University of
South-Eastern Norway

Simulated Marine Engineering:

Immersive Virtual Reality in Maritime Education and Training

Candidate name: Simen Hjellvik

University of South-Eastern Norway
Faculty of Technology, Natural Sciences and Maritime Sciences

MASTER THESIS

May 2019

Abstract

For decades simulators have been embedded in the formal education in the field of marine engineering, for training and assessment of competences and proficiencies. A new era of simulator technology, by the use of head mounted display virtual reality, is emerging to the field of maritime education and is currently unexplored in the context of marine engineer training. This is an experimental study with the latest technological increment which was conducted with a prototype immersive virtual reality simulator and a commercial 3D virtual reality desktop simulator. By means of this novel head mounted display virtual reality and the more familiar desktop option, the purpose of this study was to explore these technologies through the potential end user.

A classic between-groups experimental design was developed with a simulation exercise of starting a fuel oil separator for the treatment and tested with two marine engineering student groups and one group of professional engineer officers. The recruited sample frame was assigned either to the (*i*) 3D virtual reality desktop group (n=5), (*ii*) immersive virtual reality novice group (n=6) or the (*iii*) immersive virtual reality expert group (n=6). Instruments of declarative knowledge tests were constructed for measuring prior knowledge prerequisite to the study, and for measuring accuracy and accessibility of retaining knowledge acquired in the treatment. The results gained a significant difference in knowledge acquisition between the two technologies (P=0.005) and between the group competence levels (P=0.008). Instruments for measuring mental workload and the flow state were adopted from the original frameworks to describe the experience. No technology discrimination could be observed, though the group level experience measures indicated some difference and yielded subservient effect sizes and significance.

Acknowledgements

This study was conducted at the University of South-Eastern Norway, campus Vestfold, in collaboration with the “Innovating maritime training simulators using Virtual and Augmented Reality” research project (InnoTraining) of the “Training Assessment and Research Group” (TARG). The 4-year project, proprietary of the university, is a 13 Million NOK effort jointly funded by the government and industry.

This study was supported by the Norwegian Union of Marine Engineers, who supplied gratifications to the recruited participants. The union display great effort and engagement in maritime education and it´s development, and comprise of over 4100 active marine engineer officers.

Table of Contents

List of figures

List of tables

Definitions

2D desktop	Basic desktop simulator with process only interaction where the user interacts with the system processes through an allocentric view of a monitor.
3D Virtual Reality desktop	Non-immersive desktop simulator with first-person egocentric projection displayed on a monitor which the user interacts with through an allocentric view of the environment.
Allocentric	Interaction through a monitor or similar medium that allows the user to view the environment from the outside with object-to-object spatial processing and navigation.
Big View	An enhancement of the 2D desktop simulator setup with multiple monitors where the user interacts with the system processes through an allocentric view of a monitor.
CAVE	Cave Automatic Virtual Environment systems provides an immersive environment by projecting on the walls of a physical room, not being fully immersive this system has an allocentric interaction.
Egocentric	Interaction with an environment through a first-person view that allows the environment to completely surround the user with a self-to-object spatial processing and navigation.
Expert	Professional marine engineer officers with commission to serve in senior positions such as in capacity of Chief Engineer and 2nd Engineer, or in a junior position in capacity of 3rd Engineer, onboard a vessel of any flag and unlimited size.
Flow state	The construct of optimal experience in task performance by Csikszentmihalyi (1975).
Full Mission	Simulator environment that consists of multiple monitors and dummy equipment in one or more physical rooms to replicate an environment through allocentric monitor interaction.
Head mounted display	Goggles for immersive virtual reality which project an immersive environment to the user, substituting natural sensory input such as vision and audio.
Human element	A definition by IMO of the human as more than an agent within a system, where the human is valued as an active element within the sociotechnical systems of the maritime industry.
Immersive Virtual Reality	First-person egocentric view and interaction where the environment completely surrounds the user and multiple means of sensory input is manipulated digitally.
Mental workload	The accumulated strain of mental computation from external sources in a single task or repeatedly over time.
Novice	Students of marine engineering at the undergraduate university level vocated to become engineer officer cadets and consecutively licenced engineer officers.
Ranking test	Non-parametric statistical test method based on an ascending ranking order of data points where the sum of ranks, mean of ranks and sample sizes are used to calculate statistical difference (Field, 2009).

Abbreviations

3D VR	3D Virtual Reality desktop simulation.
DNV	Det Norske Veritas, maritime classification society.
Flow	The Flow state.
IBM SPSS	IBM Statistical Package for Social Sciences.
IMO	International Maritime Organization
ISO	International Organization of Standardization.
IVR	Immersive Virtual Reality simulation.
MWL	Mental workload.
NASA-RTLX	The raw task load index by Byers, Bittner, and Hill (1989)
NASA-TLX	The original task load index by Hart and Staveland (1988)
S FSS-2	The Short Flow State Scale by Jackson, Eklund, and Martin (2010)
STCW	The International Convention on Standardization of Training, Certification and Watchkeeping for Seafarers, with associated code, guidelines and amendments.

1 Introduction

1.1 Research background

The human element as an agent within the maritime industry is developing along with technology towards a vivid complexity of human-machine interaction. Interdependency between new technology and the human element drive the demand for both progressive technology development and the transcendence of the human capital to a new state of knowledge. Through a lifecycle from the first embarkation to retirement, the human element evolves constantly under an accretive pressure from the knowledge-society through academics and the profession. Dwindling with the improved GPS connectivity, here are only a few unconquered spaces left on the globe where the high seas provide a solitude from the modern world traditionally associated with seaborne freight. Maritime vessels are required to have the ability to operate independent of satellite connectivity and shore communication. With the increasing aid and dependence of modern technology, such as GPS, communication and automation, the decreasing number of personnel on board face complex responsibilities demanding a new and unprecedented sets of knowledge for safe vessel operation, in normal and emergency conditions. Technology holds a diverse definition, applied in progressive human endeavours and evident in most achievements. Through accretive mundane interaction technology has become a part of the seafarer´s knowledge-base, and the complex sociotechnical systems that a modern vessel now comprise of tend to put technical requirements in centre of design, engineering and operation, rendering the human element to adapt and cope with the rest through their interaction (Norman & Stappers, 2015). Towards higher degrees of automation, regulations and cost mitigates the sentiment for implementation (Mallam, Nazir, Sharma, & Veie, 2019), giving the human element a chance to develop trust and comfort in the technology increments. Training competencies requirement and technical regulations need to be bridged in the statutory framework (Mallam & Lundh, 2013), to ensure confidence in the human competence that supervise and share the conduct of duty with the fiduciary sociotechnical systems.

This study investigates state-of-the-art technology in the context of the present marine engineering education.

1.1.1 Simulators in maritime training and immersive virtual reality

With decades of 2D desktop simulators, the field of marine engineering education is now provided with 3D Full Mission (Figure 1) simulators as the established commercial standard. Such systems are mainly provided from Kongsberg Maritime of Norway, Unitest of Poland and Transas Marine of England (Shen, Zhang, Yang, & Jia, 2019). 3D Full Mission is a simulator type replicating the full engine control room and an engine room by touch screens and replicate console modules where the interaction with the systems are visually and audibly animated in 3D.

Figure 1: K-SIM Engine full mission simulator, with consent from copyright owner Kongsberg Digital.

Virtual reality is an emergent technology developing with increasing momentum, and a walkthrough application of a 3D virtual reality desktop simulator is the latest commercial addition to portfolio of engine room simulator applications (Kongsberg Maritime, 2018).

Expediting development and one step further from the 3D virtual reality desktop environment is the enhanced experience of immersion with head mounted display virtual reality. Though not available commercially, immersive virtual reality simulators are on the agenda of the developers. With immersive virtual reality technology, the environment surrounds the user with a true egocentric view, which discriminates the immersive experience from the 3D virtual reality. Figure 2 presents the egocentric vision of a 3D virtual reality desktop; however, interaction is allocentric as the user view the environment through a monitor.

Figure 2: Egocentric view of 3D virtual reality through a desktop monitor (Kongsberg Maritime, 2019a)

1.1.2 Purpose of research

This study investigates the prototype of an immersive virtual reality engine room simulator and commercially available a 3D virtual reality desktop simulator, both developed by Kongsberg Digital. As the immersive virtual reality development is not yet commercially available, it is a privilege to experiment with the technology in its intended context; maritime education. As other industries innovatively progress with the implementation of this new technology, the maritime industry needs to make up for this gap recombinantly. Maritime education can adopt this state-of-the-art technology and train the students immersed in a safe replicate of the environment they meet through their profession, given that there are beneficial improvements with the simulator technology. Scientific research such as this, is necessary for identifying outcome benefits, if there are any, and identifying present limitations for further development to succeed as an inevitable intermediate stage between development and dissemination. Measuring performance by the extremes of professional proficiency, simulator exercises can be adapted to enhance learning outcomes at various levels of expertise. Not only adding knowledge of applicability, this study exhibit that the technology is operable even at early stages of development and welcomed with ovation by the end users.

1.2 Research Questions

Simulator training in maritime education aims to train knowledge and skills for the consecutive professional service. This study attempts to describe the present and future state of simulators, and address the effect of immersive technology in maritime education. Cognitive skills by the constructs of declarative knowledge, mental workload and flow state are emphasized as the scope of this study.

RQ1: Is immersive virtual reality simulator training the better technological option for training declarative knowledge in

(a) maritime education?

(b) expert competence maintenance?

RQ2: Is there a difference in mental workload and flow state with immersive virtual reality

(a) compared to 3D Virtual Reality desktop simulator?

(b) between novice and expert groups?

1.2.1 Hypotheses

Hypotheses are stated as follows, and presented in connection to research questions in Table 1.

Hypothesis 1:

(a) The novice immersive virtual reality group (*ii*) will have a better score on declarative knowledge accuracy than the novice 3D virtual reality group (*i*).

(b) The expert immersive virtual reality group (*iii*) will have the highest declarative knowledge accuracy.

Hypothesis 2: The novice Immersive Virtual Reality group (*ii*) will have

(a) lower mental workload and,

(b) a higher flow state than the novice 3D Virtual Reality group (*i*).

Hypothesis 3: The expert Immersive Virtual Reality group (*iii*) will have

(a) a lower mental workload and,

(b) a higher flow state than the novice group (*ii*).

Assumptions for hypothesis 1 is that a) immersion will bring greater vigilance and attentiveness to the simulation which should be measurable in retention of memory after the treatment, and b) that the expert´s long-term memory and tacit knowledge frees up capacity for mental computation and render more accurate working memory internalizing to their long-term memory.

To directly answer the research question 1b, if immersive virtual reality simulator is beneficial to train expert competence maintenance, additional research including more expert groups is necessary, and falls beyond the scope of this study. However, performance and experience measurements of the expert immersive virtual reality group could give guiding indications for further design of exercise complexity to challenge this user group.

Hypothesis 2 holds the assumption that high mental workload correlates as disruptive to the flow state, and that a lower mental workload will facilitate the flow state. Defining borders between what is a high or low range of these constructs cannot be inferred by this study, it can only assume that there are such ranges by comparing group scores and perhaps find ranges that correlate with the learning outcomes measured. Hypothesis 2 a) assumes that the immersive experience is less mentally demanding than the 3D desktop experience, and that b) the flow state has a greater presence in the immersed experience than in the 3D desktop experience.

Hypothesis 3 assumes that the experts will perceive a) a lower mental workload and b) a higher flow state based on their professional experience.

Table 1: Summary of thesis research questions and hypotheses

RP: State-of-the-art technology in the context of the present marine engineering education			
RQ1: Is immersive virtual reality simulator training the better technological option for training declarative knowledge in		RQ2: Is there a difference in mental workload and flow state with Immersive Virtual Reality	
a) Maritime Education?	b) Expert competence maintenance?	a) Compared to 3D Virtual Reality desktop simulator?	b) Between novice and expert groups?
H1a: Novice immersive Virtual Reality group will have a better score on declarative knowledge accuracy than the novice 3D Virtual Reality group	H1b: Expert Immersive Virtual Reality group will have the highest declarative knowledge accuracy	H2: Novice Immersive Virtual Reality group will have lower mental workload and a higher flow state than the novice 3D Virtual Reality group	H3: Expert Immersive Virtual Reality group will have a lower mental workload and a higher flow state than the novice group
H1a: Power of memory IVR > 3D VR	H1b: Power of memory Expert > Novice	H2a: Experienced MWL IVR < 3D VR H2b: Experienced flow IVR > 3D VR	H3a: Experienced MWL Expert < Novice H3b Experienced flow Expert > Novice

1.3 Significance to the field

With emerging technologies of enhanced environments, higher order learning outcomes might become within reach. This study can hopefully inspire further research and development on maritime education simulators, and nurture interest for simulator design development, training design effectiveness and end-user acceptance. Performance and safety are the ultimate goals of professional education and training in high risk environments. If immersive virtual reality proves to be beneficiary to training effectiveness in the field of maritime education, it can easily be adopted to strengthen the educational programmes. Other fields have successfully utilized virtual reality simulators for training technical skills. It is also a possibility that it can be used for expert competence maintenance or exploring unfamiliar and new segments of the field. By focusing on training design, exercise complexity can be tailored to the individual, and differences in mental workload and perceived flow might connect to the intended learning outcomes of the design.

1.4 Thesis structure

Chapter 1 outlines the elements leading to the research study. Chapter 2 introduces the theoretical background for the research, and chapter 3 describes how the study was conducted. Chapter 4 presents the findings, which chapter 5 reflects on, before chapter 6 offer a conclusion. Chapter 7 holds the references and chapter 8 supplement all additional information of relevance to the reader.

2 Literature Review

2.1 Simulator Training

Games are discriminated from simulators as they are enjoyable and voluntary, often segregated from the real world with unproductive values and rules (Garris, Ahlers, & Driskell, 2002). Interim games and simulators are educational games which are game-based educational programmes using the same hardware as computer games. Simulators are used to train real life proficiencies in an environment safe from errors and play scenarios of extremes to practice performance of real life environments (Sellberg, 2017). Through a broad spectre of professions, simulators are used to generate specific learning outcomes and attributes needed in the real-life work of those professions.

2.1.1 Simulator training in maritime education

Commercial simulators designed for maritime training emerged in Norway in the late 1970´s and developed to be embedded in the education of both marine engineer officers and nautical officers. Norcontrol which later would be merged with Kongsberg Maritime, delivered their first analogue engine room simulator (Figure 3) to the maritime college of Trondheim in 1978.

Figure 3: The Norcontrol diesel engine simulator delivered to Trondheim Maritime College in 1978 (Kongsberg Maritime, 2019b)

The International Convention on Standards of Training, Certification and Watchkeeping for Seafarers by the International Maritime Organization (2016) sets the governing requirements for simulators and discriminates between the purpose of training and the purpose of assessing competence. This because the convention allows simulators to be used for training and assessment of novice seafarers in education, on-board training, and in revalidation of certificates for professional seafarers (A-I/11 & A-I/12, International Maritime Organization, 2016). The convention structures the whole industry in this way; not only the main competences of discipline and rank holds a corresponding certificate of competence, each additional and specialized formal proficiency requires a certificate of proficiency, some of which can be trained and assessed with simulators.

By ratification of the convention, the flag state´s responsibility with issuing and control of personal certificates lies with the national maritime authority. For quality control with simulators used for training and assessment in the approved education courses, the national maritime authority can require that the simulator equipment has been classified by a registered organization on their behalf, usually a classification society with a standard such as ST-0033 by DNVGL (2018) to preserve the convention requirements.

2.1.2 Immersive virtual reality

Virtual reality has been discussed for decades to revolutionize simulator-based education, where new skills can be practiced through correction, repetition and safe failure in an inexpensive environment representing reality (Jensen & Konradsen, 2018). Immersive virtual reality differs from non-immersive virtual reality where the user looks into the environment from an outside position, e.g. through a desktop display. Immersive technology exchanges the sensory input with digitally generated sound and vision, enabling the user´s brain and nervous system to behave as if present in a real environment (Jensen & Konradsen, 2018). With immersive virtual reality the user is surrounded by the environment by means of a head mounted display or with a CAVE system.

2.2 Verbal knowledge

Traditionally the scientific training field has focused on changes in verbal knowledge or behavioural capacities as learning outcomes (Kraiger, Ford, & Salas, 1993). Bloom (1956) proposed that there is cognitive learning outcomes beyond recollection and recognition of verbal knowledge in his taxonomy of learning. Gagné (1984) criticised and argued that it should include various cognitive, skill-oriented, and affective learning outcomes. Adapting and refining this, Kraiger et al. (1993) proposed their new framework for training evaluation and the assessment tools needed to capture the various learning outcomes. Confining to the cognitive learning outcomes of the framework, this category is built with a taxonomy of verbal knowledge, knowledge organization and cognitive strategies. As the cognitive learning outcomes are not only a static state of knowledge, evaluation and training evaluation also have to consider the dynamic process of knowledge acquisition, organization and application. Kraiger et al. (1993, p. 313) explain that *"Cognition refers to a class of variables related to the quantity and type of knowledge and the relationship among knowledge elements"*. Knowledge organization and Cognitive strategies, which underlying learning constructs are mental models and metacognitive skills falls beyond the scope of this study and are not further explored.

2.2.1 Declarative knowledge

As shown in Figure 4, verbal knowledge comprises of declarative knowledge, procedural knowledge and strategic or tacit knowledge (Kraiger et al., 1993). Declarative knowledge is information about facts, semantics and rules, and is easy to write, teach or test (Norman, 2013). Knowledge of rules doesn't ensure people will abide them and knowledge about facts don't have to be true, we only store sufficient knowledge to do tasks and don't need further precision in our judgements (Norman, 2013). Procedural knowledge is information about how to do things, and can be difficult or impossible to write down or teach in the same manner as declarative knowledge; it is best demonstrated and learned through practice (Norman, 2013). Through practice knowledge is converted from declarative form to procedural form in which it is applied, and gradually applied more appropriately and efficiently (Anderson, 1982). Wagner (1987) describes strategic or tacit knowledge in the contents of oneself, others and the task itself, and stated that knowledge about managing oneself is knowing how to best overcome procrastination. Tacit knowledge about managing tasks is knowing how to best perform specific work-related tasks, and tacit knowledge about others refers to managing others and one´s interaction with others (Wagner, 1987), i.e. leadership.

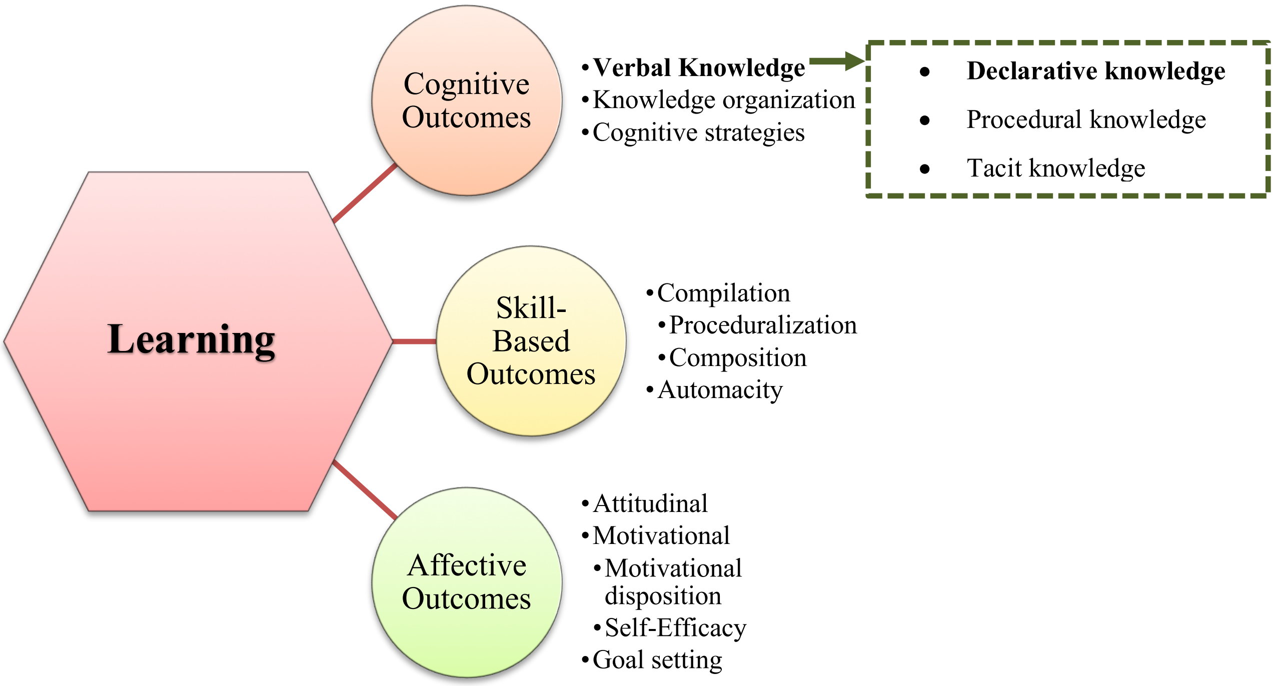


Figure 4: Classification scheme of learning outcomes adapted from Kraiger et al. (1993, p. 312)

2.2.2 Measuring declarative knowledge

Evaluating declarative knowledge is in line with how institutions today evaluate their subjects, where their acquisition of declarative knowledge is examined through multiple-choice, true-false, free recall or recognition tests (Kraiger et al., 1993). At a higher level of evaluation, speed tests measure within a given time, and power test measure correctly answered items given unlimited time (Kraiger et al., 1993). Power tests measure accuracy of stored information from memory and have traditionally ignored errors and focused on correct items answered (Ackerman & Ellingsen, 2016), these tests should be used when the consequences of errors are high and accuracy is valued (Kraiger et al., 1993). Speed tests will measure the speed of processing information and is hard to correct for guessing, to account for this, speed tests to measure fluid intelligence are designed incrementally harder for each item to discriminate at which level consistent answering disrupts. When forming a knowledge test, one should be particular in designing the format, as different tests measure different underlying constructs of cognition and knowledge.

