Learning Seriously with VR: An evaluation on the effect of a low-fidelity pre-training on cognitive workload and task performance in a virtual reality sailing task

Leann Edith Leung

Student number: 2669757

MSc. Applied Cognitive Psychology

Thesis (27.5 ECTS)

UU Supervisor: Chris Klink (p.c.klink@uu.nl)

UU Auditor: Chris Janssen (c.p.janssen@uu.nl)

External Supervisor: Heike Diepeveen (H.Diepeveen@marin.nl)

External Company: Maritime Research Institute Netherlands (MARIN)

Date: June 30, 2024

Abstract

The rise of virtual reality-based (VR) technology has prompted the gradual adoption of VR-based simulation training in the maritime industry. However, there is limited research on strategies to improve the effectiveness of VR training through cognitive workload reduction, especially with the pre-training strategy that is rooted in Cognitive Load Theory (CLT). This study investigates whether a low-fidelity mock-up pre-training prior to a VR sailing simulation improves task performance and reduces cognitive workload. Two groups of participants (n = 16) completed a VR sailing task. The experimental group received a low structural fidelity mock-up pre-training before performing the VR sailing task, while the control group did not. Task performance was measured by a performance index including time, distance, and error rate. Cognitive workload was subjectively measured with the NASA Task Load Index (TLX) and Klepsch et al.'s (2017) differentiated cognitive load (DCL) questionnaire. Results indicated that pre-training did not significantly improve performance, though performance increased significantly over the three VR sailing trials. NASA TLX revealed no significant reduction in cognitive workload, and the DCL measure indicated that pre-training increases intrinsic cognitive load. We attribute these findings to limited skill transferability between pre-training and VR simulation, individual differences in spatial and psycho-motor abilities, factors related to pre-training design, target cognitive workload type, and other factors. On a positive note, results show the benefit of repeated exposure to the task leading to automaticity-based performance improvement. Several limitations and future directions are also discussed.

Keywords: Virtual reality, pre-training, cognitive workload, cognitive load theory

Table of Contents

Table of Figures

1. Introduction

Learning often extends beyond the confines of traditional classrooms. Learning strategies are constantly evolving to keep pace with changes in the technological landscape. In recent years, mixed-reality (MR) technology, including virtual reality (VR) devices has rapidly advanced both in quality and commercial availability. This has prompted the rise of using VR as a tool for "Serious Games", in which the primary goal of a game is education and training, rather than entertainment (Breuer & Bente, 2010; Checa & Bustillo, 2020). VR devices have been increasingly incorporated as a tool to offer immersive and hands-on learning experiences in high-risk domains as they can offer a realistic, interactive, but controlled environment for learning, without exposing learners to real-world risks (Jensen & Conradson, 2018). This is particularly appealing for safety-critical domains such as aviation, military, medicine, emergency response, energy, and the maritime sector, where the cost of human errors can be catastrophic, and where training in real-life operational settings such as complex control rooms isn't always a viable option (Buttussi & Chittaro, 2017; Xie et al., 2021).

The rise of VR-based interactive simulation training warrants further research to determine the optimal approach to effectively leverage this technology as an educational tool. Cognitive workload has been a fundamental research area in human factors research in safety-critical sectors (Hollands et al., 2019). Understanding cognitive workload in high-risk and complex workplaces is crucial to ensure operators can maintain an optimal level of attentional vigilance to manage dynamic work situations. Similarly, from a training point of view, investigating learners' cognitive workload allows training and instructional designers to better tailor training material to maximize productive learning within the limited time and resources in the training period.

While the use of VR technology has gained widespread popularity in maritime training and education in recent years, there has been minimal research done to explore its educational potential from the perspective of cognitive load optimization (Mallam et al., 2019; Miyusov et al., 2022). In light of the brand new Maritime eXperience Lab (MX-Lab) constructed at MARIN, there is a need to tap into the potential of using MR and VR technology for training and concept development purposes. This study employs principles in Cognitive Load Theory (CLT), with a focus on utilizing pre-training – a cognitive load reduction strategy rooted in the well-established Cognitive Theory of Multimedia Learning (CTML), to attempt to explore ways to optimally design and utilize VR technology for educational purposes in the maritime industry. The effectiveness of the low-fidelity (LF) pre-training is determined by, first, its ability to reduce cognitive workload, and second, the extent to which it enhances procedural knowledge acquisition, measured by task performance.

2. Literature Review

2.1. VR for Procedural Knowledge Acquisition

VR technology is commonly used for learners to acquire different types of knowledge. Knowledge is generally divided into two types, procedural knowledge, the knowledge of ‘knowing how to do something’, such as sailing a boat; and declarative knowledge, the ‘knowledge of facts’, such as knowing what wind speed is (McCormick, 1997; Anderson, 1990, p. 219). VR has become increasingly integrated into various learning environments that are aimed at procedural knowledge acquisition, which is highly relevant to maritime operational tasks. The application of VR has also been widely studied in the field of educational science, mainly due to its highly graphical, interactive, and adaptable characteristics, which are essential for effective knowledge transfer (Meyer et al., 2019; Cooper et al., 2021). Along with simulators such as desktop-based simulators, full-mission simulators, and cloud-based simulators, the maritime industry is increasingly adopting VR simulators teaching various skills such as marine communication and navigation equipment maintenance (Bingchan et al., 2018; Kim et al., 2021). This indicates a growing acceptance of VR technology, warranting the need to further research how to best use VR for maritime educational training purposes.

2.2. Cognitive Workload

"Cognitive workload" is frequently employed as a crucial research topic in human factors and human-computer interaction (HCI) domains. However, as suggested in a recent review by Kosch et al. (2023) on the state of cognitive workload assessment, a universally accepted definition and standardized measurement method for cognitive workload remains a subject of controversy in psychology. HCI research frequently employs the definition of “mental workload” as the main analysis approach, which, as Kosch et al. (2023) note, comes from a task demand and human-computer interface design perspective. However, this approach alone doesn't fully address the separate types of cognitive resources needed for effective learning and skill acquisition. To provide a comprehensive view of cognitive workload and to take human working memory capacity into account, this study adopts Sweller et al.’s CLT (1998) as the main conceptual framework to investigate cognitive workload in the VR learning environment, along with support from the task resource demand perspective. To ensure consistency, terms such as "cognitive workload," "cognitive load," "mental workload," and "mental demand" are used interchangeably in this study.

2.3. Cognitive Load Theory

2.3.1. Cognitive Load Theory Fundamentals

CLT has a pedagogic origin, and it emphasizes that human working memory (WM) has an inherently limited capacity and can easily be overwhelmed by a large amount of information being processed (Sweller et al., 1998; Paas et al., 2010; Klepsch et al., 2017). *Figure 1* shows a visual breakdown of the WM model – when new visual, auditory, or other sensory information enters the WM as sensory inputs, they are temporarily stored in the WM, which has a limited capacity. WM can process the information inputs effectively within the WM limits, but contrarily, if the amount of information exceeds these limits, an overload situation may happen and hinder information processing (Orru & Longo, 2019). In an optimal task-learning situation, information is processed by WM, and then stored in the form of schemas in long-term memory (LTM) in ways they will be retrieved and used in the future (Klepsch et al., 2017). LTM does not only have the ability to store a significantly greater amount of information, but it can also host highly complex interactions between schemas, allowing for complex reasoning and skill development processes to take place (Swellers et al., 1998) (See *Figure 1*).

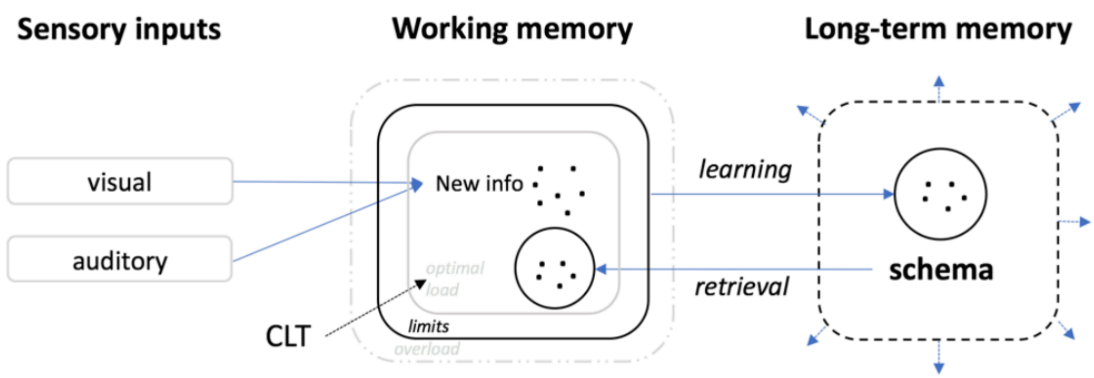

Figure 1 A diagram representing the mental architecture and the role of CLT in relation to WM and schema construction (Orru & Longo, 2019).

2.3.2. Types of Cognitive Workloads

CLT suggests that there are three separate types of cognitive load – intrinsic cognitive load (ICL), extraneous cognitive load (ECL), and germane cognitive load (GCL). ICL is the inherent difficulty associated with the learning task (Sweller et al., Reedy, 2015). ECL is the difficulty presented by the learning environment (Sweller et al., 1998; Reedy, 2015). GCL represents the effort that contributes to the construction of schemas, which is an important component of a learning task's intrinsic difficulty that is beneficial to learning (Sweller et al., 1998; Reedy, 2015). In the context of designing learning material, design based on CLT principles aims to minimize the cognitive workload presented by the learning environment (ECL) and maximize productive learning (GCL).

