460.

Magee, L.E., Smith, J.J.E., Billard, R., & Patterson, A. (2016). Simulator training for lifeboat maneuvers. Proceedings, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 16030.

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. Journal of Open Source Software, 4(40), 1541. https://doi.org/10.21105/joss.01541

Millán, E., Perez-De-La-Cruz, (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction 12: 281-330, Kluwer Academic Publishers, Netherlands

Musharraf, M., Moyle, A., Khan, F., Veitch B. (2019) Using simulator data to facilitate human reliability analysis. Journal of Offshore Mechanics and Arctic Engineering. Vol. 141(2).

Power-MacDonald, S.P., MacKinnon, S., Simões Ré, A., Power, J., & Baker, A. (2011), Effects of simulator training on novice operators' performance in simulated ice covered waters. National Research Council Canada, TR-2011-15.

Robson, J. K., (2007). Overview of TEMPSC performance Standards. Health and Safety Executive Research Report RR599. Norwich.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU Journal of Maritime Affairs, 16(2), 247-263.

Stewart, J., Johnson, D., Howse, W. (2008). Fidelity requirements for army aviation training devices, Army Research Institute for the Behavioral and Social Sciences, Research. Report 1887, U.S. Army Research Institute.

Thistle, R., Veitch, B, (2019). An evidence-based method of training to targeted levels of performance. The Society of Naval Architects and Marine Engineers. Document SNAME-SMC-2019-030.

4.0 CHAPTER 4: USING BAYESIAN NETWORKS TO MODEL COMPETENCE OF LIFEBOAT COXSWAINS

Randy Billard[1], Jennifer Smith[2], Mashrura Masharraf[2], and Brian Veitch[2]

[1] Virtual Marine, [2] Memorial University of Newfoundland

4.1 Co-authorship Statement

This manuscript was been published in the Transnav International Journal of Marine Navigation and Safety of Sea Transportation (2020). Writing was led by Randy Billard, with assistance on developing and presenting the methodology and model outcomes provided by Mashrura Musharraf. Jennifer Smith participated in the experiment to collect data used in the models. Brian Veitch, Mashrura Musharraf, and Jennifer Smith provided guidance in writing and revising the paper.

4.2 Abstract

The assessment of lifeboat coxswain performance in operational scenarios representing offshore emergencies has been prohibitive due to risk. For this reason, human performance in plausible emergencies is difficult to predict due to the limited data that is available. The advent of lifeboat simulation provides a means to practice in weather conditions representative of an offshore emergency. In this paper, we present a methodology to create probabilistic models to study this new problem space using Bayesian Networks (BNs) to formulate a model of competence. We combine expert input and simulator data to create a BN model of the competence of slow-speed maneuvering (SSM). We demonstrate how the model is improved using data collected in an

experiment designed to measure performance of coxswains in an emergency scenario. We illustrate how this model can be used to predict performance and diagnose background information about the student. The methodology demonstrates the use of simulation and probabilistic methods to increase domain awareness where limited data is available. We discuss how the methodology can be applied to improve predictions and adapt training using machine learning.

4.3 Introduction

Lifeboat training is normally performed in controlled conditions to minimize the risk to trainees and equipment. Trainees are given limited or no opportunity to practice skills in operational scenarios that represent offshore emergencies. For this reason, human performance in emergencies is difficult to predict due to the limited data that is available. Forecasts of coxswains' skill transfer to real-life operational scenarios have relied on experts' opinion. Even so, there is limited information on how much skills learned in lifeboat training transfer to adverse weather conditions. The modeling of human performance in harsh environments has not been possible due to the scarcity of human performance data.

With the advent of lifeboat simulator technology, it is now possible for trainees to practice in weather conditions typical of their location of operation and to apply their skills in realistic emergency scenarios. Simulation provides the possibility to apply knowledge in applications in highly contextualized environments that are representative of plausible emergencies. Research has shown that practice in realistic scenarios helps development of mental models to improve performance (Klein, 2008). The study of human performance using simulation is evident in other operations including flight (McClernon et al. 2011), medical (Stefandis et al. 2007) and marine

(Sellberg, 2017) training. Lifeboat training data can now be collected to assess the amount of practice needed to acquire skills and to evaluate how skills learned in practice transfer to new scenarios (Billard, 2019).

Data collected from a lifeboat simulator allow us to assess performance on tasks that were prohibitive to do, even in calm water training. This new data can be used to model learning and skill acquisition using probabilistic methods. We can study the interaction between tasks using Bayesian Networks (BN) to derive models of student competence (Millán and Pérez De-la-Cruz, 2002). These models can be used to study the relationship between training factors and to examine how practice on related tasks impacts performance. Due to scarcity of human performance data, initial models of competence can be formed with expert input (Groth et al., 2014). Performance data collected from simulator studies can provide evidence to inform models of trainee competence and validate their predictive accuracy. Bayesian methods have been used to model performance on lifeboat launch and maneuvering tasks in initial training in calm weather conditions (Billard et al., 2020). Similar approaches can be applied to model performance in more adverse weather conditions.

In this paper, we present a methodology to form probabilistic models of human performance that can be used to study this new problem space. We use a BN to define a model of the competence of slow-speed maneuvering (SSM) based on tasks performed in adverse weather conditions during an offshore emergency. The model is derived from a combination of expert prediction and data collected from an experimental study.

The methodology is used to investigate the following research goals:

- how to formulate a BN model of competency using knowledge of task type and available performance measures; and,
- how to combine expert knowledge and data collected from simulator exercises to improve the model's predictive accuracy.

We evaluate the model using available data sets from a simulator study on lifeboat coxswain performance. We demonstrate how this model can be used to 1) predict performance as trainees practice skills in simulator scenarios, and 2) diagnose background information about the student.

The paper presents an approach that is relevant to training providers and researchers. We discuss how to apply the methodology and resultant models to study performance, improve expert assumptions, and extend to training applications where new data sets are being created. The models can be used to improve training programs, adapt training exercises to individual needs, and investigate human performance in new scenarios.

4.4 Background

4.4.1 Competence – Slow-speed Maneuvering

We demonstrate the methodology of creating a BN model of competence using evidence captured in an experiment designed to study lifeboat training.

We must first frame our definition of competence considering our research goals and the objective measures that can be made. The concept of competence is a diverse topic that has diverse definitions. For our purposes, we consider how competence is normally measured in marine training through completion of demonstrable tasks specific to learning objectives (IMO 2014, STCW 2010). We consider competence the “existence of learnable cognitive abilities and skills which are needed for problem solving” as identified in research on skill acquisition (Weinert, 2001). We assume that completing tasks of a similar cognitive or physical skill form demonstrates competence.

We construct a model of competence for the skill of Slow-speed Maneuvering (SSM), as demonstrated by the ability to complete tasks related to stopping a lifeboat next to an object in the water. It is expected that trained lifeboat operators have this required competence to perform in an emergency. The completion of tasks in an emergency scenario can include stopping next to a number of objects including a life raft, a person in the water (PIW), a small vessel for transfer of personnel, or a large vessel for securing the lifeboat for recovery. All tasks considered under the competence of SSM require a similar application of skills and similar performance measures.

We assume there is a relationship between the SSM tasks based on the type of skill needed to perform the task. The maneuvering and stopping of a lifeboat is primarily a physical task and requires application of psychomotor skills to control the lifeboat, including manipulation of lifeboat throttle, steering, and making visual observations. There are also cognitive skills, including deciding angles of approach and judging distance from a target object. Practice on SSM tasks within a practice scenario is expected to improve performance on related SSM tasks based on the similarity of the tasks and type of skill that is applied.

4.4.2 Simulator Exercise and Experiment

We use data collected from a simulator scenario to formulate our model and provide evidence that can be used to inform and evaluate our methodology.

Data was taken from an experiment that used a lifeboat simulator to study skill acquisition and transfer in lifeboat coxswains. The experiment was designed to evaluate how skills acquired in different training programs transferred to a plausible emergency event that required the launch and maneuvering of a lifeboat in weather conditions typical of offshore operations. Participants completed training using different approaches over a year long period and then participated in a new simulator exercise for assessment purposes. The assessment scenario included a combination of launch tasks and on-water tasks. Details of the scenario are provided in Figure 4-1. Additional details on the experimental test plan and simulator used in the study can be found in Billard et al. (2019).

In real scenarios or in simulator exercises, SSM tasks form a part of the whole training exercise. Other tasks may need to be completed, including inspecting the lifeboat, launching the lifeboat, and navigating the lifeboat. These tasks require application of different skills and have different measures, as described in previous research (Billard et al. 2018, Billard et al. 2020). As such, these tasks are not related to competence of SSM and are excluded from the BN model creation as practice on these tasks is predicted to not affect SSM competence.

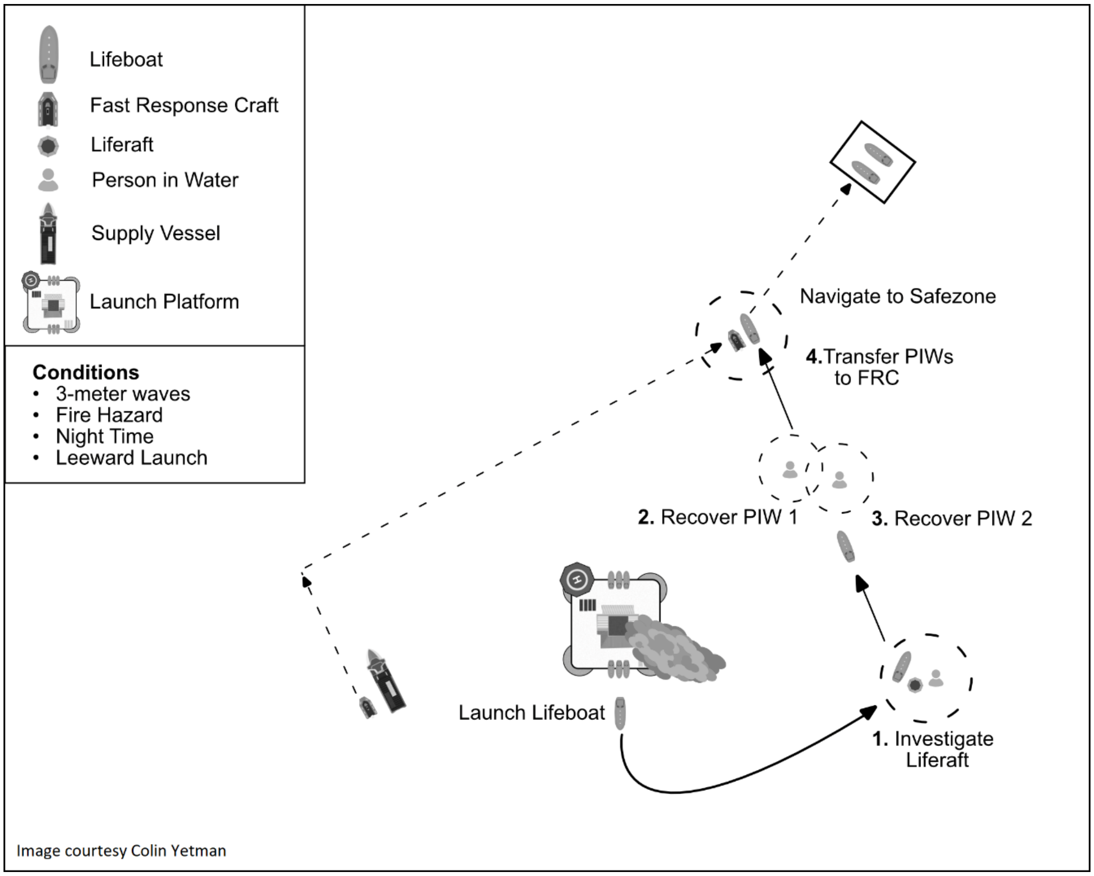

Figure 4-1: Simulator Assessment Scenario with SSM Tasks

The data collected from the assessment scenario provided evidence to evaluate SSM competence modeled in a BN. The scenario contained 4 slow-speed maneuvering tasks including, in order, stopping next to a Life Raft for inspection (LR), picking up two persons in the water (PIW1, PIW2), and stopping next to a Fast Rescue Craft (FRC) for transfer of personnel. These tasks provide evidence for the assessment of the SSM competence.

All participants completed the scenario at least two times and data was collected for the maneuvering tasks for each attempt. Tasks were completed in the same order with each attempt. A total of 39 participants completed the study.

4.4.3 *Measuring Performance*

The rubric used to define completion of the SSM task was derived from recognized training standards and is based on expected performance identified by Subject Matter Experts (SMEs). Each task requires approaching an object from a preferred direction, stopping close to the target, and maintaining a stopping speed. The specific parameters used to measure success differed slightly for each task (i.e. light contact with a vessel is acceptable for coming alongside a vessel, but not allowed for a PIW). Table 4-1 provides an outline of task objectives and the corresponding measures used in the simulator exercise. Completion of tasks was based on several simultaneous measures captured by the simulator, each of which had to be performed correctly to be considered a successful completion. Additional details on the scoring measures and rubric has been presented previously (Billard et al. 2018).

Table 4-1: Slow-Speed Maneuvering Competence Tasks

Task Identifier	Task Description	Task Objective	Measures
LR	Stop at a Life Raft	Approach a static object accounting for wind and wave direction. Use a speed to allow stopping. Stop close to Life Raft (2-3 boat lengths) and maintain position	direction of approach speed at stop time stopped contact speed heading at stop number of attempts
PIW	Recover a Person in the Water (PIW)	Approach a drifting PIW accounting for wind and waves to minimize chance of contact. Use a speed to allow stopping. Stop close enough to PIW to allow pickup and maintain position in waves	
FRC	Come Alongside a Fast Response Craft (FRC)	Approach a FRC accounting for wind and wave direction. Use a speed to allow stopping. Stop close to vessel (less than 0.5 meters) and at an angle to allow personnel transfer and maintain position	

4.4.4 *Bayesian Network Modeling*

Bayesian Networks (BN) use a graphical structure to represent the relationship between several random variables as represented in a directed acyclic graph (DAG). A sample BN DAG is provided in Figure 4-2. Nodes (a,b,c,d,e) represent the variables and arcs (arrows) represent the probabilistic relationship between the variables. Bayesian inference algorithms create a relationship between latent variables, which are inferred, based on the state of observed variables.

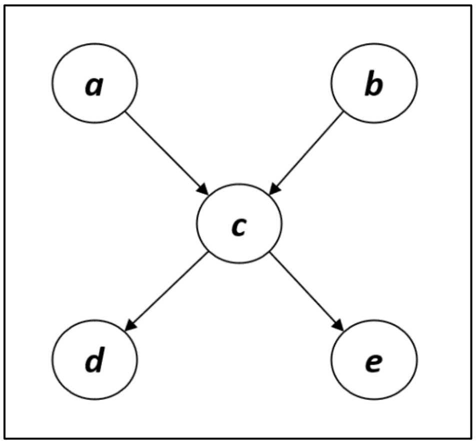

Figure 4-2: Sample Bayesian Network DAG

Building a BN includes the following steps:

1. Defining the variables that are being studied, both latent and observable, creating the nodes of the BN.
2. Defining the relationships between variables using arcs. The arcs represent a causal influence between the variables. Variables in the network that are not graphically connected are conditionally independent of each other (i.e. a and b are conditionally independent).

3. For each of the variables, defining the probability conditions with parent variables through Conditional Probability Tables (CPTs). The probabilities can be learned from real data or defined by experts.

Detailed description of BNs and how they are created is provided in other literature (S. de Klerk et al., 2013, Millán et al., 2010).

Creating a BN to use observable evidence to study an inherent competence has applications in training frameworks including Intelligent Tutoring Systems (ITS) (Millán and Perez-De-La-Cruz., 2002, Käser et al. 2017) and Evidence Centered Design (ECD) (Mislevy et al., 2004). In these frameworks, the BN forms a model of the competency that is being investigated (the student model) and identifies the relationships to the performance measures (the evidence) in the practice scenario (the activity). The relationships form a construct of competence, a latent variable, that can be measured through the collection of performance data, an observable variable.

In our case, we use the observable completion of SSM tasks to quantify the latent variable of SSM competence using evidence collected through a simulation study.

4.5 Methodology

We use a BN methodology to model competence and predict the performance of lifeboat operators as they apply skills learned in training to a new scenario. We create a BN model using observable measures from a simulation scenario designed to evaluate coxswain performance in a plausible emergency. We use a combination of expert prediction and simulator data to create and revise our model. The methodology creates a student model of SSM competence that can be used for the prediction of performance on tasks and the diagnostic study of causal relationships between model variables.

The steps in the methodology include the following, as outlined in Figure 4-3:

1. Defining a generic BN student model of competence - based on completion of tasks that are considered similar in the type of skill applied
2. Characterizing the BN model as a SSM competence student model - based on the evidence gathered in a simulator practice exercise
3. Creating the initial CPTs of the model nodes based on expert estimates
4. Refining the CPTs based on experimental data - using the simulator experimental data to tune the model parameters
5. Validating the model accuracy for predictive and diagnostic use cases using simulator data

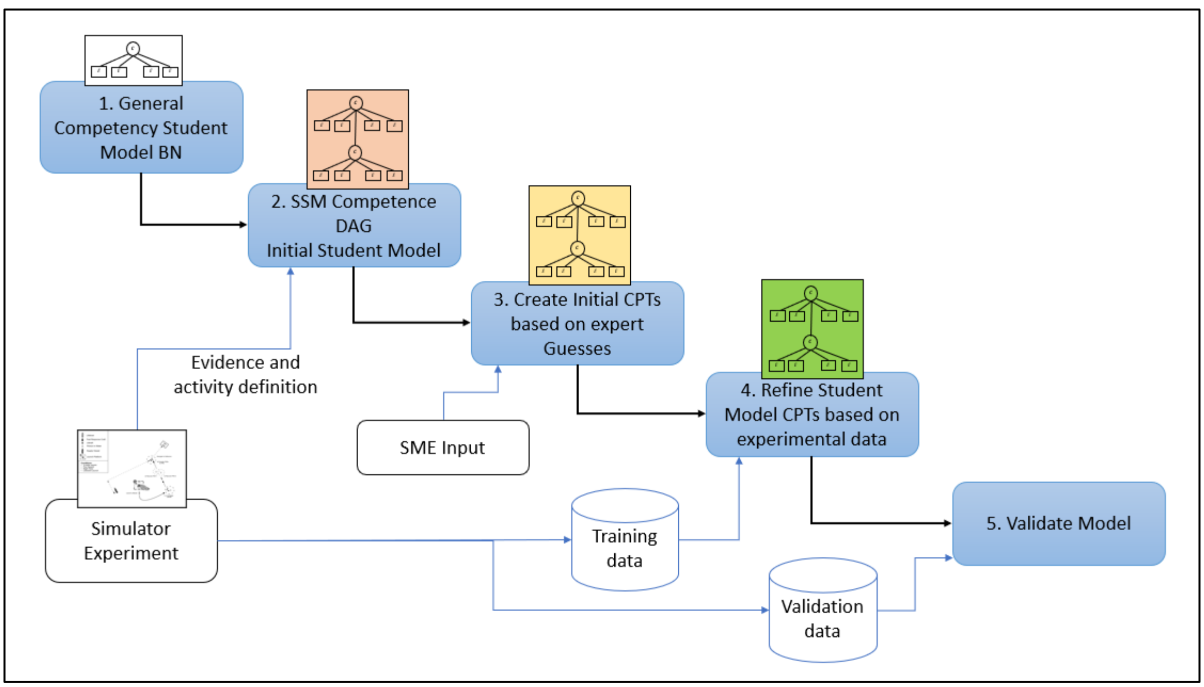

Figure 4-3: Methodology of Creating and Validating a SMM Competence BN

We perform two validation cases to show how the BN model can be applied and how the model changes with new data or variables. We first demonstrate how the predictive accuracy of the model changes as the methodology is applied. We evaluate the predictive accuracy of the model first formed with expert estimates and then re-evaluate the predictive accuracy after data have been used to refine the CPTs. We then present an example of how new variables can be added to the model and show how the model can be applied to diagnose the relationship between the new variable and observable evidence. The validation of models is discussed in Section 4.