Naturally individual differences will affect and form a group score. The underlying constructs measured by these various knowledge tests are influenced by differences connected to individual general intelligence, which can be decomposed into abilities such as fluid intelligence, crystalized intelligence, spatial abilities, perceptual speed abilities, psychomotor abilities and more (Ackerman, 2014). As general intelligence factors seem to be critical for novel task performance, trainees competent at inferring relations and memorizing information will show success in early training. Through further exercise and experience this between-subject gap will close towards a stage of procedural knowledge as behaviours become internalized and psychomotor differences affect performance as much as intellectual capabilities in task performance (Ackerman, 2014; Kraiger et al., 1993).

On measuring declarative knowledge in its traditional form during training, Kraiger et al. (1993) argue that these tests should be given at an early stage in the training, as the feedback is necessary to identify the knowledge gap that might inhibit the consecutive higher order learning, such as converting to procedural knowledge and developing tacit knowledge unbiased of false knowledge and expectations. Further implications for repeated measurement is that since variance in declarative knowledge will be greater at the beginning of training than at the end, higher scores measured early is more beneficial for predicting other learning outcomes (Kraiger et al., 1993).

2.2.3 Effect on immersive virtual simulator training

Webster (2016) investigated declarative knowledge acquisition with immersive virtual reality on soldiers, and not surprisingly in accordance with other similar studies, he finds that the immersion has a positive effect on the learning outcomes compared to lecture-based instruction. In their review of studies on immersive virtual reality training, Jensen and Konradsen (2018) find that lecture-based instruction is better for remembering facts while an immersive learning environment is better for spatial and visual knowledge, further they found no research that have examined training of higher order cognitive skills with immersive virtual reality.

While crystalized intelligence, i.e. knowledge and skills, is subject to improvement by training, the fluid intelligence is not; at least not trainable beyond the individual´s genetic boundaries. Fluid intelligence, i.e. the rate of solving novel problems, also indicates the individual´s ability to perform in other intelligence factors. Through the career studies of James Flynn, an average improvement of 17 points on the Wechsler Adult Intelligence Scale (WAIS) and Wechsler Intelligence Scale for Children (WISC) tests through the second half of the last century show a trend in intelligence scores megalomaniacally known as the Flynn Effect (Flynn, 2013). This have attracted massive attention and discussion in the field of cognitive science as explanations differ. Flynn (2018) keeps to his explanation of social change caused by technological development which stimulate a sentimental shift in the way people think while developing through their life, and not an actual improvement of the individual´s cognitive capacity as fluid intelligence is more or less static to a healthy human.

Available to only a privileged few through the last centuries, complex and abstract ideas are becoming public domain as technology and social progression allow groups of very intelligent people to create enhanced environments which can be disseminated for the general public to immerse into. With this Flynn (2018) suppose that the performance level of genes are utilized when people are given the opportunity to be expressed and immersed in these enhanced environments of cognitive stimulation.

Summarizing a decade long team effort where Passig (2015) investigates immersive virtual reality as training medium of cognitive skills, they can conclude that while some cognitive skills deteriorate in the population over time, others emerge. Though some research now find average IQ scores to decline, we might be in an erratic evolutionary process we simply cannot comprehend or measure at this time (Passig, 2015), or ever given the fluctuation of confounding cognitive factors. In summary, they conclude that human mental capabilities in fact are improving, though it is not absolute certain they do so solely through advanced technology, by stimulus-filled environments or evolution. Not only does advanced technology such as immersive virtual reality improve abstract cognitive skills as supposed by the Flynn Effect, concrete cognitive skills improves as well according to Passig (2015).

2.3 Mental Workload

Complex sociotechnical systems should be designed and developed with the human and the human-machine interaction in centre (Norman, 2013). Human operators of complex systems with an increasing level of automation face the responsibility, more as a supervisor than a manual labourer. Though excellent integrated automation systems are designed to share the workload with the operator, manual interaction with the physical systems is essential in marine engineering. The maritime industry is evolving towards a greater level of automation, and perhaps even a state of full automation where the human element´s physical presence is superfluous. With the present technology and it´s regulation, vessels granted the DNV GL "E0" or "ECO" additional class notion can operate in a state of unmanned engine room for shorter periods of time (DNV GL, 2018).This renders the officers and ratings of the department higher flexibility to perform maintenance during the workday hours, recreation in off-duty hours and rest during the night, opposed to the usual continuous three shift seawatch system with 4 hours on / 8 hours off with maintenance responsibilities during the off-duty period. The automation system shares the workload with the on-duty officers, day and night for weeks, but the accountability for safe operation falls entirely to the conduct of the human as principal to the fiduciary automation. Fatigue is a known abstract phenomenon in the maritime industry and an embedded consideration in regulatory requirements (A-VIII/1, International Maritime Organization, 2016), though mental workload is not explicitly mentioned it is a directly contributing factor to fatigue. Mental workload is a multidiscipline phenomenon, faceted of definitions across converging scientific approaches, Hart (2006, p. 904) describe it as *"... a term that represents the cost of accomplishing mission requirements for the human operator."*

2.3.1 Mental Workload and connected constructs

Mental workload is an accumulated strain, represented by proportions of cognitive and physical resources demanded of the human by the external environment while performing a task (International Organization for Standardization, 2017). A task demand superior to the individual aptitudes typically result in performance degradation, and possibly human error (Stanton et al., 2013), which arguably is facilitated by improper system design (Norman, 2013; Reason, 2000).

Reviewing 550 screened and selected articles on the usage of her NASA-TLX framework, Hart (2006) finds that situational awareness was cited as a covariate factor of mental workload in 7% of the studies. Regardless of the possible correlation it is suggested that situational awareness is a consequence of workload (Endsley, 2012) and that the two are not independent constructs (Hendy, 1996; Parasuraman, Sheridan, & Wickens, 2008).

Individual differences, such as working- and long-term memory is influencing both mental workload and situational awareness. The novice operator rely more on working memory for mental labelling of objects, gaining high mental workload and low situational awareness, opposed to the expert operator whom possesses a greater long-term memory able to free the mental computation required by the working memory to conduct a task (Scholtz et al., 2006). Figure 5 shows how humans process information through two channels of limited capacity where learning occurs by active engagement of these cognitive processes through organizing combined with integration of prior knowledge (Mayer, 2010). Sensory memory holds an exact replicate of information for less than 0.25 seconds, both sensory memory and long-term memory is capable of unlimited capacity (Mayer, 2010). The mental computation takes place in the working memory and stores processed information for up to 30 seconds, this processing has a limited capacity and act as the bottleneck of the model in Figure 5.

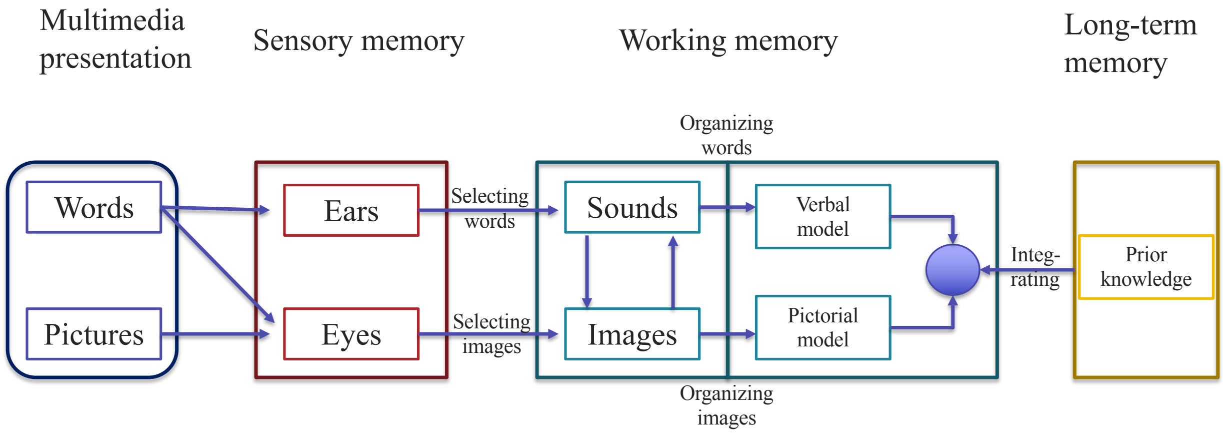


Figure 5: A cognitive model of multimedia learning adapted from Mayer (2010, p. 545)

2.3.2 Measuring mental workload

As sources of workload are numerous and vary for every single task and operator, Hart and Staveland (1988) designed a framework to reduce this subjective between-subject variability that is experimentally irrelevant, offering a scale that include and emphasize contributions of other sources of variability that is experimentally relevant for the human-machine interaction. The NASA Task Load Index (NASA-TLX) conceptual framework of Hart and Staveland (1988) defines workload as a human centered concept rather than a task oriented one where there is no objective standard to compare against when people evaluate the workload of a task they have performed (Hart & Staveland, 1988).

Since the emergence of the NASA-TLX, science has produced several other measuring instruments and techniques to supplement rating scales or to be used independently. Stanton et al. (2013) categorise both objective and subjective measurements as shown in Table 2.

Table 2: Categories of mental workload measurements

Primary and secondary task performance measures	Performance and reaction times, and the ability to perform embedded secondary tasks
Physiological measures	Measuring the physiological aspects affected such as eye movement, brain activity, heart rate (HR) and heart rate variability (HRV)
Subjective rating techniques	Among a set of generic and tailored techniques the Subjective Workload Assessment Technique (SWAT) and NASA-TLX are the most commonly self-assessments used
Quantitative evaluation of task demands	Predicting the level of mental workload with analytical tools that are used during design when an operational system is not available for empirical testing

Subjective measures often correlate with perceived performance and are administered either during or after the task. The NASA-TLX is a multidimensional subjective self-assessment tool, scaling the operators experience of the task with a global score from 0 to 100. Byers et al. (1989) developed a “raw” adoption (NASA-RTLX) of the original framework which excludes the weighting of the dimensions found in the original framework. The author of the original framework credits the NASA-RTLX and her review finds it to be more, less or equally as sensitive as the original NASA-TLX (Hart, 2006). It is likely to assume that definition or confusion of the term across scientific disciplines and their experts, also apply to the rating subject, thus the six dimensions in Table 3 are sub-scaled to represent independent clusters of variables and accumulated to the global score as an average of these (Byers et al., 1989; Hart, 2006; Hart & Staveland, 1988).

Table 3: Dimensions of the task load index (Hart & Staveland, 1988)

Mental Demand	The mental and perceptual activity demanded
Physical Demand	The physical activity required
Temporal Demand	The time pressure felt or the pace at which task elements occurred.
Effort	Mental and physical work required to accomplish the performance
Performance	The success of accomplishing the goals
Frustration level	Balance between insecurity, discourage, irritation, stress and security, contempt, relaxation and complacency.

2.3.3 Effect on immersive virtual simulator training

If the operator´s workload is too high and exceeding individual aptitudes, the operator will not have the time to collect and process the information needed to perform the task and may have less effective response to consecutive events, and likewise, if the workload is too low, the operator´s vigilance decrements and lead to boredom (Parasuraman & Hancock, 2008). Aligned with this, Endsley (2017) propose that workload covary with situational awareness in an inverted curvilinear regression with detrimental extremes to safety and performance. This is a key issue when designing automation systems in regard of distributing the operator´s workload in order to facilitate sufficient situational awareness between the human-machine interaction, allowing the operator to conduct safe and efficient decision making (Endsley, 2018). High mental workload can cause memory failure and inhibit perception of information or cause failure to comprehend information due to working memory limitations, which decreases situational awareness and thus enables errors to occur (Endsley, 1995). Though low workload can allow for a sentimental shift to secondary task performance and multitasking (Cullen, Rogers, & Fisk, 2013), this is reliant on an adaptive balance of workload between the human and the interacting system (Stowers et al., 2017) as high or fluctuating workload might disrupt task prosecution and inhibit the operator to return to and recover the interrupted task (Chisholm, Weaver, Whenmouth, & Giles, 2011).

Studies of performance, stress and workload in immersive virtual reality simulation scenarios and the subsequent transfer to live training exercises of soldiers (Lackey, Salcedo, Szalma, & Hancock, 2016), show that simulator workload can indicate the imposed workload of the live task but not the live performance. Those soldiers who reported a positive experience in terms of the flow state reported a lower level of stress and workload when engaged in the live training exercise, indicating a relationship between simulator learning, experience and workload with immersive virtual reality. The research of vigilance, workload and stress on soldiers by Warm, Matthews, and Finomore (2008) suggest that an active regulation of task demands tend to stimulate task engagement, whereas more constrained task configurations can lead to task disengagement. Recent studies suggest that task complexity of simulator training should be lesser in early training, and incrementally more complicated as the training progress, adapted to the capability of the trainee (Hjelmervik, Nazir, & Myhrvold, 2018).

2.4 Perceived Flow

Flow represent moments when everything comes together as a psychological state for the task performer. When experiencing the flow state, one feels strong, positive and disregard the fear of failure and self-consciousness. The experience is perceived as rewarding and better than usual to the context as the recipient becomes immersed into and absorbed by the task.

2.4.1 The Flow State

Csikszentmihalyi (1975) framed the concept as a special and psychological state that brings the recipient enjoyment thorough high performance in a positive experience. Flow occurs when one is totally engaged with a task which creates an intrinsic reward regardless of the task complexity. Presenting itself on relatively rare occasions (Jackson, Martin, & Eklund, 2008) it can be experienced at a lower degree in mundane tasks from everyday life, or at a higher degree as a result of greater demand and complexity where *"being totally connected to the task in which one is engaged epitomizes the flow state."* (Jackson et al., 2010, p. 8).

In their study on performance, immersion and flow of soldiers, Lackey et al. (2016) finds that for the live training exercise Sense of Control is related to lower stress and even stronger among those with better performance, while in the virtual reality training exercise this relation was only found among those whom scored higher in performance. The dimensions of Challenge/Skill Balance and Unambiguous Feedback showed a statistically significant regression with the global workload score of the highest performers, concluding that there is a relation between a high flow state and low mental workload.

2.4.2 Measuring the flow state

Through the conceptualization, Csikszentmihalyi (1990) built the theory with nine dimensions which are profiled through the LONG Flow scales or indicated through SHORT Flow scales. Jackson et al. (2010) describe the nine-faceted construct as shown in Table 4.

Table 4: Dimensions of the flow state scales (Jackson et al., 2010).

Challenge-Skill balance	During a task, the opportunities for action and goals face a subjective balance with the capacities possessed by the performer to produce the desired outcomes. The influence of perception drives this balance and the subjectivity of the performer is more important than any objective skill level to achieving a state of flow. As this balance is dynamic, challenge- and skill levels can be manipulated to create flow across all domains of task performance.
Action-Awareness merging	Feeling at one with the task in total absorption creates a harmony in the activity and brings peace to the engagement. Action-Awareness merging can be associated with a sense of effortlessness and spontaneity where automaticity of routines enables subconscious information processing and render more attention to actions.
Clear Goals	The process of goalsetting can facilitate the flow state if done successfully. Task performers in the flow state have described a clarity of purpose, occurring on a momentary base, connecting the performer to the task objective and responsive to emerging cues. Strategy cues and predefined action allow the performer to shift more attention to immediate tasks.
Unambiguous Feedback	Information processing and comprehension of feedback is necessary for determining whether one is on track towards the task outcome. In the flow state, information is received clearly and feedback interpret unambiguously with less effort, keeping the performance within the desired projectory. Sources of feedback are both internal, the movement and displacement of one´s body in the environment, and external, the given information from the environment itself. Not always positive, feedback helps the performer adjust actions to resurrect lost flow or increase its level. It is not necessary to freeze the task for reflection as the feedback processing is integrated in the performance.
Total Connection on the task at hand	Focus on the task is the clearest indicator of the flow state. Without digressing thoughts, mental clarity and sentiment on the task offer satisfaction, which in turn stimulate to increase the complexity of the task (Csikszentmihalyi, 1990). Interestingly, concentration experienced in the flow state is complete, intense and spontaneous, in contrast to usual task experiences where more effort is required to keep the concentration on the task.
Sense of Control	Like the challenge-skill dimension, sense of control has a fine balance through perception. With control comes a sense of infallibility which frees the performer from the fear of failure, though total control does not exist. The perception of total control can inhibit the flow state experience as it induces boredom and relaxation in the performance, same as if skills overbalance the challenge.
Loss of self-consciousness	People tend to constantly evaluate how they are performing in the eyes of others, especially in situations they perceive as important. Loosing self-consciousness and mitigating one´s ego by disrupting this evaluation is necessary for the flow state to occur.
Transformation of time	When nothing is entering our awareness during intense concentration, time might surprisingly fly, slow down or even stop. Transformation of time is thought to be linked to concentration and can only be experienced when the flow state experience is very deep.
Autotelic experience	Autotelic is defined as something that have an ending or purpose of itself. This experience was termed to describe the enjoyment of the flow state as a result of the other eight dimensions (Csikszentmihalyi, 1990). The experience motivates the performer to push for higher limits and further engagement, after completion and reflection on the performance of the task.

The situation-specific SHORT Flow State Scale is abbreviated S FSS-2 (Jackson et al., 2010) and consist of only one question for each dimension, providing a global sum instead of a profile for each dimension. In comparison to its antecedent the LONG Flow State Scale (FSS-2), the S FSS-2 provide a sufficient measure, though naturally the relative novelty to the field makes it less validated. The S FSS-2 was developed to capture the presence of flow while not constraining participants when other constructs was central to the research. Though the S FSS-2 is an useful indicator, the flow state construct holds some mystique to scientific research as it cannot be fully captured on a questionnaire (Jackson & Marsh, 1996), with the antecedent experience sampling methods (Csikszentmihalyi & Larson, 1987), or with in-depth interviews.

2.4.3 Effect on immersive virtual reality simulator training

The optimal experience researchers who authored the Flow scales (Jackson et al., 2010) base the research on a variety of disciplines and fields. This empirical assessment of the quality of experience and performance can be applied to more than the domains of the initial book on the flow concept (Csikszentmihalyi, 1975). Hamari et al. (2016) investigate the impact of flow, engagement and immersion in game-based learning and find that increased engagement due to flow improved the learning effect, whereas immersion did not. Further they analyse the challenge facet of flow to be the strongest predictor of learning outcomes, with the skill balance only providing a mediating effect on perceived learning. Lackey et al. (2016) found that the dimension Skill-Challenge balance was strongly related to mental workload, and that the dimensions Feedback and Sense of Control was contributing drivers. Further they connect the high flow state relation that lower mental workload, to bias and mitigate accurate perception of performance in poor performers. For the better performers, the Skill-Challenge balance mitigates the Frustration dimension of mental workload (Lackey et al., 2016). On both good and poor performance, one might expect a high flow to give low mental workload, although the flow state is conceptually associated with high performance and efficiency.

2.5 Summary of theory

Evident in the regulatory framework, simulator training has been embedded in maritime education and training for decades, from the first analogue simulators with gauges, buttons and lightbulbs, to computer-based digital simulators with different interface options. A new era of simulator environments is emerging in the field of maritime education and opens a portal for new research to benefit the advancement of both technology and the human operator.

With two simulator technologies, this study compares the effects of immersive virtual reality with 3D virtual reality desktop, and cognitive factors as verbal knowledge, mental workload and flow state are addressed with groups at polarized extremes of professional experience. These constructs are interlinked and correlated as facets of cognition aimed to be measured in the context of the human within a sociotechnical system, which lofty goal is performance and safety.

Verbal knowledge is the basis of education and is a prerequisite for professional proficiency, formed through practice and experience. As a supplement to on-the-job experience, simulator training might be a medium for transforming declarative knowledge into procedural and strategic knowledge. Different levels of prior verbal knowledge will have implication on early simulator training, and through repeated training psychomotor skills develop to render intellectual indifferences less incumbent for performance (Kraiger et al., 1993).

Mental workload is a result of mental computation while performing a task, and is a critical factor for automation and system design as safety is ultimately dependent on the human element´s fitness to perform. In the right context subjective measures of mental workload holds a value when sampled from multiple subjects as individual differences is mitigated by the numbers. Mental workload is found to describe the performer´s success in the task, and indicate the task complexity relative to the individual (Warm et al., 2008), as it is a result or a covariate of several cognitive constructs from theory. In the simulator training context, mental workload has to be adapted in the task design to facilitate learning, as excessive mental workload is disruptive performance and response (Parasuraman et al., 2008), and suboptimal mental workload is detrimental to vigilance (Parasuraman & Hancock, 2008).