2.4. Pre-training

2.4.1. Pre-training as a Form of Scaffolding

Rooted in the Cognitive Theory of Multimedia Learning (CTML), an extension of CLT in the context of multimedia learning, pre-training is an effective learning strategy to reduce cognitive load and improve learning outcomes. In general, the purpose of pre-training is to familiarize the learner with certain content before the to-be-learn system to reduce intrinsic cognitive load (Nelson & Erlandson, 2008; Mayer & Moreno, 2010). The concept of pre-training aligns with the concept of scaffolding, in essence, means adjusting instructional delivery methods or material to guide learners through a gradual skill acquisition process toward full competency (Salcedo et al., 2015). Pre-training can be presented in different forms with different target knowledge or skills. One way to present it is to show the names and characteristics of the components of a to-be-learn material. This could help learners to make necessary connections between components to construct a mental model prior to the task (Mayer & Pilegard, 2005; Mayer & Moreno, 2010; Klepsch & Seufert, 2020). Meyer et al. (2019) found that pre-training had a positive effect on knowledge acquisition, transfer, and self-efficacy. Jung et al.'s study (2021) found that pre-training is effective in reducing unnecessary intrinsic cognitive load in early phases of collaborative learning, particularly when pre-training is guided and contains definitions of key elements, to explain element interactivity. This forms the basis of the pre-training for this study, which will be covered more extensively in the *Pre-training Design and Set-up* in the *Method* section.

2.4.2. Low-fidelity Pre-training

Fidelity level is an important consideration for the pre-training design. Fidelity generally refers to the degree to which a simulation replicates real-world conditions and practices (Dieckmann, 2007; Liu et al. 2008). It is a common belief that higher fidelity results in better learning outcomes, but CLT research indicates that high-fidelity (HF) immersive simulation can lead to cognitive overload and can make learning more difficult (Dieckmann, 2007; Reedy, 2015). Cognitive overload in HF learning environments and its negative effects have been extensively studied in the field of clinical medical education (Dieckmann, 2007; Reedy, 2015; Chen et al., 2015; Say et al., 2019; Rogers & Franklin, 2021). A highly immersive learning environment may present an overwhelming number of sensory stimuli, which may overload the learner's WM and prevent them from processing the essential information they need to learn (Reedy, 2015). This is particularly true for novice learners. Chen et al.'s study (2015) comparing the use of LF and HF clinical simulation reveals that too much fidelity increases unnecessary cognitive load on novice learners, leading to ineffective learning. Studies also found that nursing students who perform their tasks using LF simulations outperform students in the HF simulation group, indicating that learning in HF settings does not always lead to higher performance (Say et al., 2019; Rogers & Franklin, 2021). Fidelity level is considered in the pre-training used in this study, more detail in the *Pre-training Design and Set-up* in the *Method* section.

2.5. CLT & CTML Application in the Maritime Sector

Studying cognitive load in the maritime industry, particularly focusing on cognitive workload management is essential as the cost of cognitive overload in both operations and training could lead to potentially catastrophic outcomes that could compromise the safety, efficiency, and overall systems performance. Cognitive overload can decrease operators' situational awareness, impair their decision-making, and increase the likelihood of accidents caused by human errors (McWilliams & Ward, 2021). In the context of safety-critical training and education, CLT provides a systematic framework for understanding how cognitive resources are allocated during learning and task execution, making it a valuable conceptual framework for designing training programs tailored to the cognitive demands of the operational task, with the consideration of the limitations of human working memory.

2.5.1. Maritime and Naval Training

Understanding cognitive workload in maritime training can help to optimize the design of training methodologies, such as the use of immersive VR simulations, that are tailored to the needs of maritime personnel. For example, exploring cognitive load-reduction strategies during complex maritime-specific training tasks can help training developers to better design their training curriculum to enhance student performance. By integrating CLT and CTML principles into the design of VR training programs, developers can better control and analyze task difficulty, feedback mechanisms, and instructional designs.

2.5.2. Concept Development

VR technology can be used to help with the visualization of abstract and future-oriented concepts in the maritime and military sectors. It is important to determine optimal ways to introduce novel concepts to seafarers without overwhelming them with excessive and irrelevant information. Applying CLT and load-reduction strategies in CTML to the design of VR resources can help users to better comprehend and internalize novel concepts, which may lead to an increase in the level of acceptance, and can transform their attitude towards innovative concepts and technology.

3. This Study

3.1. Main Research Question

Synthesizing theories and past research, the present study is an exploratory study to investigate the effectiveness of a low-fidelity pre-training in facilitating procedural knowledge acquisition and reducing cognitive workload in a VR learn-to-sail task. The main research question is: *To what extent is a low-fidelity pre-training prior to a VR learning task an effective cognitive workload-reduction strategy?*

3.2. Hypotheses

Effectiveness in this study is defined by two main metrics – task performance and cognitive workload, which lead to the following sub-questions:

Sub-question 1: How does pre-training affect task performance?
Hypothesis: Pre-training helps to improve task performance in the VR simulation.

Sub-question 2: How does pre-training affect cognitive workload?
Hypothesis: Pre-training helps to reduce cognitive workload in the VR simulation, specifically on intrinsic cognitive load.

A more detailed version of the hypotheses, including null and alternative hypotheses of each sub-measure can be found in *Appendix 6*.

4. Method

4.1. Participants

This experiment was conducted following the ethical guidelines stipulated by the Ethics Review Board of the Faculty of Social and Behavioral Sciences. Using a between-subject design, all participants (n = 16) were randomly assigned into two groups, the experimental group with pre-training (n = 8), and the control group with no pre-training (n = 8). Participants are adults with no prior experience operating a rigid-hulled inflatable boat (RHIB). Participants were recruited within MARIN via email and through word of mouth. 63% of participants are male and 38% are female. Participants' ages ranged from 20-49, the mean age for the pre-training group was 27.8 years, and

25.9 years for the no-pre-training group. 50% of them are Dutch, the other 50% are made up of other nationalities (Turkish, German, Polish, Indonesian, Taiwanese, and Chinese). 63% of them have never used a VR device before, 31% have used it 1-3 times before, and 6% have used it 7+ times before, they are equally distributed between two experiment groups. No one has operated a RHIB before, which is an eligibility requirement for this study. All participants passed the Landolt C visual acuity test before the experiment and obtained above equal or greater than 1 with their normal or corrected vision. None of the participants are colour-blinded.

4.2. Materials

4.2.1. VR Devices

The Steam Valve Index VR kit was used in this experiment. This set-up is equipped with a VR head-mounted display (HMD), two controllers, left and right (see *Figure 2*), two SteamVR 2.0 base stations, a desktop computer to launch the Steam application, and a large standing display monitor. The HMD device is equipped with dual 1440 x 1600 liquid-crystal displays (LCDs), providing high-resolution visuals with full red, green, and blue (RGB) display per pixel (Steam, n.d.). Inter-pupillary distance (IPD) was adjusted per participant within a range of 58mm to 70mm (Steam, n.d.). In terms of audio, the HMD comes equipped with built-in 37.5mm off-ear balanced mode radiators (Steam, n.d.). These audio components offer a frequency response ranging from 40Hz to 24KHz (Steam, n.d.). In VR optical systems, the visual perception of the environment is determined by various factors – including the human pupil location (which encompasses eye relief and IPD position); the aperture of the HMD lens; the focal length of the lens; the size of the display; and the binocular relationship between the two eyes (Steam, n.d.). The Steam application was used to launch the VR content. Another desktop computer was used for participants to fill out the questionnaires prepared in Qualtrics.

Figure 2 A photo of the Steam Valve Index VR HMD and the controllers, left and right.

4.2.2. VR Scenario

In the VR scenario, participants were asked to complete a sailing slalom task set within a gated square area at sea using a RHIB. Within this environment, there are a total of 14 obstacles placed throughout the course. These obstacles are identified by green laser projections with two white poles on the sides, and participants must navigate their virtual RHIB to pass each obstacle and deactivate the laser (see *Figure 3*, left). They will use a steering wheel to control the direction and the throttle to control the speed of the boat (see *Figure 3*, right*)*. The course layout comprises 12 obstacles of uniform width and height. They are equally spread out across the course, alternating between the left and right sides. The remaining two obstacles indicate the start and finish obstacles (see *Figure 6,* narrower obstacles near *Start* and *Finish*). This setup challenges participants' virtual maneuvering skills on a fast boat and requires precise control to navigate through the course effectively.

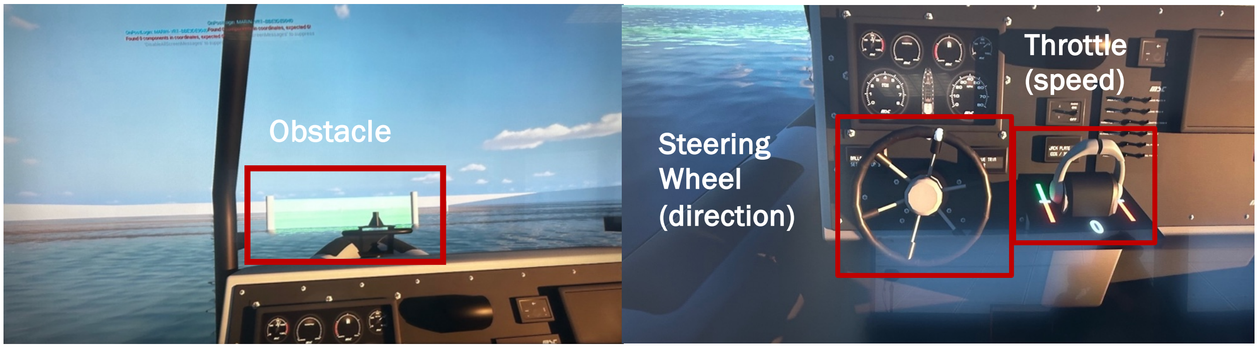

Figure 3 Screenshots of the VR scenario, and the various components (obstacles, steering wheel, and the throttle in the RHIB.