4.5.1 Step 1 – Defining a Generic BN Student Model of Competence

We first describe the types of variables and relationship assumptions for the BN student model.

We assume a latent variable of competence *(C)* and relate to task evidence nodes *(E_i)*, which can be measured or observed in a scenario. The tasks are related by the type of skills needed to complete the tasks successfully.

To create the DAG, we assume a structure where observable evidence of completing tasks changes the probability of the competence, as described in previous research (Millán and Pérez De-la-Cruz, 2002). The generic model is presented in Figure 4-4. In the model structure, we assume a causal relationship where the latent variable *(C)* causes the evidence $E_1, E_2, E_3, \ldots E_i$. In this relationship, evidence about mastering a task changes the probability of the latent parent. Consequently, evidence about mastering *C* changes the probability of its children *(E_i)* and evidence about mastering a task affects the probability of mastering the rest of the tasks on the same level. This models assumes conditional independence of the E_i given *C* (for each $i = 1, \ldots n$). In this DAG, the

CPT parameters that need to be identified are the prior probability of the competence, $P(C)$, and the conditional probabilities of the evidence nodes $\{P(E_i|C), i = 1, \ldots n\}$.

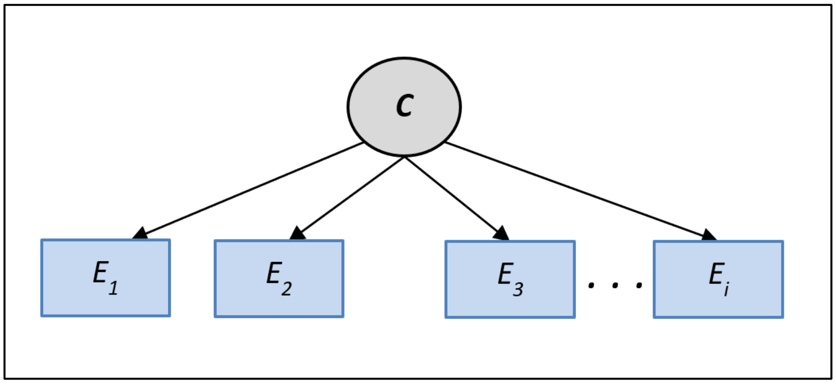

Figure 4-4: Competence Model BN DAG

4.5.2 Step 2 – Characterizing the BN Competence Model as a SSM Competence Student Model

We design the BN model to match the activity, in this case the slow-speed maneuvering exercises performed in the simulator study.

Figure 4-5 shows the DAG for the experimental study consisting of two scenarios, each having 4 evidence nodes. In the simulator study, the trainee practiced the same scenario twice, creating two sets of evidential nodes, as the trainee completed the same tasks with each attempt. As an input of evidence in the BN, the task was either considered to be completed (Yes) or not completed (No) based on the performance requirements set by SMEs to measure successful completion of task.

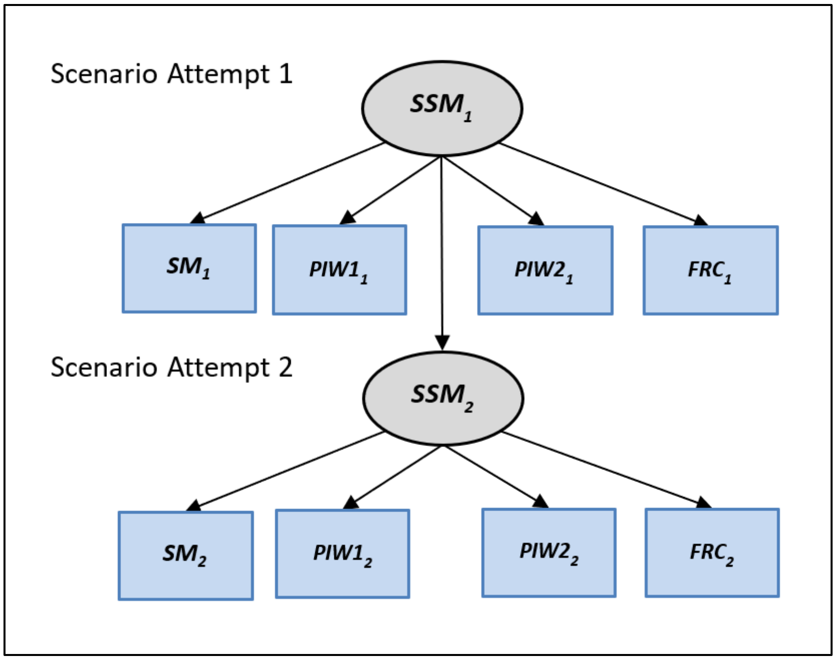

Figure 4-5: Bayesian network DAG – Simulator Assessment Scenario

The structure of the model assumes a learning effect with tasks practiced in a training session consisting of multiple simulation exercises. We use a dynamic model indicating the trainee's competence can be measured with each simulator exercise attempt. We define a relationship between the measure of competence in the first attempt *(SSM_1)* and the measure of competence on the second attempt *(SSM_2)*. The relationship assumes the measure of competence in the first attempt impacts the probability of the second attempt through a defined *CPT* $\{P(SSM_2|SSM_1)\}$. Based on the similarity of the task types it is expected that practice on any of the task types can improve the performance on other tasks, including future attempts at the same task using the same scenario.

4.5.3 *Step 3 – Creating Initial CPTs Based on Expert Estimates*

The structure of the BN requires the definition of CPTs including the prior probabilities of the SSM competence and the conditional probability of completing the evidence nodes (tasks) given the competence.

For each of the tasks, we make predictions on the relationship between having the SSM competence and the ability to complete tasks. As defined in modeling of human performance (Millán et al, 2002), we use estimates of slip and guess to define the conditional probabilities. In our context, a *slip* is the probability of not being able to complete the task successfully despite having the competence. The probability of completing the task successfully when having the competence $\{P(Task_i|SSM_i)\}$ is therefore *1 – s*, where s is the *slip* factor. A *guess (g)* is the probability of completing the tasks successfully without having the competence. The CPTs require definition of the probability of completing the task whilst having the competence *(1 - s)* and the probability of completing the task while not having the competence *(g)*.

We estimate the CPT parameters for each of the evidence nodes and the conditional probabilities for each of the competence variables. The probabilities of slip and guess were estimated by SMEs and took into the account the following:

1. The participants in the study had received initial training and refreshed skills over a one-year period. It was expected that some participants had acquired enough skill to achieve competence.
2. The simulator scenario in the study had not been practiced before and had challenging weather conditions (moderate sea states). These factors impact the probability of

completing tasks that had been practiced in previous training events in less adverse weather.

3. The task of stopping next to a PIW is more difficult to complete than stopping next to a life raft or stopping next to an FRC (Billard et al. 2020). We assume the probability of a slip is higher and the probability of a guess is lower for the PIW task.
4. The performance of tasks in the simulator, either successfully or unsuccessfully, is considered practice. Competence is expected to increase as the scenario is repeated. The probability of slip on tasks is expected to reduce and the probability of a guess is expected to increase.

In considering the type of task and the environmental conditions, SMEs estimated that there is a reasonable chance of slip given the difficulty of the task and the expectation that people could make errors despite having the competence. The irregularity of wind, wave, and propulsion forces create some variability in performance. Environmental forces could have a sudden negative impact (i.e. causing the vessel to overshoot position) resulting in slip. The environmental forces can also increase the chance of success of an inexperienced driver (e.g. helping slow and stop a vessel that is approaching too fast) creating a successful guess.

Table 4-2 provides a breakdown of the probabilities used in the BN. These are considered an initial estimate of the probabilities based on an expert prediction. The assumed initial probability of having the competence of SSM is estimated to be 60%, and increases in probability in the second scenario. For the evidence nodes, the probability of a successful completion of task is assumed to be lower for tasks that are more difficult. The assumed probability of completing LR and FRC tasks was assumed to be 70%. The probability of completing the PIW task was estimated as 60%

due to the increase in slip factor as the task is more challenging. Similarly, the assumed probability of a guess for the tasks of LR and FRC was assumed to be 30% and the estimated probability of a guess for the PIW task was estimated as 20%. To account for the effect of practice, the SSM competence is expected to increase for the second scenario. The assumed probability of a successful completion for each task was increased by an increment of 10% and the guess rate for each task was also assumed to increase by an increment of 10%.

These estimates are an initial guess of expected outcomes provided by subject matter experts. The estimates are based on expert prediction as they could not be derived from data. The next step in the methodology uses experimental data to refine the CPTs used in the BN.

Table 4-2: Inputs to BN - Expert Estimates

Scenario Attempt 1				
		$P(SSM_1)$		
		60.0%		
$\boldsymbol{SSM_1}$	$P(LR_1 \vert SSM_1)$	$P(PIW1_1 \vert SSM_1)$	$P(PIW2_1 \vert SSM_1)$	$P(FRC_1 \vert SSM_1)$
Y (1 - s)	70.0%	60.0%	60.0%	70.0%
N (g)	30.0%	20.0%	20.0%	30.0%
Scenario Attempt 2				
		$\boldsymbol{SSM_1}$	$P(SMM_2 \vert SSM_1)$	
		Y (1 - s)	70.0%	
		N (g)	30.0%	
$\boldsymbol{SSM_2}$	$P(LR_2 \vert SSM_2)$	$P(PIW1_2 \vert SSM_2)$	$P(PIW2_2 \vert SSM_1)$	$P(FRC_2 \vert SSM_2)$
Y (1 - s)	80.0%	70.0%	70.0%	80.0%
N (g)	40.0%	30.0%	30.0%	40.0%

4.5.4 Step 4 – Refine CPTs Based on Experimental Data

The BN model was created in modeling software, GeNIe, developed by Decision Systems Laboratory of the University of Pittsburgh. The DAG was based on the relationship diagram provided in Section 3.2, and the probabilities outlined in Section 3.3 were used to create the CPTs for each of the nodes.

Data were collected in a simulator exercise, with evidence collected for each of the 39 participants who completed the two scenarios. The data set was split randomly into two groups: a *learning* data set and a *validation* data set. One set of the data (19 records) was used to adjust the parameters of the BN (the learning data) model and the second data set (20 records) was used to predict the accuracy of the model (the validation data).

Conducting parameter learning in the Bayesian Network is often termed *training the BN*. In this exercise, the parameters of the BN CPTs are adjusted in an effort to match the BN model predictions to the outcomes of the learning data set. This exercise is performed in the GeNIe modeling software, which uses an EM algorithm to learn parameters from data (Dempster, 1977). In our use case, we start training the BN with the probabilities set by the experts. As we have a small data set, we assume a low level of confidence in the parameters (20%) to allow the parameters to be flexible to change.

We are now able to make comparisons between the original BN model, based on expert predictions, and the updated model, trained with experimental data.

4.6 Validation Cases

4.6.1 Validation Case 1 – Evaluating Model Predictive Capability Using Task Evidence

The validation data set is used measure the predictive accuracy of the BNs. The initial models developed by expert prediction and the trained models are applied to a new data set (the validation data) to compare each model's predicted outcomes with evidence provided in the data set.

Two validation steps are performed to show how the methodology resulted in an improved BN model:

1. Testing the predictive accuracy of the BN with initial expert predictions of CPT – this step evaluates the suitability of the probabilities estimated by the SMEs.
2. Testing the predictive accuracy of the BN after using the simulation data – this validation shows the impact of using additional simulator data to revise the model parameters.

The validation demonstrates the use of BN for prediction, as the model attempts to identify the most likely occurrence of the evidence nodes. For each of the validation exercises we consider the model's ability to predict the outcome of the final two tasks in the simulation exercise ($PIW2_2$ and $FRC2_2$). These two evidence nodes are selected as they are the last two tasks performed in the simulator exercise. Performance on these tasks is expected to be more likely a result of competence gained through practice than due to a random slip or guess. We compare the predicted outcome of the evidence nodes from the BN model to the actual outcome from the data set.

A benchmark comparison is made with a BN that uses a uniform distribution for initial CPT parameters for all latent and observable nodes. We use this BN to make a comparison with a model that is formed with no expert input and driven only by available data. This approach disregards the expert predictions and assumes an equal probability (50%) for completing or not completing tasks, and related slip and guess probabilities. The parameters are adjusted using the same learning data and using the same learning algorithm as in the expert prediction.

Table 4-3 shows the differences in prediction accuracy of the BN models that were investigated. The Table indicates the number of times the model and validation set had a common outcome on successful completion of task (Yes) or when tasks were not successfully completed (No) for the 20 records in the set. The predictive accuracy of the BN based on expert guesses was 75%, indicating the expert informed probabilities were reasonable. The predictive accuracy of the model increased slightly to 78% when trained with experimental data. The approach of using expert input showed a much higher predictive accuracy than a model trained from uniform parameters. This outcome suggests that the expert guess was needed to generate a suitable model given the amount of available data.

Table 4-3: BN Model Predictions and Comparisons

	Initial Expert Estimate	**Expert Estimate Trained**	**Uniform Trained**
Overall	75% (30/40)	78% (31/40)	48% (19/40)
$PIW2_2$			
Combined	80% (16/20)	80% (16/20)	50% (10/20)
Yes	80% (8/10)	80% (8/10)	0% (0/10)
No	80% (8/10)	80% (8/10)	100% (10/10)
FRC2			
Combined	70% (14/20)	75% (15/20)	45% (9/20)
Yes	100% (11/11)	73% (8/11)	0% (0/11)
No	33% (3/9)	78% (7/9)	100% (9/9)

The method also allows us to investigate how the data set changed the BN CPTs from the initial expert estimates. These changes provide insights on the predicted competence and task difficulty, as a refinement to the estimates initially made by the SMEs. Table 4-4 presents the change in CPT from the initial estimates provided in Table 4-2. The outcomes show the initial probability of SSM competence (SSM_1) was lowered by 13%, indicating the initial estimate of competence was too high. The outcomes also show that most of the probability parameters for successful PIW pickup for each attempt had to be lowered, suggesting this task was more difficult than predicted. The probabilities for stopping at a life raft were increased for each attempt.

Table 4-4: Change in BN Probabilities – Trained model

Scenario Attempt 1				
		$P(SSM_1)$		
		47% (-13%)		
$\boldsymbol{SSM_1}$	$P(LR_1 \mid SSM_1)$	$P(PIW1_1 \mid SSM_1)$	$P(PIW2_1 \mid SSM_1)$	$P(FRC_1 \mid SSM_1)$
Y (s)	76.1% (+ 6.1%)	57.4% (- 2.6%)	50.1% (-9.9%)	63.7% (- 6.3%)
N (g)	41.5% (+11.5%)	16.6% (- 3.4%)	13.4% (- 6.6%)	23.8% (- 6.2%)
Scenario Attempt 2				
		$\boldsymbol{SSM_1}$	$P(SMM_2 \mid SSM_1)$	
		Y (1 - s)	67.7% (- 2.3%)	
		N (g)	25.6% (- 4.4%)	
$\boldsymbol{SSM_2}$	$P(LR_2 \mid SSM_2)$	$P(PIW1_2 \mid SSM_2)$	$P(PIW2_2 \mid SSM_1)$	$P(FRC_2 \mid SSM_2)$
Y (1 - s)	83.8% (+ 3.8%)	69.3% (- 0.7%)	70.4% (+ 0.4%)	81.2% (+ 1.2%)
N (g)	48.4% (+ 8.4%)	26.4% (- 3.6%)	28.6% (- 1.4%)	32.1% (+ 2.1%)

Given the limited amount of data that is available, it is difficult to make conclusive remarks about the final probabilities of the BN model. Additional data are expected to further change the CPTs and increase the predictive accuracy of the BNs.

4.6.2 Validation Case 2 – Investigate Diagnostic Causal Relationship of Background Training

In this section we discuss how the BN can be used as a diagnostic tool and identify causes given a set of observations. We incorporate additional information about the test participants and show

how the model can be used to associate performance to the new information. We introduce a new evidence node, *Background Training (BT)*, to indicate whether the participants received hands-on training during their regular practice prior to performing the simulator exercise. Participants who received hands-on training in regular practice sessions were more likely to be able to complete on-water tasks compared to those who did not (Billard et al. 2019). This information is known for all participants who completed the simulator scenario and the related validation data sets. 26 of 39 participants received hands-on training; 13 did not.

The updated BN for this model is provided in Figure 4-6. The *BT* node is introduced and forms a causal relationship having an influence on the starting competence of the trainee (SSM_1).

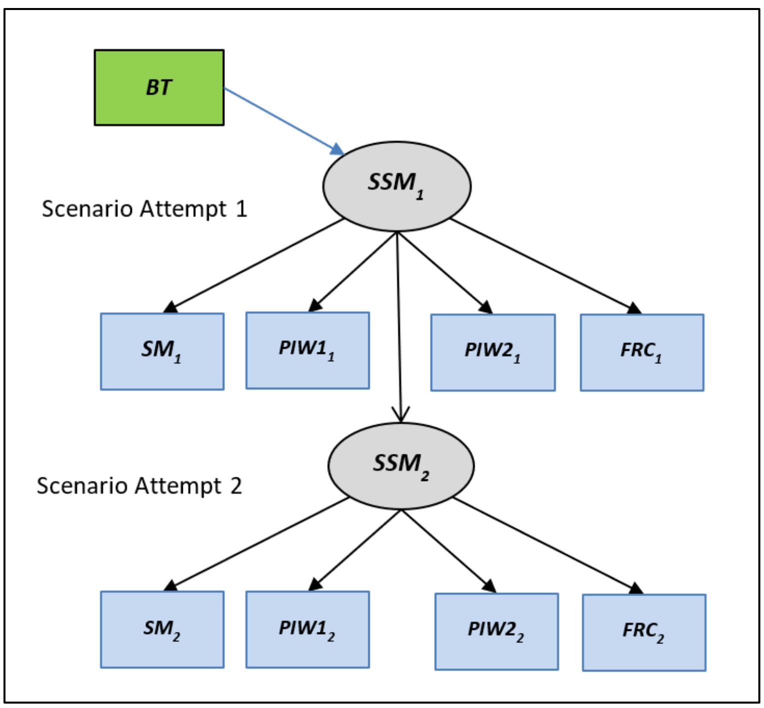

Figure 4-6: BN with Training Evidence Introduced

We again define the conditional probabilities for the influence of training on competence using an expert estimate as there were no existing data available. It is assumed that those who received hands-on training had a higher probability of having the competence, but not greater than 60% as training had not been received in the weather conditions used in the assessment scenario. It was assumed the participants who had not received hands-on training had a lower probability of having the competence, having not received any scenario-based practice. The probability of having received initial training was set to 50%, making the initial probability random. This allows the model to predict the causal affect based on the evidence nodes from the simulator experiments and inherent relationships. Table 4-5 shows the new CPT values defined in the BN.

Table 4-5: Background Training (BT) Conditional Probabilities

	P(BT)
	50%
$\boldsymbol{BT}$	$P(SSM_1 \mid Training)$
Y (1-s)	60%
N (g)	40%

We perform a similar validation procedure outlined in section 4.1. We compare the BN model prediction of BT to the evidence from the validation data set. The evidence in this case is knowledge of the trainee's background in terms of having received hands-on training (Yes) or not (No).