Flow state measures provide an indication of how optimal the experience was for the performers and give another angle to the mental workload measures. Presence of the flow state is found to have connection to immersion and performance (Hamari et al., 2016), and some of the dimensions of flow have been found to be significantly related to mental workload (Lackey et al., 2016).

A connection between the concepts described in this study follows in Figure 6. The model depicts the constructs as the author´s comprehension of the selected theory and represent one individual´s cognitive processing. Situational awareness and mental workload are suggested to covary, as are mental workload with the flow state. Here they are proposed to change in magnitude as their state fluctuate together. Sensory input and prior knowledge are filtrated and organized in the working memory as mental computation exchange and correspond with the other cognitive constructs in the process of enacting.

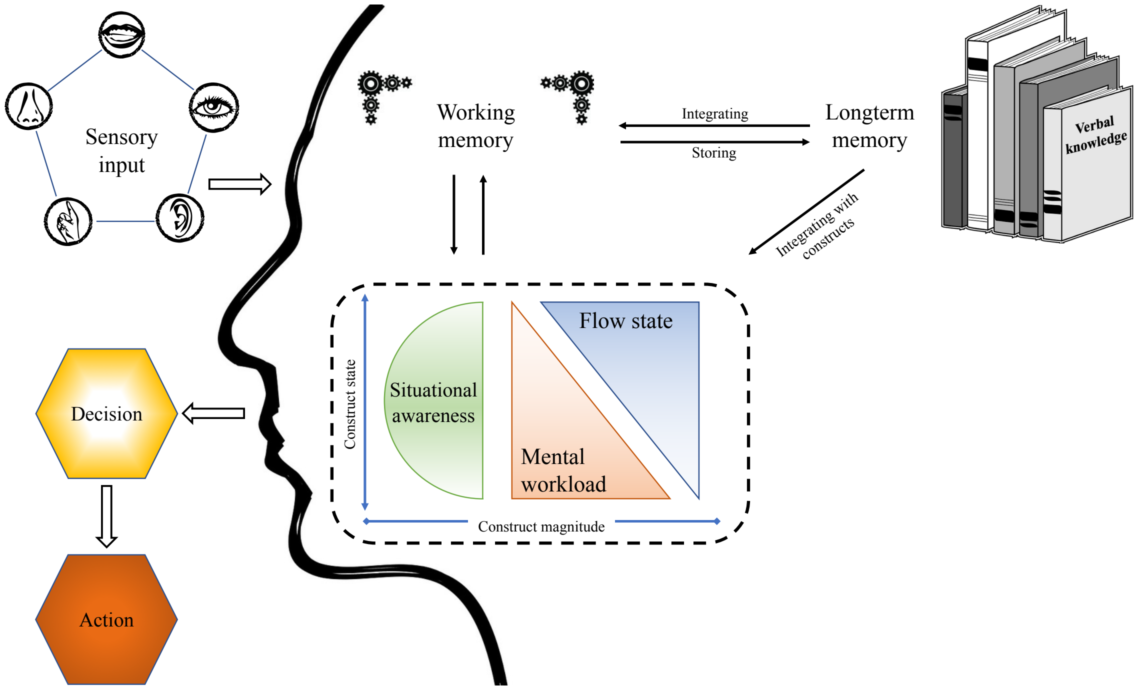


Figure 6: Author´s hypothetical model of the constructs

The prerogative of this study is to introduce the novel technology of immersive virtual reality by the means of head mounted display to the small field of marine engineering education and find measurements to describe it´s position relative to the extremes of professional proficiency.

3 Methodology

3.1 Introduction

This study investigates state-of-the-art technology in the context of the present marine engineering education. This chapter elaborates the process conducted to answer the intensions and premonitions of table 1.

3.1.1 Ethical Implications

Measures to avoid simulator sickness during the experiment was taken. The lab was quiet, with constant lighting, mechanical ventilation and had a stabile temperature of 18°C. The equipment was rigorously quality tested and the risk of simulator sickness was considered relatively low. Time exposed to immersion was designed to be brief and intermissions between the immersed sessions was designed for the participants to recover. No formal measurements of simulator sickness were included; however, the researchers present had attentiveness towards signs of discomfort and were acute to immediately help abort the immersed sessions if necessary.

An approval with the file number 188181(Appendix B) from the Norwegian Centre of Research Data was granted after commencement of the novice groups data collection. Due to the two exiting questions on the demographic questionnaire (Appendix D), this study is considered to collect personal health information and thus are under strict legal regulation concerning prosecution of that data. The approval was granted on the final edition of the information and consent form (Appendix C). The novice groups participants were given the updated version of the form by email once approved. The author decided to write an endorsement on the old information and consent forms, legally binding them to the approved form through the file number 188181, instead of collecting new signatures. All expert group participants signed the correct form.

On these two questions, the participants must answer if their eyesight is normal or corrected to normal, and if they have any history with or a diagnosis of epilepsy. Having epilepsy was considered as an excluding factor, and would result in abortion of the experiment. No participants were excluded or wished to withdraw from the study, before or after the experiments.

Although the author holds an unlimited chief engineer certificate of competence and an assessor certificate (STCW model course 1.30: On-board Assessment), he does not hold the STCW model course 6.10: Train the Simulator Trainer and Assessor. Being uncertified as a simulator trainer for maritime education was not considered as an implication for the author to act the function of simulator instructor in the context of this research study.

3.1.2 Limitations

During the initial process of this research the The Directorate of ICT and Joint Services in Higher Education where in negotiation with 11 publishers in their process of restructuring access and licence agreements regarding scientific publications. In effect, this limited or denied the author´s access to Elsevier, Wiley, Taylor & Francis and SpringerNature, enduring the first months of the research. In the transition to the OpenAccess platform, the only option to obtain these inaccessible records was to inquire the authors through ResearchGate, corollary ineffective. As negotiations progressed, access to publications was reinstated, and refinement of the theory chapter prolonged throughout the study.

Norway is a relatively small country with a mere population of 5.3 million citizens (Statistisk Sentralbyrå, 2019) dissipated over 385.000 km^2 land area and a 2.532km coastal baseline. Of approximately 20.000 national professional seafarers, there are only about 4.100 marine engineer officers active. The University of South-Eastern Norway holds the only national marine engineering programme at a university under-graduate level. This offer some challenge if conducting a large sample studies in this specific field. The challenge was resolved in the between-groups design; a whole class at the program was conveniently recruited and a cross-section of the professional population was purposely recruited locally to the expert group.

Both simulators used are developed by Kongsberg Digital. Though the underlying simulator programme used comprises a high level of details in the simulated system, both simulators had some elements visually missing in the environment. These elements are more or less critical for operating the real-life equipment. Affecting limitations was adopted into the design of the study with benefit to the memory power test, and measurements was aggregated with the limitations (8 items) of the immersive virtual reality simulator as index.

3.2 Research Design

This study followed a quantitative research strategy and classic experimental design with three groups for between-group pre-test/post-test measures as shown in Figure 7. The design was chosen regardless of the assumption that the focus groups of the sample frame would be limited due to access of professional marine engineer officers and students. Initial intentions were to have group sizes at the higher end of a small sample study. The treatment was designed to be as similar as possible in both the immersive virtual reality simulator and the 3D virtual reality desktop simulator.

The pre-test was developed by the author to capture the subject's semantics and system knowledge of the machinery operated in the treatment. The post-test consisted of a memory test developed by the author and questionnaires adopted from original frameworks.

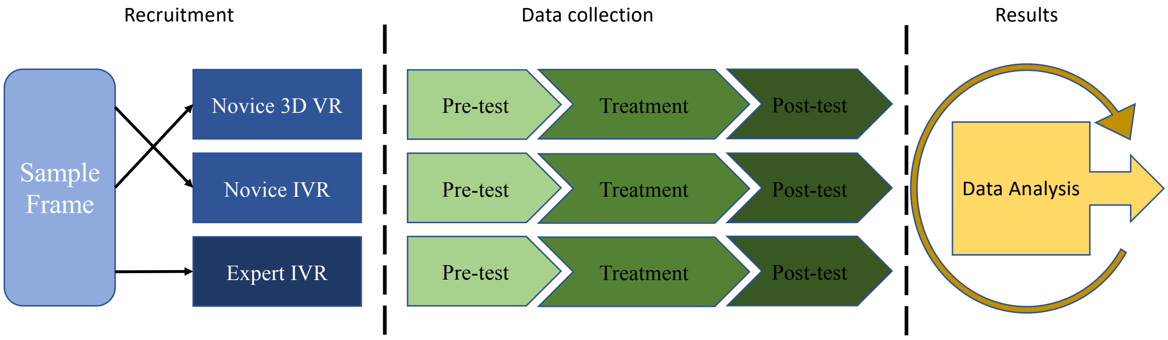


Figure 7: Research stages and underlying processes

Hypothesis 1a, measure the power of memory in the two novice groups with different simulator types, the group assignment is the independent variable and the power of memory after the treatment is the dependent variable. Hypothesis 1b compares novice and expert power of memory, the group level is the independent variable and the power of memory after the treatment is the dependent variable.

When comparing the two simulators in hypothesis 2a and hypothesis 2b with the post-test measurements of mental workload and flow state in the two novice groups, the group assignment is the independent variable and the experience is the dependent variable.

When comparing the novice and expert immersive virtual reality groups in hypothesis 3a and hypothesis 3b with the post-test measures of mental workload and flow state, the group level is the independent variable and the treatment is the dependent variable. Table 5 summarizes hypotheses and variables.

Table 5: Variables of the hypotheses investigated

Hypothesis	**Construct**	**Prediction**	**Dependent variable**	**Independent variable**
H1a	Declarative Knowledge	IVR > 3D VR	Power of memory	Group assignment
H1b	Declarative Knowledge	Expert > Novice	Power of memory	Group level
H2a	Mental workload	IVR < 3D VR	Experienced MWL	Group assignment
H2b	Flow state	IVR > 3D VR	Experienced flow	Group assignment
H3a	Mental Workload	Expert < Novice	Experienced MWL	Group level
H3b	Flow State	Expert > Novice	Experienced flow	Group level

3.3 Setting

All experiments and data collection were conducted in the virtual reality lab at the University of South-Eastern Norway. The lab was set up with both the immersive virtual reality simulator and the 3D virtual reality desktop simulator in the same room. For the immersive virtual reality simulation, an instructor station was set up as shown in Figure 8 where the instructor could monitor the process, the participant´s actions in the environment and the participant physically. In the 3D virtual reality desktop simulation, the instructor station only includes the 2D process control interface, though the lab was set up for the instructor to also have vision of the 3D virtual reality desktop monitor used by the participants.

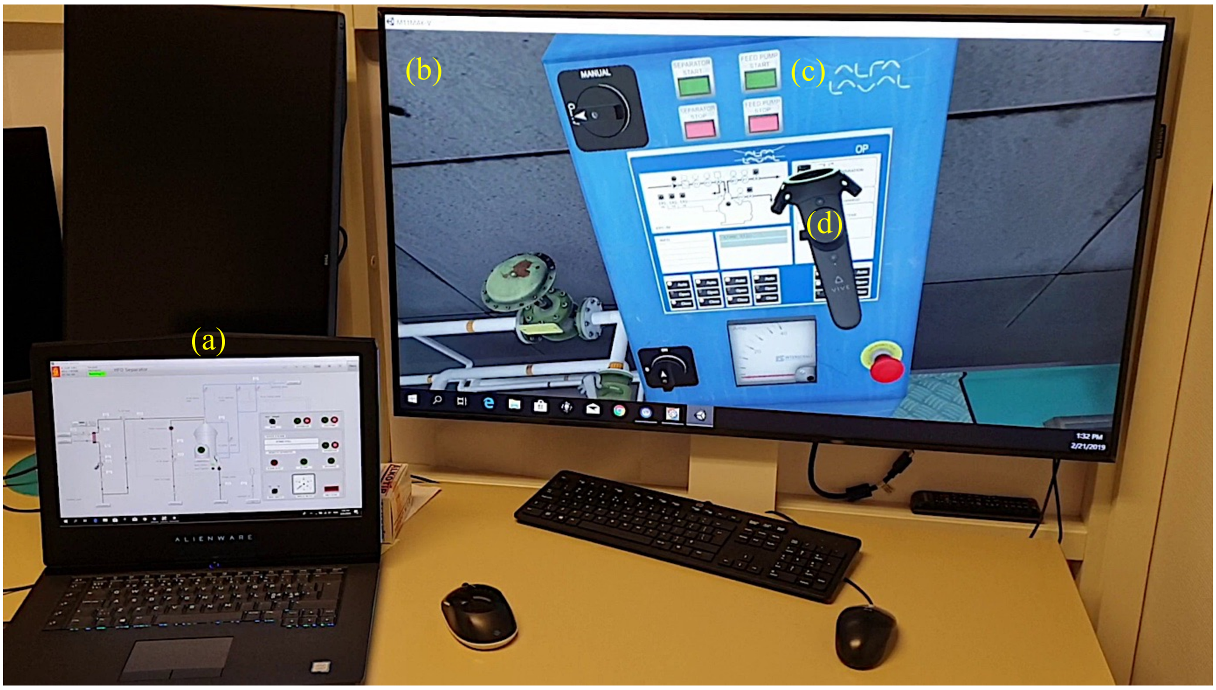


Figure 8: Instructor station of the immersive virtual reality simulator with; (a) 2D process control monitor, and (b) simulator projection from the head mounted display, as a participant is operating the (c) local control panel of the separator with the (d) hand controller.

3.4 Participants

The populations of the sample frame comprehensively represent a cohort of the marine engineer education and an exclusive cross-section of the professional population. The sample frame was organized in three groups; two novice groups and one expert group. The criteria for being classified as a novice is given by the definition (page 7), in addition the participants had to be formally enrolled in the educational programme and had to be recruited prior to any sea service time as engineer officer cadets which is possible during the summer semesters after the second year. The novice groups were conveniently recruited from the 2nd year class of the undergraduate programme in marine engineering at The University of South-Eastern Norway. The students were randomly assigned to either a (*i*) 3D virtual reality desktop group (*n*=5) or an (*ii*) immersive virtual reality novice group (*n*=6) based on their voluntary booking time for the experiment. The (*iii*) immersive virtual reality expert group (*n*=6) comprised of professional marine engineer officers with strong experience, purposely recruited through the author´s network. The expert group recruitment criteria, in addition to the definition (page 7), were a substantial competence as a senior engineer officer or equivalent. This means the expert group participants should have at least ten years of professional experience as evident in Table 6, hold a senior engineer officer certificate of competence, and have service time on board vessel with large machinery both in size and complexity. In a practical measure this would ensure that the participants had a professional experience from vessels least as large and complex as the container feeder vessel used in the simulators. Incidentally all participants were males, and recruited individually by the author. The student groups were addressed in plenum prior to formal invitation (Appendix A) and the expert group was directly invited formally (Appendix A).

Table 6: Participant demographics

Group		**(i) 3D VR Novice**	**(ii) IVR Novice**	**(iii) IVR Expert**
	n=	5	6	6
Age	Mean	28,40	22,67	42,50
	SD	12,66	0,82	5,01
Professional experience	Mean	0,60	0,17	17,17
	SD	0,89	0,41	8,01

3.5 Intervention and materials

An exercise of starting up a fuel oil separator was chosen for the treatment, which is an important machinery system both in education and during sea service. At the time of the research, this machinery system had been covered in the marine engineering programme through lecture-based instruction, 2D desktop simulator training and Full Mission simulator training. Though the immersive virtual reality simulator and the 3D virtual reality desktop simulator had slightly different limitations in their replication of the real-life equipment, the exercise description (Appendix E) was formulated to match both conditions. Due to these divergencies the post-test declarative measurements would have a different range between the two simulators as elaborated in the next section.

3.5.1 Pilot study

After defining the treatment and the measurements, a pilot-study was conducted with two participants performing the experiment in the immersive virtual reality simulator. Both participants are professional marine engineer officers with substantial first-hand knowledge of student simulator training, but without any prior experience with immersive virtual reality or this specific simulator. After the pilot study feedback, minor adjustments completed the experiment design with the final procedure (Appendix F) as shown in Figure 9.

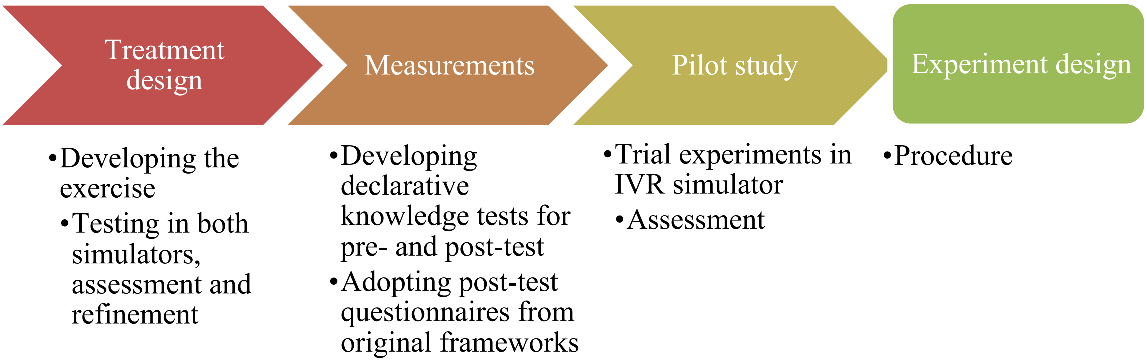


Figure 9: Process deriving at the experiment design

3.5.2 Simulator equipment

In the immersive virtual reality simulator, the participants interacted through a HTC VIVE VR system with a head mounted display and hand controllers (Figure 10). The simulator and instructor station were set up in the virtual reality lab as displayed in Figure 8.

In the 3D virtual reality simulator, the participants interacted with the simulator through a 27” desktop monitor and a wireless Bluetooth XBOX controller (Figure 10:a). The instructor station was set up with a 2D process control monitor (Figure 8:a) and a partition wall between the participant and the instructor where the instructor had view of the participant´s desktop monitor.

Figure 10: Simulator equipment. (a)Wireless XBOX controller for the 3D virtual reality simulator (Microsoft, 2019), and HTC VIVE VR system with (b) head mounted display and (c) hand controllers (VIVE, 2019) for the immersive virtual reality simulator.

3.5.3 Treatment

For both simulator interfaces, an Alienware 15 R3 laptop computer with the K-SIM ENGINE software was used. The laptop computer had a 2.9GHz Intel Core i7 processor and 16GB RAM memory with the Windows 10 Pro operating system. Kongsberg Digital used the Mak 8M43C M11 Container simulator software to develop the immersive virtual reality interface (Figure 11). The 3D virtual reality interface (Figure 14) of the simulator is commercially available. The Mak 8M43C M11 Container software is stated to replicate a 120m long reefer container vessel, with 1100 tons deadweight capacity, an 8-cylinder 43cm bore four-stroke Mak main engine and capable of 17 knots, which is equivalent to a small 1000 TEU container feeder vessel.

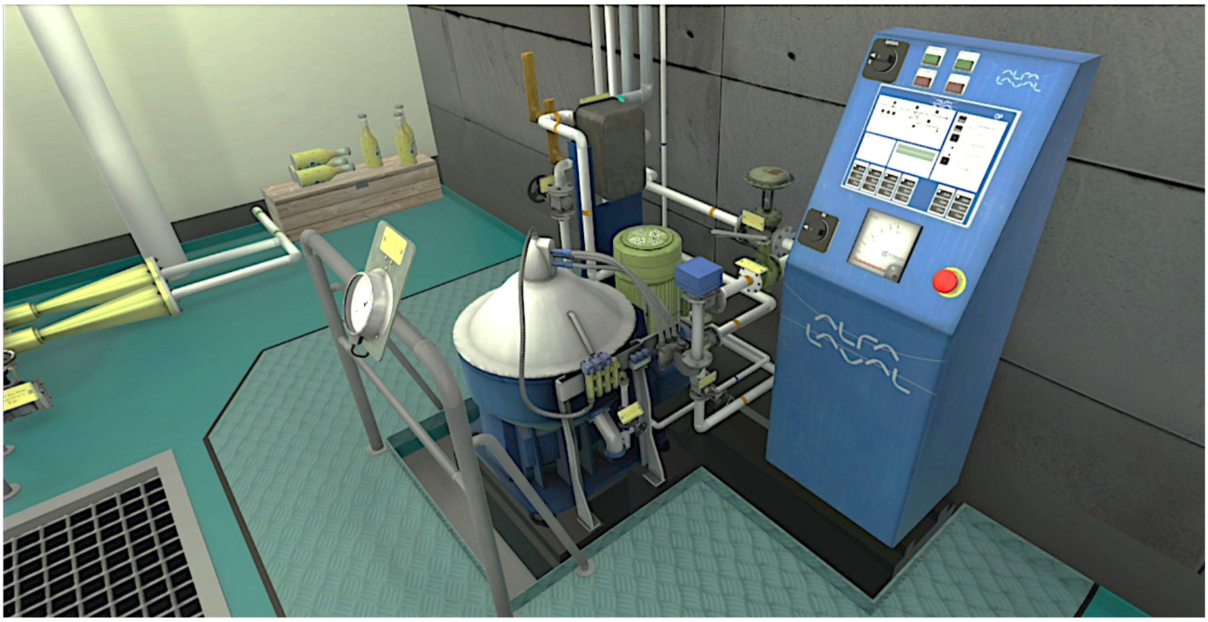


Figure 11: Fuel oil separator as viewed through the head mounted display of the immersive virtual reality simulator

Compared to the 2D process control interface (Figure 12) of the instructor station, both simulator environments had divergencies with missing elements. These elements were only visually missing in the simulator environments, and had no critical implication to the treatment as their function were included in the actual simulator programme. The 2D process control monitor is the equivalent interface to the ship´s Integrated Automation System were the duty engineer officer would operate and monitor the machinery from in the engine control room.