4.3. Pre-training Design and Set-up

The design of the pre-training set-up draws on literature from both fidelity level and cognitive workload. First, this pre-training focuses on reducing the structural fidelity (how the simulator appears) and replicating the functional fidelity (what the simulator does) of the VR task (Hamstra et al., 2014). Hamstra et al. (2014) suggest that the structural or physical fidelity of a learning setting can be reduced without the loss of educational effectiveness of the training, given that there is a sufficient relationship between the functional aspects of both settings.

Further, upon evaluation of the VR task, it was concluded that pre-training that aims to lower the intrinsic cognitive load of the task would be beneficial. In line with CLT, the goal of this pre-training is to reduce intrinsic cognitive load by familiarizing learners with the essential functional information, such as task components, the interactions between the components to highlight causal relationships (Fraser et al., 2015), and tips about sailing to further enhance the understanding of interactions between components before beginning the actual VR task.

Utilizing concepts from Fraser et al.'s study (2015), this LF mock-up pre-training design attempts to execute the following goals:

1. Display individual task components
2. Highlight the causal relationship between components
3. Provide multi-sensory (visual, auditory, tactile) learning for modality effect

In the mock-up set-up (see *Figure 4*), The RHIB is represented by the black and white dice. The multiple obstacles along the slalom course, are represented by the coloured dice. The start and finish lines are also labeled. Participants were tasked to go between the two dice with the same colour.

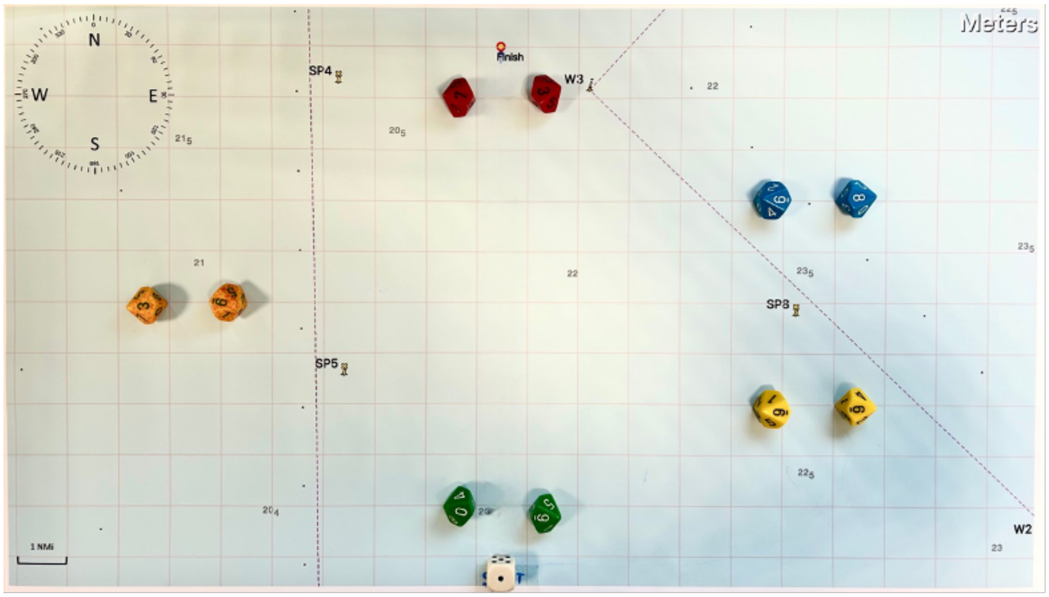

Figure 4 A photo of the mock-up pre-training set-up.

Additionally, some instructional tips in the pre-training include:

- *Slow down before obstacles – Allows for more precise turns, helps to prevent loss of control and overshooting, and enhances situational awareness.*
- *Plan turns in advance – Anticipating maneuvers helps to reduce the risk of sharp turns and maintains stability.*
- *Maintain steady boat speed – Helps to minimize instability, motion sickness, and VR-induced cybersickness.*
- *Sail as close to the inner pole as possible – Helps to minimize the length.*
- *Back up if you miss a turn – Helps to save time and avoid unnecessary detours.*

The goal of these tips is to depict the causal relationships between the components. The full pre-training script can be found in *Appendix 4*.

4.4. Procedure

When a participant came in for the experiment, they were first given a two-minute short PowerPoint introduction about the experiment, including an introduction to the tasks and measures. Then they read and signed the informed consent form, and continued with the demographic questionnaire prepared on the computer in Qualtrics. They were then randomly assigned to either the control or experimental group, as seen in *Figure 5* below.

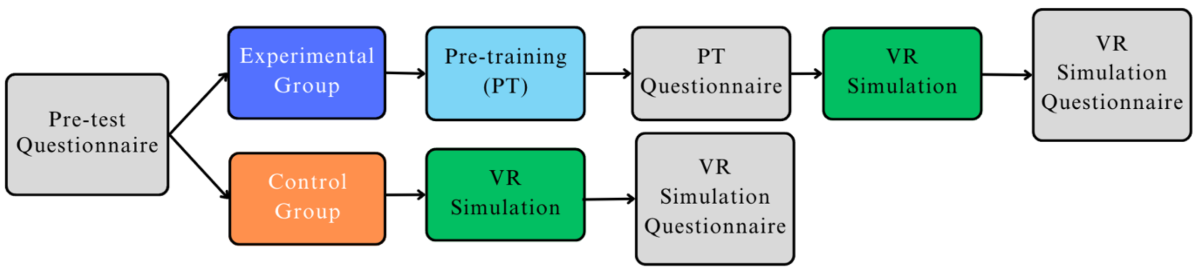

Figure 5 Experimental process flow.

4.4.1. Mock-up Pre-training

For the experimental group, after completing the pre-test questionnaires, participants were asked to complete the mock-up pre-training. First, they will get an introduction to the purpose, context, task, and rules of the pre-training. The purpose of the mock-up pre-training is to help them visualize the VR environment and get familiarized with the VR simulation task components and the interaction between them before the actual VR task (see *Figure 6* for mock-up set-up). The context is to imagine they are the driver of a RHIB and they are tasked to operate the RHIB through a slalom course. The task was they have to navigate through the course from start point to finish point as fast as possible, take the shortest route possible, and miss as few obstacles as possible. The course was marked by obstacles, indicating a zigzag path that they must follow. There were five obstacles in total in the mock-up, but they were told that the VR task contained fourteen in total. They were also pre-briefed on the layout of the mock-up and the VR task, and the discrepancy between their layouts (the first obstacle is in the middle, then right, left, right, then the last one (14th) is at the center), that the layout discrepancy is a result of the attempt to test if a LF, simple mock-up design that contains all necessary information is already sufficient in producing an cognitive workload off-loading effect. One of the rules in the mock-up was that hitting an obstacle is considered an error. This helps to build the concept of the error for the VR task. They were also warned that the boat piece should be touching the map the whole time, they cannot lift it when approaching an obstacle. Next, participants received the instructional tips. Following that, they were instructed to navigate the sailing task three times, using their fingers to maneuver the boat dice through the obstacles, while adhering to all rules and instructional tips provided previously. After the mock-up task, participants were asked to complete the *NASA Task Load Index* questionnaire on an iPhone (Hart and Staveland, 1988), the *Differentiated Load* questionnaire (Klepsch et al., 2017), the *Igroup Presence Questionnaire* (IPQ), and the *Simulator Sickness Questionnaire* (SSQ) on the desktop computer. More details about the questionnaires can be found in the *Subjective Cognitive Workload Measures* and *Other Measures* sections.

4.4.2. VR Simulation

All participants in both the experimental (with pre-training) and control group (no pre-training) were tasked to complete a VR slalom task. The control group immediately entered this step after completing the demographic questionnaire. First, they were instructed to stand still between the yellow lines for optimal signal detection of the base stations. Then, IPD was measured for each participant using the EyeMeasure 1.22 version application on an iPhone. All participants' IPD fell within the 58mm to 70mm range of the VR HMD. The IPD measurement was used to adjust the focal points in the HMD before putting it on the participant. The researcher then helped the participant put on the HMD and the controllers and adjusted the size appropriately. The researcher then explained to the participant how the VR set-up and the simulation task work. For example, they had to make a grabbing motion (by opening and closing their hand) with the controllers to hold the steering wheel for direction on the left and the throttle for adjusting speed on the right (see *Figure 3,* right). There is a minor technical limitation to the steering wheel in the VR simulation – there is no sensory feedback when the user oversteers the wheel to either left or right passed around 120 degrees from the center top, unlike with physical steering wheels. This absence could cause users to unintentionally oversteer, resulting in a sudden 180-degree change in course direction. A white marking was present on the wheel to indicate the exact position that could trigger the change in direction, and participants were warned that should avoid steering past this marking (see *Figure 3,* right, white mark at the top of the steering wheel). This technical limitation was explained during the task introduction.

Similar to the pre-training, in the VR simulation task, the goal is to sail from beginning to end as fast as possible. Participants were asked to zigzag through the entire course and go through all of the green obstacles, they had to pass through each of them to activate the next one. Each participant was asked to complete the task three times, given their physical condition allowed them to do so.