Table 4-6 indicates the model correctly guessed if background training had been received for 65% of the records in the data set. This outcome suggests that additional data or a revised estimate is needed to refine the model and increase the predictive accuracy for this evidence node. As highlighted in Table 4-7, the conditional probabilities of having the SSM_1 competence decreased for both cases (with or without having received background training) when data were used to train the model. These changes in probability can be used to refine the expert estimate or initial CPT for new data sets.

Table 4-6: Diagnostic Accuracy - Background Training

	Expert Estimate Trained
BT	
Overall	65% (13/20)
Yes	54% (7/13)
No	86% (6/7)

Table 4-7: Change in SSM_1 CPTs

$\boldsymbol{BT}$	$P(SSM_1 \mid Training)$
Y (1-s)	55.4% (-4.6%)
N (g)	35.3% (-4.7%)

4.7 Discussion

The methodology in this paper presents an approach to use available information and background expert experience to create probabilistic models of human performance in scenarios for which there is limited available data. This approach can be applied to training applications where the desire is to investigate how observable measures of performance impact skills acquisition and competence. We chose lifeboat coxswain training as the use of simulation has extended training capabilities, and data from new scenarios are available to study this problem area.

We presented a method to develop a student model of lifeboat competence that integrates expert prediction and evidence from a simulator experiment. We derived the BN model for SSM competence using a framework that has been applied in ITS and ECD to use observable evidence from a simulation assessment to design the model. We demonstrated how the BN model can be used to predict performance and diagnose causal relationships, illustrating how the model can be applied to investigate relationships between latent and observable variables.

The validation examples indicate that embedding expertise in the model can result in a high initial predictive accuracy, despite using a small data set. The model's predictive accuracy was further increased as simulator data were used to inform the BN probabilities. This outcome indicates that domain knowledge is valuable in initializing probabilistic models in cases where there is limited data. It is expected that the model's predictive accuracy would improve further if the CPTs are trained with a large data set derived from user performance data.

The scalability of the BN model is a strength that can be further explored. We presented a model of lifeboat coxswain competence that is very narrow (a single competence) and derived from a scenario with fixed weather and tasks. For this study, the modeling of competency is specific to the environmental conditions used in the scenario. In a training program involving multiple practice exercises, the number and order of task types can be varied, and the level of difficulty can change with environmental conditions (i.e. increase in wave height or wind, day or night). The probabilities are expected to be different in scenarios that are easier or more difficult. Additional background information can also be considered, including time between training events and student training experience. The relationship between other competencies can also be established (e.g. practice in maintaining heading seakeeping exercises may improve control of the vessel in SSM).

Figure 4-7 shows an example of how the BN could be expanded to explore causal relationships between variables as more information on the student is known and as evidence is gathered through a training program. These BNs can become complex as they form a detailed model of student competence. These models can be used to investigate factors that affect performance while gaining insights on human performance limitations.

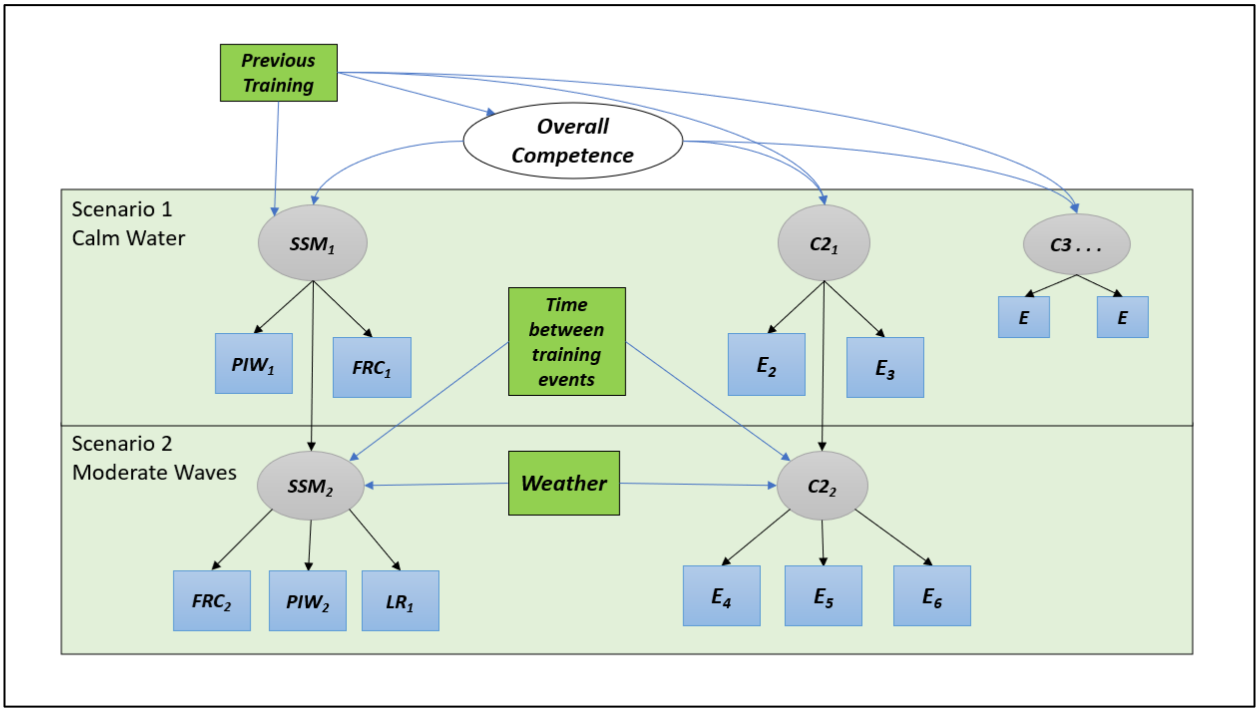

Figure 4-7: Sample BN with Expanded Relationships Representing a Lifeboat Training Program

The formation of a student model using BNs offers additional means to apply probabilistic models to improve training. We have presented a model to study performance based solely on assessment of task performance (i.e. was the task completed successfully or not). The model can be expanded to investigate the specific behaviours performed by the participant in completing the task to study which actions result in the highest probability of success. This type of model tracing is possible given the measures identified in the rubric. The outcomes can be used to model novice and expert performance as inputs to ITS (Millán et al, 2011). The probabilistic modeling of the BN can be integrated with machine learning algorithms to build adaptive training applications to customize training material to an individual's strengths and weaknesses based on evidence gathered in training.

To conclude the discussion, we make four recommendations to researchers who wish to use the methodology to study human performance and training for situations that have limited data. First, we advise the student model to be built as early as practicable to allow for the student BN to be informed with evidence that will be collected. This approach will allow for alignment between the student model with research objectives, and scenarios can be designed to study relationships of interest. Second, we recommend a balance of expert and data-driven input in the probabilistic models. As demonstrated, the modeling of CPTs using expert input can provide a model with suitable predictive accuracy. In cases where data are being collected for scenarios with limited initial data, the expert prediction is a guess. Probabilistic models derived from large data sets are expected to have a higher predictive accuracy. We also suggest that users consider the extended uses of relationship modeling of the BN approach. The BN models can be restructured, and new variables added (latent or observable) to investigate causal relationships and influence of new information. Finally, we suggest the use of simulation to perform assessments and collect data for situations that are normally prohibitive due to risk. Simulation scenarios extend studies to new operating conditions and provide a consistent measure of performance. Digital measures from a simulator exercise can input directly into probabilistic models such as BNs to apply machine learning and adapt training in real time.

4.8 Acknowledgements

We thank Petroleum Research Newfoundland and Labrador and the Industrial Research Assistance Program of the National Research Council who sponsored the study. The authors acknowledge with gratitude the support of the NSERC/Husky Energy Industrial Research Chair in Safety at Sea.

4.9 References

Billard, R., Smith, J.J.E. (2018). Using simulation to assess performance in emergency lifeboat launches. Proceedings, e Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 19179.

Billard, R., Smith, J., Veitch B., (2019) Assessing lifeboat coxswain training Alternatives using a simulator. The Journal of Navigation, Published online by Cambridge University Press: 19 September 2019.

Billard, R., Musharraf, M., Smith, J., Veitch B., (2020), Using Bayesian methods and simulator data to model lifeboat coxswain performance. WMU Journal of Maritime Affairs. Published May 2020. https://doi.org/10.1007/s13437-020-00204-0

de Klerk, S., Veldkamp, B.P., Eggen, T., (2015). Psychometric analysis of the performance data of simulation-based assessment: A systematic review and a Bayesian network example. Computers & Education 85 (2015), 23-34.

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977), Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1. (1977), pp.1-38.

Groth K., Smith, C., Swiler, L. (2014). A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability and System Safety 128 (2014), 32-40

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). STCW including 2010 Manila Amendments, 2017 Edition.

International Maritime Organization. (2014). International Convention for the Safety of Life at Sea (SOLAS), Consolidated Edition. London: International Maritime Organization.

Käser, T., Klingler, S., Schwing, A., Gross, M. (2017). Dynamic Bayesian Networks for student modelling. IEEE Transactions on Learning Technologies, Vol. 10, No. 4. Oct.-Dec. 1 2017.

Klein, G., (2008), Naturalistic decision making. Human Factors: The Journal of Human Factors and Ergonomic Society, 50(3), 456-460.

McClernon, C. K., McCauley, M. E., O'Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(3), 207-218.

Millán, E., Perez-De-La-Cruz, J.L., (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction 12: 281-330, Kluwer Academic Publishers, Netherlands

Millán , E., Loboda, T., Perez-de-la-Cruz, J.L. (2010). Bayesian networks for student model engineering. Computers and Education, 55, 1663-1683

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). A brief introduction to evidence-centered design. CSE technical Report. Los Angeles: The National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved from http://www.cse.ucla.edu/products/reports/r632.pdf.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU Journal of Maritime Affairs, 16(2), 247-263.

Stefanidis, D., Korndorffer, J.R., Markley, S., Sierra, R., Heniford, B.T., & Scott, D.J. (2007). Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training? *Journal of the American College of Surgeons*, 205(2), 307-313.

Weinert, F. E. (2001): Competencies and Key Competencies: Educational Perspective. International Encyclopedia of the Social and Behavioral Sciences, vol. 4, Elsevier, 2433–2436.

5.0 Chapter 5: USE OF SIMULATIONS TO PREDICT LIFEBOAT SURVIVABILITY IN EXTREME WAVES AND THE EFFECTIVENESS OF COXSWAIN PERFORMED ACTIONS

Randy Billard[1], Robert Rees[1], Brian Veitch[2], and Antonio Simões Ré [3]

[1] Virtual Marine, [2]Memorial University of Newfoundland, [3]Xataan Consulting

5.1 Co-authorship Statement

This manuscript was submitted to the Royal Institute of Naval Architects International Journal of Maritime Engineering in 2020. Writing was led by Randy Billard. Robert Rees assisted in developing and running the numerical simulations used to create data sets. Brian Veitch and Antonio Simões Ré assisted in validation of the simulator outcomes with previous experimental work and provided guidance in writing and revising the paper.

5.2 Summary

Simulations were used to investigate the performance of lifeboats in high sea states using a virtual wave tank. Numerical simulations were performed in regular and irregular waves to study launch performance in extreme weather conditions. Limitations in launch equipment and the role of the timing of coxswains' actions were investigated. The study indicated that the lifeboat may not be able to successfully launch when significant wave heights are above 8 m and the lifeboat is

launched near the trough of a wave. High initial setback and continuous wave forces result in the vessel being unable to clear away from the launch platform. As wave heights increase, the amount of setback and time to exit the launch area increases. Over 35% of launches resulted in the lifeboat being unable to clear from the launch area when significant wave heights were 10 m or above. The study also identified that delay in completion of actions performed by the coxswain, such as releasing the lifeboat hooks and applying throttle, can increase setback and time to exit the launch area.

5.3 Introduction

The successful launch and clear away of a lifeboat in high sea states is affected by both the capabilities of the lifeboat and the actions taken by the coxswains. The effects of coxswain actions on the ability to complete a successful launch and sail away have not been fully investigated, nor have the limitations of the launching equipment in high sea states been fully explored. This paper investigates both.

Previous scale model experiments were performed to evaluate the factors affecting a successful lifeboat launch and sail away (Simões Ré et al., 2002, Simões Ré & Veitch, 2004, Simões Ré et al., 2008). These experiments investigated the limitations in launching considering factors related to wave height, launch configuration, and lowering speeds from the davit. The experimental studies used regular waves which is a simplification of real conditions where wave shapes are irregular. These studies also did not include the full range of sea states that are possible in offshore operations as wave heights were limited to 10 m. Additional studies used numerical simulations to study similar factors and explored the effect of timing of hook release and application of propulsion (Gabrielson et al., 2011).

Industry studies have identified that coxswain skill has an impact on a successful lifeboat launch, although benchmarking of skills is difficult due to limitations in training (Robson, 2007). Evaluating the impact of human performance and skills on a successful launch in high sea states is not practical. Due to the perilous nature of launching lifeboats in rough conditions, the role of the operator (coxswain) is not something that can be ethically investigated in field trials or experiments, nor practiced in realistic (rough) wave conditions. Due to the nature of model experiments, specifically the scaling of time, it is difficult to use model tests as a means to investigate time dependent human factors.

In this research simulations were used to explore the lifeboat performance in wave heights not previously tested in scale model experiments and field trials. The simulator is also used to study how the timing of actions performed by the coxswain, including applying the throttle and releasing the hooks, affect launch performance.

Details are first presented on the launch procedure and the performance measures discussed in the paper.

5.4 Background

5.4.1 Launch Procedure

As summarized in previous research (Simões Ré et al., 2002, Gabrielson et al., 2011), there are multiple phases of a lifeboat launch. They are as follows:

- Lowering phase: lowering the lifeboat from the davit system to the water surface.
- Water entry: starts when the vessel enters the water and the lifeboat becomes buoyant. During this phase, water fills the vessel hydrostatic release unit, and hydrostatic pressure moves a cable link to allow the hook release handle to open.
- Release: starts when the hook is released and the vessel is free from the fall wires.
- Sail away: vessel propulsion (throttle) is applied and the operator maneuvers the vessel to a safe area away from the launch platform.

The launch starts when a brake wire is pulled and the vessel begins lowering from the davit. The vessel continues to lower until the vessel is in the water and a time count begins on filling the hydrostatic release. The hydrostatic release activates when the vessel remains buoyant for 3 seconds or longer. If the wave falls away from the vessel before it is buoyant, the hydrostatic release drains and the time restarts. Once the vessel is buoyant for three seconds, the hydrostatic release system allows the hook system to operate. A hydrostatic indicator on the hook release system moves providing a visual cue to the coxswain. The standard procedure is to release the hooks and then apply full throttle as quickly as possible and drive away from the launch platform.

This paper focuses primarily on the water entry, release, and sail away and considers the relationship between the actions performed by the coxswain and the timing of transition between the phases.

5.4.2 Setback

The experimental studies (Simões Ré et al., 2002, Simões Ré & Veitch, 2004, Simões Ré et al., 2008), identified that the amount of setback, or backwards trajectory of the vessel, increases in higher sea states when the launch position on the wave is near the trough on the lower part of the wave upslope. Wave and wind forces impact the vessel on water entry and can push the vessel backwards towards the launch platform if the waves are against the evacuation direction (a head sea). In head seas with wave heights of 6 m and above, setback can result in the vessel being pushed back to within critical safety zones of launch platforms. The amount of setback and likelihood of occurrence of setback increases with wave height.

Total setback can be a result of a single wave or multiple wave encounters may cause progressive setback before the vessel begins to move forward (Simões Ré et al., 2002). Figure 5-1 shows a sample trajectory plot of a launch to illustrate the case where a vessel experiences initial setback (SB), progressive setback, and is then able to progress forward. The vertical position is plotted on the Y axis, with vessel lowering to the still water line at y = 0. The horizontal displacement positive is the distance in the direction away from the launch platform. X = 0 is the starting horizontal position of the vessel when lowered. In this sample the vessel is setback (-ve x direction) on the first wave encounter, and the second wave encounter results in higher, or progressive wave setback. The vessel is then able to progress forward. The subsequent waves create some backwards displacement with each wave encounter, but the overall movement is in the +ve x direction away from the launch platform. Initial setback, progressive setback, and forward progress are used to describe results in this paper.

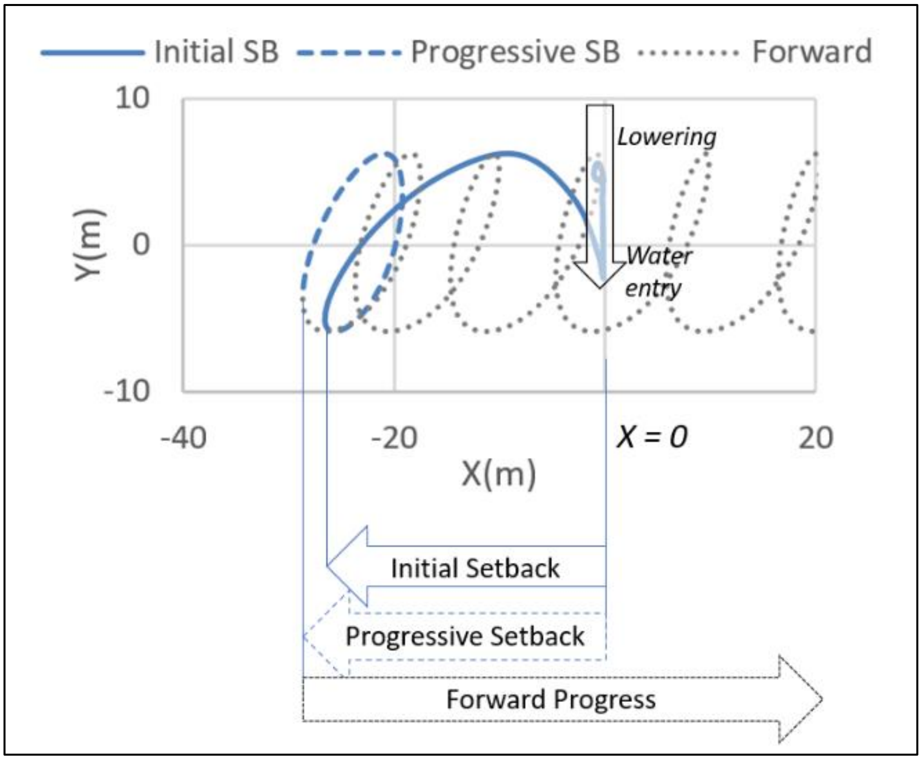

Figure 5-1: Progressive Setback

5.4.3 Impact of Launch Position on Wave

Studies identified the impact of timing of release of the lifeboat at different points on the wave in a head sea (Simões Ré et al., 2002, Simões Ré & Veitch, 2004, Simões Ré et al., 2008). Launching on the trough of the wave can result in significant setback and launching on the crest of a wave results in minimal, or no, setback. The experiments also showed the effect of "wave shadowing," whereby the lowering speed of the vessel and the wave speed resulted in launches on the leeward side of the wave. With reference to Figure 5-2, most launches occurred between -60 and 60 degrees, on the wave upslope. In effect, it is difficult to launch on downslopes, which are favourable to good launches, and launches are more likely to occur on upslopes which result in large setbacks. As wave height or wave steepness increases, the zone of possible launch positions tightens.

Taken together, these findings indicated that the timing of the launch relative to the wave is very important. It should not be left to chance as it is something that is within the operator's ability to control, at least to some extent.