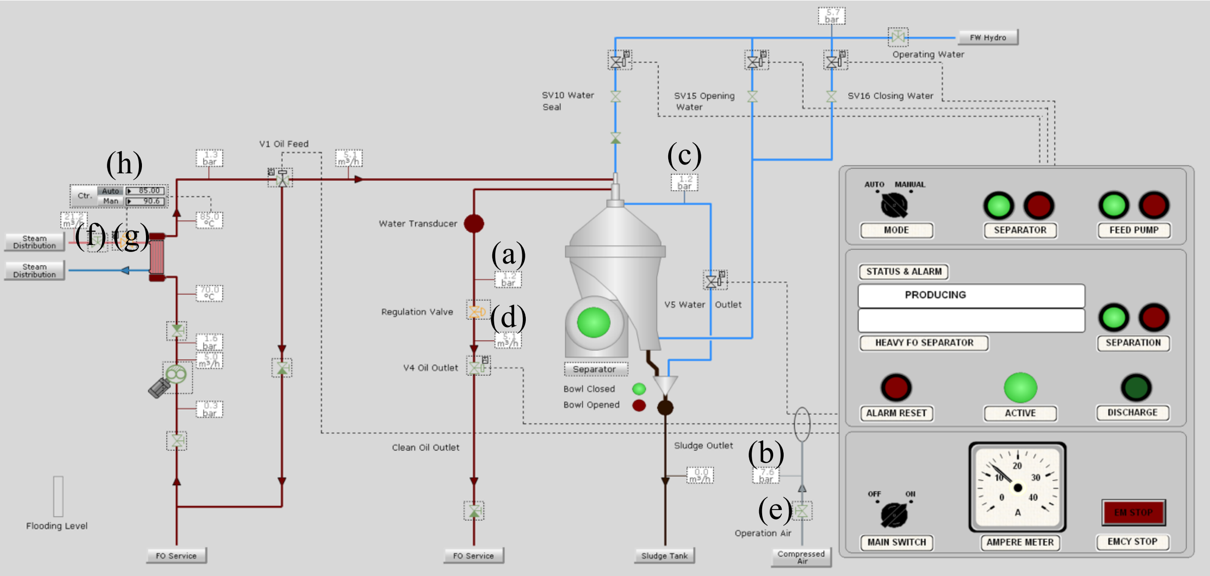


Figure 12: 2D process control interface. Missing elements in both environments; (a) back pressure gauge, (b) operating air pressure, (c) water outlet pressure, (d) clean oil outlet flow. In addition, the immersive virtual reality environment missed; (e) operating air inlet valve, (f) steam shut-off valve, (g) temperature-controlled steam valve, (h) feed temperature controller.

In the immersive virtual reality simulation, the environment was confined to one room with the fuel oil separator (Figure 11), a fresh water generator, a hydrophore unit and a ballast water transfer section. The participants could move freely in the immersed environment, only bounded by the physical walls of the virtual reality lab. Physical displacement within the virtual reality lab, enacted an equivalent movement in the virtual reality environment. All interaction with the simulator systems was administered through movement of the hand controls (Figure 13). As the simulator environment was larger than the lab, a locomotion technique called teleportation was used in addition to physical walking. The teleportation technique is a common function for distant movement in immersive virtual reality with head mounted displays, where the user project an illuminated beam with the hand controller towards the position one wants to teleport on to.

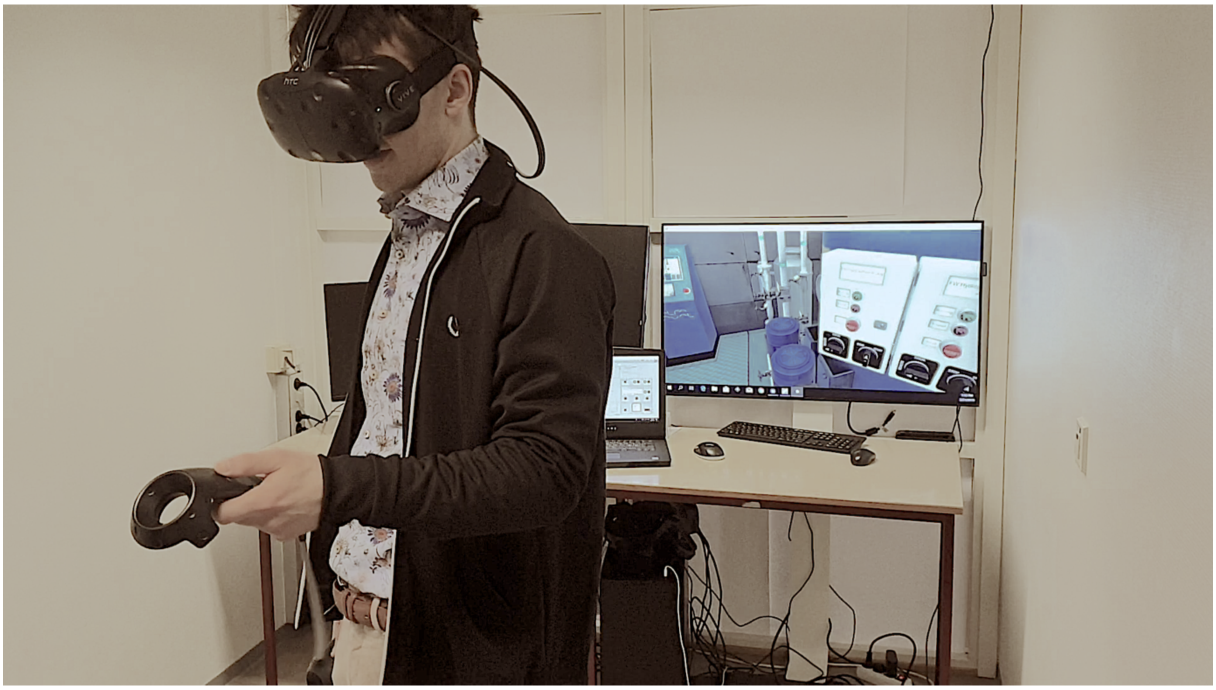


Figure 13: Pilot study participant engaged with the immersive virtual reality simulator after the trails

In the 3D virtual reality desktop simulation, the environment (Figure 14) comprised of the full engine room of the container vessel with all machinery systems operable. The participants sat in front of the desktop monitor, enacted movement in the environment and interacted with the simulator systems through the XBOX hand controller.

Figure 14: Fuel oil separator as viewed through the monitor of the 3D virtual reality desktop simulator

The task in the treatment was to conduct a starting operation of the fuel oil separator system from a shut-down condition with a simplified procedure (Table 7) created by the author. The system description with a flow chart diagram copied from the 2D process control interface, and the task description with the procedure was given on paper (Appendix E) for 10 minutes to the participants, to review and internalize before withdrawn again and commencement of the treatment began. Within this 10-minute review, the participants could ask the instructor to clarify eventual uncertainties found. The treatment was timed, observations on task sequencing and performance was noted by the instructor. If participants felt stuck between procedural steps or uncertain about system statuses which could not be read in the environment, they were allowed to ask the instructor for help which also was documented. These performance measures and observations was collected for future use, and not relevant for the hypotheses in this study.

Table 7: Experiment fuel oil separator start-up procedure (Appendix E)

	Task description
1	Switch on electricity and set local operating panel in Manual mode.
2	Line up all valves on the oil system. Open valves for heating steam, operating air and operating water.
3	Start oil feed pump and check that heat regulation and three-way oil feed return valve is ready.
4	Start separator. Monitor amperemeter during speed ramp up.
5	When amperemeter drop, switch local operating panel control to Auto.
6	Adjust throughput by throttling back pressure valve from fully open position and ensure correct production.

3.6 Measurement Instruments

3.6.1 Declarative knowledge

The pre-test consisted of an assessment of initial learning (Kraiger et al., 1993) through a recognition test developed by the author (Appendix G) to also disclose prerequisite system knowledge. On cognitive skill acquisition, Anderson (1982) states that at least 100 hours is required to gain any significant degree of proficiency, more than the students would have spent on this specific system but way less than time spent on learning and training with machinery systems in general.

In the pre-test, the separator system´s process flow chart was assigned 20 numbers to the system´s main elements, including elements which was visually missing from the simulator environments. A table with the label names of these 20 elements was included and the participants were asked to assign the correct element number from the included process flow chart to the corresponding label name of the table. The task was tailored to be at a difficult level, though without a time limit means of logic reasoning should provide an adequate score. To close any knowledge gap and mitigate the effect of individual intelligence aptitudes, feedback was given on the incorrect answers. This was an important design feature as identification and awareness of the main elements was considered essential for performance in the treatment and on the post-test.

The post-test consisted a memory power test created by the author to measure accuracy and accessibility of retaining knowledge acquired in the treatment (Appendix G). Focusing on the treatment rather than the individual, no time limit was set to ensure the test was measuring accuracy and not processing speed of mental computation. Accuracy of memory is a more valid construct to train and test when the consequences of error is high (Kraiger et al., 1993). Memory power tests usually focus on correct items answered and neglect the incorrect; to give the power of memory an additional descriptive dimension, incorrect items answered was also considered in accordance with Ackerman and Ellingsen (2016). The alternative, if not disregarded, would be to subtract the incorrect score from the correct score, and was not considered optional as this would dilute the construct investigated.

The post-test was on the same paper sheet as the pre-test; in the next column to the element´s name label, the participants was asked to mark off the elements they recall were missing in the simulator environment. As shown in Figure 12, the 3D virtual reality simulator had 4 elements missing and the immersive virtual reality simulator had 4 additional elements missing. Correct items answered gave a range of 4 and 8 respectively, and incorrect items answered a range of 16 and 12. Items failed to be identified were not considered a relevant measure as it is the antithetical value of correct items answered within the given range. Elements with practical function for the procedure, such as valves which where only missing in the immersive environments (Figure 12: e and f), was intentionally kept open in both conditions prior to the treatment to not induce a practical difference for the task design between the two simulators. No admonitory indication of the post-test was given prior to the treatment. The measurement score was aggregated with range of the immersive virtual reality simulator as index, i.e. the 3D desktop post-test scores was multiplied with 2.

3.6.2 Mental Workload

The Raw TLX questionnaire (Appendix G) from Byers et al. (1989) was used as is. Administration of the questionnaire followed the S FSS-2 form, and the participants was given with a brief explanation of instructions. The participants had to read the self-explanatory instructions on the questionnaire before answering and was given the opportunity to ask for clarifications. On the performance dimension the participants were made aware that the scale was inverted, if not explicitly ascertained by themselves.

3.6.3 The Flow State

The self-assessment is recommended to be collected within an hour after the activity (Jackson et al., 2010), and thus the S FSS-2 form (Appendix G) reproduced by the author was administered as close to the completion of the experiment as possible. The reproduction was an exact replicate of the original questionnaire of Jackson et al. (2010). The sum of the nine items on the five-point Likert scale are divided by 9 to produce the global SHORT Flow score, which can still be considered valid if one item is missing and the score is an average of the remaining eight. Administration of the questionnaire was given the participant following the declarative knowledge post-test, with a brief explanation of the instruction test. The participants had to read the self-explanatory instructions on the questionnaire before answering and was given the opportunity to ask for clarifications. Upon answering the transformation of time dimension the, participants were given their time performance after reflecting subjectively on their own time perception first.

3.6.4 Validity and Reliability

Content validity for the recognition pre-test is relatively strong through its face validity as it was constructed for the purpose of measuring declarative knowledge specific for a machinery system, but generic enough to represent any separator layout or make. The test was confined to 20 items which probes the main items of the main and auxiliary systems of the machinery. The test was credited construct validity by both experts in the pilot study as a difficult test encompassing the system, but fit to the purpose of measuring system knowledge without an aim of discriminating students from professionals. An observation of the pre-test incorrect answers show that most errors were due to conflation of similar label names, perhaps as the assigned item numbers and the label names were randomized intentionally.

This is considered to support empirical validity as incorrect answers were relatively evenly dispersed across the groups and varied between the individuals. Construct validity is also considered to be high as recognition tests are common in education for lower order declarative knowledge examination, though the design of the test is entirely the author´s.

The author constructed the memory power post-test and designed it to leverage the missing elements of the environments, ascertained in the pre-test and included in the treatment procedure, to give a measure of both accuracy and inaccuracy of memory. Designed with intention to measure accuracy and accessibility of retaining knowledge acquired in the treatment, face validity is argued to be present, though content validity can be refuted as elements included in the test was not excluded from the environments by the author himself. The test was credited by both experts in the pilot study to be very difficult as it required them to visualize the whole system and reenact the experience, item by item mentally. Though simple in design, the post-test could be subject to the influence of confounding factors as the treatment focused on executing the task and not memorizing the experience. On the empirical note, the two declarative knowledge tests could be argued to hold some predictive validity between them, considering the results of groups (*ii*) and (*iii*). Though the recognition pre-test scores as a prediction for the memory power test scores of groups (*ii*) and (*iii*) are purely incidental by design due to the feedback element of the pre-test. Like with recognition tests, the concept of power tests holds construct validity as they are recognized in theoretical frameworks in the fields of education and training, though rarely accompanied with an advocated design (Kraiger et al., 1993).

By administrating the NASA-RTLX questionnaire as is, the measurements hold the full construct validity of the original. Sampling validity is prone to error from the administration of the test by the author, and the fact that the sample frame was mainly recruited by convenience sampling can facilitate the grouping of measurements scores. The "raw" NASA-RTLX (Byers et al., 1989) as a simplified version of the NASA-TLX (Hart & Staveland, 1988), have been widely used and found to be highly correlated to the original framework with a reliability of 0.96-0.98 (Stanton et al., 2013).

The Flow State scale have been subject of validity studies by Jackson et al. (2010) and others for decades. The SHORT flow scales, and the S FSS-2 questionnaire used as an instrument for this study, was validated with a confirmatory factory analysis (Jackson et al., 2008), and the S FSS-2 was found to have a coefficient alpha estimate of reliability of 0.77-

0.78 (Jackson et al., 2010). By replicating the S FSS-2 questionnaire, the construct validity is considered maintained, though the full integrity can be argued against. The pre-test, the post-test and both questionnaires used in the study was given in English, and there was no control with the participants linguistic affluence of their second language. The pre-test and the post-test utilized specific terminology while the questionnaires were worded in their original form.

The declarative knowledge pre-test and post-test was considered reliable as the feedback would ensure an equal standard of knowledge before the treatment and the post-test. With subjective self-assessment however, perception is more prone to individual differences. As the grouped experience itself is under investigation, and not the task performance, the subjective measures did not need to be triangulated by equivalent objective instruments. Subjectivity have shown that groups of lower competence tend to overestimate themselves compared to groups of higher competence (Dunning, 2018), if a within-subject or repeated measure design was used, this would induce a necessity for additional triangulation by objective measures. As prosecution of collecting the data was according to the manuscript, no distortion of the measures is assumed, though there could be no control with the stability of the participants response with the between-subject design. Stability reliability of participant response can only be identified through repeated response. No repeated trails were conducted to test the reliability of the instruments in this experiment design, though the sample frame itself should provide representative reliability. Equivalence reliability was only optional for the flow state measures were the LONG flow state scale offers four correlated questions on each of the nine dimensions, the SHORT flow state scale was favoured with only one question per dimension due to the total endurance of the sampling.

3.7 Procedure

The experiments with the two novice groups was part of a joint data collection where they performed an additional task, as elaborated in the full manuscript (Appendix F). All the data collection for this study was prosecuted as summarized in Table 8, maintaining integrity of the design and the internal validity of the study. The expert group performed only the experiment of this study. In the novice groups experiments there were two researchers present, whereas the author functioned as the simulator instructor. In the expert group experiments the author was alone as instructor and the only researcher with the participants.

Table 8: Experiment procedure (appendix F)

Phase	Manuscript	Reference	Time
Welcome	Presenting research team, research project and the lab. Describing the experiment phases and timeline.		5 min.
Information and consent form	Explaining the purpose of informed consent, data storage and data protection rights according to the approval from the Norwegian Centre of Research Data. The participant reads the form before signing.	Appendix C	5 min.
Demographic questionnaire	Participant completes questionnaire.	Appendix D	5 min.
Familiarization with the simulator	Instruction on the hardware and interaction. The participant tests the simulator freely for 10 minutes.		15 min.
Pre-test	Declarative knowledge recognition test.	Appendix G	10 min.
Briefing of task	The task description sheet is given and explained.	Appendix E	10 min. absolute
Treatment	Participant conduct the task in the simulator.		10 min. approximately
Post-test and questionnaires	Declarative knowledge memory power test. S FSS-2 NASA-RTLX	Appendix G	10 min.
Finish	Experiment concluded.		Approximately 70 min. in total

3.8 Data Analysis

When choosing a statistical test for hypothesis testing, Hinton (2014) explain that if testing differences between conditions, the sample design must be discriminated as either independent and unrelated samples, or as paired samples from repeated measures. Next the samples must be tested for normality, to which parametric samples should use an independent *t* test if an independent sample design, and a paired samples *t* test if a paired samples design. The samples collected to be analysed has one independent variable and one dependent variable, and is found to have an unpaired sample design and thus are independent according to Table 9. Depending on the normality of the samples, the correct statistical test would be the independent *t* test or the Mann-Whitney *U* test. The independent *t* test holds the assumption of normal distribution, while the Mann-Whitney *U* test is a ranking test and does not.

Table 9: Choosing a statistical test adapted from Hinton (2014, p. 1)

Objective	**Design**	**Normality**	**Statistical test**
Look for differences between conditions where there is one independent variable, one dependent variable, and two conditions	Independent (unrelated samples)	Parametric	Independent *t* test
		Nonparametric	Mann-Whitney *U* test
	Paired samples (repeated measures or related samples)	Parametric	Paired samples *t* test
		Nonparametric	Wilcoxon signed-rank test

The data was analysed with IBM SPSS version 25. The collected data set was explored with descriptive statistics to identify normality and eventual outlier data points. As the samples sizes are considered to be low, the Shapiro-Wilk test is the preferred test of normality instead of the Kolmogorov-Smirnov test, which is better for larger sample sizes (Field, 2016). Some of the unpaired samples was found to be significantly non-normally distributed by the Shapiro-Wilk test (Appendix H: Table 13), and the non-parametric Mann-Whitney *U* test was chosen as the main null hypothesis significance test. The Mann-Whitney *U* test is a ranking test which is nearly as effective as the independent *t* test on normally distributed samples (Fay & Proschan, 2010), but does not depend on this assumption (Hinton, 2014).

The "Exact test" option in SPSS was included to supply the statistical Mann-Whitney U test with an additional significance test based on simulated data created from the normality of the original data points. As the sample sizes are relatively small, SPSS supply the statistical test with an simulation by the exact method instead of the Monte Carlo method (Field, 2009). In addition, an independent *t* test was conducted on samples found to be parametric by the Shapiro-Wilk test.

Effect sizes was calculated manually and classified according to Cohen (1992); $r=\frac{Z}{\sqrt{n}}$ was chosen for the Mann-Whitney *U* test (Field, 2009; Fritz, Morris, & Richler, 2012), and $d_Z=\frac{t}{\sqrt{n}}$ was chosen for the independent *t* test (Lakens, 2013), where r is a large effect size at 0,5 and d_z is a large effect size at 0,8 (Cohen, 1992).

By the interquartile method of outlier labelling (Hoaglin, Iglewicz, & Tukey, 1986), SPSS label data points outside 1.5 and 3.0 times the interquartile range. Five outlier data points was identified (Table 10) and as the post-test incorrect score was not a variable critical for hypothesis testing, these two data points was included. Reflecting on the method, Hoaglin and Iglewicz (1987) find that 2.2 would be the better criterion for excluding data points. None of the outliers influenced the group normality distribution (Appendix H: Table 13), and ranking tests are less distorted by heavy tails or skewedness than parametric tests (Fay & Proschan, 2010). With only 5 participants in the (*i*) 3D VR group, removing the 3.0 outlier data point could influence validity of the Mann-Whitney *U* test since the groups should be greater than 4 (Fay & Proschan, 2010). The group was also tested with the outlier removed and as no significant difference in the Mann-Whitney *U* test (Appendix H: Table 20) was found, all data points was included in the final data set for hypothesis testing.