After putting the HMD on, participants were asked if they could see their surrounding environment clearly, including whether the environment appeared to be smooth and if there were delays in the visual. All participants were given two minutes of familiarization time. During this time, they were asked to try accelerating and steering to get through the obstacles. After that, the VR simulation starts when the researcher starts the timer. Performance data such as participants' time to complete the course, length of course sailed, and errors were extracted from the system log files afterward, more on it in the next section, *Task Performance Measures*.

4.5. Measures

4.5.1. Task Performance Measures

At the end of each run, the VR system produced a log file in JSON format with the position coordinates of each frame recorded and the start and end times of each run. The coordinates of the absolute position of the boat's bow were used for calculating time and length. Python was used for plotting, visualization, and calculation for the following performance metrics.

Time Calculation

A framerate, average frames per second was calculated by dividing the total time lapsed by the number of frames in the log file. Then time was calculated by dividing the total number of frames by the generated framerate. Frames before the start line and after the finish line were discarded.

Length Calculation

The following distance formula was used to calculate the Euclidian distance between each of the x and y coordinates along the entire length of the course in Python.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

d = Distance

x = *x* coordinate

y = *y* coordinate

Error Calculation

Error was counted manually by using plotted graphs. There were three criteria for an error – first, the boat enters the obstacle from behind (red circle in *Figure 6*); second, the boat passes the same obstacle twice (green circle in *Figure 6*); third, hitting one of the obstacle poles (when sailing course fully covers the obstacle pole in the plot). The third error was not made by any participant.

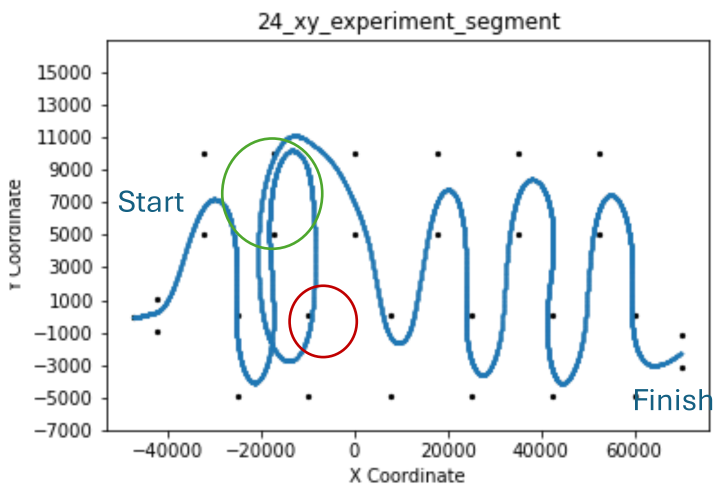

Figure 6 A plotted route from one of the participants.

Task Performance Index

$$Task\ Performance\ Index\ (TPI) = \frac{HitT}{TotalT} \times \frac{MinTime}{Time} \times \frac{MinLength}{Length}$$

Adapted from Marucci et al.'s formula (2021), the Task Performance Index (TPI) was used to normalize the scores of the three performance metrics. The TPI is a combination of the three measurements into a single score. It is calculated using the three measurements – Hit Target (HitT), participants' time to complete the course (Time), and participants' length of the course (Length). Minimum time (MinTime) and length (MinLength) were the lowest time and length within the sampled participants, yielding a relative performance ratio for this particular sample, rather than an absolute ratio. The higher the TPI value, the better the performance.

4.5.2. Subjective Cognitive Workload Measures

Differentiated Cognitive Load Questionnaire

Klepsch et al. (2017) introduced a novel method for assessing three separate types of cognitive load suggested in CLT. In their study, a series of reliability and validity analyses were conducted and confirmed that the three cognitive load categories could be measured separately. Their study yielded a version of differentiated cognitive load with Cronbach's $\alpha = 0.81$ for the ICL scale, $\alpha = 0.86$ for the ECL scale, and $\alpha = 0.67$ for the GCL scale (Klepsch et al., 2017). This final version of the questionnaire from Klepsch et al.'s study (2017) is used to differentiate the three types of cognitive load in this study. It contains 8 items on a 7-point Likert scale, ranging from "Absolutely Wrong" to "Absolutely Right". The full questionnaire can be found in *Appendix 1*.

NASA Task Load Index Questionnaire

Developed by Hart and Staveland (1988), the National Aeronautics and Space Administration Task Load Index (NASA TLX) is a multi-dimensional rating scale that consists of 6 subscales of workload to assess the level of perceived workload from a task demand perspective. The NASA TLX 1.0.3 version iPhone application was used in this study. The NASA TLX dimensions each contain a 0-100 rating scale, with increments ranging from "Very Low to "Very High". Users rate the contribution of each of the six dimensions of workload to identify the intensity of the perceived workload. The subscales represent six independent clusters of variables: Mental Demand, Physical Demand, Temporal Demand, Frustration, Effort, and Performance (Hart, 2006). While only mental demand falls squarely into the cognitive workload definition, the other five dimensions are also closely associated with cognitive workload. Therefore all dimensions are included in the analysis. The full questionnaire can be found in *Appendix 1*.

4.5.3. Other Measures

The following two measures are considered supplemental metrics for this study:

Igroup Presence Questionnaire

Given the mixed opinion on fidelity measurement in simulation literature, this study employs Sense of Presence (SOP) as a proxy to measure the subjective level of immersion, primarily to confirm that mock-up pre-training indeed offers a low level of immersion, and the VR simulation offers a higher level of immersion (Liu et al. 2008; Hamstra et al., 2014).

This study used the *Igroup Presence Questionnaire* (IPQ) to measure the subjective SOP experienced by participants in both the LF mock-up and the HF VR simulation. SOP is the subjective sense of being in a virtual environment (Schubert et al., 2001). The IPQ consists of 14 items on a 7-point Likert scale, including four sub-constructs – spatial presence (SP), the sense of being physically present in the environment; involvement (INV), the attention devoted to the environment; and experienced realism (REAL), the subjective experience of realism in the environment, and one question on general presence (G), a sense of being there in the environment (Schubert et al., 2001). The Cronbach's alpha coefficients for each construct in the IPQ are as follows: $\alpha = 0.77$ for SP, $\alpha = 0.76$ for INV, $\alpha = 0.70$ for REAL, and $\alpha = 0.87$ for all IPQ items (Schubert et al., 2001). The original questionnaire was designed to measure SOP in virtual environments only, therefore the wording in the questions was modified to adapt to both the mock-up pre-training and VR simulation. The full modified questionnaire can be found in *Appendix 1*.

Simulator Sickness Questionnaire

The *Simulator Sickness Questionnaire* (SSQ) is a widely adopted self-report tool that is commonly used to assess participants' cybersickness symptoms in VR simulations (Kennedy et al., 1993). The SSQ consists of 16 items that measure the severity of simulator-induced sickness symptoms in three sub-scales – nausea, oculomotor, and disorientation. The SSQ was administered both before and after the VR simulation, and the pre and post-VR simulation score difference was analyzed. Each symptom is rated by participants on a from 0 (none) to 3 (severe). The full questionnaire can be found in *Appendix 1*.

5. Results

Python was used to calculate the time and length, and to plot the coordinates to count the errors made by each participant in each trial. IBM SPSS Statistics 29 was used for all statistical analyses. Results are reported based on research questions on task performance, measured by TPI, and cognitive workload, measured by NASA TLX and DCL questionnaires. Other supplemental measures such as IPQ and SSQ are also reported.

5.1. Task Performance Results

5.1.1. Task Performance Index Mean Results

It was hypothesized that pre-training would lead to an increase in the TPI mean score. Since we had a small sample size ($n = 16$), we needed to test the assumptions of equal variance and normality to determine whether to use parametric or non-parametric statistical tests. Upon running a Levene's test for equal variance and a Shapiro-Wilk test for normal distribution, despite the small sample size, the mean TPI scores met both of these assumptions (see *Appendix 3* for results).

An independent samples t-test on the mean TPI scores indicated a very marginal mean score difference between the no pre-training group ($M = 0.398$, $SD = 0.227$, $SE = 0.080$) and the pre-training group ($M = 0.399$, $SD = 0.115$, $SE = 0.041$). There was no significant difference in mean TPI scores between the pre-training and no pre-training groups ($t(14) = -0.011$, $p = .496$). The effect size, as assessed by Cohen's d, was found to be $d = 0.180$, $d < 0.300$, suggesting a small effect size. Therefore, the null hypothesis, stating that pre-training does not increase mean TPI scores, was accepted.

5.1.2. Task Performance Index 1st Trial Results

Pre-training could have a greater effect on the 1st trial of the VR simulation than the mean score of the three trials as the cognitive load reduction effect from the pre-training is expected to be the most pronounced immediately after its completion, and it does not suffer from the learning effect. It was hypothesized that pre-training would lead to an increase in the TPI of the 1st trial. Levene's and Shapiro-Wilk tests showed that the TPI 1st trial scores met both the equal variance and normality assumptions (see *Appendix 3* for test results). An independent samples t-test indicated a small mean score difference between the first trial TPI scores of the no pre-training group (M = 0.315, SD = 0.180, SE = 0.064) and the pre-training group (M = 0.282, SD = 0.124, SE = 0.044). There was no significant difference in first trial TPI scores between pre-training and no pre-training group ($t(14) = -0.425$, $p = .677$). The effect size, as assessed by Cohen's d, was found to be d = 0.212, suggesting a small effect size. Therefore, the null hypothesis, stating that pre-training does not increase 1st trial TPI scores, was retained.