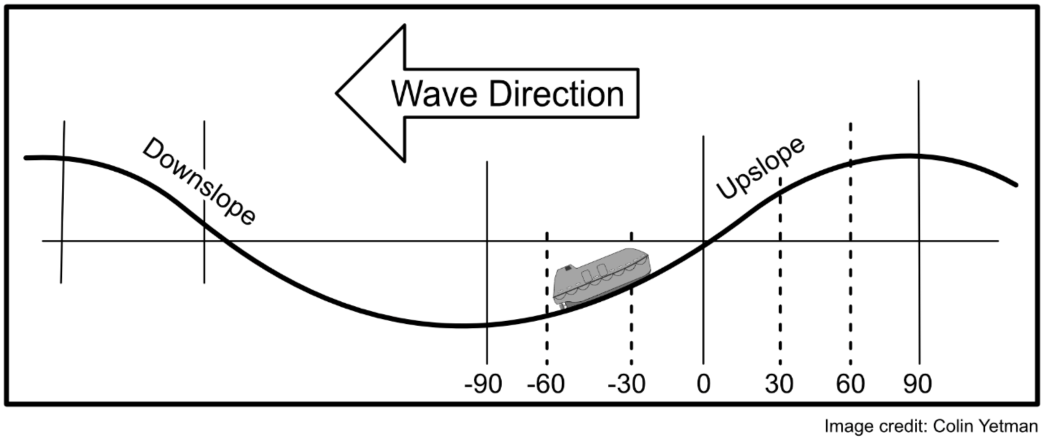

Figure 5-2: Launch Positions for Lifeboat Water Entry

5.4.4 Performance Measures

The primary measure of performance in the study is setback. Additional measures are defined for this study based on target operational outcomes. These new measures are as follows, with reference made to Figure 5-3.

The first additional measure identifies launches which may result in contact with the launch platform and consequently may result in damage to the vessel or harm to the crew. Examination of launch platforms and launch configurations indicates that davit systems are placed to provide 20 to 40 m of clearance from the base of the platform. Setback greater than 20 m may result in impact with the launch platform or result in vessel being within a zone of high risk of impact. In Figure 5-3, X = 0 is the position directly below the launch area and X = -20 m is the distance travelled towards the platform, opposite the target evacuation direction. To evaluate performance for a given set of launches, the percentage of outcomes with greater than 20 m in setback (%Setbacks>20m) was calculated.

Another measure was introduced to evaluate whether the lifeboat is able to evacuate from the launch platform quickly. Clearing time is defined as the time required for the vessel to leave the splashdown area in the evacuation direction and reach a target distance, which is defined as 20 m from the launch position (X = 20 m in Figure 5-3). Timely clearance of the lifeboat from the launch area is desired to escape harm from hazards near the launch platform and to permit the launch of other vessels from adjacent davits. The percentage of occurrences with greater than 60 s of time needed to reach 20 m is calculated for a set of launches (%Clearance Times>60s). The amount of setback and clearing time to exit the splashdown area are related, as the vessel must travel a longer distance to exit the splashdown area if it is setback farther.

If no measure was recorded for this performance criteria, the vessel was unable to reach the escape line at X = 20 m. This outcome is defined as a failed clearance. A measure of failed clearance identifies a performance limit as the vessel is not able to progress forward in the wave environment tested. This limit is further discussed in the results.

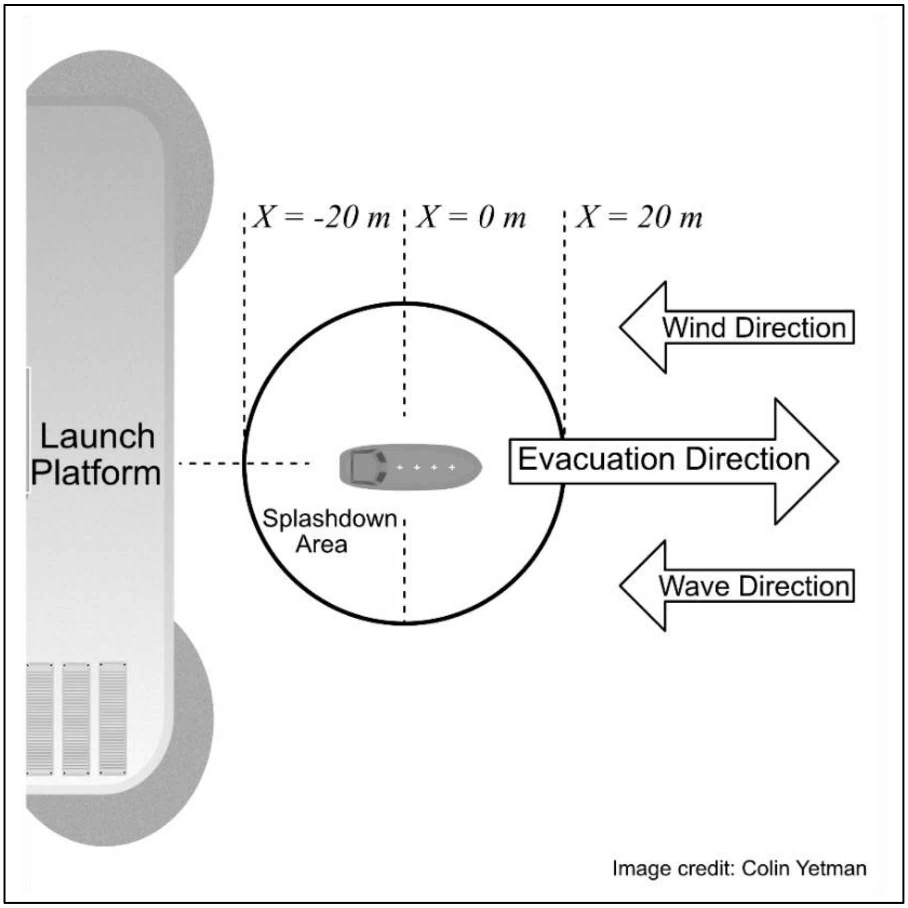

Figure 5-3: Performance Measures: Setback>20m and Clearance Time>60s

5.5 Scope

With the advent of simulator technology, it is now possible to explore lifeboat and operator performance in weather conditions typical of their location of operation. Lifeboat simulators are designed with accurate numerical behaviour of vessel motions and wave environments. Trainees interact with realistic lifeboat equipment and perform actions as they would in a real vessel. Studies performed with a lifeboat simulator have evaluated how skills transfer from simulator training to real vessels (Magee et al., 2016) and how skills are acquired in initial training (Billard et al. 2019). Recent studies have focused on how training affects coxswain skill acquisition and launch performance in moderate weather environments (Billard et al., 2018, Billard et al., 2020). These studies have not investigated the impact of human factors in high sea states.

Simulators are increasingly used to train lifeboat coxswains. Trainees can practice and improve timing of actions, such as releasing the hooks and applying the vessel throttle. There is increased knowledge of the times taken to complete tasks in a lifeboat launch via data collected through simulator training programs. The timely or delayed performance of these actions is expected to impact the amount of lifeboat setback and the time to clear from the launch area. Evaluating how the timing of these actions affects the launch outcomes will help to define training objectives.

As in other studies, simulators can explore scenarios where data is scarce or difficult to obtain (Groth et al. 2014) and can specifically extend knowledge of coxswain and lifeboat performance to high sea states. The study of human factors using simulation to evaluate performance is evident in other operations including flight (McClernon et al. 2011), medical (Stefandis et al. 2007) and marine (Sellberg, 2017) training. This research shows an example of how simulations can be used to evaluate how operator actions impact the ability to successfully launch a lifeboat.

The purpose of the research was to use numerical simulations to 1) assess the performance and limitations of lifeboat launch systems in extreme seas and to 2) study the impact of the timing of human actions on the launch and sail-away of the lifeboat.

A numerical simulator, a Virtual Wave Tank (VWT), was designed to emulate the lifeboat and wave conditions performed in previous research (Simões Ré et al., 2002, Simões Ré & Veitch, 2004, Simões Ré et al., 2008). Validation was performed to ensure the measured setback is comparable between the numerical simulator and experimental studies performed in a wave tank.

Comparisons were made for multiple wave heights and launch positions on the waves. The kinematics of the vessel in the VWT were also compared with results from the experimental studies.

After validation, we performed three investigations with the VWT to study the effect of wave height and timing of coxswain-performed tasks on launch performance. The first investigation built on the outcomes of the experimental tests (Simões Ré et al., 2002) and extended the regular wave conditions to waves up to 16 m. Simulations were then performed in irregular sea states with significant wave heights of 6 to 12 m to investigate launch performance in irregular waves with 100-year return period extreme wind speeds based on historical data of weather conditions in the North Atlantic (C-Core, 2015). The third investigation studied how the timing of human actions affected the likelihood of a successful launch. The time taken to apply propulsion (throttle) is varied. The time to release the lifeboat hooks once the vessel is buoyant in the water and able to be released is also varied. The impact of delayed response in applying the throttle and hook release is studied. Comparison are also made between cases where throttle is applied prior to release of the hook to investigate how applying an initial propulsion force influenced the ability to complete a successful evacuation.

Performance is evaluated using the measures identified in the previous section. The investigations focused on performance in head seas.

The following research questions are investigated:

- What is the expected setback of a lifeboat in extreme regular waves and irregular waves?
- How is the time to clear the lifeboat from the launch structure affected by sea state?

- How does delay in lifeboat throttle and hook release affect launch and evacuation of a lifeboat?

5.5.1 Virtual Wave Tank (VWT)

The VWT simulation environment used in the study is a 3D physics-based engine that was specifically created to model the motion of small crafts in marine environments. Physics models were derived from studies of vessel motions in waves, including scale model and full-scale testing of the vessels (Simões Ré et al., 2002, Simões Ré & Veitch, 2004, Simões Ré et al., 2008, Magee et, al. 2016). Numerical models for all phases discussed in the launch procedure were included in the simulation environment.

Numerical models were implemented to provide physic-based responses and timings during the launch phases. The vessel behaviours at water entry were modeled to include the tension in the lowering wires before the hook is released, the dynamic behaviour of the lifeboat as it interacted with the water surface, and the release of the vessel. The propulsion and hydrodynamic behaviours of the vessel once in the water were also modeled. Previous studies have validated the maneuvering and performance characteristics of the vessel modeled in this study are representative of the behavior of the real lifeboat (Billard et al. 2020). Models are resolved on the computer GPU to allow for high-speed and high-resolution wave meshes to calculate hydrostatic and hydrodynamic forces.

The vessel was modeled with dimensions, weight, propulsion and steering to study the lifeboat's ability to maneuver in the environmental conditions considered. The vessel modeled in the simulator was a fully loaded lifeboat with length, weight, and displacement parameters closely matched to the vessel used in experiments performed by Simões Ré et al. (2002).

A twin fall davit was modeled with fall wires attached to the fore and aft hooks of the vessel during lowering. The lowering speed of the vessel was kept constant at 1.0 m/s. The launch height of the davit was 35 m. The lowering of the lifeboat is normally controlled by pulling a brake release from within the lifeboat to extend the fall wires. In the simulations, the fall wires continued to extend until the hook is released. This is a normal procedure to make sure the vessel begins to float, and to reduce the likelihood the vessel is only temporarily buoyant if the wave falls away from the lifeboat.

A virtual agent was used to perform the actions of the coxswain in the simulator. The virtual coxswain could be programmed to release the hook, manipulate the throttle and attempt to steer the vessel to desired headings. Timings could be set to perform actions instantaneously or with delays, or in different orders (i.e. applying throttle before hook release). The resultant behaviour of the vessel was determined by the physics engine which applies and resolves forces depending on the actions taken by the coxswain. As an example, a delay in moving the throttle ahead after the hook was released resulted in a delay in the propulsion and the vessel was free to drift until propulsion was applied. When maneuvering, the virtual coxswain attempted to maintain a constant heading and used corrective steering to come back to a heading if the vessel veered off course. The study assumed steering was maintained to target a heading directly into the waves and away from the launch platform.

5.6 Study Methodology

The study included two stages: 1) validation of the simulator measures with data from experimental studies, and 2) using the simulator environment to perform new studies. Three studies, or investigations, were performed with the VWT to study the lifeboat performance in higher sea states and to consider the timing of coxswain actions. The investigations varied wave shape, wind speed, and coxswain timings to study the effect of these variables. Comparisons are made between each study to illustrate results.

5.6.1 Validation – Simulator and Scale Model Experiment

Comparisons are made between the outcomes of scale model testing performed in previous research to validate the simulation. Data sets were created using the VWT using a stokes regular wave, with wave heights from 2 to 10 m, and wind speeds matching the scale model experimental tests (Simões Ré et al. 2002). Validation is performed using the following comparisons between the simulator outcomes and the scale model tests:

1. maximum setback for each wave height;
2. setback for multiple launch positions on the wave; and,
3. checking the trajectory of the vessel during water entry and sail away.

5.6.2 Investigation 1 – Study of Individual Wave Setback in High Sea States, Regular Waves

The first set of test cases investigated the impact of environmental conditions on lifeboat setback with testing extended to higher sea states and wind speeds representing storm and hurricane conditions. Test cases were performed with a stokes regular wave shape with wave heights ranging from 2 m to 16 m. The approximate wave steepness in each case was 1/20. The simulation used wind speeds and wave heights similar to the parameters used in the experimental studies (Simões Ré et al. 2002) for wave heights up to 10 m. The wave heights and wind speeds were extended to wave heights of 12, 14, and 16 m, using average wind speeds for observed wave conditions (C-Core, 2015). The parameters for each wave tested is provided in Table 5-1.

48 launches were performed for each wave height. For each launch, the starting time of the launch was varied resulting in a different launch position on the wave, with launches covering a full wave cycle of one wave period. The maximum time permitted was 240 s (4 minutes).

Table 5-1: Series 1 - Regular Wave Parameters

Wave Height (H_w)	Wave Period (T)	Mean Wind speed
[m]	[s]	[m/s]
2	5	10
4	7	12
5	8	16
6	9	17
7	9	18
8	10	19
10	11	22
12	12	28
14	13	30
16	14	33

5.6.3 Investigation 2 – Study of Lifeboat Performance in Irregular 100 YR Seas

The second set of simulations investigated the launch and sail away phase of the lifeboat in irregular shaped head seas and high wave heights. The lifeboat was lowered and launched into a sea state with a defined significant wave height (H_s) and irregular wave pattern. The irregular wave shape included dominant waves and lower frequency minor waves. Waves were generated from a fast Fourier transform to generate the desired H_s, as measured by the mean wave height of the highest 1/3 of the waves. Individual wave heights could exceed H_s. The maximum wave heights in the test cases are presented in Table 2. The peak period (T_p) is the dominant wave with the highest energy. Wave heights of 6 m to 12 m were selected to study vessel performance where high setback is likely. Wind speeds were taken to be representative of 100-yr occurrences in the North Atlantic (C-Core, 2015) and are higher than the winds used in the regular waves.

For each wave height, simulations were performed with three different wave patterns. Each wave shape had the characteristic parameters identified in Table 5-2. 48 launches were performed for each wave pattern to cover a full cycle of a dominant wave. The data for each wave pattern was combined for analysis, resulting in 144 launches for each combination of wave height and wind studied.

Table 5-2: Series 2 - Irregular Wave Parameters

Significant Wave Height (H_s)	Max Wave Height	Peak Wave Period (T_p)	Mean Wind Speed
[m]	[m]	[s]	[m/s]
6.0	8.7	9.0	20
8.0	11.5	10	25
10	13.2	11	30
12	15.8	12	33

5.6.4 Investigation 3 – Study of Human Performance on Evacuation Performance in Irregular 100 YR Seas

The third set of simulations varied the time to complete actions performed by the coxswain in the lifeboat launch and clear away. The virtual coxswain in the VWT simulation was programmed to perform the hook release and to move the throttle from neutral to full propulsion at controlled times. Data collected from training courses performed by Virtual Marine has indicated that the timing of release of hooks can vary from 1 to 5 seconds following an indication that the hydrostatic bladder has filled, and the hook release system can be operated. This delay can be caused by a combination of human reaction time, difficulty in operating the hook release handle, or time taken to perform other tasks. Training records have also identified the time to apply full throttle can vary between coxswains. Delay in application of throttle following hook release means the propulsion of the vessel is delayed, and the vessel is free to drift if the hooks have been released.

The study first investigated the application of throttle and delay in hook release separately. Throttle delay cases assumed the hook was immediately released when the hydrostatic bladder had filled, and times presented are relative to the time of hook release. The time to throttle (TT) is the amount of time the vessel is untethered by the fall walls and free to drift before throttle is applied. For the hook release cases, time to hook release (TR) was relative to the instant the hydrostatic bladder has filled (t = 0), and the vessel remained tethered until release of the hook. In these cases, throttle was applied immediately on hook release. The timings are summarized in Table 5-3.

Table 5-3: Delayed Throttle and Hook Release Cases

Label	Hydrostatic Ready	Time to Throttle (TT)	Time to Hook Release (TR)
TT2	t = 0 s	t = 2 s	t = 0 s
TT4	t = 0 s	t = 4 s	t = 0 s
TR2	t = 0 s	t = 2 s	t = 2 s
TR4	t = 0 s	t = 4 s	t = 4 s

These initial cases studied the delayed performance of actions normally taken in a launch sequence where the typical launch procedure is 1) wait until the vessel is buoyant, 2) release the hook and 3) apply throttle.

An additional series of tests was performed to investigate the impact of early application of throttle, prior to release of the hooks. This emulates an operator decision to apply propulsion before the lifeboat is released from the fall wires. This procedure has been suggested by experienced operators as a means to give the vessel initial thrust to combat wave forces, albeit not a standard operating procedure.

In these cases, the virtual coxswain applied the throttle fully when the vessel was buoyant (i.e. hydrostatic interlock had filled, t = 0), and remained tethered. Four use cases with different combinations of time to throttle (TT) and time to hook release (TR) are identified in Table 5-4. Early throttle provided a propulsion force before the vessel becomes untethered, and the hook was released at a time following the throttle.

Simulations were performed for the irregular waves identified in Table 5-2, with H_s from 6 to12 m. Data sets were again acquired for three wave patterns and combined for analysis, resulting in 144 launches for each case and wave studied.

Table 5-4: Early Throttle Cases

Label	Hydrostatic Ready	Time to Throttle (TT)	Time to Release (TR)
TT1-TR2	t = 0 s	t = 1 s	t = 2 s
TT1-TR3	t = 0 s	t = 1 s	t = 3 s
TT2-TR3	t = 0 s	t = 2 s	t = 3 s
TT2-TR4	t = 0 s	t = 2 s	t = 4 s

5.7 Results

In this section summarize the outcomes of the investigations are summarized and discussed. Comparisons are made between outcomes of the studies to illustrate the effect of the variables studied. Multiple measures are discussed to provide insights on how the outcomes are related and to make comparisons between the individual investigations.

5.7.1 Results – Validation, Simulator and Scale Model Experiment

Comparisons were made between the simulator measures and the experimental studies performed by Simões Ré et al. (2002) to validate the measures and behaviors observed in the simulator are similar to the experimental studies. A sample of the validation cases are discussed.

Figures 5-4 and 5-5 show the measured setback for various launch positions on a regular wave, with 90 degrees being the wave crest and -90 degrees being the wave trough. The comparisons show the observed behavior is the same in the simulator (Simulator) compared to the scale model experiment (Experiment), with setback increasing as the vessel is launched closer to the trough of the wave. Of note, the setback in the experiment was limited at approximately 11 m due to the experimental setup, with the model impacting the launch structure at this point. The dashed line on Figures 5-4 to 5-6 indicates this limit for the experimental trials. The setback in the simulator trials was not limited. Some differences in the setback measures are observed on the upslope near the trough of the wave (30 to 60 degrees) when the wave height is 6 m and there is a close match with most phase angles when the wave height is 10 m. As indicated in Figure 5-5, the measured setback for the simulator continued to increase above 11 m as the vessel was launched closer to the trough (0 to 30 degrees) as the launces were not limited by collisions.

