Immersive Virtual Reality in Marine Engineer Education

Simen Hjellvik1, Sathiya Kumar Renganayagalu1,2, Steven C. Mallam1, Salman Nazir1
Training and Assessment Research Group
1Department of Maritime Operations, Faculty of Technology, Natural Sciences and Maritime Sciences, University of South-Eastern Norway, Horten, Norway
2Institute for Energy Technology, Halden, Norway

Abstract - As simulation and computing technologies 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 Head Mounted Display (HMD) Virtual Reality (VR) and Non-immersive 3D Desktop 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) Non-Immersive 3D Desktop VR (n=5), and (ii) Immersive HMD VR (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. A significant difference was found in accuracy of memory which diverges between the groups with the Non-Immersive 3D Desktop VR group scoring lower than the Immersive HMD VR group. These results provide empirical evidence for the value of Immersive HMD VR 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 progressive technology development and require the human capital to obtain a new state of knowledge.

The complex sociotechnical systems that a modern vessel now comprise of tend to put technical requirements in the centre of design, engineering and operation, rendering the human element to adapt and cope with the rest through interaction (Norman & Stappers, 2015). As operations move towards higher degrees of automation, regulatory complexity and cost (Mallam, Nazir, Sharma, & Veie, 2019), training and education for personnel must adapt.

Purpose of research
The research question investigated in this study is if Immersive Head Mounted Display (HMD) Virtual Reality (VR) simulator training is be the better technological solution 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 Immersive HMD VR simulator, shown in Figure 1, than with the Non-immersive 3D Desktop Virtual Reality (3D VR) simulator, shown in Figure 2.

Figure 1: Participant engaged in the immersive HMD VR simulator condition. The screen in the background display a projection of the participant’s view.

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
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. The convention allows simulators to be used for training and assessment of novice seafarers in education and on-board training, and in revalidation of certificates for professional seafarers (A-I/11 & A-I/12, International Maritime Organization, 2016). This convention structures the industry by defining the main competences required for each discipline and rank, some of which partially can be trained and assessed with simulators.

**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 training solution. 3D Full Mission is a simulator type replicating both the full engine control room and engine room using monitors, touch screens and equipment where the interaction with the environment is visually and audibly animated in 3D. VR is an emergent technology developing with increasing momentum, for example Kongsberg Marine’s K-SIM Engine simulators as is used in this study. VR has been advertised 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 HMD VR differs from all non-Immersive VR where the user views the simulated environment from an outside position, e.g. through a traditional desktop display. Immersive HMD VR technology exchanges the natural 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).

**Immersion**

Advancing the development one step further from the non-Immersive 3D VR environment, introduce the enhanced experience of full immersion. Though not currently available commercially, Immersive HMD VR simulators are of interest to maritime simulator developers. With Immersive HMD VR technology, the environment surrounds the user with an egocentric self-to-object view, which discriminates the immersive experience from the non-immersive 3D VR and all previous generations of simulators. The non-immersive 3D VR desktop depicts an egocentric vision; however, interaction is allocentric object-to-object as the user view the environment through a monitor. This allocentric interaction is also the denominator with previous generations such as CAVE systems, Big View desktop and 2D desktop. With stereoscopic graphics through the HMD, visual updating by the user's head movement and direct interaction through hand controllers makes the experience more immersive and realistic (Freina & Ott, 2015).
Learning outcomes
Traditionally, the fields training and education have focused on changes in verbal knowledge or behavioural capacities as learning outcomes (Kraiger, Ford, & Salas, 1993). Bloom (1956) proposed that there are cognitive learning outcomes beyond recollection and recognition of verbal knowledge in his taxonomy of learning. Gagné (1984) later argued that this taxonomy 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. Knowledge organization and Cognitive strategies, which underlying learning constructs are mental models and metacognitive skills falls beyond the scope of this study.

Declarative knowledge
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).

Measurements of 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 Figure 4.

![Figure 4: Model of learning outcomes adapted from (Kraiger et al., 1993).](image)

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 (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 level 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).
Effects on training
Webster (2016) investigate declarative knowledge acquisition with Immersive HMD VR on soldiers, and in accordance with similar studies, he found that the immersion has a positive effect on the learning outcome compared to lecture-based instruction. In their review of studies on immersive HMD VR training, Jensen and Konradsen (2018) found 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-level cognitive skills with immersive HMD VR.