5.1.3. Task Performance Index by Trial Sequence Results

A between-subject factorial analysis of variance (ANOVA) was conducted to examine the effects of pre-training and trial sequence (first, versus second, versus third) on TPI scores (see *Figure 7*). *Table 1* in *Appendix 5* displays the descriptive statistics of the ANOVA test. The analysis revealed a significant main effect of the trial sequence on TPI scores ($F(2, 38) = 5.836$, $p = .006$). However, results showed that the presence of pre-training did not significantly affect TPI scores ($F(1, 38) = 0.248$, $p = .621$), meaning no main effect is found. There was also no significant interaction between pre-training and trial sequence ($F(2, 38) = 0.396$, $p = .675$). Additionally, the Bonferroni post hoc tests showed a significant difference in TPI scores between the first and third trial ($p = .006$), but not between the first and second trial ($p = .115$) and the second and third trial ($p = .542$). Overall, the factorial ANOVA suggested that a learning effect significantly influenced TPI scores over the three trials, particularly from the first to the third trial, with a small to medium effect size

($\eta^2 = .249$, adjusted $\eta^2 = .151$). However, the presence of pre-training did not significantly moderate this effect, therefore, the null hypothesis, was retained. (as seen in *Figure 7*).

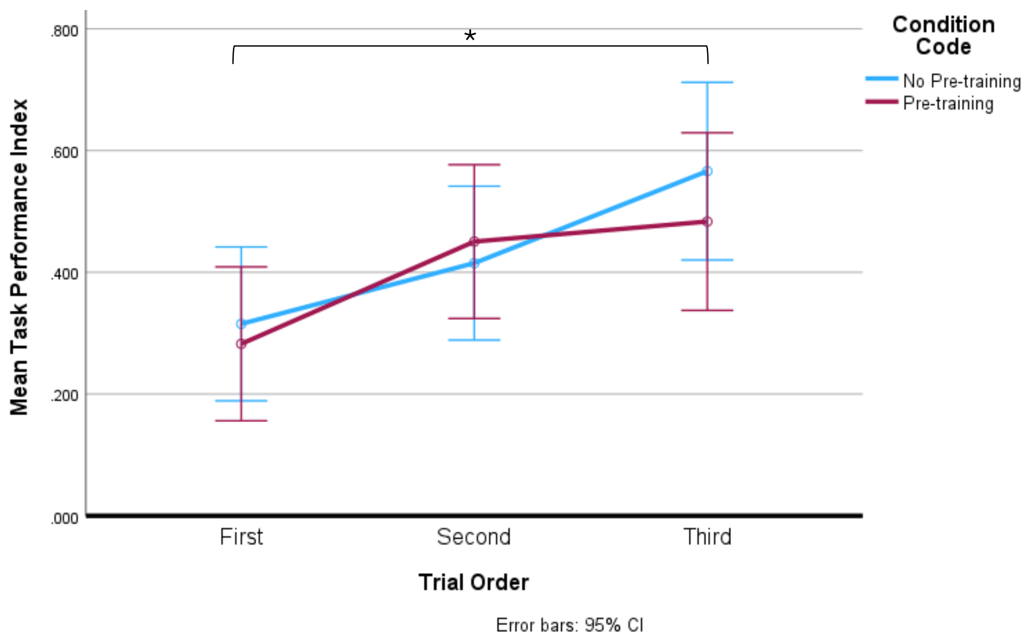

Figure 7 A line graph depicting TPI mean by trial sequence (order) and by condition (pre-training and no pre-training). The error bars represent the 95% confidence interval (* = $p < 0.05$).

5.2. Cognitive Workload Results

5.2.1. Differentiated Cognitive Load Questionnaire Results

The DCL measure met the assumption of both equal variance and normality, meaning a t-test can be used for the DCL measure (see *Appendix 3* for Levene's and Shapiro-Wilk test results). Descriptive statistics for ICL, ECL, and GCL are shown in *Figure 8* and *Table 3* in *Appendix 5*.

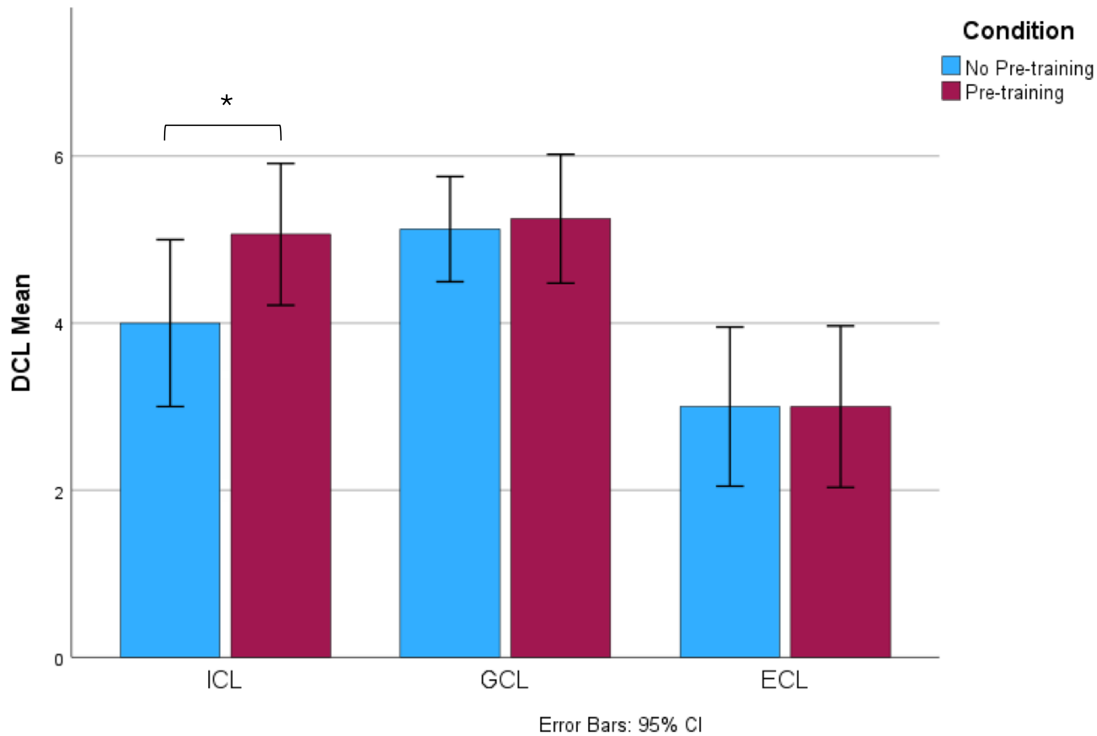

Figure 8 A bar chart showing the mean (bars) and standard deviation (error bars) of DCL scores of both experimental and control groups (* = $p < 0.05$).

An independent samples t-test revealed that there was a statistically significant difference between the two groups for ICL, ($t(14) = -1.916$, $p = .038$), but surprisingly indicating higher ICL scores in the pre-training group, which contradicts the ICL hypothesis which hypothesized that exposure to pre-training will decrease the ICL scores. The effect size estimates indicated large effects for ICL, with $d = 1.10901$.

For GCL, the t-test did not reveal a statistically significant difference ($t(14) = -0.297$, $p = .771$), suggesting similar GCL scores between groups. The effect size estimates showed large effects, with *d* = 0.84221. Similarly, for ECL, the t-test did not reveal a statistically significant difference ($t(14) = 0.000$, $p = 1.000$), indicating comparable ECL scores between groups. The effect size estimates showed large effects for ECL, with *d* = 1.14781. In summary, DCL results suggested that while there were significant differences in ICL scores between the pre-training and no pre-training groups, no significant differences were observed in GCL and ECL scores. However, as shown in *Figure 8,* it was evident that both groups exhibited higher ICL (no pre-training *M* = 4.000; pre-training *M* = 5.063) and GCL scores (no pre-training *M* = 5.125; pre-training *M* = 5.250) in comparison to ECL (no pre-training *M* = 3.000; pre-training *M* = 3.000).

5.2.2. NASA Task Load Index Questionnaire Results

Upon testing the assumption of equal variance and normality for NASA TLX raw scores, results from the Levene's and Shapiro-Wilk tests showed most but not all dimensions meet these assumptions (see *Appendix 3* for test results). To maintain test consistency across all TLX dimensions, a series of non-parametric independent-samples Mann-Whitney U tests were used for the analyses. Descriptive statistics are shown in *Figure 9* below and *Table 2* in *Appendix 5*.

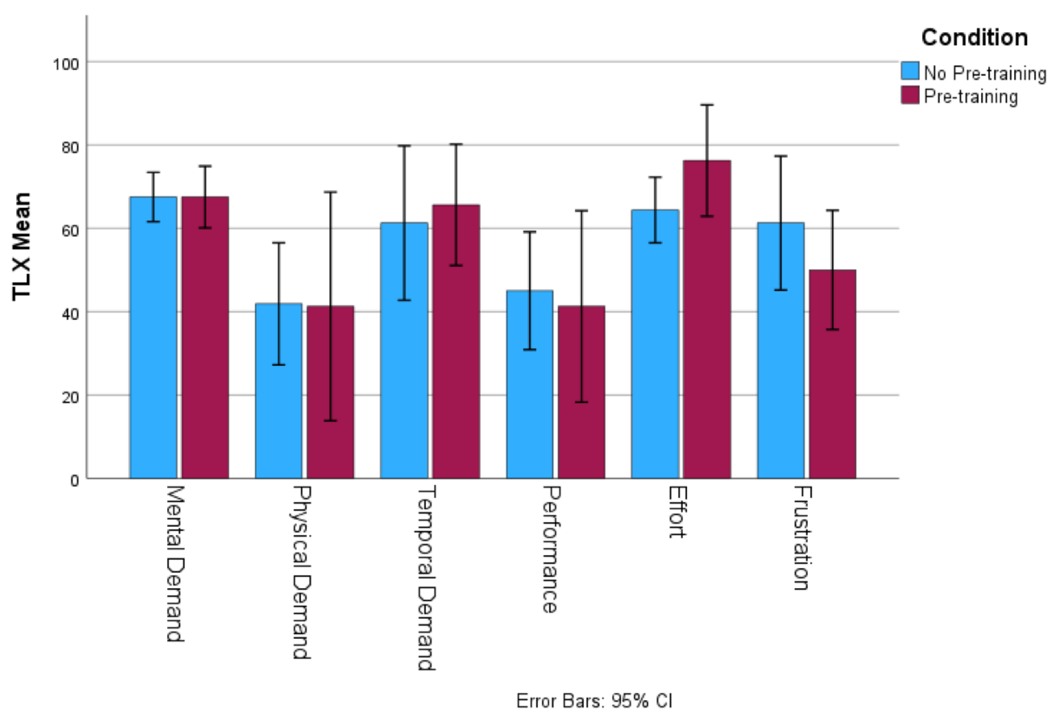

Figure 9 A bar graph showing the mean (bars) and standard deviation (error bars) of NASA TLX scores of both experimental and control groups.