Table 10: Outlier data points

Variable	**Data point group**	**Rule**	**Value**	**Inference**
Post-test incorrect	(*i*) 3D VR	3.0	-4	Included
	(*ii*) IVR	1.5	-4	Included
S FSS-2	(*i*) 3D VR	1.5	2.78	Included
NASA-RTLX	(*i*) 3D VR	3.0	26.67	Included
	(*i*) 3D VR	1.5	45.83	Included

4 Results

4.1 Declarative knowledge accuracy

4.1.1 Novice groups with different simulators

Hypothesis 1a predicted that declarative knowledge accuracy, by measurement of the memory power post-test, would be larger in the immersive virtual reality group (*ii*) than in the 3D virtual reality desktop group (*i*).

The descriptive statistics show a quite even prerequisite knowledge from the recognition pre-test, where the groups score (*i*) $\bar{x}=19.00$, Mdn=20.00, SEM=0.63 and (*ii*) $\bar{x}=18.17$, Mdn=18.00, SEM=0.65 respectively.

The post-test correct scores resulted as predicted were the 3D virtual reality desktop group (*i*) $\bar{x}=0.80$, Mdn=0.78, SEM=0.49 scored lower that the immersive virtual reality group (*ii*) $\bar{x}=5.33$, Mdn=5.00, SEM=0.42, and the Mann-Whitney U test (Appendix H: Table 14) had a difference in medians $U=0$, $Z=-2.796$, $P=0.005$, $r=-0.843$.

The post-test incorrect scores gave a Mann-Whitney U test with difference in medians $U=7.5$, $Z=-1.447$, $P=0.148$, $r=-0.436$.

By these results the post-test correct scores are significantly different with a large effect size. The post-test incorrect score is insignificant and with a medium effect size.

4.1.2 Expert and novice groups with the immersive virtual reality simulator

Hypothesis 1b predicted that declarative knowledge accuracy, by measurement of the memory power post-test, would with the immersive virtual reality simulator be larger in the expert group (*iii*) than in the novice group (*ii*).

The descriptive statistics show slightly more confident prerequisite knowledge from the recognition pre-test in favour of the (*iii*) expert group, where the groups score (*ii*) $\bar{x}=18.17$, Mdn=18.00, SEM=0.65 and (*iii*) $\bar{x}=19.50$, Mdn=20.00, SEM=0.34 respectively.

The post-test correct scores resulted as predicted were the immersive virtual reality group (*ii*) $\bar{x}=5.33$, Mdn=5.00, SEM=0.42 scored lower that the immersive virtual reality group (*iii*) $\bar{x}=7.50$, Mdn=8.00, SEM=0.34, and the Mann-Whitney U test (Appendix H: Table 15) had a difference in medians $U=2$, $Z=-2.636$, $P=0.008$, $r=-0.761$.

The post-test incorrect scores gave the Mann-Whitney *U* test with difference in medians *U*=17.5, Z=-0.082, *P*=0.934, r=-0.024.

By these results the post-test correct scores are significantly different with a large effect size. The post-test incorrect score is insignificant and with a small effect size.

4.2 Novice groups experience with different simulators

4.2.1 Mental workload with different simulators

Hypothesis 2a predicted that the experienced mental workload would be larger in the 3D virtual reality desktop group (*i*) than in the immersive virtual reality group (*ii*). The descriptive statistics show the global NASA-RTLX scores respectively (*i*) $\bar{x}$=37.33, Mdn=37.50, SEM=3.11 and (*ii*) $\bar{x}$=42.77, Mdn=40.83, SEM=5.33.

The Mann-Whitney *U* test (Appendix H: Table 16) gave a difference in medians *U*=13.5, Z= -0.275, *P*=0.783, r=-0.082.

As these samples are normally distributed according to the Shapiro-Wilk test (Appendix H: Table 13), an independent *t* test was also performed (Appendix H: Table 17) with *t*=-0.834, df=9, *P*=0.426, difference in means -5.444, 95% CI -20.211 to 9.322, d_z=-0.251.

The result point to a difference without statistical significance and a small effect size in the opposite direction than predicted in the hypothesis.

4.2.2 The Flow state with different simulators

Hypothesis 2b predicted that experienced flow state would be more present in the immersive virtual reality group (*ii*) than in the 3D virtual reality desktop group (*i*). The descriptive statistics show the Flow state global scores respectively (*i*) $\bar{x}$=3.49, Mdn=3.55, SEM=0.20 and (*ii*) $\bar{x}$=3.63, Mdn=3.61, SEM=0.17.

The Mann-Whitney *U* test (Appendix H: Table 16) gave a difference in medians *U*=13, Z= -0.367, *P*=0.713, r=-0.110.

As these samples are normally distributed according to the Shapiro-Wilk test (Appendix H: Table 13), an independent t test was also performed (Appendix H Table 17) with t=-0.541, df=9, P=0.602, difference in means -0.140, 95% CI -0.729 to 0.447, d_z=-0.163.

The result point to a difference without statistical significance and a small effect size in the direction predicted in the hypothesis.

4.3 Expert and novice groups experience with immersive virtual reality

4.3.1 Mental workload with the immersive virtual reality simulator

Hypothesis 3a predicts that the novice immersive virtual reality group (*ii*) would experience larger mental workload than the expert immersive virtual reality group (*iii*). The descriptive statistics show the global NASA-RTLX scores respectively (*ii*) $\bar{x}$=42.77, Mdn=40.83, SEM=5.33 and (*iii*) $\bar{x}$=29.72, Mdn=30.83, SEM=4.84.

The Mann-Whitney U test (Appendix H: Table 18) and gave a difference in medians U=8.5, Z=-1.526, P=0.127, r=-0.440.

As these samples are normally distributed according to the Shapiro-Wilk test (Appendix H: Table 13), an independent t test was also performed (Appendix H: Table 19) with t=1.813, df=10, P=0.100, difference in means 13.056, 95% CI -2.911 to 29.104, d_z=0.523.

The result point to a difference without statistical significance and with a medium effect size in the direction predicted in the hypothesis.

4.3.2 The Flow state with the immersive virtual reality simulator

Hypothesis 3b predicts that the expert immersive virtual reality group (*iii*) would have a larger presence of the flow state than the novice immersive virtual reality group (*ii*). The descriptive statistics show the global S FSS-2 scores respectively (*ii*) $\bar{x}$=3.63, Mdn=3.61, SEM=0.17 and (*iii*) $\bar{x}$=4.04, Mdn=4.05, SEM=0.22.

The Mann-Whitney U test (Appendix H: Table 18) and gave a difference in medians U=10, Z=-1.285, P=0.198, r=-0.371.

As these samples are normally distributed according to the Shapiro-Wilk test (Appendix H: Table 13), an independent t test was also performed (Appendix H: Table 19) with *t*=-1.462, df=10, *P*=0.174, difference in means -0.407, 95% CI -1.028 to 0.213, d_z=-0.422.

The result point to a difference without statistical significance and with a medium effect size in the direction predicted in the hypothesis.

4.4 Summary

The data analysis offered no means of triangulation beyond the additional *t* test of the parametric samples. The inferential statistics was found to be significant, and thus infers acceptance of hypothesis 1a and 1b. For the retaining hypotheses, the effects found are not significant and the hypotheses have to be rejected as summarized in table 11.

Table 11: Hypotheses inferences

Hypothesis	**Prediction**	**Significance**	**Effect size**	**H_1**	**H_0**
1a	Power of memory IVR > 3D VR	$P_{(U)}$=0.005	r=-0.843	Accepted	Rejected
1b	Power of memory Expert > Novice	$P_{(U)}$=0.008	r=-0.761	Accepted	Rejected
2a	Experienced MWL IVR < 3D VR	$P_{(U)}$=0.783	r=-0.082	Rejected	Accepted
		$P_{(t)}$=0.426	d_z=-0.251		
2b	Experienced flow IVR > 3D VR	$P_{(U)}$=0.713	r=-0.110	Rejected	Accepted
		$P_{(t)}$=0.602	d_z=-0.163		
3a	Experienced MWL Expert < Novice	$P_{(U)}$=0.127	r=-0.440	Rejected	Accepted
		$P_{(t)}$=0.100	d_z=0.523		
3b	Experienced flow Expert > Novice	$P_{(U)}$=0.198	r=-0.371	Rejected	Accepted
		$P_{(t)}$=0.174	d_z=-0.422		

5 Discussion

The purpose of this study was to investigate state-of-the-art technology in the context of the present marine engineering education. This chapter reflects on the theory, methods and results related to the hypotheses, before addressing experienced limitations and future research.

5.1 Declarative knowledge

The pre-test was a recognition test on identification of system items. As the pre-test scores were relatively even between the groups, this strengthens the post-test results, but might induce a question of necessity regarding the pre-test feedback element in the experiment design. If given a larger sample frame one might not need the feedback element, though it was highly uncertain at what level the participants would score which would be perilous to the pre-test. While designing the experiment, the pre-test was expected to give a larger margin of error, rendering the feedback element necessary for a standardized commencement of the treatment. Regardless of the feedback on the pre-test, the scores of groups (*ii*) and (*iii*) could hold a prediction of consecutive learning outcomes measured on the post-test, or simply be two instruments that capture some of the same underlying cognitive constructs which cofounding factor has a larger presence in the latter group.

The two types of simulators show an effect of different knowledge acquisition with the same population and the same knowledge base. Immersion has shown to be a positive factor for knowledge acquisition (Webster, 2016), and the hypothesis 1a might hold evidence accordingly. The difference found between the simulators might descend from the wholeness of their environments, whereas the 3D virtual reality desktop simulator is encompassing and the immersive virtual reality simulator is relatively confined. Observations during the novice groups experiments led the author to note an incidental tendency to digress from the task in both simulators. As the encompassing simulator hold a precedented amount of details and other systems, it is easy to consider the effect of seductive details (Towler & Kraiger, 2008) as an influencing factor on the lower score of the (*i*) 3D virtual reality desktop group. It is likely to believe that the two different environments, or their means of interaction, require or facilitate a different level of mental computation. One observation that is difficult to leave unmentioned; all (*i*) 3D virtual reality desktop group participants that scored 0 on the post-test forfeited the attempt to recall the experience effortlessly.

Mental computation and it´s bottle-neck within the working memory was an assumption for hypothesis 1b, whereas the expert (*iii*) immersive virtual reality group brought their procedural and tacit knowledge to the comparison of group levels. Though all expert participants had practical experience with fuel oil separator systems, for most of them this experience was and felt distant of age. Working memory surely seems spacious to (*iii*) immersive virtual reality expert group compared to the (*ii*) immersive virtual reality novice group, by the memory power post-test. Reflecting elusively, the memory power post-test might also capture the presence of a broader spectre of cognitive constructs, with higher order learning constructs such as mental models and metacognition (Kraiger et al., 1993) being present with the expert group. Regardless, the expert group results can display a benchmark for the task design and the training of the students enrolled in the marine engineering programme.

Accepting hypotheses 1a acknowledges the prototype immersive virtual reality simulator as superior to the commercial 3D virtual reality desktop simulator for the novice groups in this treatment, statutory to the design of this research study. Accepting hypothesis 1b acknowledges that there is a skill gap between novices and experts that can be identified with the immersive virtual reality simulator; and if it can be identified, it can be approximated by training in the education programme.

The post-test incorrect scores from the memory power test show that with a greater correct score performance, the error element of memory also increase (Figure 15). This effect between the two simulators could simply be due to the lack of effort to response from the (*i*) 3D virtual reality desktop group on the post-test. Interestingly however, this error element of human cognition seems static through the professional experience and age that discriminated the two immersive virtual reality groups. If true, this would be an important factor to consider in training design and competence assessment at all complexity levels and professional stages, as a margin of error could be expected, at least on the first training session.

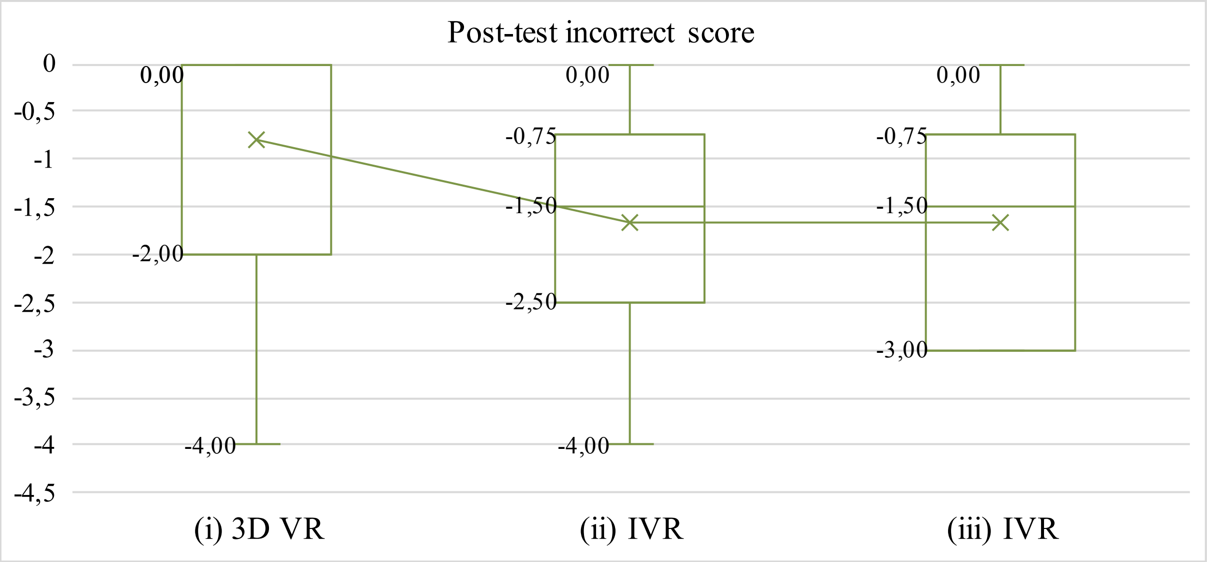


Figure 15: Post-test incorrect score

The first research question investigated in this study was if immersive virtual reality simulator training could be the better technological development for training declarative knowledge in maritime education, and for expert competence maintenance. As far as this study can disclose, there are upsides to implementing immersive virtual reality in the educational programmes. An appropriately designed simulator and training programme should achieve to supplement students with facets of knowledge and skills the present 2D desktop simulators and the 3D virtual reality desktop simulators cannot offer, at least with further development of the design in this study. If supported in further developed training designs, the author would argue the immersive virtual reality simulator to be evocative, if not evident, positioned interim both the 2D desktop and the 3D virtual reality desktop experience, and the real-life experience. For professional marine engineers, immersive virtual reality could be used to recap and refresh ancient knowledge and skills, though purely exploratory designs without a challenging task complexity or objective might not endure their interest for long. For both extremes of the profession and it´s intermediaries, immersive virtual reality simulator exercises designed with a wide range of complexity options to draw from, could hold some commercial interest to professional and private dissemination.

5.2 Mental workload

With no significant effect of the global score evident between simulators, one might say the two simulator exercises were equally difficult. The presence of mental workload is evident, though not with saliency able to discriminate the two simulators. Hoping to disclose and describe ranges of mental workload that would be significant for learning, is outside the gains of this study. The small effect of higher mental workload with the immersive virtual reality simulator, contrary to prediction, might indicate that the environments are differently absorbed by the users, and perhaps better captured through instruments of other constructs.

The medium size effect found between the novice (*ii*) immersive virtual reality group and the expert (*iii*) immersive virtual reality group show that there is a difference, not yet significant. Although some of the novice participants disrupted their task sequencing with divagating focus to other elements of the environment. This observation is argued to be in the search for information and is more similar to task disengagement due to high mental workload as found by Warm et al. (2008), than due to the phenomena connected with low mental workload; degradation of vigilance. A majority of the expert participants explicitly stated that the experience was challenging, enjoyable and not difficult, supporting their low mental workload score.

For this exercise design the expert scores state a lower benchmark for the student's education to aim for. Mental workload can affect memory and cause lapses in working memory, even with experienced operators (Endsley, 1995). The high mental workload of the two novice groups, depicted by the mental demand dimension in Figure 16, might have had an influence on their performance on the memory power test, however it does not provide any clarity to the difference between the two simulators. With repeated training and consecutive real-life skill transfer, the concept of mental workload might give a better description and guidance of training design and complexity for the better learning outcome and training effectiveness. The size or design of this research might simply not be worthy of a significant result in terms of mental workload, but may display a presence that can be investigated further.

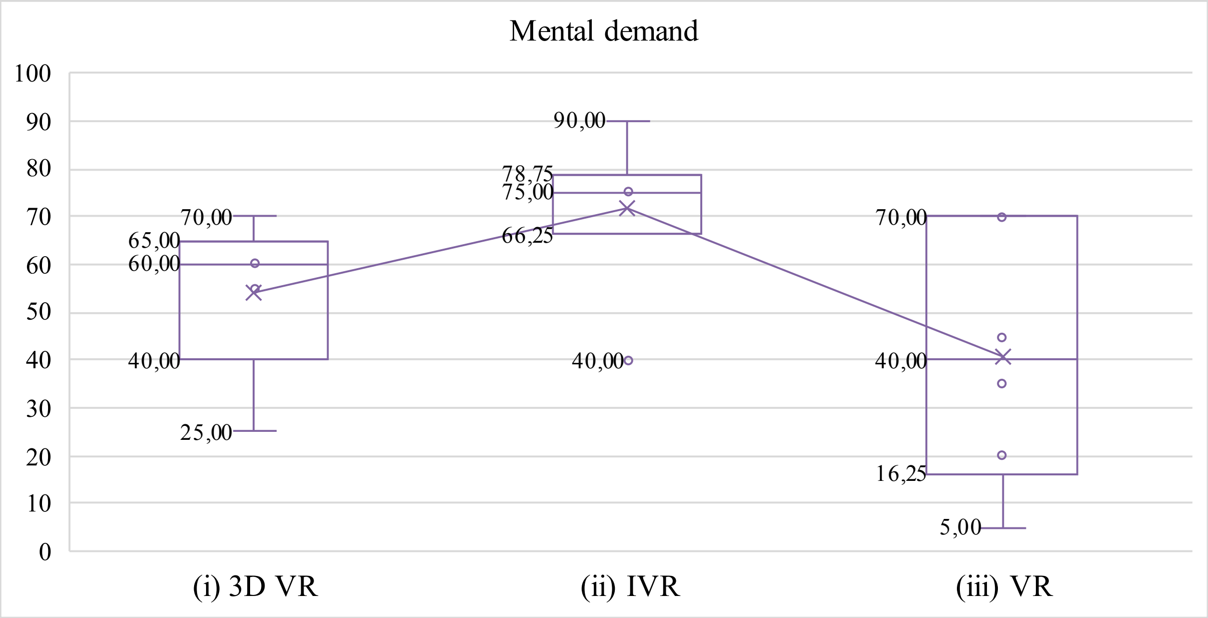


Figure 16: Mental demand

5.3 The Flow state

Experienced flow shows no sizable effect between the two simulators, again supporting that the exercise was perceived as equally difficult with both simulator interfaces. As with mental workload, no objective guidance exists to categorize if these scores are evidence that the flow state was present or not. It is still assumed the novice groups experienced some state of the construct, and it is argued to be present in the expert group.

The medium size effect between the novice (*ii*) immersive virtual reality group and the expert (*iii*) immersive virtual reality group show that there is a difference, though not significant. The higher global flow state score of the expert group indicate that the experience was perceived more strenuous to the novices. This effect is supporting the prediction that a high flow state renders low mental workload according to the studies of Lackey et al. (2016), and to the studies of Hamari et al. (2016) where a high flow state is positive for learning.

The second research question investigated in this study was if there would be observable differences in mental workload and flow state between the two simulators, and between the two group levels. This study finds an effect of both constructs between the novice group and the expert group, while no notable effect is found between the two simulators. Not surprisingly, this provide some indication that there are cognitive differences with seniority to be identified. The novices described the two environments slightly different; the immersive as more intense and with a closer to real-life sense of presence, and the non-immersive as highly detailed with a close to real-life replication of systems. As only two of the students had experience from a real engine room and none of them had actually operated fuel oil separator machinery, both simulators were credited as valuable to their experience as an epiphany of how the specific machinery would be presented in real life. The close to equal scores between the simulators help interpret the exercise design as applicable for both simulators, and complex enough for the two novice groups to focus on the task and not be too influenced by the different means of interaction.

5.4 Limitations

The novice groups were recruited conveniently, but without any omission bias as the whole population of the 2nd year class volunteered and all were included. Purposely recruiting a generalizable cross section of the professional population might be practically impossible, but the recruitment criteria for the group was satisfied. Though not a generalizable cross-section, the author argues that the educational standard should be measured against the higher end of professional competence at concrete proficiencies such as specific technical tasks, and perhaps more leniently measured against the general population on abstract proficiencies such as higher order learning outcomes which would be developed through practical application.

Some limitations to the experimental design are inevitable. As the two novice groups experiments were administered by two researchers and the expert group experiments only had the one researcher present, the latter is more prone to the possibility of experimenter bias. Researchers in pairs might exhibit a better self-conscious behaviour than the sole researcher, and especially a less seasoned one cannot prosecute an experiment while being present without some countertransference occurring. Prosecution according to the scrip was maintained to not induce the experiments with error from researcher behaviour, at least not consciously.

If the measurements of the study could fit into a within-subject design, there would be more statistical power to the data. Between-subjects design seemed initially to be the best fit for competence level comparison, and the design offer a probe for further repeated measures. As a within-subject design induce a carry-over effect of learning, this complicates the frame of the declarative knowledge tests and other machinery units of the simulators might have to be included, prolonging the experiments of the study.