Passig (2015) investigates immersive HMD VR 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, as now measured, to decline, we might be in an erratic evolutionary process we simply cannot comprehend or measure at this time (Passig, 2015). 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 both. Not only does advanced technology such as Immersive Virtual Reality improve abstract cognitive skills as supposed by Flynn (2018), concrete cognitive skills improves as well according to Passig (2015).

METHODS
This study followed a classic experimental design with two groups for between group measures. The design was chosen as the participants could not conduct both conditions due to the potential carry-over effect. The treatment was designed to be as similar as possible in both the immersive HMD VR simulator and the non-immersive 3D VR simulator.

The pre-test was developed by the authors to capture the subject’s semantics and system knowledge of the machinery operated in the treatment. The post-test consisted of a memory power test developed by the authors to capture accuracy of the students’ knowledge acquisition.

The hypothesis stated that a measure of the post-test with the two student groups would be different with the two simulator types. The random group assignment is the independent variable as the two different conditions could cause different outcomes. The power of memory by our post-test after the treatment is the dependent variable, measured as a retention of knowledge acquired.

The study was approved by the Norwegian Centre For Research Data (NSD), project file number 188181.

Experimental setup
All experiments and data collection were conducted in the quiet and ventilated VR Lab at The University of South-Eastern Norway. For both simulator conditions, the instructor station was assigned to a monitor in the lab with the 2D process interface. In the non-immersive 3D VR condition (Figure 2), the instructor also had view of the monitor used by the participants, while a partial wall inhibited the participants from viewing the instructor and his monitor. The instructor administered the same station for the immersive HMD VR condition (Figure 1) where another monitor show the visuals from the HMD as the participants were immersed in the scene.

Participants
The sample frame consisted of two student groups recruited from the second year of the marine engineer programme. The students were randomly assigned to either a (i) Non-immersive 3D VR desktop group (n=5) or an (ii) Immersive HMD VR group (n=6) based on their voluntary booking time for the experiment. All participants were males.

Table 1: Group demographics

<table>
<thead>
<tr>
<th>Group</th>
<th>(i) 3D VR</th>
<th>(ii) IVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Mean 28,40</td>
<td>22,67</td>
</tr>
<tr>
<td></td>
<td>SD 12,66</td>
<td>0,82</td>
</tr>
<tr>
<td>Prof</td>
<td>Mean 0,60</td>
<td>0,17</td>
</tr>
<tr>
<td>Exper</td>
<td>SD 0,89</td>
<td>0,41</td>
</tr>
</tbody>
</table>

Intervention
An exercise of starting up a fuel oil separator, as shown in Figure 5, was chosen for the treatment task. This was chosen as it is an important machinery system focused on in the education programme, as well as during sea service. A fuel oil separator produces a purified quality fuel oil to the daytanks and is often designed with redundancy for safe operation, as the lack of fuel oil transferring options or clogged fuel supply can induce main engine shut-down situations. At the time of the research, this machinery had been covered in the marine engineering programme through lecture-based instruction. Though the immersive HMD VR simulator and the non-immersive 3D VR desktop simulator had slightly different limitations in their replication of the real-life equipment, the exercise was formulated to match both simulator conditions.
Figure 5: The fuel oil separator system of the immersive VR simulator

Treatment

Compared to the 2D process control interface of the instructor station as shown in Figure 6, both simulator conditions had missing elements, such as valves or gauges. These elements were only visually missing in the participants' simulator environment, and had no implication to the treatment procedure as their function were included in the actual simulator programme of the 2D process control interface controlled by the instructor.

Figure 6: 2D process control of the fuel oil separator of the K-SIM Engine simulator as viewed from the 2D Desktop option.

The 2D process control monitor is the equivalent to the ship’s Integrated Automation System (IAS) were the duty engineer officer remotely operate and monitor the machinery from in the engine control room. This is also the interface of the simulators that both students and instructor would use when operating the simulator system as a 2D desktop option.