The results showed that TLX cognitive load scores between the pre-training and the no pretraining group were not statistically significant across all of the NASA TLX dimensions. Results were shown as following: Mental Demand (U = 33.500, z = 0.166, p = .878), Physical Demand (U = 28.000, z = -0.422, p = .721), Temporal Demand (U = 33.500, z = 0.161, p = .878), Performance (U = 29.000, z = -0.317, p = .798), Effort (U = 48.000, z = 1.700, p = .105) and Frustration (U = 22.500, z = -1.018, p = .328). Therefore, all null hypotheses, stating that there is no significant difference between participants who received pre-training and those who did not receive pre-training in all cognitive workload dimensions in the NASA TLX, were accepted.

The effect size calculations for all NASA TLX dimensions were conducted to assess the magnitude of differences between the pre-training and no pre-training conditions. The estimated effect size was calculated using a z-score, with the following formula, as suggested in Field (2013, p. 227).

$$r = \frac{z}{\sqrt{N}}$$

The results revealed a small effect size in Mental Demand (r = 0.042), Physical Demand (r = -0.106), Temporal Demand (r = 0.040), Performance (r = -0.079), and Frustration (r = -0.255). However, notably, the dimension of Effort presented a medium effect size (r = 0.425). This indicates that although the statistical test did not yield significant results (p = .105), the difference in Effort between the pre-training and no pre-training conditions was relatively more substantial compared to other dimensions.

5.3. Other Measures Results

5.3.1. Igroup Presence Questionnaire Results

Upon running a one-sample t-test for the experimental group who experienced both the pre-training and the VR simulation, the results showed that the SOP constructs in the IPQ questionnaire – general immersion (t(7) = 16.523, p < .001), involvement (t(7) = 15.513, p < .001), spatial presence (t(7) = 18.158, p < .001), and realism (t(7) = 14.634, p < .001) – all demonstrate statistically significant mean differences (see *Figure 10*). These results indicated that there is indeed a significant difference between the level of presence experienced by experimental group participants in the LF mock-up compared to the HF VR simulation.

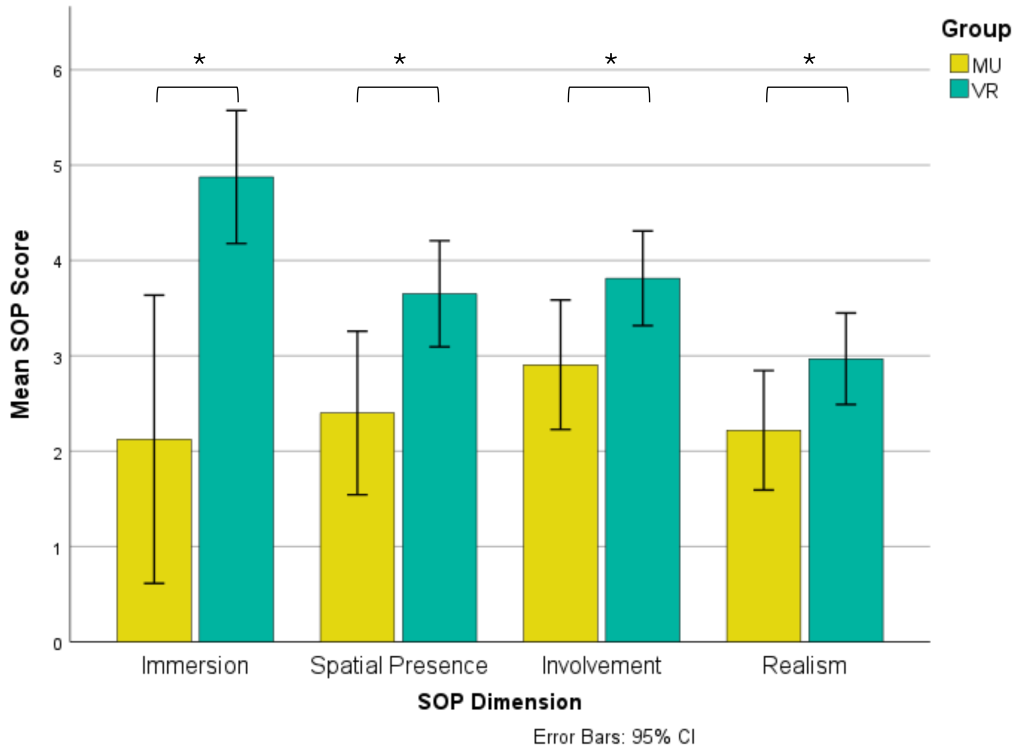

Figure 10 A bar graph displaying IPQ results, comparing presence experienced by participants in mock-up pre-training (MU) vs VR simulation. Four sub-dimensions, general immersion, spatial presence*,* involvement, and realism, are shown (* = $p < 0.05$).

5.3.2. Simulator Sickness Questionnaire Results

A correlation test was performed to investigate the relationship between cyber-sickness level, measured by the SSQ and the TPI of the last trial, when participants should feel the highest level of sickness induced by the VR simulation. The result was found to be not statistically significant (r = -.029, p = .458). These results indicate that there was no significant relationship between cybersickness symptoms experienced during the last trial of the VR sailing task and task performance of the last trial, meaning the null hypothesis, cyber-sickness does not affect the task performance of the last trial, is accepted.

5.4. Other Qualitative Results

Out of all the participants in the pre-training group (n = 8), 50% of them responded on a 5-point Likert scale stating the pre-training is "Somewhat Helpful", 25% of the participants responded "Very Helpful", and 25% responded "Slightly Helpful". When participants were asked to describe how the mock-up pre-training helped them to understand the task more effectively, they indicated that the layout was helpful in multiple ways. It reminded participants to slow down when making turns, gave them a basic understanding of the course, and boosted their confidence by providing a clear idea of what to expect. Some mentioned it offered a basic understanding of the course, emphasizing the tips such as staying close to the inner side of the course helped them to complete the route in a short time. Full qualitative feedback received can be found in *Appendix 2*.

6. Discussion

Pre-training's effect on task performance?

The first aim of the study is to explore the effect of pre-training on task performance. The results suggest that pre-training has no statistically significant effect on the performance index, including mean, 1st trial, and moderating the performance score over the three trials. One possible explanation for the lack of effect of mock-up pre-training on VR sailing task performance could be the limited transfer of procedural skills between the mock-up design and the VR environment. Although both the VR and mock-up pre-training broadly require similar procedural skills related to fine motor skills and hand-eye coordination, they also require distinct sets of psycho-motor abilities. The visual, auditory, and haptic cues inherent to each setting differ, leading to limited transferable learning effects. For instance, in the VR environment, participants manipulate controllers to grab onto the virtual steering wheel, steer the virtual boat, adjust the speed with the virtual throttle with precision, and engage in various arm extension motions. Quite distinctively, in mock-up scenarios, individuals handle small dice, calibrating and controlling finger movement to avoid obstacles and move the dice across a map board with precision. While the goal of the mock-up is strictly to provide a mental model of the VR environment, the procedural motor and spatial skill learning acquired from the mock-up pre-training may have been rendered irrelevant to the VR task.

Also, the error bars in *Figure 8* and the standard deviations (SD) in *Table 1* in *Appendix 5* show that all SDs are over 0.1 out of a 0-1 TPI ratio value, indicating a high level of variability between individual's task performance. The individual differences in cognitive mechanisms that underlie learning in VR such as spatial and psycho-motor abilities were not systematically controlled. The only skill-related eligibility requirement was that the participant must not have experience in operating the specific type of boat used in the VR task. Their inherent cognitive and motor skill level was not measured for baseline considerations.

In the case that certain participants have higher spatial and psycho-motor abilities, an expertise reversal effect could be present. The expertise reversal effect suggests that the effectiveness of a learning task varies depending on learners' prior skill levels (Kalyuga et al., 1998). In other words, what may be beneficial and non-redundant for novice learners could become redundant for those who are experts in a specific domain. In this study, for participants with higher levels of spatial and psycho-motor abilities, their existing skills may override the effectiveness of instructions and tips provided in the pre-training. Especially for skills as intuitive as steering and spatial navigation, practice on a mock-up paper board is not expected to produce any considerable effect in addition to the skills they already possess. The additional tips provided could all be tacit knowledge to participants who frequently partake in activities that require these skills, such as operating a vehicle in real life.

Pre-training increases cognitive load?