Figure 5-6 shows the setback vs. wave height (H_w) for specific waves for both the simulator and experimental measures. The solid line indicates the values where setback is double the maximum wave height. The experimental outcomes showed maximum setback is approximately double the wave height up to a wave height of 6 m (Simões Ré et al., 2002). At higher sea states this could not be confirmed in the experimental results due to the impacts of the evacuation craft with the structure. The increase in setback from the simulator tests followed a similar trend line, with some occurrences of setback above the prediction for the 6 m wave height. The trend of increasing setback and variability in setback with increased wave height is consistent between the simulator and experimental measures.

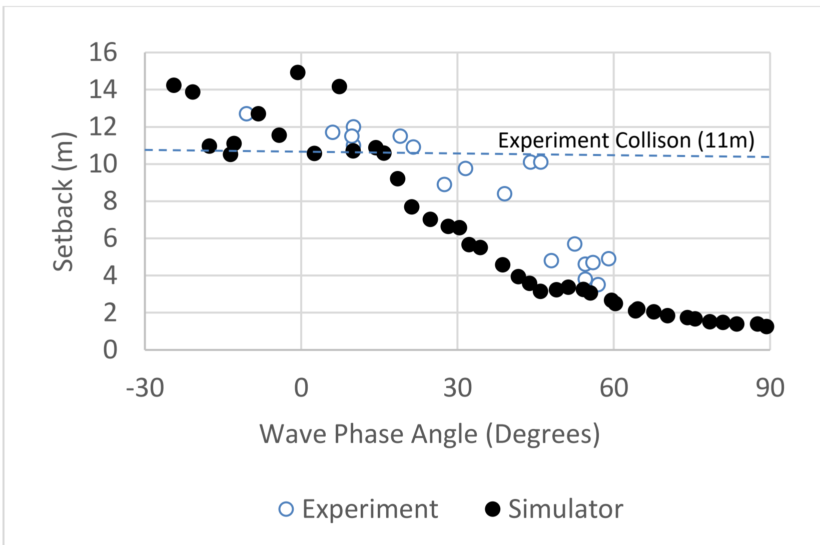

Figure 5-4: Setback vs. Wave Phase Angle, H_w = 6 m

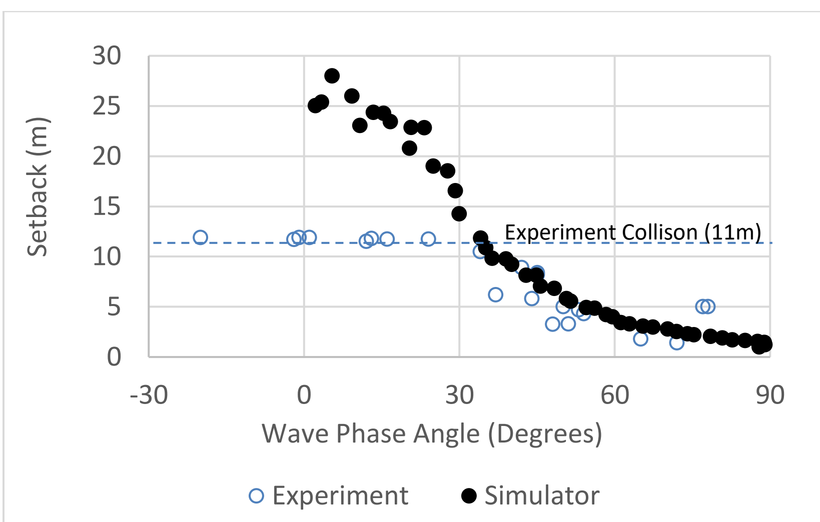

Figure 5-5: Setback vs. Wave Phase Angle, H_w = 10 m

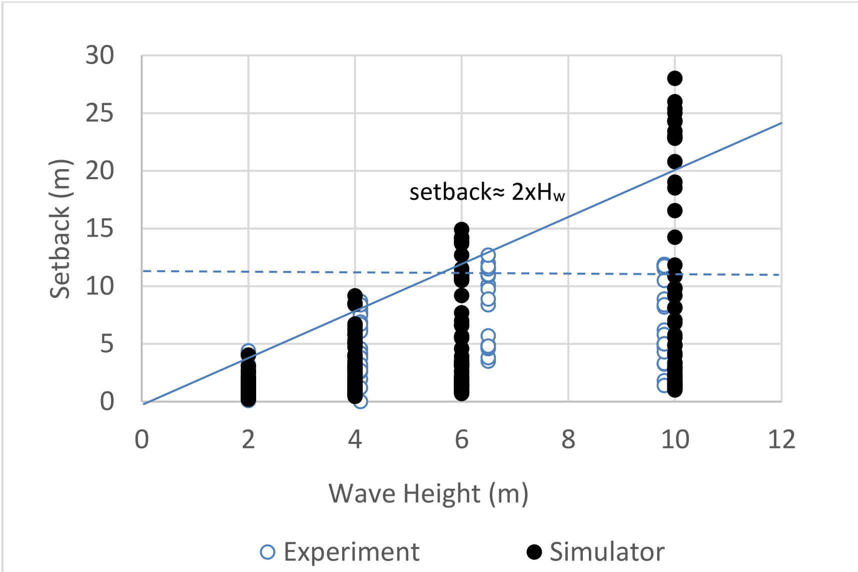

Figure 5-6: Setback vs. Wave Height

Trajectory comparisons were made between the experimental cases and the simulator to ensure vessel kinematics were similar. A key focus was the observed behaviour of the vessel when it was launched on different positions between the crest or trough of a wave. A sample of the validation case is discussed. Figures 5-7 and 5-8 are sample runs from the simulator showing the trajectory of the vessel on launch and sail away. Figure 5-7 shows the vessel setback was lower when the vessel was launched near the crest of the wave, and the vessel was able to continue forward steadily with each wave encounter. With large vessel setback, as in Figure 5-8, the vessel had to first overcome the backwards motion. The vessel progressed more slowly, with some progressive setback on the initial wave encounters, and then was able to continue forward with additional wave encounters. Similar behaviour was observed in the experimental studies.

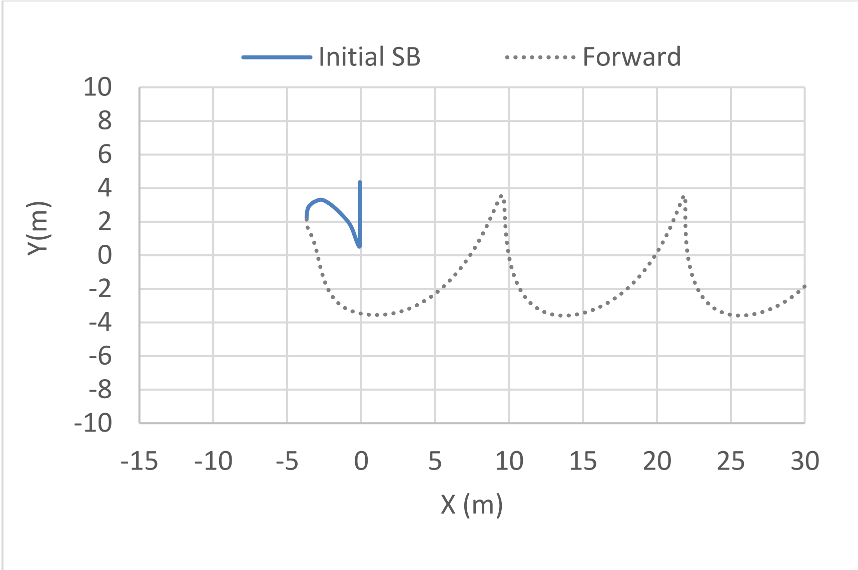

Figure 5-7: Simulator XY Trajectory - Launch near wave crest: H_w = 7m, T = 9m

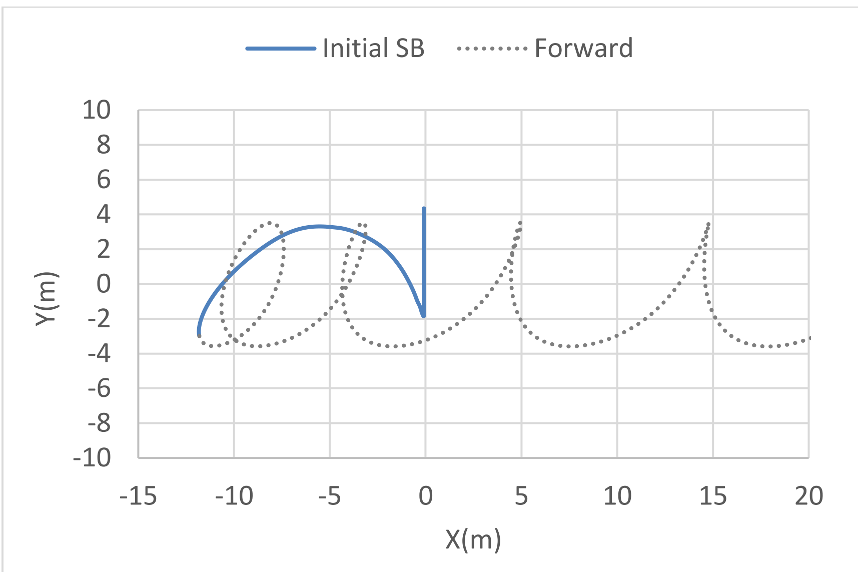

Figure 5-8: Simulator XY Trajectory – Launch near wave trough H_w = 7 m, T = 9

In summary, the comparisons indicate the virtual wave tank provides measures that are representative of the vessel and wave interactions seen in the experimental studies. The amount of setback with wave height and the change in setback with position of launch on the wave (crest or trough) were consistent with the experimental studies. The motion of the vessel on water entry and sail away in the simulator was also representative. Differences between the measures can be attributed to differences in scaling, variability in physical observations compared to numerical simulations, and differences in limitations between the experimental test setup and the simulator.

5.7.2 Results: Investigation 1 – Study of Individual Wave Setback in High Sea States, Regular Waves

A summary of the setback measures for each set of launches is provided for each of the regular waves studied. The measured setback for each launch is also related to the launch position on the wave (phase angle). In effect, this data set provides an extension to the outcomes presented in the scale model experiments, with the outcomes extended to higher wave heights.

Table 5-5 provides a summary of the setback measures for launches performed for each wave height tested, from 2 m to 14 m. Summary data includes the average setback (Avg. SB), the median of the measured setback (Med.), and standard Deviation (SD) for each set of 48 launches performed for each wave height. The 90th percentile (90th PER.) of measured setback is provided to indicate the higher measures in the data set for each sea state.

The outcomes indicate increasing setback with increase in wave height, with the average setback increasing from 3.23 m in a 4 m wave height to over 30 m in a 16 m wave. The measures indicate setback as high as 65.9 m in a 16 m wave. There is also higher variability in the setback as wave heights increase, which is consistent with previous studies. For all wave heights, the median was lower than the average setback, indicating there were a higher number of low setback measures for each set of launches. Figure 5-9 shows a graphical summary of this information in a box plot.

Table 5-5: Setback Summary - Regular Waves

H_w (m)	4	6	8	10	12	14	16
Avg. SB(m)	3.23	5.82	8.47	10.1	14.1	20.0	30.4
Med.	2.85	3.75	6.05	6.31	9.11	14.0	28.2
SD	2.99	4.68	6.83	8.98	13.1	17.0	22.4
90^{th}PER.	7.23	13.0	17.8	24.3	36.3	45.6	62.4
Max SB(m)	9.1	14.9	25.1	28.0	42.7	54.5	65.9

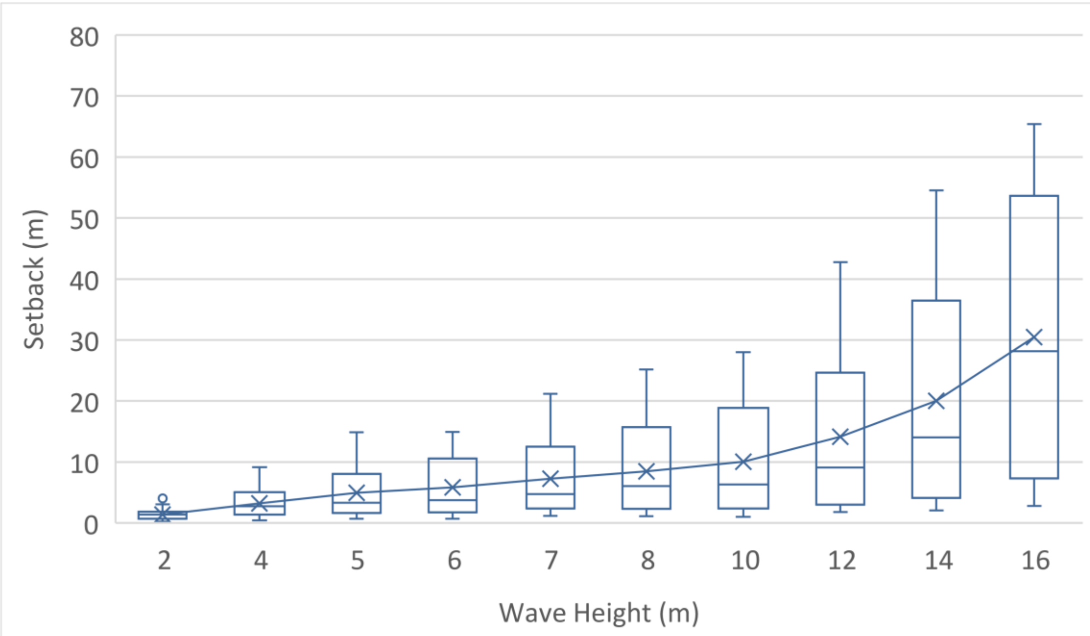

Figure 5-9: Vessel Setback, Regular Waves

Figure 5-10 shows the setback values and phase angles for each set of simulated launches in the higher wave heights, from 10 m to 16 m. The results indicate that the splashdown occurs most frequently between 0 and 90 degrees, with few occurrences of launches outside of this range when the wave height is greater than 8 m. Analysis of the setback and wave angle for higher wave heights shows the maximum setback increased significantly when the vessel was launched closer to the trough of the wave (0 to 30 degrees). Low setback values are possible when the boat is released closer to the crest of the wave (60 to 90 degrees). The results are similar to the outcomes of previous research (Simões Ré et al., 2002).

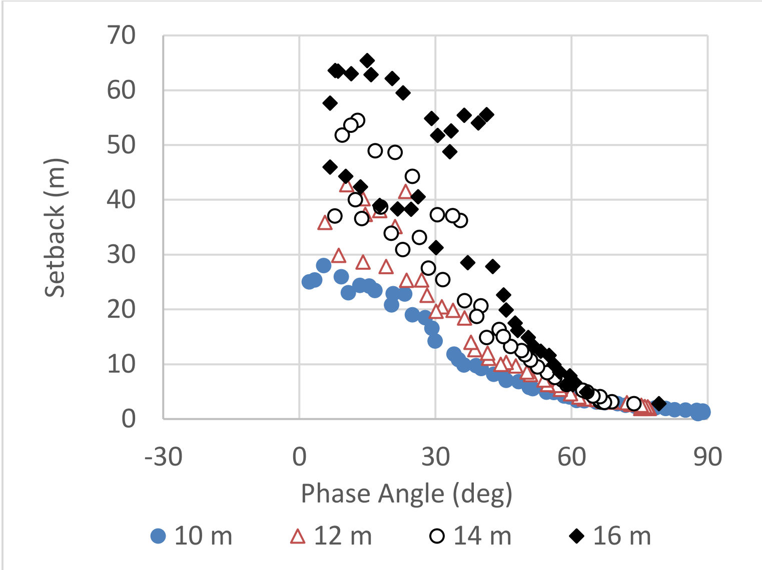

Figure 5-10: Setback vs. Phase Angle, Regular Waves

Figure 5-11 shows the percentage of launches with greater than 20 m setback. As indicated in this figure, in wave heights of 10 m or greater, the number of occurrences increases with wave height, with over 50% of the launches in a 14 m wave meeting this criterion. Figure 5-12 shows the percentage of launches that required greater than 60 s clearance time (%Cleartimes>60s). For wave heights of 10 m and above there were observed cases where the vessel could not exit the launch location in less than 60 s, with occurrences increasing as wave height increases. The vessel was able to evacuate and reach 20 m from the launch position in less than 60 s for all launches performed in 8 m wave height or less.

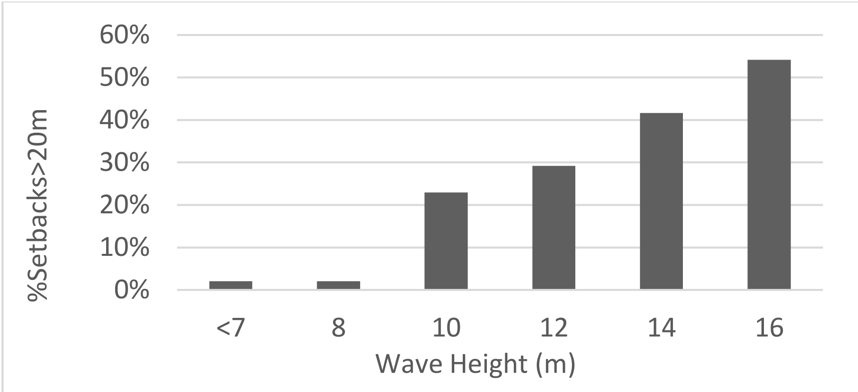

Figure 5-11: Setback Occurrences Greater Than 20m, Regular Waves

The results show that in wave heights of 12 m or above the number of cases where the vessel was not able to exit the evacuation zone increased, as indicated by the Failed Clearances series in Figure 5-12. For 50% of the launches performed in a 16 m wave height the vessel was unable to exit the evacuation zone. This outcome indicates a limit of the lifeboat in this high sea state. The outcomes again showed an increase of occurrences with increasing wave height.

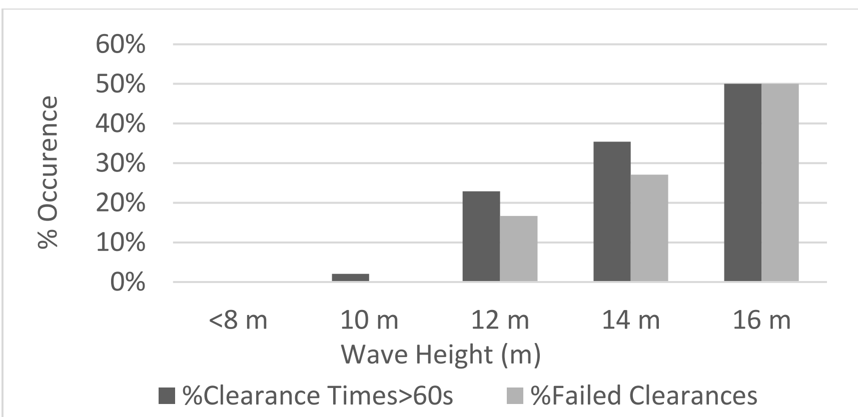

Figure 5-12: Clearance times Greater Than 60 s and Failed Clearances, Regular Waves

Investigation of the trajectories highlights that the maximum setback in high sea states can be a result of continued progressive setback. Figures 5-13 to 5-15 present samples of the XY trajectory for a vessel launch in the three highest wave heights. For each plot, the launch position on the wave is the same for each wave height and is near the trough of the wave. In the 12 m wave, the vessel was setback initially and was able to progress forward after 2 wave encounters. In a 14 m wave, the initial and progressive setback set the vessel back further and the vessel was still able to start moving forward after two wave encounters. For the 16 m wave, the wave and wind forces continued to push the vessel backwards, and the lifeboat was unable to move forward. This outcome indicates a limit has been reached, and there is not enough propulsion force to overcome the wave and wind forces. As noted, there were cases in the data sets for both 12 and 14 m waves where the vessel was not able to exit the evacuation zone, indicating that the combination of initial setback and continuous wave and wind forces resulted in a limit being reached in these sea states. These cases relate to the launches with high setback shown in Figure 5-9, which occurred when the vessel was launched near the trough of the wave (0 to 30 degrees).