In the Immersive HMD VR simulator (Figure 1), participants could move freely by walking in the simulator environment, only bounded by the physical walls of the VR Lab. Physical displacement within the VR Lab, enacted an equivalent movement in the virtual reality environment. Since the virtual environment was larger than the laboratory a locomotion technique called teleportation was used for moving further in VR environment. All interaction with the simulator systems was administered through the hand controllers (Figure 7).

In the non-immersive 3D VR desktop simulation (Figure 2), the environment comprised of the full engine room of a container. The participants sat in front of the desktop monitor, enacted movement and interaction through the Microsoft XBOX hand controller (Figure 8).

The task in the treatment was to conduct a starting operation of the fuel oil separator system from a shut-down condition with a procedure created by the authors (Table 2). A brief system description, a flow chart diagram copied from the 2D process control interface (Figure 6), and the task description with procedure (Table 2) was given on paper for 10 minutes to the participants to review and internalize before withdrawn again and commencement of the treatment. Within this 10-minute review prior to the treatment, 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 procedure steps or uncertain about system statuses which could not be read in the environment, they were allowed to ask the instructor for help.

Table 2: Task Procedure

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

Materials

Equipment

For both immersive VR and desktop 3D VR 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. An engine room simulator of a reefer container vessel (Kongsberg Digital Mak 8M43CM11) was used in the study.
For the immersive HMD VR simulator, a HTC VIVE VR system was used (Figure 7). For the non-immersive 3D VR simulator, the participants interacted with the simulator through a 27” desktop monitor and a wireless Microsoft Bluetooth XBOX controller (Figure 8). The instructor station was set up with a 2D process control (Figure 6) monitor and a partition wall between the participant and the instructor where the instructor had view of the participant’s desktop monitor as shown in Figure 2.

**RESULTS**

The hypothesis predicted that declarative knowledge accuracy, by measurement of the power test, would be larger in the Immersive HMD VR group (ii) than in the Non-immersive 3D VR desktop group (i). The Table 3: Descriptive statistics show a quite even prerequisite knowledge from the pre-test.

The post-test correct scores resulted as predicted were the Non-immersive 3D VR desktop group (i) scored lower than the Immersive HMD VR. For the inferential statistics the Mann-Whitney U test was used and had a significant difference in medians $U=0$, $Z=-2.796$, $P=0.005$, $r=-0.843$.

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

**Measurement Instruments**

The pre-test consisted of an assessment of initial learning through a recognition test developed by the author. 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 machinery systems in general.

In the pre-test, a process flow chart from the Figure 6 display was assigned 20 numbers to the systems main elements. A table with the label names of these 20 elements was included and the participants was asked to assign the correct element number from the 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 logic reasoning should provide a score. To close any knowledge gap and mitigate the effect of individual 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 to assess accuracy and accessibility of retaining knowledge acquired in the treatment. 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 test usually focus on correct items answered and neglect the incorrect; to give the power of memory an additional dimension, incorrect items answered was also considered in accordance with Ackerman and Ellingsen (2016).

The post-test was on the same paper sheet as the 20-item pre-test; the participants were asked to mark off the elements they recall that were missing from the simulator environment. The non-immersive 3D VR simulator had 4 elements missing and the immersive HMD VR simulator had 8 elements missing. These elements could be valves or gauges expected to be present in a live system, and was present on the 2D flow chart given prior to the treatment. Correct items answered gave a score range of 4 and 8 respectively, and thus, incorrect items answered a range of 16 and 12. No admonitory indication of the post-test was given prior to the treatment. The measurement score was aggregated with range of the immersive HMD VR simulator as index, i.e. the 3D VR scores was multiplied with 2.
The gains the post-test correct scores are significantly different with a large effect size and the post-test incorrect score is insignificant and with a medium effect size.