The results of the experiment contradict the hypotheses on cognitive load, revealing that 1) pre-training does not affect any of the task load dimensions in the NASA TLX, as well as ECL and GCL; and 2) pre-training did not lead to a decrease in ICL, but on the contrary, participants in the pre-training condition exhibits higher ICL than the no pre-training group. While the results challenge the theoretical assumptions backed by CLT, upon evaluation, this finding can likely be attributed to the design of the study. There are a few reasons that can potentially explain these findings – pre-training design, pre-training's target cognitive workload type, and other factors such as longer exposure time and the self-reporting subjective measures.

Pre-training design

In addition to the limited skill transfer between the pre-training and the VR simulation, the discrepancy between the design of the pre-training and the actual VR environment may have led to the increase in cognitive workload. Literature suggests that when novice learners are faced with a problem-solving task, they will first engage in solving problems that require a higher mental effort (Sweller & Cooper, 1985). When participants in the pre-training group entered the VR environment, it is possible that they spent additional cognitive workload to reconcile with, first, the visual discrepancy between the pre-training mock-up, and the VR simulation environment; and second, the tactile discrepancy between maneuvering with a dice and maneuvering with VR controllers, resulting in higher ICL than participants in the control group. In the qualitative data, some participants also mentioned their recommendation for improving the pre-training, including providing the exact same layout of the course in the pre-training, and implementing a very simple console-based game, e.g. on PlayStation or Xbox to allow for a more similar sensory experience as the VR simulation (see *Appendix 2*). Future studies can consider that to reduce the ICL of a novel VR simulation task.

Pre-training's target cognitive workload type

Another potential reason behind the increase in ICL is that the pre-training did not target the most cognitive workload-intensive task component. The design of the pre-training targets task difficulty, which is a form of ICL. However, since most of the participants have never, or have only used a VR device 1-3 times before, the most difficult part of the experience could be adapting to the VR environment itself (a form of ECL) rather than the task. The sensory stimuli present in this relatively high-fidelity VR environment, including visual, auditory, and tactile cues, can be overwhelming for novices, overloading their working memory. Additionally, learning how to effectively use the handheld controllers to navigate within the virtual environment can be complex for individuals with limited experience with interactive gaming technologies, which presents a high degree of ECL. Relating to literature, both Reedy's (2015) and Say et al.'s (2019) studies suggest that high-fidelity simulation does not always produce high performance, and that LF simulations with simplified interfaces and interactions are easier for novice learners to form necessary new schemas in early stages of the learning process. This is also confirmed by the data as there is no significant difference in ECL between the two groups (see *Figure 8*), as participants in both groups are likely experiencing the same level of ECL to adapt to the VR device functionality and learn to navigate in the virtual environment.

Longer experiment time

Another potential factor is that the pre-training group session is simply longer than the control group session. The experimental group session lasted for around 1.25-1.5 hours on average. Longer experiment duration may increase fatigue and decrease attentional resources, affecting working memory capacity. Participants may misattribute the fatigue experienced due to prolonged exposure to the VR task, to the cognitive workload experienced.

Self-reporting Subjective Measures only

Using self-reporting subjective measures exclusively for cognitive load assessment also introduces a considerable degree of subjectivity to the scoring measures. Participants' interpretations of their cognitive load levels may differ significantly, and it requires learners to be highly aware of the state of their cognitive workload, which varies based on individual knowledge, perceptions, and subjective experiences (Klepsch et al., 2017).

Other Findings

Repeat Exposure is a favourable strategy

Despite the lack of effect of pre-training on task performance, we found that there was a significant improvement in performance over three trials. The substantial performance improvement can be attributed to the learning effect as a result of repeated exposure to the task. This phenomenon is also called automaticity, where repeated practices lead to increased performance speed and reduced the necessary cognitive load to perform the task (Haith & Krakauer, 2018). This finding also aligns with Jensen and Konradsen's study (2018), where they found that repeating the target task in the same VR environment can positively affect memory retention, as well as visual-spatial and psychomotor skills.

This result sheds a positive light on the practical implications of automaticity in the context of VR-based learning. Repeated exposure, or practices, with VR devices and environment, is an intuitive and highly desirable way to increase proficiency, decrease cognitive load, and increase performance (Haith & Krakauer, 2018). This relationship can help to inform the design of training for high-risk and safety-critical tasks, where performance accuracy is crucial, and the cost of human errors is high. Future research can be done to determine the optimal number of repeated trials required to meet a certain performance threshold with the potential consideration of individual baseline differences.

Correlation between TLX's Effort and DCL's ICL dimensions

To my knowledge, this is the first study to use both the NASA TLX and Klepsch et al.'s (2017) DCL questionnaires to conceptualize and measure cognitive workload. While the goal of this study is not to evaluate the construct validity of these two questionnaires, our results showed that the Effort dimension in the TLX, although not significantly different between the control and experimental groups, had a larger effect size than other TLX dimensions. This dimension showed a similar level of sensitivity to pre-training as the ICL in Klepsch et al.'s DCL questionnaire (2017). This result differs from the findings of Naismith et al. (2015), who compared TLX dimensions with a cognitive load component (CLC) questionnaire they developed for simulation-based procedural training. Instead of Effort, Naismith et al. found correlations between the TLX Mental Demand dimension and the ICL dimension in their CLC questionnaire. The encouraging aspect is that these dimensions—TLX Mental Demand, TLX Effort, and ICL in both Naismith et al. (2015) and Klepsch et al.'s (2017) studies—are all direct or proxy measures for the inherent difficulty and element interactivity of a task. Future studies should further cross-reference the performance and the construct validity of the subjective cognitive workload measures used in this study.

Moderate intrinsic load and high germane load indicate productive learning

Also, as seen in *Figure 8.*, GCL and ICL mean scores are higher in comparison to ECL mean scores. Through this comparison, it is possible that a moderately difficult task, indicated by medium mean ICL scores and high mean GCL scores in both groups, is an ideal learning environment that leads to productive learning. GCL is high when learners effectively employ cognitive strategies to process information in the working memory and construct schemas in their LTM, which is essential for deep learning (Klepsch & Seufert, 2020). Kosch et al. (2023) also suggest that maintaining ICL at a "sweet spot" helps to foster task engagement while reducing frustration on the task. For these reasons, moderate ICL scores and high GCL scores indicate that participants are optimally challenged, and are not overwhelmed by the VR simulation. Thus, although pre-training had led to slightly higher ICL scores, both groups appear to be exhibiting a high degree of productive learning through the VR simulation alone, indicating that this high-fidelity VR simulation activity is effective in engaging learners in this particular learn-to-sail task.

7. Limitations

Small sample size, low power, high variability

It is important to highlight that this study has a small sample size (n = 16), with a between-subject experimental design. This could be the underlying reason for all the insignificant results. A small sample size decreases statistical power, making it difficult to detect true effects between conditions. Furthermore, the variability within a small sample can be higher as the effect of individual data points can be disproportionately large. Several mitigating measures were in place to reduce the impact of these outliers, such as the normality check to ensure data follows a normal distribution, and the TPI calculation, to normalize task performance scores, making them less sensitive to extreme values. Despite these measures, the low statistical power and high variability could still be the main reasons for the lack of statistical effect, and should be taken into account when interpreting the results.

No skill transfer is measured (no real-life testing of sailing)

The study did not evaluate skill transfer to real-world scenarios, such as actual sailing tasks, which might have offered insight into the practical applicability of the training. Instead, the assessment focused only on performance within the virtual simulation. Therefore, it cannot be used to validate the efficacy of training interventions beyond the simulated environment (Jensen & Konradsen, 2018). Consequently, it remains plausible that the training design purely enhanced participants' proficiency in navigating the virtual environment rather than developing skills transferable to sailing in real life.

8. Future Research

Exploration of learning transfer – Application of procedural knowledge acquired from virtual learning environments to real-world scenarios remains a promising direction for future research.

Learning transfer is a crucial factor in measuring the acquisition of procedural knowledge. In Cooper et al.'s study (2021), they found that VR training has a beneficial effect on learning transfer and improving real task performance. In the context of using VR as a training tool, this particular VR simulation can be used as a training tool to teach learners how to operate a RHIB performance in real life. In terms of using VR as a concept development tool, VR simulation can also be used as a visual aid, where the transfer can be measured by operators' understanding of the concept in the real-life environment to determine if the VR simulation is effective in introducing a novel concept compared to introducing the same novel concept through other methods.

Target reducing extrinsic load rather than intrinsic – as discussed, the insignificant results of this study may be an indication that extrinsic load, adapting to the functionality of the VR device, may be higher than the intrinsic load of the task. This is especially relevant for a relatively new technology like VR, which many people do not have prior experience with. This could mean implementing a practice session with the VR device on a simplified version of the learning task, or simply a pre-session on VR familiarization could work. We also recommend future studies to first analyze the target task difficulty to determine which task component(s) present a higher cognitive workload, and subsequently decide which task to pre-train learners on.

Integrate in-task cognitive workload measures – We did not employ in-task subjective cognitive load measure, but upon further research, in-task subjective mental state measures such as the Instantaneous Self-Assessment (ISA) technique (Jordan & Brennen, 1992) would be appropriate for this study, and should be considered for future research. In-task cognitive workload measure allows for comparison between trials, and determines if automaticity-based improvement reduces cognitive workload over the trials. We also did not employ in-task physiological measures because of the cybersickness risk associated with the VR scenario used, as the task was initially designed to induce cybersickness in a different experiment. Future studies with an appropriate VR scenario could employ physiological measures such as gaze behaviour tracking using an eye tracker, heart-rate tracking using an electrocardiogram (ECG) device, and skin conductance response using a galvanic skin response (GSR) device, or brain electrical activity using electroencephalography (EEG) to complement behavioural or subjective measures of cognitive load. If resources allow, researchers can use VR HMDs with built-in or add-on eye-tracking capabilities, or incorporate wearable eye-tracking devices into mixed reality (XR) environments to measure users' cognitive load in a learning task in a VR environment.