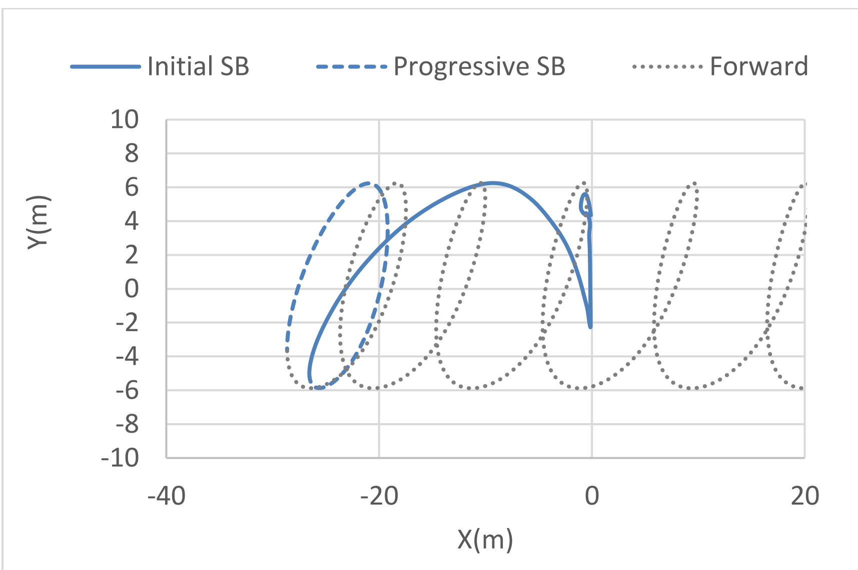

Figure 5-13: Vessel Trajectory H_w = 12 m, Regular Waves

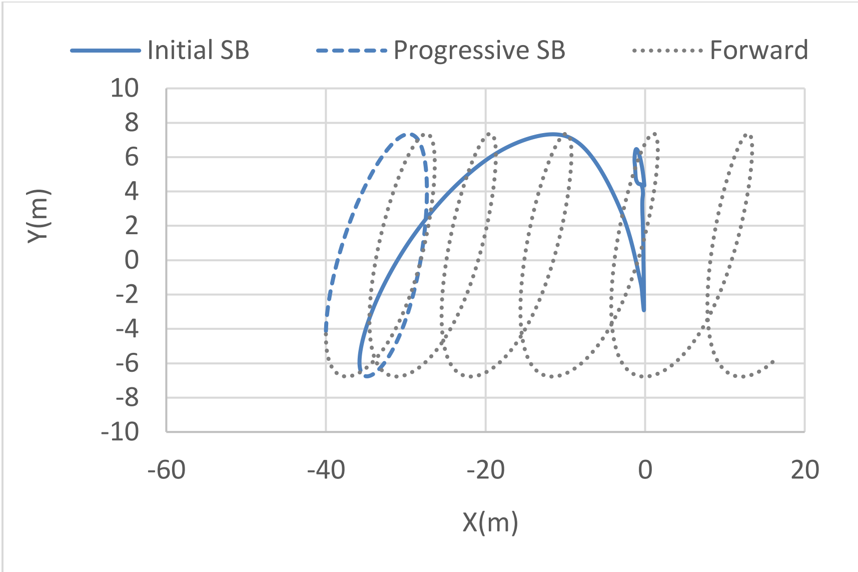

Figure 5-14: Vessel Trajectory H_w = 14 m, Regular Waves

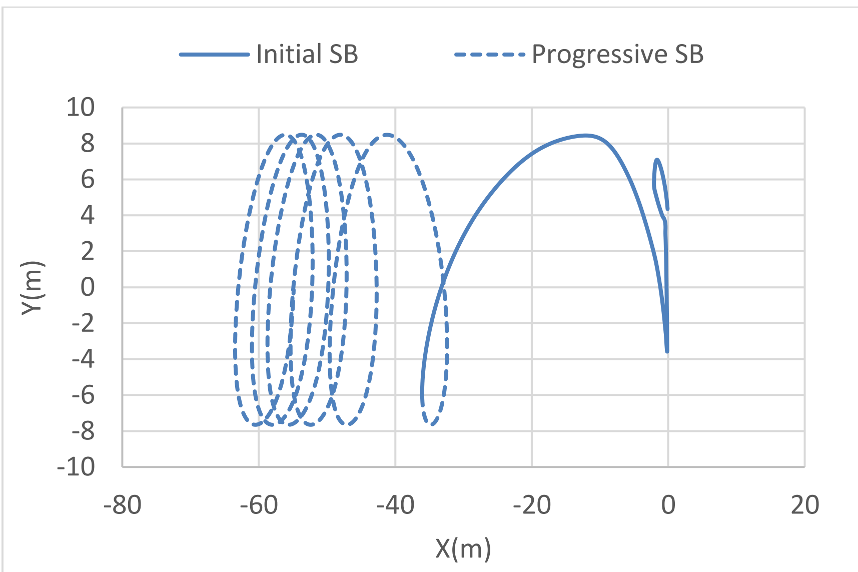

Figure 5-15: Vessel Trajectory H_w = 16 m, Regular Waves

5.7.3 Results: Investigation 2 – Study of Lifeboat Performance in Irregular 100 YR Seas

For irregular seas, setback is again analysed for the sets of launches for each wave height. The percentage of occurrences with clearance time greater than 60 s and failed clearances is also discussed.

A summary of the setback measures (Avg. SB, Med., SD, 90th PER.) is provided for each set of the 144 launches performed for each wave height. Table 5-6 summarizes the setback measures of the lifeboat in the irregular seas tested, with H_s from 6m to 12 m. The average measured setback for each set of launches increases with increasing sea sate. The 90th percentile is again provided to indicate the higher measures in the data set.

The 90th percentile indicates there were occurrences with setback above 20 m for an 8 m wave height, with setback values above 37 m and 50 m in 10 m and 12 m waves, respectively. The standard deviation of the data increased with wave height indicating higher variability in the measured setback as wave height increases. The median of the measured setback for each sea state remained low and below the mean, with a skew towards lower values. This outcome indicates that there were still a higher number of low setback values for each set of launches, similar to the tests performed in regular seas.

Table 5-6: Setback Summary - Irregular Seas

Hs (m)	**6**	**8**	**10**	**12**
Avg. SB (m)	6.36	8.94	12.99	17.16
Med.	4.60	5.80	5.84	7.07
SD	5.31	8.40	15.35	21.49
90th PER	6.23	21.85	37.01	51.20

Figure 5-16 provides a breakdown showing the percentage of occurrences for measured setback. Ranges of setback values are grouped to summarize the data. The figure indicates the over 50% of launches resulted in less than 10 m setback for each of the wave heights tested. Impact with the launch structure is unlikely in these cases. The percentage of launches with setback less than 10 m decreased from 78% in a 6 m wave height to 59% in a 12 m wave height. Over 74% of all test cases resulted in less than 20 m setback. Above 20 m, contact with the launch platform is more likely, as discussed in the performance measures. The percentage of launches with greater than 20 m setback was 16% in a 10 m wave height and 25% in a 12 m wave height. In 10 m waves, setback greater than 40 m occurred in 8% of test cases, increasing to 16% in a 12 m sea. This result indicates high setback values are possible in these extreme seas.

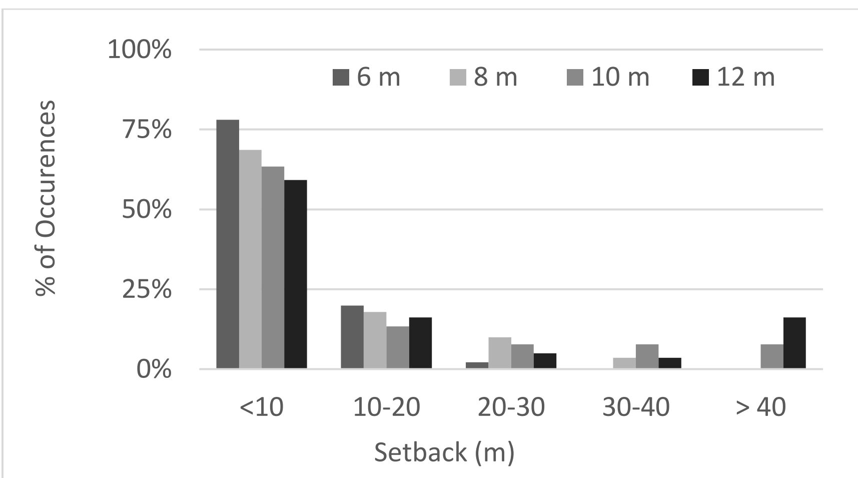

Figure 5-16: Setback Occurrences, Irregular Waves

Figure 5-17 shows the breakdown of the times to reach the target 20 m distance for a clearance. In most cases, the vessel was able to reach the target distance in less than 60 s. The results also indicate that in 8, 10 and 12 m wave heights there were several cases where the vessel was unable to reach the target distance of 20 m required for clearance, as indicated by the Fail series in Figure

5-17. In an 8 m significant wave height, in 13% of the simulations resulted in a failed clearance. 35% of cases performed in a 10 m sea resulted in a failed clearance, increasing to 41% in a 12 m wave.

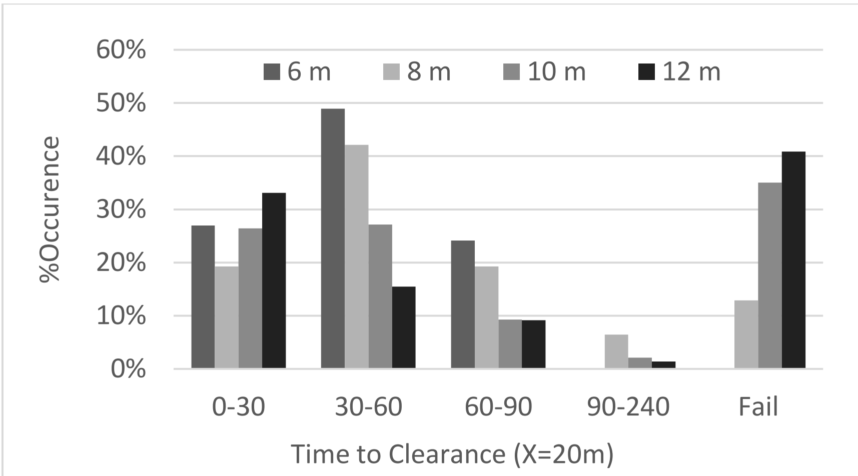

Figure 5-17: Time to Clearance, Irregular Waves

Figure 5-18 shows the sample trajectory of the lifeboat in a 10 m H_s where initial launch position close to a trough results in high initial setback. The lifeboat was initially setback over 20 m, and experienced progressive setback for 2 wave encounters resulting in a further setback of $\approx$ 8 m. The vessel was then able to progress forward. An additional 5 wave encounters occurred before the vessel could return to the launch position. For this case it took approximately 56 seconds for the vessel to progress from the maximum setback point to the original launch position. Figure 5-19 shows the vessel trajectory in a 12 m H_s and a launch near the trough of the wave. The vessel experienced initial setback of approximately 25 m and additional wave encounters set the vessel back further to close to 50 m. The vessel was not able to start forward progress. This result indicates a limit has been reached. These examples are provided to show a case where high initial setback occurred due to location of launch on the wave and was then not able to progress forward and another where the vessel could not overcome the environmental forces. As noted, cases were observed in both 10 m and 12 m wave heights where the vessel was unable to exit the escape zone due to progressive setback.

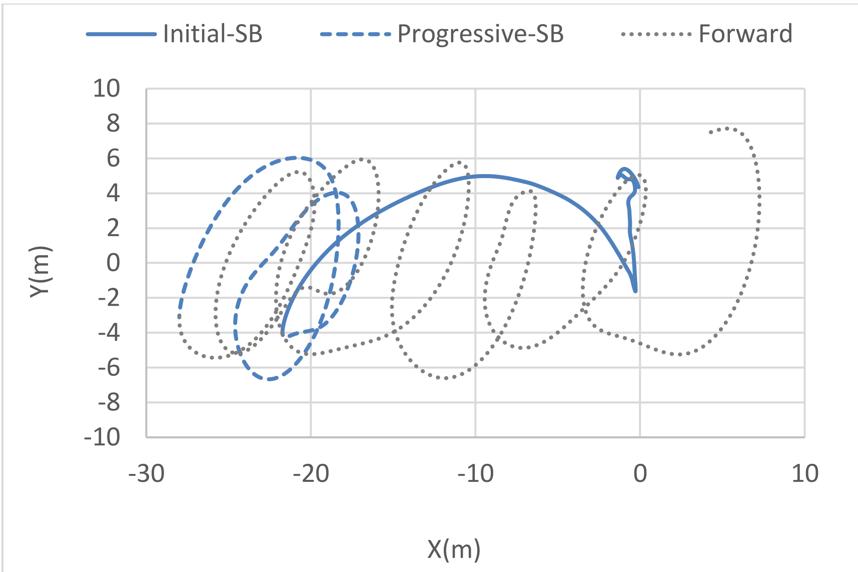

Figure 5-18: Vessel Trajectory, H_s = 10 m, Irregular Waves

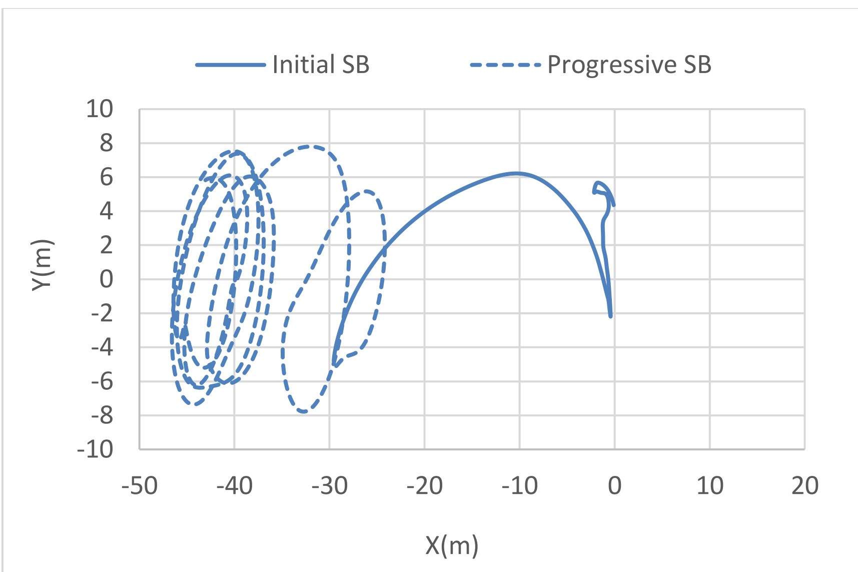

Figure 5-19: Vessel Trajectory, H_s =12 m, Irregular Waves

These outcomes show there is a higher likelihood of encountering a hazard if sea states are higher than 10 m. The combination of high setback and progressive setback can result in possible impact of the vessel with the launch structure or the inability to exit to a safe area. In wave heights of 8 m or less, the setback was reduced but not eliminated.

The results also indicated that most of the launches resulted in low setback even in higher sea states. For all the wave heights tested the median of the setback measures is less than 7 m and most of the launches resulted in setback less than 10 m. For the highest sea state tested (H_s = 12 m), the time to evacuate was less than 60 s for 48% of the launches. This percentage was higher for lower sea states. This result shows that successful launches can occur in the highest waves tested if the vessel avoids launching on a wave position that results in high initial setback.

5.7.4 Results: Investigation 3 – Study of Human Performance on Evacuation Performance in Irregular 100 YR Seas

This section discusses the impact of 1) a delay in throttle, 2) a delay in hook release and 3) cases where the throttle is applied prior to hook release. For each of these cases, 144 launches were performed for each wave height tested. A summary of the setback measures for each set of launches performed for each wave height is provided. The percentage of occurrences with greater than 20 m setback, clearance times greater than 60s, and failed clearances are discussed. Comparisons are made to the data from the second investigation where there was no delay in throttle or time to hook release.

5.7.4.1 Delay in Throttle

Table 5-7 presents the summary of the setback measures for sets of launches performed with a 2 second delay (TT2) and a 4 second delay (TT4) in time to applying throttle after hook release. Figure 5-20 shows a comparison of average setback measures for each sea state, with comparison made to no throttle delay (TT0).

The results show that there was an increase in average setback of approximately 17% over all wave heights when time to throttle is delayed by 2 s, compared to the set of launches when there was no throttle delay. There was an average setback increase of 35% when the time to throttle was delayed by 4 s. Similar to the previous investigations, the median of the setback measures was below the average setback for each wave height, indicating a high number of low setback cases for each set of launches. The increase in the 90th percentile of measured setback for each of the wave heights shows the increased throttle delays resulted in higher setback measures.

Table 5-7: Setback Summary - Delayed Throttle, Irregular Waves

Average Setback (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	6.36	8.94	12.99	17.16
TT2	7.12	10.75	15.37	19.91
TT4	7.95	15.77	15.94	20.75
Median (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	4.56	5.73	5.73	7.06
TT2	4.59	5.91	5.62	7.39
TT4	4.72	6.03	6.30	8.88
90th Percentile (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	14.74	21.86	37.01	51.20
TT2	17.18	25.61	38.64	54.91
TT4	19.31	27.24	39.77	54.17

Figure 5-21 shows the percentage of occurrences of setbacks greater than 20 m increased an average of 4% for TT2 and 7% for TT4, compared to no throttle delay. The percentages increased to over 20% in an 8 m wave and to over 30% in a 10 or 12 m wave when throttle was delayed 4s. As shown in Figure 5-22, with a delay in throttle of 2 s, the measured occurrences with clearance times greater than 60 s increased by 11% in 10 m waves and by 12% in 12 m waves. Related to this outcome, the increased throttle delays resulted in more occurrences of the vessel being unable to leave the clearance zone, as shown by the Failed Clearances in Figure 5-23. In a 12 m wave, delayed throttle by 2 or 4 s resulted in the vessel not being able to exit the evacuation zone in over 40% of the launches.

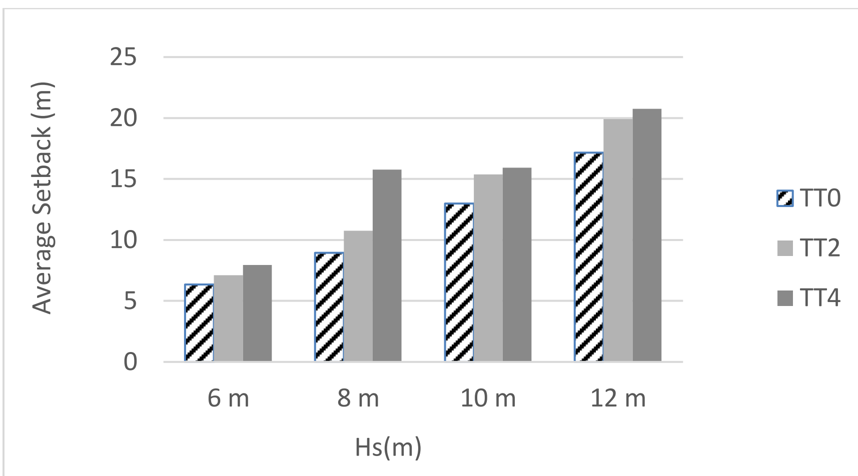

Figure 5-20: Average setback, Delayed throttle, Irregular Waves

Relating this to operational objectives, the results suggest the target is to apply throttle as quicky as possible following hook release. It is unrealistic to assume a coxswain will be able to release the hook with no delay though applying the throttle in less than 2 s is achievable, as observed in training courses. Training should provide sufficient practice for trainees to learn to operate the hook release as quickly as possible. Training scenarios can also incorporate plausible outcomes identified in this study. If during training coxswains are observed to take a long time to apply throttle then there is a possibility of a collision or inability to evacuate the launch area. These outcomes can be built into simulator scenarios to provide feedback.