Table 3: Descriptive statistics

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (i) 3D VR</th>
<th>Mean (ii) IVR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-test score</strong></td>
<td>19,00</td>
<td>18,17</td>
</tr>
<tr>
<td>SEM</td>
<td>0,63</td>
<td>0,65</td>
</tr>
<tr>
<td><strong>Post-test correct score</strong></td>
<td>0,80</td>
<td>5,33</td>
</tr>
<tr>
<td>Median</td>
<td>0,78</td>
<td>5,00</td>
</tr>
<tr>
<td>SEM</td>
<td>0,49</td>
<td>0,42</td>
</tr>
<tr>
<td><strong>Post-test incorrect score</strong></td>
<td>-0,80</td>
<td>-1,67</td>
</tr>
<tr>
<td>Median</td>
<td>0,00</td>
<td>1,50</td>
</tr>
<tr>
<td>SEM</td>
<td>0,80</td>
<td>0,56</td>
</tr>
</tbody>
</table>

DISCUSSION

The pre-test was a recognition test on identification of system items. As the pre-test scores were relatively even between the groups, this strengthen the post-test results as independent of the former, but might induce a question of necessity regarding the pre-test feedback element in the experiment design. 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. Still evaluated as valuable to the design, the pre-test holds no prediction of consecutive learning outcomes, only a probe of the prior knowledge base and the feedback a uniform standard before the experiment commencement.

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 might hold evidence accordingly. The difference found between the simulators might be a factor from their environments, whereas the non-immersive 3D VR desktop simulator is encompassing with a complete engine room environment and the immersive HMD VR simulator is relatively confined to a single room. Observations during the experiments led the author to note an incidental tendency to digress from the task in both simulator conditions. As the more encompassing environment of the non-immersive 3D VR simulator has higher level of details and other systems, its effects (Towler & Kramar, 2008) are possibly an influencing factor on the lower score of the (i) non-immersive 3D VR 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.

Accepting the hypotheses acknowledges the prototype immersive HMD VR simulator as superior to the commercial non-immersive 3D VR desktop simulator, based on the design of this study.

The research question investigated in this study was if immersive VR simulator training was a more effective technological solution for training declarative knowledge in maritime education for a specific marine engineering task. As far as this study’s result show, there are advantaged to implementing immersive VR in the educational programs. An appropriately designed simulator and training programme should achieve to supplement students with facets of knowledge and skills the present 2D desktop simulations and the Non-immersive 3D Virtual Reality desktop simulators cannot offer.

Future Research

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. Training higher-order cognitive skills (Figure 4) such as mental models and metacognition are unexplored with immersive HMD VR, and there are opportunities in maritime education for designing training programs approximating these constructs with resources management, safety training and team exercises. With new technology and new simulators there is always the opportunity for training effectiveness studies of the established training programs, and development of new ones as the governing regulations regarding training are quite flexible to exercise designs. Regulations of assessment schemes are more explicit, thus studies of competence assessment with Immersive Head Mounted Display Virtual Reality is necessary both before and after a quality standard approval of the simulator.

With the two simulators as is, the consecutive response to this preliminary study would be to focus effort on a training scheme with repeated measures with an untainted new cohort of second year students. Task specific exercises with marine engineering students over a semester or two and a final assessment of competence could be designed with training in the non-immersive 3D Virtual Reality desktop simulator and assessment in the Immersive Head Mounted Display Virtual Reality simulator. Another group could be trained in the Immersive Head Mounted Display Virtual Reality simulator and assessed on real life equipment. As the participants of this study
has advances to their final year of the marine engineering program, a repeated measure could be taken to measure knowledge retention, or a new and more complex task design with performance indicators to evaluate the technology as an assessment tool.

**Limitations**
The clinical environment of the Virtual Reality Lab and presence of researchers in the room during data collection may have influenced the participant’s performance and results. A segregated instructor station or remote supervision of the experiments could mitigate these effects. However, this was not a practical option for this study of this experimental design.

As there is only one undergraduate programme in marine engineering within Norway, the only solution to strengthen the sample frame would be to recruit students enrolled in the marine engineer programs at the vocational college level, or expand the research to international institutions.

**CONCLUSION**
This study found a difference between the accuracy of memory between the non-immersive 3D Virtual Reality desktop simulator and the immersive Head Mounted Display Virtual Reality simulator. These results provide empirical evidence for the value of immersive Virtual Reality simulators for marine engineering education.

**ACKNOWLEDGEMENTS**
The authors would like to thank the Research Council of Norway and Kongsberg Digital for financial support of the research project Innovating Maritime Training Simulators using Virtual and Augmented Reality, InnoTraining (project number: 269424). Thank you to Kongsberg Digital for providing the lab their simulators and technical support.

**REFERENCES**