With the given sample frame, no other standard design could utilize the participants in the same manner with this scope. A classic experimental design hold good internal validity, but suffer less external validity due to the pre-test element and its influence on the post-test (Frankfort-Nachmias, 2015). This study design balances both internal validity and external validity, where the simulator comparisons between the two novice groups favours external validity through generalizability to the population of marine engineering education, and the group level comparison of the immersive virtual reality simulator favour internal validity through the causal relationship of professional experience to memory power.

If a larger sample frame was obtainable post-test only groups could be included for greater external validity and generalizability, however, if given this option the researcher would rather strengthen the groups sizes. A within-subject design could have been used, if altering the scope to utilize the objective performance measures and including the familiarization session in the pre-test.

All the six dimensions of the NASA-TLX and the nine dimensions of the S FSS-2 was included in the global score. Although some dimensions such as Physical Demand could have been excluded from the global scores, the integrity of the instruments was more important than altering the second research question. Exploring each dimension exhaustively fell beyond the magnitude of this study.

Opposed to a pure clinical setting, responder bias might occur more freely with researchers present. Acquiescence bias occur when the participants responds lenient towards an expectancy, self-constructed or admonitory. To what effect the researcher supervision influences the experience and response is hard to depict with this design and these instruments. Not unique to this study alone, knowing one is monitored seems uncannier when not in control of one´s surroundings as indicated by the loss of self-consciousness dimension of the flow state (Figure 17). Due to necessary assistance with the equipment and safety supervision, remote monitoring of the experiments was not conveniently possible.

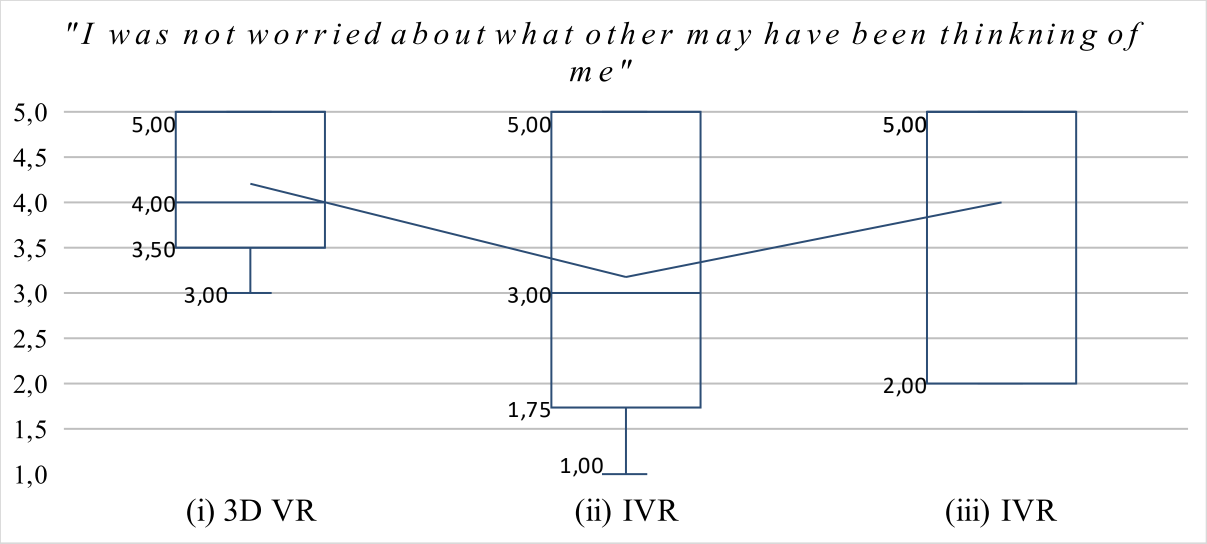


Figure 17: Loss of self-consciousness

5.5 Further research and future projection

For the benefit of marine engineer education, either if not both simulators, should be implemented for research and further developed. Further developments of the two simulators demand development of task designs that might enhance learning outcomes with both interfaces. As both simulator technologies are relatively unexplored in marine engineering education, there is a lot of uncovered ground for research. In general, for the maritime industry and the field of maritime education, areas for further research and development could be as elaborated in Table 12.

At the submission date of this thesis, a conference paper (Appendix I) based on this thesis is pending acceptance to the ErgoShip 2019 conference by Western Norway University of Applied Sciences in collaboration with World Maritime University.

This study has obtained a comprehensive data set which will be further investigated. Given the medium effect between the two groups with the immersive virtual reality simulator, there are fifteen dimensions with the NASA-RTLX and the S FSS-2 that might hold some interesting information paired or independently. If the global scores captured cannot find covariation between the two constructs, some of their dimensions might. Objective performance measures collected such as engagement time to complete the task can also help to describe training effectiveness of the captured learning outcomes.

Table 12: Further areas to research

Training design	Development of simulator training programmes tailored with incremental complexity (Hjelmervik et al., 2018). Training topics could be extended beyond specific machinery procedures and aim for higher learning outcomes through Resource management, safety training scenarios, team exercises, etc.
Training effectiveness	Develop a repeated measure designs applicable with all simulator types and assess with real-life skill transfer. In transition to adopting the newer simulators, repeated measures with several cohorts can be conducted to compare todays simulators with immersive virtual reality. Training designs such as the three-stage training syllabus of Nazir, Øvergård, and Yang (2015) can be utilized with real-life skill transfer evaluation to assess both design and different simulator technology.
Assessment methods	Developing an assessment scheme according to STCW regulations for 3D virtual reality and immersive virtual reality. Perhaps a model-building study that coincides established assessment practices with formal requirements in the context of immersive virtual reality. Assessment methods for new simulator technology could be developed for new performance indicators (Ernstsen & Nazir, 2018).
Knowledge evaluation	The operational maritime industry relies immensely on higher order constructs of verbal knowledge, that can be hard to capture with lower order declarative knowledge test. To supplement the task specific “show and tell” competence assessment with more concrete knowledge test and means of evaluation, model-building studies could be directed towards developing tailored knowledge test for the maritime industry to capture higher order and abstract learning outcomes.
Commercialization	Dissemination of the immersive virtual reality simulator to the professional field of marine engineering. Simulator development with a commercial goal of addressing professional officers and shipping companies. Environments can be developed for competence maintenance and knowledge acquisition in fields undiscovered or novel to the individual or for inhouse company training. The target can be both private and company use.
Implementation	When developed further to the satisfaction of the DNVGL ST-0033 standard the immersive virtual reality simulator can be implemented in the education programme to train concrete learning outcomes by tailored exercises.
Integration	Further development of the simulator software should allow for integration with the other products of the K-SIM family, to the extent that one can run team exercises where the immersive virtual reality simulator is integrated with a Full Mission engine simulator and a bridge simulator.

6 Conclusion

With the present state of the Norwegian marine engineer education, the immersive virtual reality simulator suggests a precedented learning outcome in terms of accuracy and accessibility of retaining knowledge acquired, in contrast to the 3D virtual reality simulator. The accomplished simulator technology also identifies the affluence of competence seniority, though the prosperity of sagaciousness provides no impunity from marginal error. Not abound with saliency to discriminate the two technologies, the constructs of mental workload and the flow state is indicated to be present by some extent. Medium size effects of these constructs esteem the professional marine engineering officers vocated to share their competence through the immersive virtual reality simulator.

As a prototype, the immersive virtual reality simulator might require further development before being made commercially available to the field of maritime education. This study is positive that there are different learning outcomes obtainable with both simulators that can be disclosed, and provides new insights into areas of research for the field of both simulator technology and maritime education.

7 References

Ackerman, P. L. (2014). Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence, 45*(1), 6-17. doi:10.1016/j.intell.2013.04.009

Ackerman, P. L., & Ellingsen, V. J. (2016). Speed and accuracy indicators of test performance under different instructional conditions: Intelligence correlates. *Intelligence, 56*, 1-9. doi:10.1016/j.intell.2016.02.004

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*(4), 369-406. doi:10.1037/0033-295X.89.4.369

Bloom, B. S. (1956). *Taxonomy of educational objectives: the classification of educational goals: handbook i* (Vol. 16).

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). *Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? In A* (Vol. 1).

Chisholm, C. D., Weaver, C. S., Whenmouth, L., & Giles, B. (2011). A Task Analysis of Emergency Physician Activities in Academic and Community Settings. *Annals of Emergency Medicine, 58*(2), 117-122. doi:10.1016/j.annemergmed.2010.11.026

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety* (25th anniversary ed. ed.). San Francisco: Jossey-Bass Publishers.

Csikszentmihalyi, M. (1990). *Flow : the psychology of optimal experience*. New York: HarperPerennial.

Csikszentmihalyi, M., & Larson, R. (1987). *Validity and Reliability of the Experience-Sampling Method* (Vol. 175).

Cullen, R. H., Rogers, W. A., & Fisk, A. D. (2013). Human performance in a multiple-task environment: Effects of automation reliability on visual attention allocation. *Applied Ergonomics, 44*(6), 962-968. doi:10.1016/j.apergo.2013.02.010

DNV GL rules for classification: Ships (RU-SHIP), § Chapter 2, Section 2 (2018).

DNVGL-ST-0033 - Maritime Simulator Systems, (2018).

Dunning, D. (2018). The best option illusion in self and social assessment. *Self and Identity*, 1-14. doi:10.1080/15298868.2018.1465460

Endsley, M. R. (1995). A Taxonomy of Situation Awareness Errors. *Human Factors in Aviation Operations*, 287-292.

Endsley, M. R. (2012). *Designing for situation awareness : an approach to user-centered design* (2nd ed. ed.). Boca Raton, Fla: CRC Press.

Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned From Human–Automation Research. *Human Factors: The Journal of Human Factors and Ergonomics Society, 59*(1), 5-27. doi:10.1177/0018720816681350

Endsley, M. R. (2018). Level of Automation Forms a Key Aspect of Autonomy Design. *Journal of Cognitive Engineering and Decision Making, 12*(1), 29-34. doi:10.1177/1555343417723432

Ernstsen, J., & Nazir, S. (2018). Consistency in the development of performance assessment methods in the maritime domain. *The international Journal for professionals in maritime administration, industry and education, 17*(1), 71-90. doi:10.1007/s13437-018-0136-5

Fay, M. P., & Proschan, M. A. (2010). Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statist. Surv. 4, (2010)*(1-39), 1-39. doi:10.1214/09-SS051

Field, A. (2009). *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll)* (3rd ed. ed.). Los Angeles: SAGE.

Field, A. (2016). *An adventure in statistics : the reality enigma*. Los Angeles: Sage.

Flynn, J. R. (2013). The "Flynn Effect" and Flynn's paradox. *Intelligence, 41*(6), 851-857. doi:10.1016/j.intell.2013.06.014

Flynn, J. R. (2018). Reflections about intelligence over 40 years. *Intelligence, 70*, 73-83. doi:10.1016/j.intell.2018.06.007

Frankfort-Nachmias, C. (2015). *Research methods in the social sciences* (8th ed. ed.). New York, NY: Worth publishers.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect Size Estimates: Current Use, Calculations, and Interpretation. *Journal of Experimental Psychology: General, 141*(1), 2-18. doi:10.1037/a0024338

Gagné, R. M. (1984). *Learning outcomes and their effects: Useful categories of human performance* (Vol. 39).

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming, 33*(4), 441-467. doi:10.1177/1046878102238607

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*(C), 170-179. doi:10.1016/j.chb.2015.07.045

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later (pp. 904-908).

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183): North-Holland.

Hendy, K. C. (1996). *Situation Awareness and Workload: Birds of a Feather?*

Hinton, P. R. (2014). *Statistics explained* (3rd ed. ed.). London: Routledge.

Hjelmervik, K., Nazir, S., & Myhrvold, A. (2018). Simulator training for maritime complex tasks: an experimental study. *The international Journal for professionals in maritime administration, industry and education, 17*(1), 17-30. doi:10.1007/s13437-017-0133-0

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association, 82*(400), 1147-1149. doi:10.1080/01621459.1987.10478551

Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association, 81*(396), 991-999. doi:10.1080/01621459.1986.10478363

International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW) : including 2010 Manila amendments ; STCW Convention and STCW Code, (2016).

ISO 10075-1:2017 Ergonomic principles related to mental workload, (2017).

Jackson, S. A., Eklund, B., & Martin, A. (2010). *The flow manual : the manual for the flow scales : manual, sampler set.* Menlo Park, Calif: Mind Garden.

Jackson, S. A., & Marsh, H. W. (1996). Development and validation of a scale to measure optimal experience: the flow state scale. *Journal of sport & exercise psychology, 18*(1), 17. doi:10.1123/jsep.18.1.17

Jackson, S. A., Martin, A. J., & Eklund, R. C. (2008). Long and short measures of flow: the construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of sport & exercise psychology, 30*(5), 561. doi:10.1123/jsep.30.5.561

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *The Official Journal of the IFIP Technical Committee on Education, 23*(4), 1515-1529. doi:10.1007/s10639-017-9676-0

Kongsberg Maritime. (2018). K-Sim Engine MaK 8M43C Container M11. Retrieved from https://www.kongsberg.com/globalassets/digital/maritime-simulation/k-sim-engine/docs/medium-speed-engine-datasheets/k-sim-engine_mak_8m43c_container_m11_model_datasheet_2018_lr.pdf?_t_id=1B2M2Y8AsgTpgAmY7PhCfg%3d%3d&_t_q=k%5c-sim+vr&_t_tags=language%3aen%2csiteid%3a24c9be7d-c7a0-47ff-9aff-d09ef8b15bbc&_t_ip=81.166.27.147&_t_hit.id=KongsbergMaritime_Web_Models_Media_Document/_cca5eddf-e6b2-4550-ba13-6afc0b2aa010&_t_hit.pos=2

Kongsberg Maritime. (2019a). K-Sim Safety - Kongsberg Gruppen. Retrieved from https://kongsberg.com/en/kongsberg-digital/maritime%20simulation/k-sim%20safety/

Kongsberg Maritime. (2019b). Kongsberg Maritime History. Retrieved from https://www.km.kongsberg.com/ks/web/nokbg0237.nsf/AllWeb/03E4EADBA358FC78C1256A55004648C0?OpenDocument

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of Cognitive, Skill-Based, and Affective Theories of Learning Outcomes to New Methods of Training Evaluation. *Jurnal of applied psychology, 78*(2), 311-328. doi:10.1037/0021-9010.78.2.311

Lackey, S. J., Salcedo, J. N., Szalma, J. L., & Hancock, P. A. (2016). The stress and workload of virtual reality training: the effects of presence, immersion and flow. *Ergonomics, 59*(8), 1060-1072. doi:10.1080/00140139.2015.1122234

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science : a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12.

Mallam, S. C., & Lundh, M. (2013). Ship Engine Control Room Design: Analysis of Current Human Factors & Ergonomics Regulations & Future Directions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57*(1), 521-525. doi:10.1177/1541931213571112

Mallam, S. C., Nazir, S., Sharma, A., & Veie, S. (2019). *Perspectives on Autonomy – Exploring Future Applications and Implications for Safety Critical Domains*, Cham.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education, 44*(6), 543.

Microsoft. (2019). Wireless XBOX Controller. Retrieved from https://www.microsoft.com/nb-no/p/tradl%C3%B8s-xbox-kontroller-phantom-black-special-edition/8pn2zfgvdh4r/1sss?cid=msft_web_collection&activetab=pivot%3aoverviewtab

Nazir, S., Øvergård, K. I., & Yang, Z. (2015). Towards Effective Training for Process and Maritime Industries. *Procedia Manufacturing, 3*, 1519-1526. doi:10.1016/j.promfg.2015.07.409

Norman, D. A. (2013). The design of everyday things (Rev. and exp. ed. ed.). New York: Basic Books.

Norman, D. A., & Stappers, P. J. (2015). DesignX: Complex Sociotechnical Systems. *She Ji: The Journal of Design, Economics, and Innovation, 1*(2), 83-106. doi:10.1016/j.sheji.2016.01.002

Parasuraman, R., & Hancock, P. A. (2008). Mitigating the adverse effects of workload, stress, and fatigue with adaptive automation. In P. A. H. a. J. L. Szalma (Ed.), *Performance under Stress* (pp. 45-57): Ashgate.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160. doi:10.1518/155534308X284417

Passig, D. (2015). Revisiting the Flynn Effect through 3D Immersive Virtual Reality (IVR). *Computers & Education, 88*, 327-342. doi:10.1016/j.compedu.2015.05.008

Reason, J. (2000). Human error: models and management. *The Western Journal of Medicine, 172*(6), 393.

Scholtz, J., Lewis, M., Fong, T., Steinfeld, A., Kaber, D., Schultz, A., & Goodrich, M. (2006). Common Metrics for Human-Robot Interaction: Source of Acquisition

NASA Armes Research Center

Sellberg, C. (2017). Representing and enacting movement: The body as an instructional resource in a simulator-based environment. *Education and Information Technologies, 22*(5), 2311-2332. doi:10.1007/s10639-016-9546-1

Shen, H., Zhang, J., Yang, B., & Jia, B. (2019). Development of an educational virtual reality training system for marine engineers. *Computer Applications in Engineering Education*. doi:10.1002/cae.22099

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). *Human factors methods : a practical guide for engineering and design* (2nd ed. ed.). Farnham: Ashgate.

Statistisk Sentralbyrå. (2019). Fakta om befolkningen. Retrieved from https://www.ssb.no/befolkning/faktaside/befolkningen#faktaside-sticky-menu

Stowers, K., Oglesby, J., Sonesh, S., Leyva, K., Iwig, C., & Salas, E. (2017). A Framework to Guide the Assessment of Human-Machine Systems. *Human factors, 59*(2), 172-188. doi:10.1177/0018720817695077

Towler, A., & Kraiger, K. (2008). The Effects of Seductive Details on Recognition Tests and Transfer Tasks. In I. N. Depaul Univ Greencastle (Ed.).

VIVE. (2019). VIVE VR System. Retrieved from https://www.vive.com/eu/product/#vive-spec

Wagner, R. K. (1987). Tacit Knowledge in Everyday Intelligent Behavior. *Journal of Personality and Social Psychology, 52*(6), 1236-1247. doi:10.1037/0022-3514.52.6.1236

Warm, J. S., Matthews, G., & Finomore, V. S. J. (2008). Vigilance, Workload, and Stress. In P. A. H. a. J. L. Szalma (Ed.), *Performance under Stress*: Ashgate.

Webster, R. (2016). Declarative Knowledge Acquisition in Immersive Virtual Learning Environments. *Interactive Learning Environments, 24*(6), 1319-1333. doi:10.1080/10494820.2014.994533

8 Appendix

8.1 Appendix A – Invitation

Invitation to take part in a research study

Dear all,

This is an invitation to participate in a research study at the Training and Assessment Research lab at University of South-Eastern Norway (USN).

We are currently recruiting participants for two studies that are part of a doctoral research and a master's thesis. During the experiment you will be taught and asked to operate two maritime simulators. Your participation in the research will be of great importance to understand better the maritime training simulators and develop simulators for the future.

Your participation is completely voluntary, and you may withdraw from the study at any time. You will be compensated for your participation.

Thank you for your time and participation.

Best regards,
Sathiya Kumar Renganayagalu,
Doctoral research fellow
Simen Hjellvik,
Masters student
TARG lab,
USN.