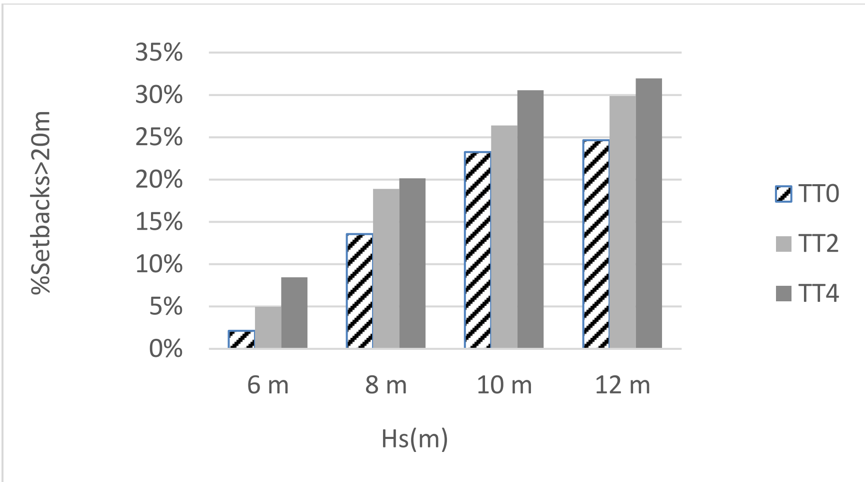

Figure 5-21: Setback Occurrences >20 m, Throttle Delays, Irregular Waves

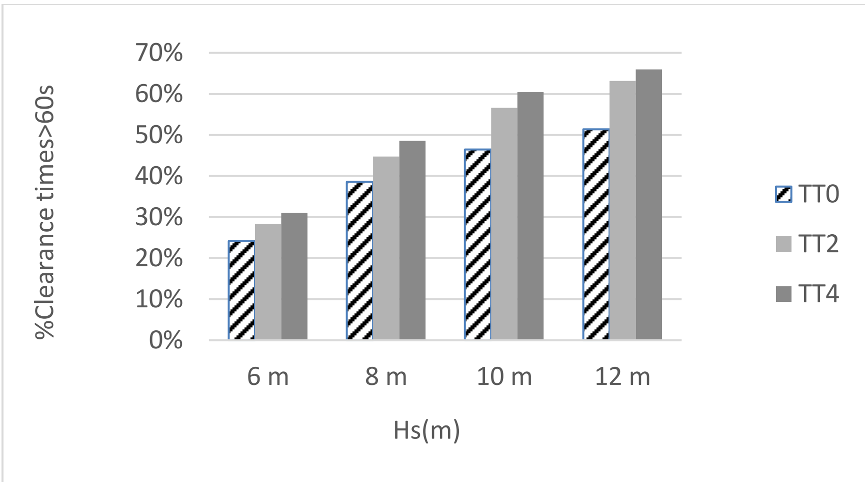

Figure 5-22: Clearance Times Greater Than 60 s, Delayed throttle, Irregular Waves

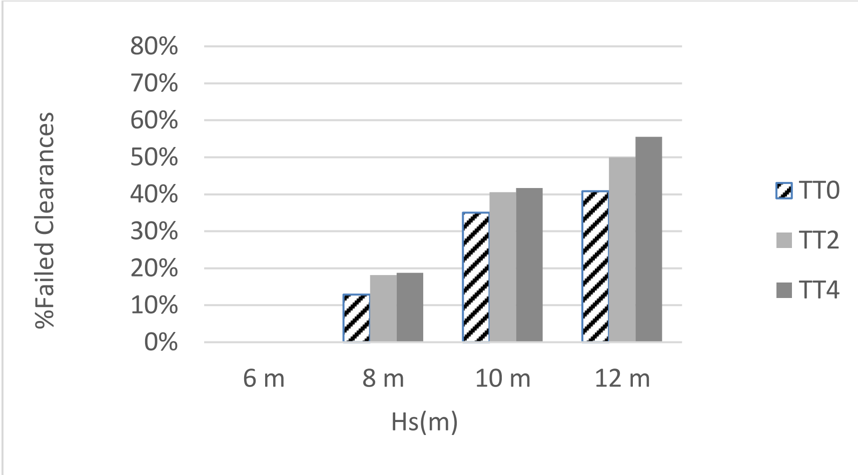

Figure 5-23: Failed Clearances, Delayed Throttle, Irregular Waves

5.7.4.2 Delay in Hook Release

Table 5-7 shows the summary of the setback measures for each of the sets of launches performed with a 2 second delay in time of hook release (TR2) and a 4 second delay in releasing the hook (TR4). Comparisons are again made to an instant time to hook release and throttle (TT0-TR0).

Table 5-7 and Figures 5-24 and 5-25 indicate an initial reduction in average setback and occurrences of setback greater than 20 m when the hook release delay was 2 s (TR2), and then an increase in these values when the throttle delay was 4 s (TR4). The occurrence of clearance times greater than 60 s also changed, with a reduction the percentage of occurrences for TR2 and an increase in TR4. A considerable increase in occurrence of clearance times greater than 60s and failed evacuations occurred in a 12 m wave height, with 74% of the cases resulting in failed clearances. These outcomes are shown in Figures 5-26 and 5-27.

Table 5-8: Setback Summary - Delayed Hook Release, Irregular Waves

Average Setback (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	6.36	8.94	12.99	17.16
TR2	5.53	6.39	8.02	8.02
TR4	9.49	11.05	17.57	17.57
Median (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	4.56	5.73	5.73	7.06
TR2	4.84	5.54	6.94	7.06
TR4	5.58	7.07	9.79	20.1
90^{th} Percentile (m)				
	6 m	**8 m**	**10 m**	**12 m**
TT0	14.7	21.9	37.0	51.2
TR2	8.59	10.3	11.4	27.3
TR4	23.3	25.4	43.5	57.0

While these outcomes seem counterintuitive, the behaviour can be explained by considering how the delay in hook release affects the position of release on the wave. Given the wave shape and slope, the vessel is likely to land on the upslope of the of the dominant wave (0 to 90 degrees), as indicated in previous research (Simões Ré et al., 2002) and Investigation 1 of this paper. The delay in hook release keeps the lifeboat in position, and the fall wires do not extend enough for the vessel to drift backwards significantly. As a result, for a small delay the vessel could release on the top or the downslope of the wave where wave forces were more favourable to reduce setback. Too long a delay resulted in both the vessel starting to drift backwards and release occurring closer to a trough of the wave where wave forces could induce more setback. In effect, delay in hook release provided a short window of benefit and the "wave shadowing" was reduced for a short time.

This window is expected to be highly dependent on wave shape (height, period, and steepness). The results are specific to the wave shapes used in this study. Further investigation is required to determine if similar outcomes are seen in different wave shapes, which is outside of the scope of this paper.

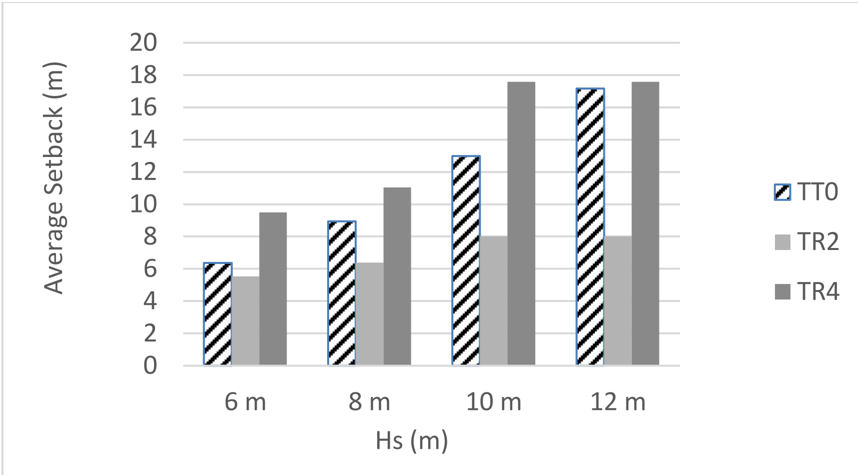

Figure 5-24: Average Setback - Delayed Hook Release, Irregular Seas

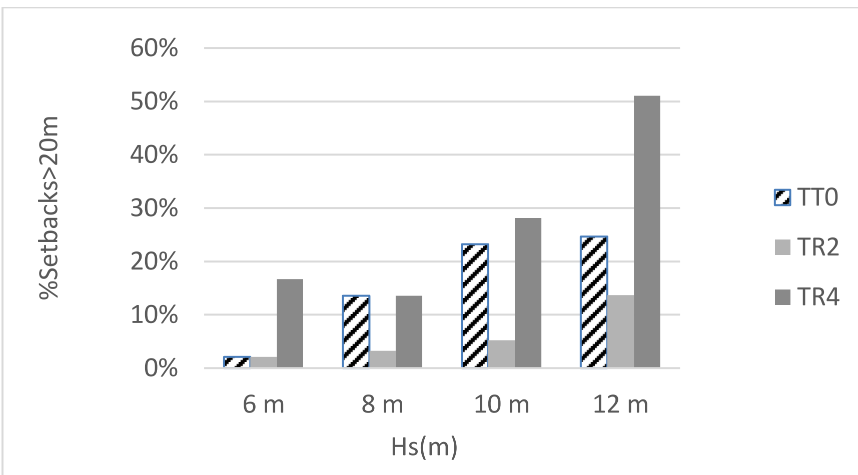

Figure 5-25: Setback Occurrences > 20 m, Hook Release Delays, Irregular Waves

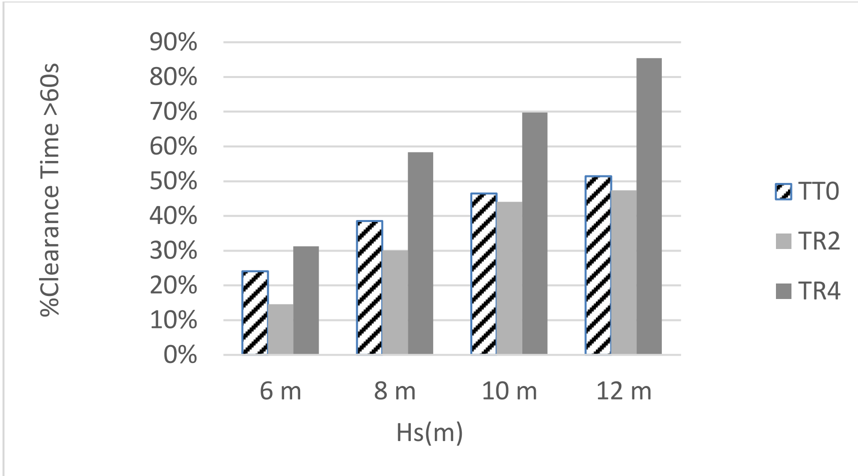

Figure 5-26: Clearance Times Greater Than 60 s, Hook Release Delays, Irregular Waves

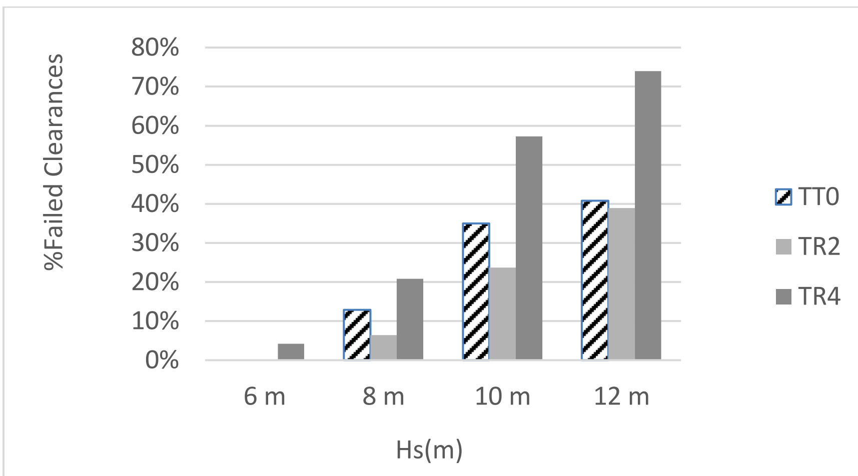

Figure 5-27: Failed Clearances, Hook Release Delays, Irregular Waves

5.6.4.3 Throttle Before Hook Release

The following cases summarize the measures from launches where the throttle is applied prior to the release of the hooks. As noted in the methodology, the times presented are relative to the time the vessel was able to be released (t = 0) and the timing was varied for the time to throttle (TT) and time to release (TR) of the hooks. In these cases, the lifeboat remained tethered and propulsion force was applied before the hook was released. Comparisons are made with the case where throttle and hook release were applied immediately (TT0-TR0).

As indicated in Figures 5-28 to 5-30, there was an improvement in most outcomes when throttle was applied early and prior to the hook release. The greatest improvement in all performance measures occurred when the throttle was applied 1 s after the vessel could be released and the hook was released 1 s later. This series is noted as TT1-TR2. As indicated in Figure 5-28, Average setback was reduced by approximately 50% and there was a reduction of setback occurrences greater than 20 m and clearance times greater than 60 s. Similar outcomes were observed for cases TT1-TR3 and TT2-TR3. For these cases, the percentage of setback occurrences greater than 20 m was reduced to 10% or less in all wave heights tested, as indicated in Figure 5-29. These results indicate that application of throttle before hook release creates enough initial propulsion to improve launch performance based on the measures discussed.

The results indicate that the timing of throttle before hook release must still be performed quickly, and hook release cannot be delayed too long. This is shown in the case where the time to throttle was performed 2 seconds after the vessel is able to be launched and time to hook release was performed two seconds following (TT2-TT4). Figure 5-28 indicates a small increase in average setback in high sea states for this case. Figure 5-29 shows there were increased occurrences of setback greater than 20 m in a 12 m wave height. Figures 5-30 and 5-31 show there was an increase of occurrence of clearance times greater than 60s and failed clearances in 10 and 12 m wave heights. This result again suggests that high throttle and hook release delays can result in reduced performance.

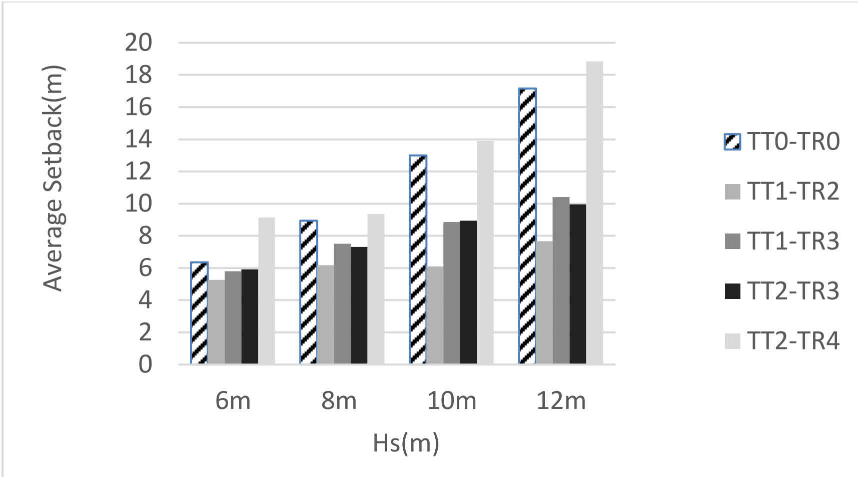

Figure 5-28: Average Setback, Throttle Before Hook Release, Irregular Waves

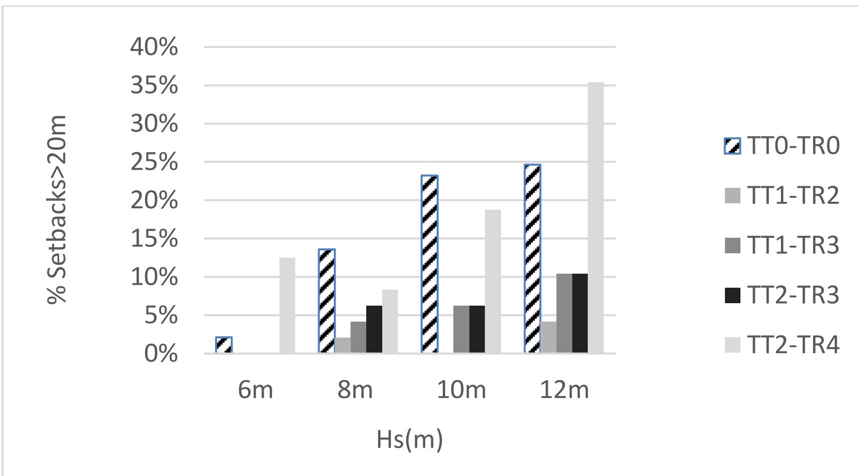

Figure 5-29: Setback Occurrences Greater Than 20 m, Throttle Before Hook Release, Irregular Waves

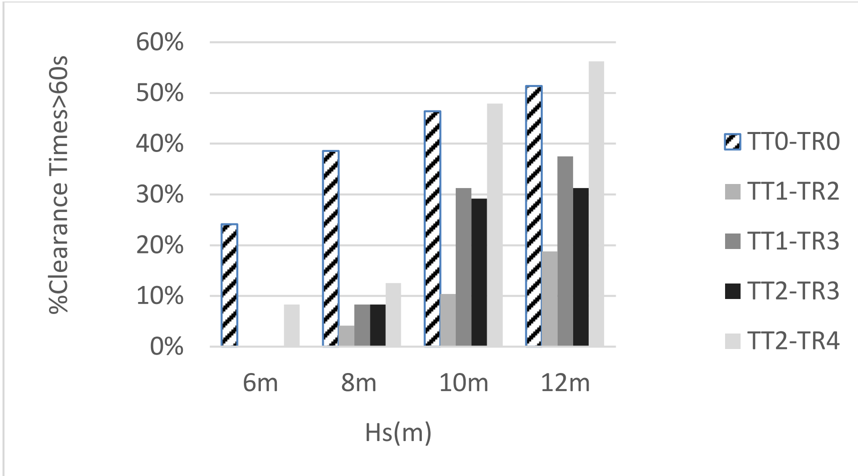

Figure 5-30: Clearance Times Greater Than 60 s, Throttle Before Hook Release, Irregular Waves

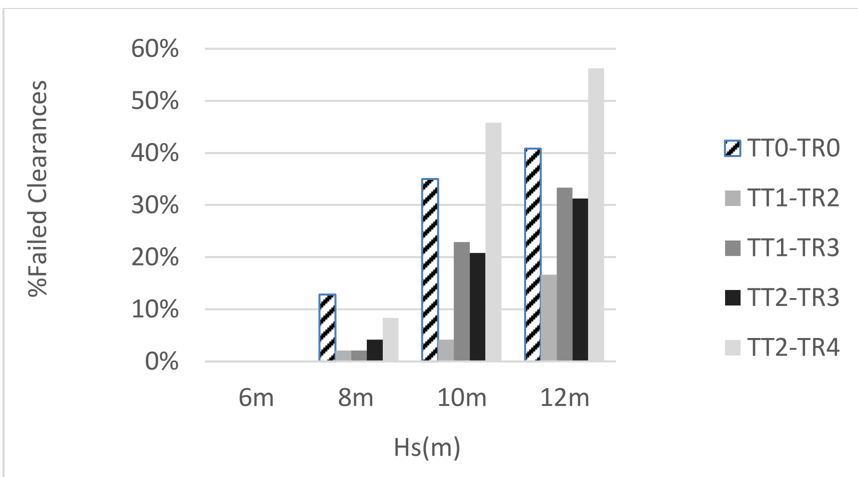

Figure 5-31: Failed Clearances, Throttle Before Hook Release, Irregular Waves

These cases show a procedure that can be performed to give the vessel initial propulsion prior to being released. The results show better launch performance when throttle is applied before the vessel is released. The results indicate a need for these actions to be performed in a timely manner.

5.8 Conclusions

The goals of the research were to use simulation to extend the knowledge of lifeboat performance in high sea states and to evaluate how human performance can affect outcomes.

The results show a strong relationship between the performance measures and wave conditions. Specifically, both setback and time to exit the launch area were both dependent on wave height and the wave phase angle at the launch point. These results are the same as found previously in experimental work up to about 10 m (Simões Ré et al., 2002), but have extended the wave heights up to 16 m in the simulation environment.

The position on the incoming wave at which the lifeboat was launched (i.e. the wave phase angle) was found to be particularly important. When launched at or very near the crest, lifeboats avoided large setback and were able to make way relatively quickly to clear the launch zone. Conversely, when launched near the trough or the upslope of the incoming wave, the lifeboats were setback immediately by the wave. The magnitude of the setback was dependent on wave height in addition to wave phase angle. Consequently, the initial setback experienced by the lifeboat during its first wave encounter made clearing the launch more difficult for two reasons: first, the lifeboat had to overcome the momentum associated with the setback action; second, its effective starting point was behind the nominal launch target (directly below the davits) by a distance equal to the setback (or progressive setback).

In practical terms, one consequence of setback is that the lifeboat can collide with the launch platform if there is insufficient clearance between the launch target and the platform. While the environmental conditions at the time an evacuation are outside the control of evacuees, the timing of the launch is not. Timing a launch requires that the coxswain can see or otherwise sense the approaching waves and has enough familiarity with the lifeboat controls (e.g. lowering, releasing the hooks, throttling) to perform the launch operation within the narrow time window required for a successful launch on a crest. For a typical large wave, the window for a crest launch is only about 5 to 7 seconds.

The studies of time of throttle delay and time of hook release timing provide insights on how human actions can affect launch performances. Interpreting the outcomes of the third investigation, we see a general trend that a quicker performance of actions results in better performance outcomes. This result has implications for training. Delays in actions can be due to inability to recognize launch cues (i.e. the hydrostatic indicator movement), improper movement of the hook release handle, or performing actions out of order. These timings can be further delayed if there are faults in the system that require additional time to remedy, such as performing a hydrostatic override procedure. The results of this research suggest training goals should target the quick performance of these actions and training to provide practice to improve these timings.

The research also indicates that new operational procedures can improve launch performance. Applying the throttle prior to hook release can reduce setback and escape times significantly, as long as these actions are performed quickly. This procedure was suggested by operators with marine experience. Operational procedures that result in improved performance can be embedded into curriculum to train coxswains.

Considerations must be given to the specificity of the wave environments and launch configuration when interpreting the research outcomes in this paper. As indicated in previous research (Simões

Ré & Veitch, 2004), the wave steepness can have a considerable effect on the amount of measured setback, although wave steepness was not varied in the current work. The simulations focused only on escape from the platform in a head sea where wave direction is directly against the desired escape path of the launch vessel. This scenario was considered as a worst case. Scenarios with oblique waves and winds would present additional operational challenges (e.g. maintaining a desired heading).

5.9 Acknowledgements

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

5.10 References

Billard, R., Smith, J.J.E. (2018) Using simulation to assess performance in emergency lifeboat launches. Proceedings, Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 19179.

Billard, R., Smith, J., Veitch B., (2019) Assessing lifeboat coxswain training Alternatives using a simulator. The Journal of Navigation Published online by Cambridge University Press: 19 September 2019.

Billard, R., Musharraf, M., Smith, J., Veitch B., (2020) Using Bayesian methods and simulator data to model lifeboat coxswain performance. WMU Journal of Maritime Affairs. Published May 2020. https://doi.org/10.1007/s13437-020-00204-0

C-Core (2015) Metocean climate study offshore Newfoundland and Labrador, Nalcor Energy Report. http://exploration.nalcorenergy.com/wp-content/uploads/2016/09/Nalcor-Metocean-Study-Final-Report-Volume-2-27-May-2015.pdf

Gabrielsen, O., Helland, B, Seines, P.O., Helland, L. R., (2011) Study of davit-launched lifeboats during lowering, water entry, release and sail away. Norweigian Oil and Gas Association website publication http://www.olf.no.

Groth K., Smith, C., Swiler, L. (2014) A Bayesian method for using simulator data to enhance human error probabilities assigned by existing HRA methods. Reliability and System Safety 128 (2014), 32-40

International Maritime Organization., & International Conference on Training and Certification of Seafarers (2010). S*TCW including 2010 Manila Amendments*, 2017 Edition.

Magee, L.E., Smith, J.J.E., Billard, R., & Patterson, A. (2016) *Simulator training for lifeboat maneuvers.* Proceedings of the Inteservice/Industry Training, Simulation, and Education Conference (I/ITSEC). Paper number 16030.

McClernon, C. K., McCauley, M. E., O’Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(3), 207-218.

Sellberg, C. (2017). Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis. WMU Journal of Maritime Affairs, 16(2), 247-263.

Stefanidis, D., Korndorffer, J.R., Markley, S., Sierra, R., Heniford, B.T., & Scott, D.J. (2007). Closing the gap in operative performance between novices and experts: does harder mean better for laparoscopic simulator training? Journal of the American College of Surgeons, 205(2), 307-313.

Robson, J. K., (2007) *Overview of TEMPSC performance Standards.* Health and Safety Executive Research Report RR599. Norwich.

Simões Ré, António & Veitch, B., Pelly, D. (2002) *Systematic investigation of lifeboat evacuation performance.* Transactions, Society of Naval Architects and Marine Engineers. 110:341-360

Simões Ré, António & Veitch, B. (2004) *Evacuation performance of davit launched lifeboats.* Proceedings, Offshore Mechanics and Offshore Engineering, June 20-25, 2004.

Simões Ré, António & Veitch, B. (2008) *A comparison of three types of evacuation system.* Transactions - The Society of Naval Architects and Marine Engineers, 115, pp. 119-139, 2008

6.0 CHAPTER 6: CONCLUSIONS

This research presented a set of studies to use data collected from experimental and numerical studies performed with simulators to investigate the problem space of learning and performance of lifeboat coxswains. The work combines experimental data sets, modeling techniques, subject matter expertise and numerical simulations to expand the knowledge of human performance and equipment limitations. Simulation is used as a safe means to collect data and to investigate scenarios that are not possible to create in real life due to risk. The thesis demonstrates how data collected from simulator studies is used to investigate research questions that were previously unable to be studied. The research focused on the performance of lifeboat coxswains and launch equipment, though the methodologies used can be applied to other research areas where data is scarce.

To summarize the outcomes of the thesis, the research questions presented in section 1.4 (Table 1-1) are revisited and key results are discussed.

The research from Chapter 2 and 3 indicates the following related to skills acquisition, learning, and transfer of skills in lifeboat coxswain training:

- There is a benefit to receiving hands-on training with a simulator or live boat, compared to users who completed only CBT training. This outcome was determined by assessing coxswain performance on tasks in a plausible emergency event. Increased performance was seen in both procedural tasks during launch, and psychomotor tasks of maneuvering the lifeboat when trainees practiced with real equipment. An incremental benefit was observed when trainees practiced in scenarios that increased in difficulty and included representative weather.

- For all types of training studied (drills in calm water, CBT, simulator) there was an overall low success rate on performance of launch and maneuvering tasks in emergency scenarios involving adverse weather. Quarterly training performed in calm water or in less difficult scenarios did not provide enough opportunity to achieve competence to complete tasks in scenarios with moderate waves and hazards.
- The research indicates that initial lifeboat training practice does not provide enough practice for trainees to build mental models needed to perform procedural tasks required to launch a lifeboat, or to master slow-speed maneuvering tasks. Trainee performance in new scenarios involving multiple tasks types, as would be required in a real lifeboat evacuation, was very low.
- Tasks involving the launching of the lifeboat and performing slow-speed maneuvers (e.g. stopping next to a person in the water) require more practice to master than other skills (e.g. navigating by compass).

These research outcomes indicate there is a benefit of performing training on real lifeboat equipment and practicing in scenarios that are representative of real emergencies. The research also indicates that initial training does not provide practice required to acquire skills needed to perform in an emergency. More frequent training events and shortened training intervals are expected to improve skill acquisition and limit skill fade. Marine education and training instructors can utilize overtraining and practice in representative environments to improve trainee performance. As shown in other research, practicing in realistic scenarios and with increased difficulty can increase preparedness for emergencies (Klein, 2008, Mclernon, 2011).

Chapters 3 and 4 illustrate how probabilistic methods were used to investigate the research questions related to human performance. The research objective was to predict and assess trainee performance using available data and expert knowledge. Models were developed to study learning and predict competence as trainees practiced tasks in simulator scenarios. Bayesian inference allowed for the creation of Human Performance Probability CDFs that were used to study initial learning in new lifeboat operators and discern tasks that require more practice to reach competence. Bayesian Networks were used to model the competence of slow-speed maneuvering and to diagnose causal relationships between practice on similar tasks types and training background. These methods used a small data set collected from a simulator study to form models and study learning. The research presented how the models could be utilized with a limited data set. Chapter 3 indicates the models created with BI were strengthened as new data was used and the models can be further improved as new data is available. The research in chapter 4 indicates the predictive accuracy of BN models can be increased through expert knowledge and demonstrates how available data and domain expertise can be combined to create models to study learning. The methodologies provided an effective way to study the problem area of learning in lifeboat and lifeboat operators using a combination of available data and expert input.

Chapter 5 investigated lifeboat and human performance in high sea states using numerical simulations. Previous experiments studied lifeboat evacuations in wave heights up to 10 m wave heights and used regular waves (Simões Ré et al., 2002). In the thesis, simulations were used to create data sets of lifeboat launches for regular waves up to 16 m wave height and irregular waves up to 12 m significant wave height. The results show a strong relationship between the performance measures (setback and time to clear the launch area) and the wave conditions. The study identified performance limitations in the lifeboat when significant wave heights were above 8 m and the lifeboat is launched near the trough of a wave. The amount of setback and the time to exit the launch area increase as wave heights increase. The numerical simulations also allowed for an investigation of how the timing of human actions impact launch success in high sea states. The results indicate that a quicker application of throttle and hook release result in better performance outcomes. The numerical studies allowed for the investigation of alternate operating procedures, such as applying the throttle before the hook is released. This procedure was suggested by subject matter experts with marine experience and subsequently investigated using simulations. The results indicate an increase in launch performance if actions are performed quickly. The study provides evidence of how simulation can be used to test procedures that would not be possible to test safely using real equipment.

The research provides outcomes that are relevant to training providers and researchers. The studies indicate that regular and frequent training is needed to prepare operators for plausible emergency events. The research also indicates that a high level of human performance is needed in adverse sea states to achieve a successful lifeboat launch. The research highlights how simulation can be applied to study performance and extend models to training applications where new data sets are being created. The outcomes and methodologies can improve training programs through improved knowledge of how practice impacts learning, identifying factors that affect performance, and the creation of adaptive training programs.

6.1 Technical Limitations and Uncertainties

Several technical challenges and uncertainties arose from the research. The following describes some of the uncertainties and provides suggestions for future studies.

- The goal of the experimental study was to study learning in trainees with no previous experience in launching or maneuvering a lifeboat. The participants selected for the study were naive and were not familiar with the lifeboat or the launch procedure. The study did not include personnel who had received regular training and the models and outcomes do not represent the performance of experienced lifeboat coxswains. The performance of lifeboat operators with a marine background or with several years of practice is expected to be higher, but is also unknown as these individuals were not included in the study. Future studies could measure the performance of experienced coxswains and make comparisons with the outcomes of the research.
- The experiment used initial training to bring participants to a baseline level of competence but was not able to consider all individual differences between trainees. Studies have indicated that individuals learn and acquire skills differently (Joe and Boring, 2014, Arthur et al. 1998) and individual differences should be considered when modeling learning in virtual environments (Musharraf et al. 2017). Participants had varying levels of marine experience, familiarity with technology, and training backgrounds. While this variability could not be controlled, these differences are expected to impact the learning rate and performance of trainees in the study. To mitigate this effect, the experiment was designed to sort trainees into groups based on knowledge of their background and initial training

performance. Future work could investigate training background and learning rate as performance shaping factors.

- The virtual environments closely matched the lifeboat equipment and visualizations of a marine environment. The simulator used in the study was certified for marine training based on how accurately the simulator matched the vessel and marine environment. However, a simulation is not an exact representation of the real world, and there are differences in the visual, audible, and motion cues. The simulator was tested to ensure suitable cueing was available to achieve task objectives. The experiment also used a controlled environment and common scenarios to make consistent and repeatable measures. Real emergencies would include additional stressors, and training conditions could be highly variable. Future studies could examine the impact of adding stressors (e.g. backstory, motion cues) or adding variability in the weather or tasks to make the scenario less repeatable. Similarly, the numerical models used in the virtual wave tank are not exact representations of the real world. The differences between the numerical models, experimental models and real world could not be quantified as performance data of real lifeboats in high waves does not exist. The simulator measures were compared to the outcomes of experimental testing to demonstrate the numerical models provided representative characteristic behavior needed for the study.
- The sample size used in the study was selected based on availability of test participants and the logistics of conducting the experiment. While some of the research considered group performance, other parts of the research considered individual performance on tasks, or sub-groups of the larger group (those trained by drills, CBT, or simulation). The sample size for these sub-groups was small. A larger sample size would likely have resulted in

less uncertainty and would increase the predictive accuracy of the models that were developed.

- Testing performed with the numerical simulations (virtual wave tank) considered specific wave shapes and directions. Considerations must be given to the specificity of the wave environments and launch configuration used in the research. As indicated in previous research (Simões Ré et al., 2002), the wave steepness can have a considerable effect on the amount of measured setback, and consequently would impact the performance measures. The studies also focused only on escape from the platform in a head sea where wave direction is directly against the desired escape path of the launch vessel.

6.2 Future Work and Recommendations

The following section describes some guidance on future and related work that can be performed with simulations or using the methodologies presented in this thesis.

- Ongoing training research – the research indicates that the type and amount of training received by the experimental participants did not provide enough practice to achieve competence on some skill types. The study also indicated that certain tasks, such as stopping next to an object in moderate sea states, are difficult to perform and required more practice than provided in the study. The thesis did not investigate the frequency or type of training that resulted in competence or provide enough data to indicate if additional practice would have resulted in a higher performance. As the goal of the training is to prepare coxswains for plausible emergencies, further research on the type and amount of training that results in competence in emergency scenarios is suggested. Investigating the impacts of overtraining, training in real scenarios, increased training frequency, and alternative training techniques are areas of possible future research. Information learned from simulation-based assessments can also be used to improve training programs. Performance data can be used to determine training targets and to evaluate an individual's performance with comparisons made to historic data sets.

- Continue to explore human and equipment limitations – the research provided insights on both trainee readiness for emergencies and the limits of launch equipment in extreme weather events. The study indicates that some tasks require a significant amount of training to master (slow-speed maneuvering) and the research did not evaluate if training could result in a high rate of success in adverse weather. The maneuverability of the lifeboat in waves must be considered as the vessel is difficult to control at slow speeds and in high wind and wave conditions. Studies of expert performance on tasks in high seas could evaluate the probability of errors made (slip) as weather increases and define human performance limitations. The research also indicated that human actions improve launch performance if actions are performed quickly. Training to improve timing of actions, to practice releasing on a favorable part of a wave, or to evaluate alternate procedures is expected to improve launch performance. Studies performed with expert personnel in high sea states could evaluate new procedures to increase the likelihood of a successful launch.
- Extend data sets to improve models – the research used a small data from an experimental study to form models of learning and to assess competence. Probabilistic models derived from large data sets are expected to have a higher predictive accuracy. The predictive accuracy and practicality of the models can be improved with additional or larger data sets. Additional data can be collected from experimental studies or training programs that use

simulations to collect and track data on task performance. Data sets collected for trainees with different training backgrounds or operational experience can be used to model and compare performance for novices and experts and to study the impact of different training alternatives.

- Extend numerical studies and consider different wave characteristics - scenarios with oblique waves and winds would present new operational challenges (i.e. maintaining a desired heading) and new measures that would again be affected by human factors, and would assess if the vessel being studied is seaworthy. The investigation of varying wave shape, wind speeds, and wave directions is recommended to assess the change in performance across a broader spectrum of weather and launch configurations.
- Incorporate models into adaptive learning applications – the study demonstrates how data collected on human performance can be used to gain insights on skill acquisition and development. Simulator-based assessments can be used to measure performance for trainees as they practice in scenarios that evaluate their ability to complete tasks in scenarios. Data sets can be created to model novice and expert performance and create inputs to ITS (Millán et al, 2011) which are used to tailor the training experience. The probabilistic BI and BN models can be integrated with machine learning algorithms to build adaptive training applications to customize training material to an individual's strengths and weaknesses based on evidence gathered in training. BN models can also be expanded to explore learning and the impact of different variables (i.e. time between training events, type of training received, background) on performance outcomes.
- Apply methodologies to new applications – the methodologies and approaches used in this thesis can be applied to other problem areas where data is scarce to gain insights on trainee

and equipment performance. Training simulators or numerical simulations can be used to perform assessments and collect data for situations that are normally prohibitive due to risk. Simulation based assessments can be used to extend studies to new operating conditions and provide measures of performance in scenarios that could not previously be tested. The approach and methodologies of assessing performance and examining learning of different skill types can be applied to other research in emergency response and has applications in other industries. The probabilistic models derived from BI and BN can be applied to new data sets and expanded to study learning in new training applications.

6.3 References

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance*, **11(1)**, 57-101.

Joe, C.J., Boring, R.L., (2014) Individual Differences in Human Reliability Analysis, Proceedings Probabilistic Safety Assessment and Management Conference, June 2014., USA.

Klein, G., (2008), Naturalistic decision making. Human Factors: The Journal of Human Factors and Ergonomic Society, 50(3), 456-460.

Millán, E., Loboda, T., Perez-de-la-Cruz, J.L. (2011). Bayesian networks for student model engineering. Computers and Education, 55, 1663-1683

Mashrura Musharraf, M., Smith, J. Khan. F., Veitch, B., Mackinnon, S. (2017). Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique, Safety Science .Volume 107, August 2018, 107, 216-223

McClernon, C. K., McCauley, M. E., O’Connor, P. E., & Warm, J. S. (2011). Stress training improves performance during a stressful flight. Human Factors: The Journal of the Human Factors and Ergonomics Society, 53(3), 207-218.

Simões Ré, António & Veitch, B., Pelly, D. (2002). Systematic investigation of lifeboat evacuation performance. Transactions, Society of Naval Architects and Marine Engineers. 110:341-360