Background Information:

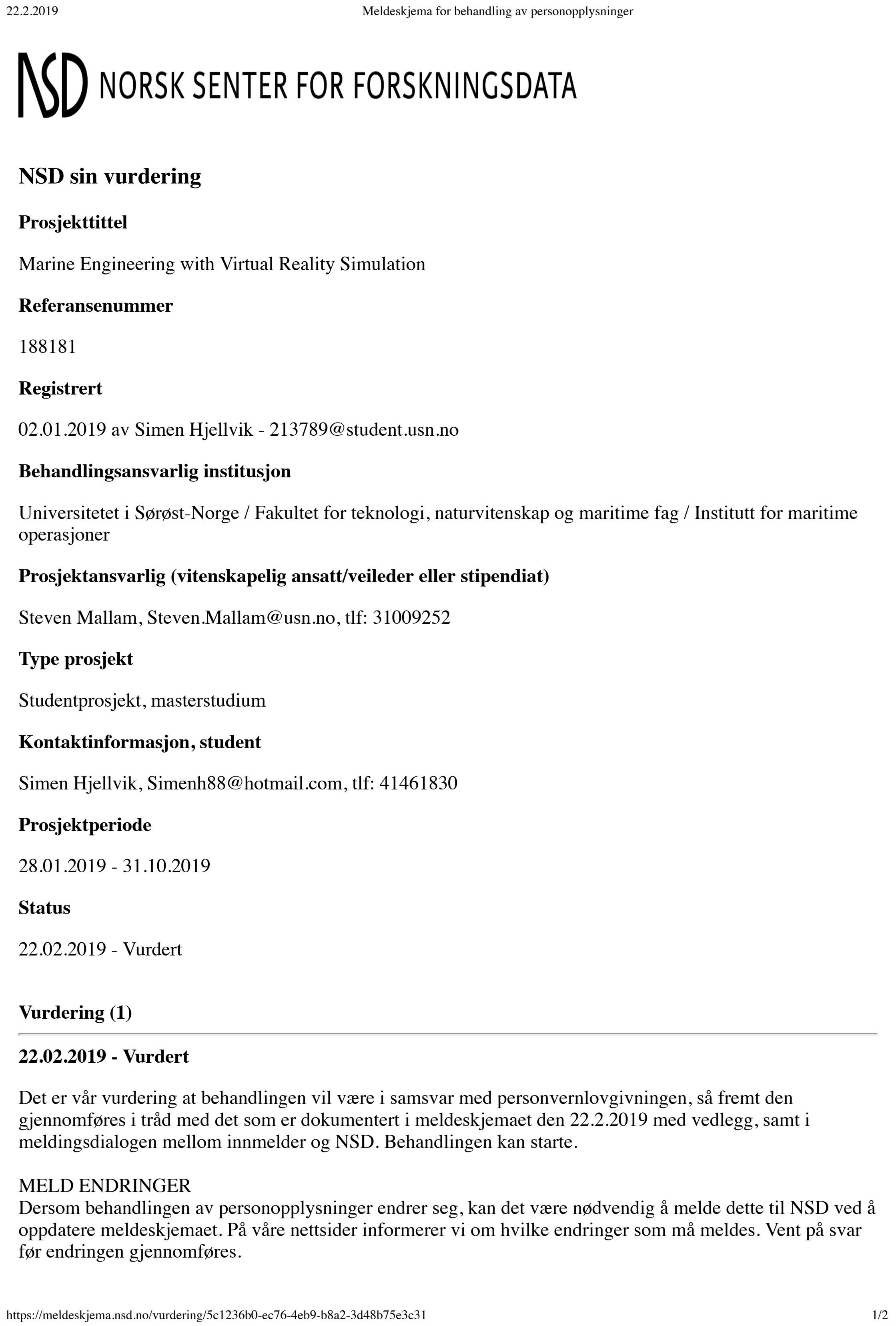
The purpose of this study is to assess the differences in training effectiveness between two maritime training simulators.

Procedures:

If you agree to be in this study, you will be asked to:

- complete a brief demographic questionnaire
- train using a simulator
- perform a task in the simulators and fill questionnaires after the using them.
- interviewed (short, two questions) by the researchers

8.2 Appendix B – NSD Approval



22.2.2019 Meldeskjema for behandling av personopplysninger

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Marine Engineering with Virtual Reality Simulation

Referansenummer

188181

Registrert

02.01.2019 av Simen Hjellvik - 213789@student.usn.no

Behandlingsansvarlig institusjon

Universitetet i Sørøst-Norge / Fakultet for teknologi, naturvitenskap og maritime fag / Institutt for maritime operasjoner

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Steven Mallam, Steven.Mallam@usn.no, tlf: 31009252

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Simen Hjellvik, Simenh88@hotmail.com, tlf: 41461830

Prosjektperiode

28.01.2019 - 31.10.2019

Status

22.02.2019 - Vurdert

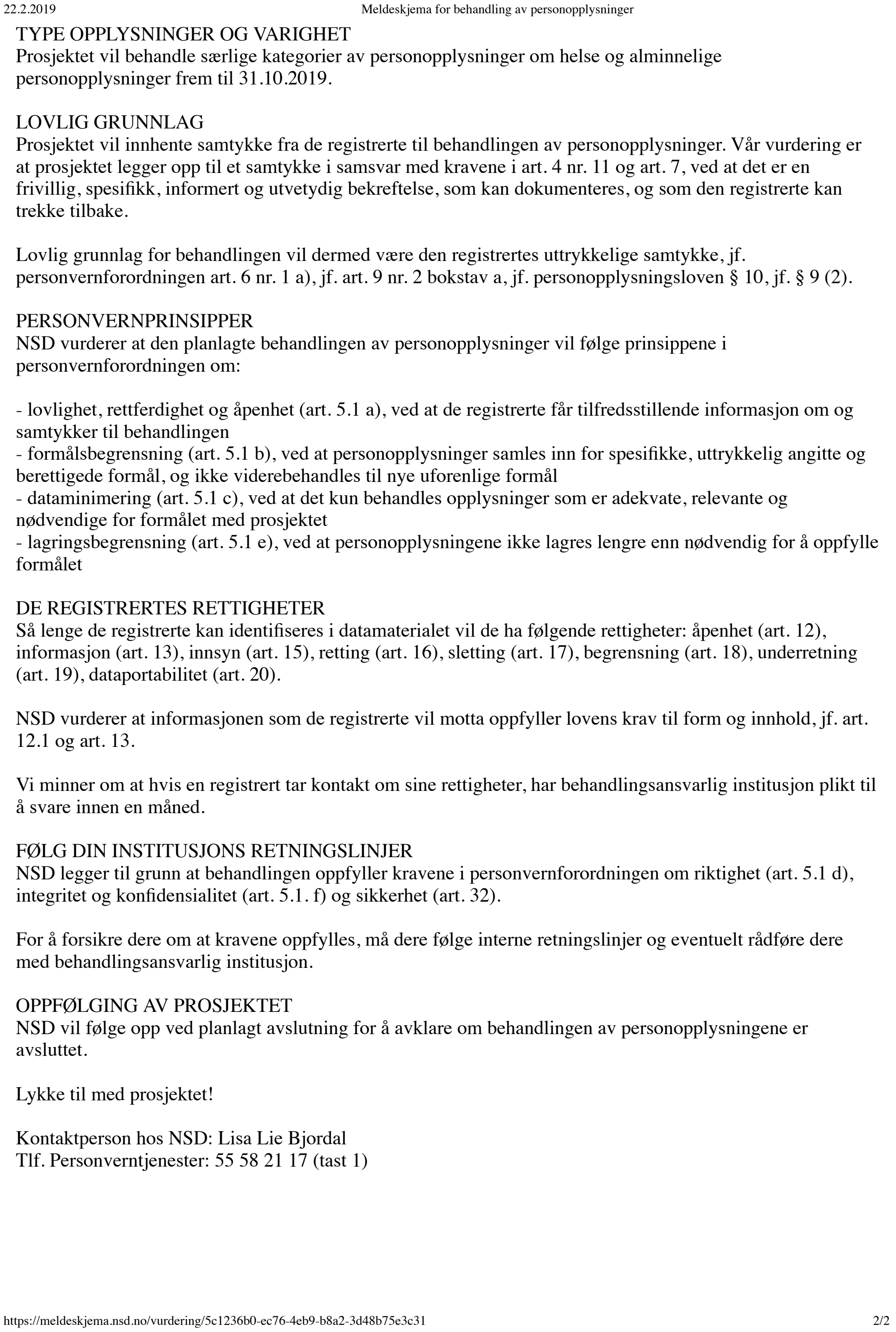
Vurdering (1)

22.02.2019 - Vurdert

Det er vår vurdering at behandlingen vil være i samsvar med personvernlovgivningen, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 22.2.2019 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD ENDRINGER
Dersom behandlingen av personopplysninger endrer seg, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringen gjennomføres.

https://meldeskjema.nsd.no/vurdering/5c1236b0-ec76-4eb9-b8a2-3d48b75e3c31 1/2



TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle særlige kategorier av personopplysninger om helse og alminnelige personopplysninger frem til 31.10.2019.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 nr. 11 og art. 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse, som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes uttrykkelige samtykke, jf. personvernforordningen art. 6 nr. 1 a), jf. art. 9 nr. 2 bokstav a, jf. personopplysningsloven § 10, jf. § 9 (2).

PERSONVERNPRINSIPPER

NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke viderebehandles til nye uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og eventuelt rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET

NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Lisa Lie Bjordal
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

8.3 Appendix C – Information and consent form

Informed Consent Form for Participation in a Research Study

Training and Assessment Research Group,
Department of Maritime Operations
Faculty of Technology, Natural Sciences & Maritime Sciences
University of South-Eastern Norway (USN)

Are you interested in taking part in the research project; ” Marine Engineering with Virtual Reality Simulation”?

This is an inquiry about participation in a research project where the main purpose is to test different types of marine engineering simulators. In this letter we will give you information about the purpose of the project and what your participation will involve.

Description of the research and your participation

You are invited to participate in a research study conducted by Sathiya kumar Renganayagalu, doctoral research fellow and Simen Hjellvik, Masters student at USN. The research is part of Norwegian Research Council funded InnoTraining project (RCN project number: 269424) led by Kongsberg Digital together with USN, Institute for Energy technology (IFE) and Politecnico di Milano (POLIMI). The purpose of this research is to study the effectiveness of Virtual Reality (VR) and desktop-based engine room simulators on training.

Your participation will involve:

- completing a brief demographic questionnaire
- training using VR and desktop simulators
- perform tasks in the simulators and fill questionnaires after the using them.
- interviews (short, two questions) by the researchers

Who is responsible for the research project and what does your participation involve??

The University of South-Eastern Norway is the institution responsible for the project.

There is a difference between confidentiality and anonymity: confidentiality is ensuring that identities of participants are accessible only to those authorized to have access (i.e. the USN researchers). Anonymity is a result of not disclosing participant's identifying characteristics (such as name or description of physical appearance). Any published material as a result of this study will ensure your name and personal information is anonymized. Your personal information, measurements and audio recording will be securely stored on the University of Southeast Norway campus and secured work computers, with access only being given to the listed researchers and members of the Training and Assessment Research Group at the Department of Maritime Operations at USN. Please visit http://targlab.com/ to know more about the research group.

Research team list:

Sathiya Kumar Renganayagalu,
Simen Hjellvik,
Dr. Salman Nazir,
Dr. Steven Mallam,
Jørgen Ernstsen.

Why are you being asked to participate?

As a student or a professional marine engineer, you are convincedly recruited to participate in the sample frame of out experiments. This study is dependent on the participation of a handful, handpicked candidates with your particular knowledge and experience.

Participation is voluntary

Your participation in this research study is voluntary. You may choose not to participate, and you may withdraw your consent to participate at any time. If you decide to withdraw from the study, all your personal data will be made anonymous or deleted.

Your rights

So long as you can be identified in the collected data, you have the right to:

- Access the personal data that is being processed about you
- Request that your personal data is deleted
- Request that incorrect personal data about you is corrected/rectified
- Receive a copy of your personal data (data portability), and send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

Your personal privacy – how we will store and use your personal data

The participant´s anonymous demographics and background data will be recorded through a questionnaire and the exit interview will be audio recorded while the interviewer may take notes. This study results will form a part of doctoral thesis, journal papers and conference papers. All information collected will remain in the TARG lab and will be accessible to the researchers involved in the project. The end date of this project will be October 31st, 2019. All data material will be made anonymous by the end of data collection. The data will be retained for maximum one year after end of project, after which it will be deleted.

We will only use your personal data for the purpose specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

The only record which will contain identifiable data is this document with your signature.

Risks and discomforts

The possible risks of this research study are low. Some participants may experience discomfort using the simulator. The equipment may affect the well-being of the participant by inducing simulator sickness. Symptoms of simulator sickness is nausea, dizziness, or similar bodily discomforts.

Potential benefits

As a participant in this study you will be contributing to the testing and development of new maritime training simulators using virtual reality technology. You will be providing information on the usability and effectiveness of such training tool, by sharing your experience and results.

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with The University of South-Eastern Norway, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Contact information

If you have any questions or concerns about this study or if any problems arise, please contact Sathiya Kumar Renganayagalu at University of South-Eastern Norway at 41347719, sr@usn.no.

The study has been notified and reviewed by the Data Protection Official for Research, NSD - Norwegian Centre for Research Data.

Where can I find out more?

If you have any questions or concerns about this study or if any problems arise, please contact:

- Sathiya Kumar Renganayagalu at The University of South-Eastern Norway at 41347719, sr@usn.no.
- Simen Hjellvik at The University of South-Eastern Norway at 41461818, 213789@usn.no.
- Paal Arne Solberg, Data Protection Officer at The University of South-Eastern Norway at 91860041, personvernombud@usn.no.
- NSD – The Norwegian Centre for Research Data AS, by email: (personverntjenester@nsd.no) or by telephone: +47 55 58 21 17.

Yours sincerely,

Sathiya Kumar Renganayagalu	Simen Hjellvik
(PhD Researcher/supervisor)	(Masters student)

Consent form

I have read this consent form and have been given the opportunity to ask questions. I give my consent to participate in this study.

I have received and understood information about the project “Marine Engineering with Virtual Reality Simulation” and have been given the opportunity to ask questions. I give consent:

☐ to participate in filling out questionnaires
☐ to participate in testing simulators through exercises
☐ to give an exit interview

Participant’s signature_______________________________ Date:_________________

A copy of this consent form should be given to you, if you request it.

Researcher’s Signature:

I have explained this study to the best of my ability. I invited questions and gave answers. I believe that the participant fully understands what is involved in being in the study, any potential risks of the study and that he or she has freely chosen to be in the study.

Researcher’s signature_______________________________ Date:_________________

8.4 Appendix D – Demographics questionnaire

Introduction Questionnaire

Participant code:

Personal demographics

1. Gender
   - Male ☐
     Female ☐

2. Age

3. Education level
   High school ☐
   Diploma ☐
   University degree ☐
   Other ☐ ________________

4. Work status (i.e. Student, employed, self-employed, etc.)

5. Do you have experience from onboard ship?
   - Yes ☐
     No ☐

- If yes,
  - Years of experience ________
  - Rank ______________

Virtual Reality and Gaming experience

1. Have you had experience with VR before?
   - Yes ☐
     No ☐

- If yes, could you explain your familiarity with VR?
-

Not at all	Slightly	Somewhat	Moderately	Extremely

-

2. Have you played video games (computer-based, console-based, other) before?
   - Yes ☐
     No ☐

   If yes, could you explain your familiarity with video games?

-

Not at all	Slightly	Somewhat	Moderately	Extremely

Simulator use

1. Have you trained using engine room simulators?

   - Yes ☐
     No ☐

- If yes,
-
  i) could you explain your familiarity with engine room simulators?
-

Not at all	Slightly	Somewhat	Moderately	Extremely

-
  ii) what type simulator are you familiar with?
   - Desktop ☐
     Big view ☐
     Full mission ☐
     Other ☐

Vision and health

1. Do you have normal/corrected to normal vision?
   - Yes ☐
     No ☐

2. Have you suffered from epileptic seizures before?
   - Yes ☐
     No ☐

8.5 Appendix E – Experiment task description

Fuel Oil Separator

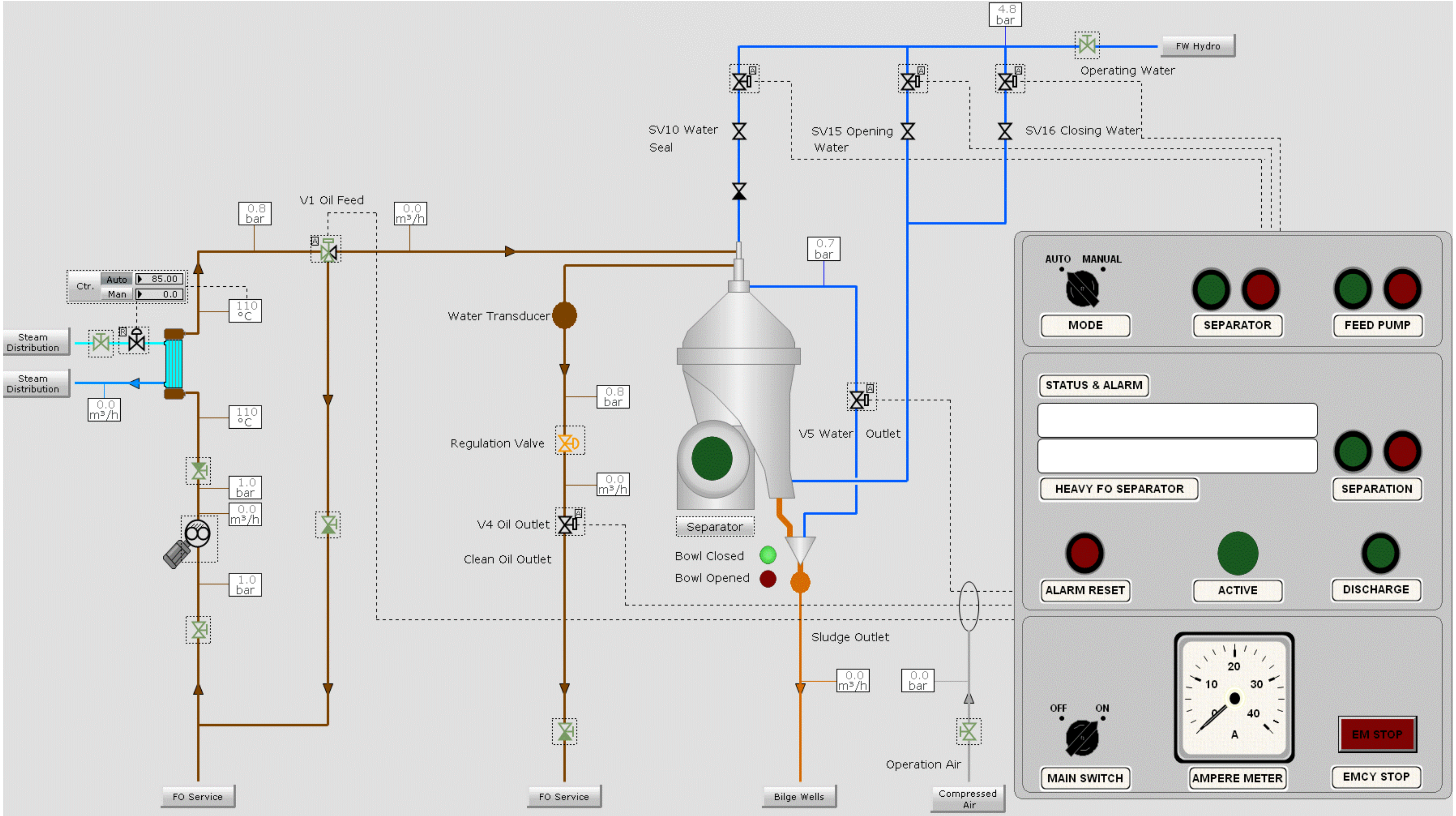
Description

The separator's operation is based on the Alcap principle, which means the separator automatically adjusts to the nature of the oil. No gravity disc is needed. A water transducer in the clean oil outlet measures capacitive resistance and signals changes to the controller. Depending on the water content, the controller either opens the drain valve or expels the water through the bowl discharge ports during sludge discharge. During normal operation, vital process parameters are monitored. These parameters, as well as alarms, are indicated by easy-to-understand text messages on the display.

Operation of the FO purifying system is automatic with control from the local control panel. When the system has been lined up and started manually from the local control panel the system operates automatically.

Task Description: START UP FROM SHUT-DOWN CONDITION

- Switch on electricity and set local operating panel in Manual mode.
- Line up all valves on the oil system. Open valves for heating steam, operating air and operating water.
- Start oil feed pump and check that heat regulation and three-way oil feed return valve is ready.
- Start separator. Monitor amperemeter during speed ramp up.
- When amperemeter drop, switch local operating panel control to Auto.
- Adjust throughput by throttling back pressure valve from fully open position and ensure correct production.


4.8 bar
FW Hydro
Operating Water
SV10 Water Seal
SV15 Opening Water
SV16 Closing Water
0.8 bar
V1 Oil Feed
0.0 m³/h
0.7 bar
Ctr.
Auto
85.00
Man
0.0
110 °C
Steam Distribution
Steam Distribution
0.0 m³/h
110 °C
Water Transducer
0.8 bar
Regulation Valve
0.0 m³/h
V4 Oil Outlet
Clean Oil Outlet
1.0 bar
0.0 m³/h
1.0 bar
V5 Water Outlet
Separator
Bowl Closed
Bowl Opened
Sludge Outlet
0.0 m³/h
0.0 bar
Operation Air
FO Service
FO Service
Bilge Wells
Compressed Air
AUTO
MANUAL
MODE
SEPARATOR
FEED PUMP
STATUS & ALARM
HEAVY FO SEPARATOR
SEPARATION
ALARM RESET
ACTIVE
DISCHARGE
OFF
ON
MAIN SWITCH
10
20
30
40
A
AMPERE METER
EM STOP
EMCY STOP

8.6 Appendix F – Experiment manuscript

Experiment Manuscript

Scenarios

Task 1: Start the fresh water generator - FWG

Task 2: Start the Fuel Oil separator – FO

Phase	Script	Time
Welcoming	Hi, Welcome to the Training and Assessment Research lab. Thank you for participating in our research study. The study will take place in this room. Please make yourself comfortable. You will participate in a simulator study. You will use two simulators: Virtual reality and desktop based. In this study, you will be asked to fill out some questionnaires before and after the tasks in simulators. You will be introduced to the training simulator to familiarize yourself for 15 mins. Then you will be briefed about the task you are going to do in the simulator. You will then be asked to perform the task in the simulator. The task will take approximately 10 mins. After you have completed the tasks in the simulator, you will fill questionnaires and there will be a short exit interview to hear more about your experience. The study will take approximately 60 minutes. Then you will get a break and you will come back for a study in another simulator with similar procedure.	5 min
Information and Consent form	Please fill out this consent form. This is to confirm that you are aware that you are participating in a scientific experiment. The data collected will be used in our theses. Your participation will be anonymous.	5 min
Demographic Questionnaire	Next is a short questionnaire to get some basic information about you. Please fill this out as correctly as possible.	5 min

Task 1		
Familiarization of the simulator	In the next step we ask you to familiarize with the simulator. In front of you, you have a desktop/VR version of the engine room simulator. There are two functional systems in the simulators: Fresh water generator and Fuel Oil separator. Please try them Fresh water generator/ Fuel Oil separator to get yourself familiar with the system. Please ask us if you have trouble interacting with them.	15 mins
Briefing of task	You will perform a task in the simulator: fuel oil separator/fresh water generator. (give the procedure sheet and brief the students)	10 min
Test run	Desktop simulator: Please startup the fresh water generator to start producing potable water. Please try to do this on your own. Only when you are stuck ask us to provide you the next step	10 min
Questionnaires	Please fill out these questionnaires	10 min
Post task questionnaires and exit interview	Please fill out the following questionnaires based on your experience in the simulator.	

Task 2		
Familiarization of the simulator	We ask you to familiarize with the simulator. In front of you, you have a desktop/VR version of the engine room simulator. There are two functional systems in the simulators: Fresh water generator and Fuel Oil separator. Please try the Fresh water generator/ Fuel Oil separator to get yourself familiar with the system. Please ask us if you have trouble interacting with them.	15 min
Pre-test	Here is a flow chart of the separator system with numbers assigned to the main elements. In the table above, you find their respective name labels. Please take 10 minutes and assign the correct numbers to the name labels in the column to the left.	10 min
Briefing of task	You will perform a task in the simulator: fuel oil separator/fresh water generator. (give the procedure sheet and brief the students)	10 min
Test run	(Help student put on the VR glasses) VR simulator: Please startup the fuel oil separator. Please try to do this on your own. Only when you are stuck ask us to provide you the next step. (Help student remove the VR glasses after the study)	10 min
Post-test and questionnaires	Please revisit the flow chart from before the task. As you might remember from the task, certain elements were missing from the environment. In the table to the right of the name labels, please max off any elements you remember where missing. Please fill out the following questionnaires	10 min
Post-task questionnaires and exit interview	Please fill out the following questionnaires based on your experience in the simulator.	
Now the study is over. thank you very much for your participation. here is a gift voucher for you as a token of appreciation for your participation.		

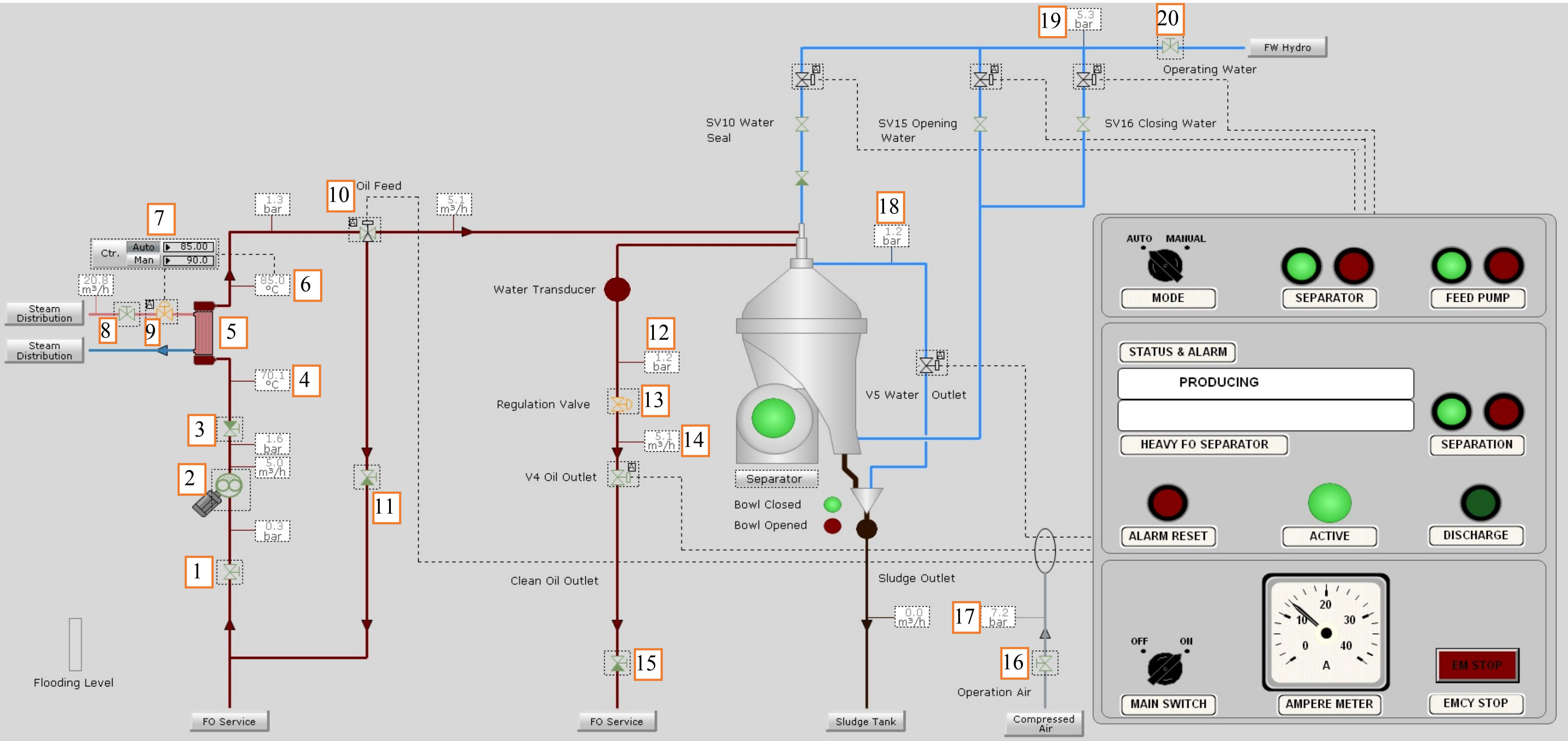
8.7 Appendix G – Pre- and Post-test

Pre- and post-test questionnaire

Subject no:____________

	Feed Pump Suction Valve	
	Clean Oil Outlet Valve To Tank	
	Feed Pump	
	Return Valve From Three-way	
	Back Pressure Regulating Valve	
	Operating Air Inlet Valve	
	Steam Shut-off Valve	
	Temperature Controlled Steam Valve	
	Three-way Feed Valve	
	Back Pressure Gauge	

	Operating Water Inlet Valve	
	Feed Pump Discharge Valve	
	Operating Water Pressure	
	Steam Heater	
	Operating Air Pressure	
	Heater Inlet Feed Temperature	
	Heater Outlet Feed Temperature	
	Water Outlet Pressure	
	Feed Temperature Controller	
	Clean Oil Outlet Flow	

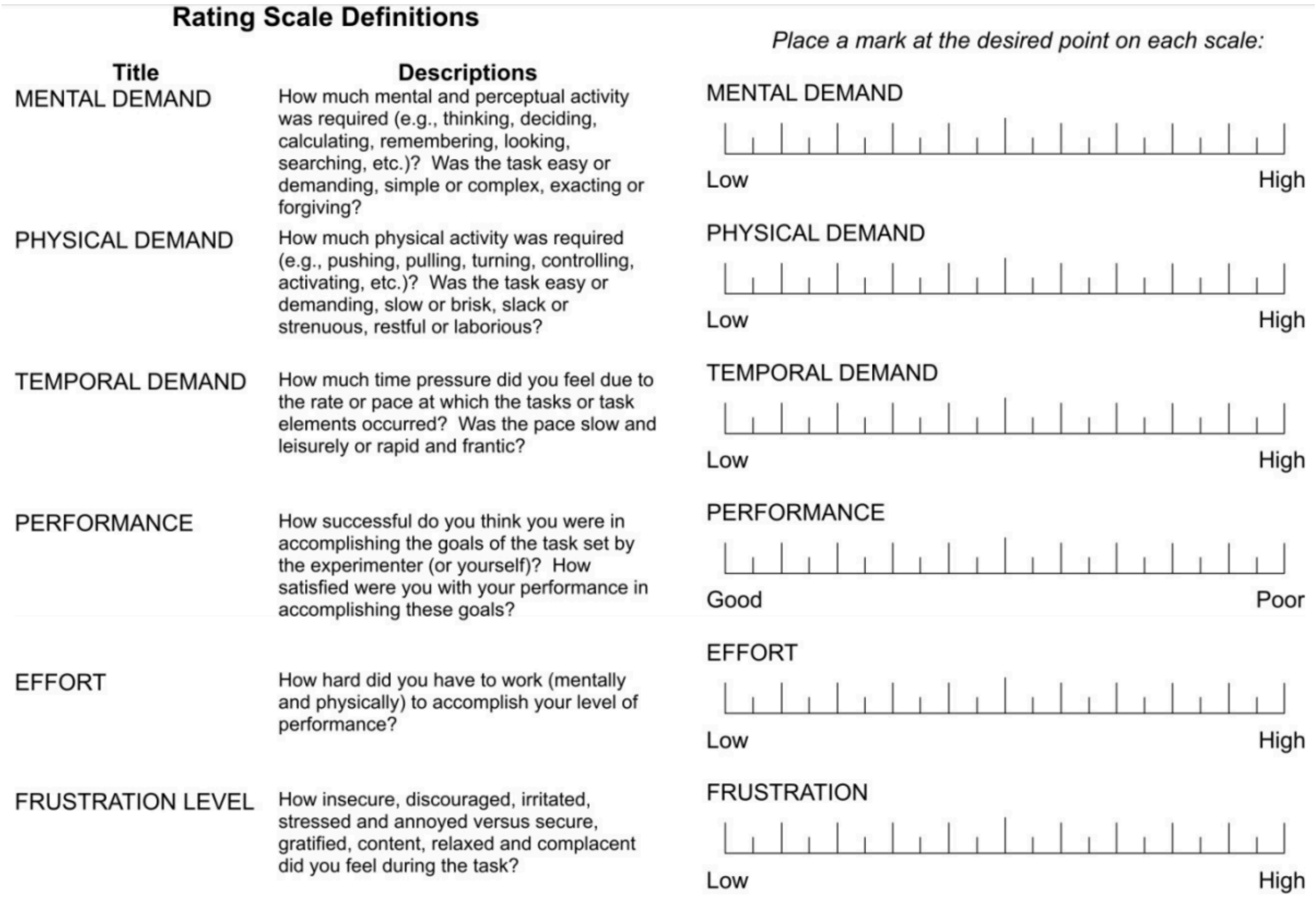


SHORT Flow State Scale (S FSS-2)

Please answer the following questions in relation to your experience in the event or activity you have just completed. These questions relate to the thoughts and feelings you may have experienced while taking part. There are no right or wrong answers. Think about how you felt during the event/activity, then answer the questions using the rating scale below. For each question, circle the number that best matches your experience.

During the event of: **Fuel Oil Separator start up**

		Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	I felt I was competent enough to meet the demands of the situation	1	2	3	4	5
2	I did things spontaneously and automatically without having to think	1	2	3	4	5
3	I had a strong sense of what I wanted to do	1	2	3	4	5
4	I had a good idea about how well I was doing while I was involved in the task/activity	1	2	3	4	5
5	I was completely focused on the task at hand	1	2	3	4	5
6	I had a feeling of total control over what I was doing	1	2	3	4	5
7	I was not worried about what others may have been thinking of me	1	2	3	4	5
8	The way time passed seemed to be different from normal	1	2	3	4	5
9	I found the experience extremely rewarding	1	2	3	4	5



Rating Scale Definitions

Title	Descriptions
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Place a mark at the desired point on each scale:

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Good Poor

EFFORT

Low High

FRUSTRATION

Low High

8.8 Appendix H – Statistics

Table 13: Test of normality

Shapiro-Wilk		**(i) 3D VR**	**(ii) IVR**	**(iii) IVR**
Age	Statistic	0,609	0,882	0,954
	Significance	**0,001**	0,091	0,775
Experience	Statistic	0,771	0,496	0,848
	Significance	**0,046**	**0,000**	0,151
Pre-test	Statistic	0,767	0,908	0,701
	Significance	**0,042**	0,425	**0,006**
Post-test correct	Statistic	0,684	0,915	0,701
	Significance	**0,006**	0,473	**0,006**
Post-test incorrect	Statistic	0,552	0,927	0,907
	Significance	**0,000**	0,554	0,415
S FSS-2	Statistic	0,932	0,905	0,925
	Significance	0,608	0,405	0,543
NASA-RTLX	Statistic	0,947	0,916	0,944
	Significance	0,719	0,480	0,692

Table 14: (*i*) 3D VR and (*ii*) IVR declarative knowledge *U* test

Ranks		**N**	**Mean rank**	**Sum of ranks**
Pre-Test score	(*i*) 3D VR	5	6,9	34,5
	(*ii*) IVR	6	5,25	31,5
Post-test correct score	(*i*) 3D VR	5	3	15
	(*ii*) IVR	6	8,5	51
Post-test incorrect score	(*i*) 3D VR	5	7,5	37,5
	(*ii*) IVR	6	4,75	28,5

Test statistics	**Pre-test score**	**Post-test correct score**	**Post-test incorrect score**
Mann-Whitney U	10,5	0	7,5
Wilcoxon W	31,5	15	28,5
Z	-0,873	-2,796	-1,447
Asymp. Sig. (2-tailed)	0,383	0,005	0,148
Exact Sig. [2*(1-tailed Sig.)]	0,429b	0,004b	0,177b
Exact Sig. (2-tailed)	0, 437	0,004	0,156
Exact Sig. (1-tailed)	0,262	0,002	0,089
Point Probability	0,130	0,002	0,022

a. Grouping Variable: Group
b. Not corrected for ties.

Table 15: (*ii*) IVR and (*iii*) IVR declarative knowledge *U* test

Ranks		N	Mean rank	Sum of ranks
Pre-Test score	(*ii*) IVR	6	5	30
	(*iii*) IVR	6	8	48
Post-test correct score	(*ii*) IVR	6	3,83	23
	(*iii*) IVR	6	9,17	55
Post-test incorrect score	(*ii*) IVR	6	6,58	39,5
	(*iii*) IVR	6	6,42	38,5

Test statistics	Pre-test score	Post-test correct score	Post-test incorrect score
Mann-Whitney U	9	2	17,5
Wilcoxon W	30	23	38,5
Z	-1,550	-2,637	-0,082
Asymp. Sig. (2-tailed)	0,121	0,008	0,934
Exact Sig. [2*(1-tailed Sig.)]	0,180b	0,009b	0,937b
Exact Sig. (2-tailed)	0,177	0,011	0,998
Exact Sig. (1-tailed)	0,089	0,005	0,499
Point Probability	0,049	0,004	0,034

a. Grouping Variable: Group
b. Not corrected for ties.

Table 16: (*i*) 3D VR and (*ii*) IVR mental workload and flow state *U* test

Ranks		N	Mean rank	Sum of ranks
Mental workload global score	(*i*) 3D VR	5	5,7	28,5
	(*ii*) IVR	6	6,25	37,5
Flow state global score	(*i*) 3D VR	5	5,6	28
	(*ii*) IVR	6	6,33	38

Test statistics	Mental workload global score	Flow state global score
Mann-Whitney U	13,5	13
Wilcoxon W	28,5	28
Z	-0,275	-0,368
Asymp. Sig. (2-tailed)	0,783	0,713
Exact Sig. [2*(1-tailed Sig.)]	0,792b	0,792b
Exact Sig. (2-tailed)	0,825	0,758
Exact Sig. (1-tailed)	0,420	0,377
Point Probability	0,048	0,028

a. Grouping Variable: Group
b. Not corrected for ties.

Table 17: (*i*) 3D VR and (*ii*) IVR mental workload and flow state *t* test

Independent samples t test	**t**	**df**	**Sig.(2-tailed)**	**Mean difference**	**SE difference**	**95% CI of the difference**	
						Lower	**Upper**
RTLX	-0,834	9	0,426	-5,444	6,527	-20,211	9,322
S FSS-2	-0,541	9	0,602	-0,140	0,260	-0,729	0,447

Table 18: (*ii*) IVR and (*iii*) IVR mental workload and flow state *U* test

Ranks		**N**	**Mean rank**	**Sum of ranks**
Mental workload global score	(*ii*) IVR	6	8,08	48,5
	(*iii*) IVR	6	4,92	29,5
Flow state global score	(*ii*) IVR	6	5,17	31
	(*iii*) IVR	6	7,83	47

Test statistics	**Mental workload global score**	**Flow state global score**
Mann-Whitney U	8,5	10
Wilcoxon W	29,5	31
Z	-1,527	-1,286
Asymp. Sig. (2-tailed)	0,127	0,198
Exact Sig. [2*(1-tailed Sig.)]	0,132b	0,240b
Exact Sig. (2-tailed)	0,145	0,216
Exact Sig. (1-tailed)	0,073	0,108
Point Probability	0,015	0,009

a. Grouping Variable: Group
b. Not corrected for ties.

Table 19: (*ii*) IVR and (*iii*) IVR mental workload and flow state *t* test

Independent samples t test	**t**	**df**	**Sig.(2-tailed)**	**Mean difference**	**SE difference**	**95% CI of the difference**	
						Lower	**Upper**
RTLX	1,813	10	0,100	13,056	7,202	-2,991	29,104
S FSS-2	-1,462	10	0,174	-0,407	0,278	-1,028	0,213

Table 20: (*i*) 3D VR and (*ii*) IVR mental workload *U* test with outlier removed

Ranks		**N**	**Mean rank**	**Sum of ranks**
Mental workload global score	(*i*) 3D VR	4	5,62	22,5
	(*ii*) IVR	6	5,42	32,5

Test statistics	**Mental workload global score**
Mann-Whitney U	11,5
Wilcoxon W	32,5
Z	-0,107
Asymp. Sig. (2-tailed)	0,914
Exact Sig. [2*(1-tailed Sig.)]	0 ,914b

a. Grouping Variable: Group
b. Not corrected for ties.

8.9 Appendix I – Conference paper first page

Immersive Virtual Reality in Marine Engineer Education

Simen Hjellvik[1], Sathiya Kumar Renganayagalu[1,2], Steven C. Mallam[1], Salman Nazir[1]
Training and Assessment Research Group

[1]Department of Maritime Operations, Faculty of Technology, Natural Sciences and Maritime Sciences,
University of South-Eastern Norway, Horten, Norway
[2] Institute for Energy Technology, Halden, Norway

Abstract - **As simulation and computing technology advance, new pedagogic opportunities are enabled which can add value to student learning outcomes. This study examines simulator training in maritime education comparing the emerging state-of-the-art technology of Immersive Virtual Reality and 3D Virtual Reality desktop simulators. Two student groups from an undergraduate marine engineering programme completed identical tasks related to starting up a fuel oil separator in one of the two conditions: (i) 3D Virtual Reality desktop computer (*n*=5), and (ii) Immersive Virtual Reality head mounted device (*n*=6). After the experimental scenario the participants were given a memory power test to address differences in memory accuracy between the two simulator types. Accuracy of memory diverges between the groups with the 3D Virtual Reality group scoring lower ($\bar{x}$=0, SD=0.71) than the Immersive Virtual Reality group ($\bar{x}$=3.67, SD=1.63). These results provide empirical evidence for the value of Immersive Virtual Reality simulators for marine engineering education.**

Keywords
Memory, Knowledge, Simulator Training, Maritime Education, Shipping.

INTRODUCTION

The human element, as an agent within the maritime industry is developing along with technology towards a vivid complexity of human-machine interaction. Interdependency between new technology and the human element drive the demand for both progressive technology development and the transcendence of the human capital to a new state of knowledge. Control and monitoring of the fiduciary duties of technology has become a part of the seafarer´s knowledge-base. The complex sociotechnical systems that a modern vessel now comprise of tend to put technical requirements in centre of design, engineering and operation, rendering the human element to adapt and cope with the rest through interaction (Norman & Stappers, 2015). Towards higher degrees of automation, regulations and cost mitigates the sentiment for implementation (Mallam, Nazir, Sharma, & Veie, 2019), giving the human element a chance to develop trust and comfort in the technology increments.

Corresponding author
Name: Simen Hjellvik
Affiliation: University of South-Eastern Norway
Address: Raveien 215
3184 Horten
Norway
Email: simen.hjellvik@usn.no
Phone: +47-41461830

Purpose of research

The research question investigated in this study is if Immersive Virtual Reality (IVR) simulator training could be the better technological development for training declarative knowledge in maritime education.

The hypothesis tested predicts that declarative knowledge accuracy, by measurement of the power test, will be larger with the IVR simulator than with the 3D Virtual Reality (3D VR) simulator.

Figure 1: K-SIM Engine 3D Virtual Reality desktop

BACKGROUND

Virtual reality

After decades of 2D desktop simulators, the field of marine engineering education is now saturated with Big View Desktop and 3D Full Mission simulators as the established commercial increment. 3D Full Mission is a simulator type replicating both the full engine control room and an engine room by monitors, touch screens and dummy equipment where the interaction with the environment is visually and audibly animated in 3D. Virtual Reality (VR) is an emergent technology developing with increasing