Investigation of other load-reducing strategies – Future studies could delve deeper into exploring other load-reducing strategies proposed by the CTML, such as off-loading – moving some essential information from the visual channel to the auditory channel, or segmenting – separating content into successive bite-size segments (Mayer & Moreno, 2010). Exploring ways to combine different load-reducing strategies for different VR content and evaluating their effectiveness in reducing cognitive load and improving the performance in the VR task could also be a viable direction.

9. Conclusion

In conclusion, this study explored the pre-training method as an attempt to reduce the cognitive load experienced in a procedural knowledge acquisition task in a VR simulation. Despite the robust theoretical considerations on principles from CLT and CTML, the pre-training designed for this task did not yield any statistically significant improvement in task performance and cognitive load reduction. Regarding task performance, the lack of significant effect can be attributed to the limited transferability of procedural skills between the mock-up pre-training and VR simulation due to modal differences in psycho-motor abilities and sensory cues. Secondly, the individual differences in spatial and psycho-motor abilities potentially cause an expertise reversal effect, where individuals' prior skills override the effect of pre-training instructions. Regarding cognitive workload, the increase in intrinsic load can be attributed to pre-training design, target cognitive workload type, and other factors such as longer exposure time and self-reporting subjective measures.

Despite these limitations, this study provides meaningful insights into the benefits of repeated exposure to a learning task in improving task performance. Future studies and training can consider allowing novice learners ample time to practice the same simulation task in a particular VR environment, especially for high-risk operational tasks that are crucial for safety. Moreover, results suggest a similar sensitivity to pre-training between the TLX Effort dimension and the ICL in Klepsch et al.'s DCL questionnaire, highlighting the need for future studies to cross-reference these constructs. Finally, a moderate level of ICL and a high level of GCL indicate that meaningful learning took place in this study, that participants are optimally challenged, and this VR scenario is a useful tool for procedural skill acquisition.

Overall, using VR as a “Serious Gaming” tool for training has tremendous potential to complement the traditional learning paradigm in the maritime sector. This study serves as a starting point to apply the pre-training cognitive load reduction strategy that is rooted in the well-established CLT and CTML to guide the design of VR-aided learning. It is hoped that this approach can open up future research directions to further optimize the use of VR technology as an educational tool in the maritime sector.

Additional Remark

This study was developed as an alternative research design of another more extensively researched topic: novel maritime concept development and mental model assessment. The former question aimed to study the effectiveness of different concept visualization tools and compare the effectiveness of VR and mock-ups in helping end-users visualize a novel operational concept. However, pilot testing revealed flaws in the original VR scenario, with experienced seafarers noting it was under-developed, and identified missing information and impractical interfaces. Also, upon evaluation with senior staff at MARIN, the original VR scenario was deemed to be too complicated for the research question and would be counterproductive to the study goal. For these reasons, the research topic was changed halfway through this thesis development process, resulting in a relatively simpler experimental design to fit into the remaining allotted time for the Master's thesis. The literature review, experiment, and completion of the first draft were completed within two months.

References

Anderson, J. R.: 1990, Cognitive Psychology and Its Implications (Third edition), W. H. Freeman and Company, New York.

Bingchan, L., Mao, B., & Cao, J. (2018). Maintenance and management of marine communication and navigation equipment based on virtual reality. *Procedia Computer Science*, *139*, 221-226.

Breuer, J., & Bente, G. (2010). Why so serious? On the relation of serious games and learning. *Journal for Computer Game Culture*, *4*, 7-24.

Buttussi, F., & Chittaro, L. (2017). Effects of different types of virtual reality display on presence and learning in a safety training scenario. *IEEE Transactions on Visualization and Computer Graphics*, 24(2), 1063–1076. https://doi.org/10.1109/TVCG.2017.2653117

Checa, D., & Bustillo, A. (2020). A review of immersive virtual reality serious games to enhance learning and training. *Multimedia Tools and Applications*, *79*(9), 5501-5527.

Chen, R., Grierson, L. E., & Norman, G. R. (2015). Evaluating the impact of high-and low-fidelity instruction in the development of auscultation skills. *Medical Education*, *49*(3), 276-285.

Cooper, N., Millela, F., Cant, I., White, M. D., & Meyer, G. (2021). Transfer of training—Virtual reality training with augmented multisensory cues improves user experience during training and task performance in the real world. *PloS one*, *16*(3), e0248225.

Dieckmann, P., Gaba, D., & Rall, M. (2007). Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare*, *2*(3), 183-193.

Fraser, K. L., Ayres, P., & Sweller, J. (2015). Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, *10*(5), 295-307.

Haith, A. M., & Krakauer, J. W. (2018). The multiple effects of practice: skill, habit and reduced cognitive load. *Current Opinion in Behavioral Sciences*, *20*, 196-201.

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, *89*(3), 387-392.

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload.* Amsterdam: North Holland Press.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage publications.

Hollands, J. G., Spivak, T., & Kramkowski, E. W. (2019). Cognitive load and situation awareness for soldiers: effects of message presentation rate and sensory modality. *Human Factors*, *61*(5), 763-773.

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*, 1515-1529.

Jung, J., Shin, Y., & Zumbach, J. (2021). The effects of pre-training types on cognitive load, collaborative knowledge construction and deep learning in a computer-supported collaborative learning environment. *Interactive Learning Environments*, *29*(7), 1163-1175.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1–17

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

Kim, T. E., Sharma, A., Bustgaard, M., Gyldensten, W. C., Nymoen, O. K., Tusher, H. M., & Nazir, S. (2021). The continuum of simulator-based maritime training and education. *WMU Journal of Maritime Affairs*, *20*(2), 135-150.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, 1997.

Klepsch, M., & Seufert, T. (2020). Understanding instructional design effects by differentiated measurement of intrinsic, extraneous, and germane cognitive load. *Instructional Science*, *48*(1), 45-77.

Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., & Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, *55*(13s), 1-39.

Liu, D., Yu, J., Macchiarella, N. D., & Vincenzi, D. A. (2008). Simulation fidelity. In *Human Factors in Simulation and Training* (pp. 91-108). CRC Press.

Marucci, M., Di Flumeri, G., Borghini, G., Sciaraffa, N., Scandola, M., Pavone, E. F., ... & Aricò, P. (2021). The impact of multisensory integration and perceptual load in virtual reality settings on performance, workload and presence. *Scientific Reports*, *11*(1), 4831.

Mayer, R. E., & Pilegard, C. (2005). Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. *The Cambridge Handbook of Multimedia Learning*, 169-182.

McCormick, R. (1997). Conceptual and procedural knowledge. *International Journal of Technology and Design Education*, *7*, 141-159.

McWilliams, T., & Ward, N. (2021). Underload on the road: measuring vigilance decrements during partially automated driving. *Frontiers in Psychology*, *12*, 631364.

Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education*, *140*, 103603.

Miyusov, M. V., Nikolaieva, L. L., & Smolets, V. V. (2022). The Future Perspectives of Immersive Learning in Maritime Education and Training. *Transactions on Maritime Science*, *11*(02), 14-14.

Moreno, R., & Mayer, R. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review*, *19*, 309-326.

Naismith, L. M., Cheung, J. J., Ringsted, C., & Cavalcanti, R. B. (2015). Limitations of subjective cognitive load measures in simulation-based procedural training. *Medical Education*, *49*(8), 805-814.

Nelson, B. C., & Erlandson, B. E. (2008). Managing cognitive load in educational multi-user virtual environments: Reflection on design practice. *Educational Technology Research and Development*, *56*, 619-641.

Orru, G., & Longo, L. (2019). The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: a review. In *Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, The Netherlands, September 20-21, 2018, Revised Selected Papers 2* (pp. 23-48). Springer International Publishing.

Paas, F., Van Gog, T., & Sweller, J. (2010). Cognitive load theory: New conceptualizations, specifications, and integrated research perspectives. *Educational Psychology Review*, *22*, 115-121.

Reedy, G. B. (2015). Using cognitive load theory to inform simulation design and practice. *Clinical Simulation in Nursing*, *11*(8), 355-360.

Rogers, B. A., & Franklin, A. E. (2021). Cognitive load experienced by nurses in simulation-based learning experiences: An integrative review. *Nurse Education Today*, *99*, 104815.

Salcedo, J., Lackey, S., & Badillo-Urquiola, K. (2015). Leveraging stress and intrinsic motivation to assess scafolding during simulation-based training. In R. Shumaker & S. Lackey (Eds.), Virtual, augmented and mixed reality. VAMR 2015 (pp. 309–320). Springer. https://doi.org/10.1007/978-3- 319-21067-4_32

Say, R., Visentin, D., Betihavas, V., & Minutillo, S. (2019). A cognitive load theory simulation design to assess and manage deteriorating patients. *International journal of nursing education scholarship*, *16*(1), 20190009.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, *10*(3), 266-281.

Steam (n.d.). Valve Index VR Kit. Retrieved from https://store.steampowered.com/sub/354231/

Steam (n.d.). Index Headset. Retrieved from https://www.valvesoftware.com/en/index/headset

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59-89.

Sweller, J., van Merriënboer, J. J. G., and Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205

Sweller, J., Ayres, P., Kalyuga, S. (2011). The Redundancy Effect. In: Cognitive Load Theory. Explorations in the Learning Sciences, *Instructional Systems and Performance Technologies, vol 1.* Springer, New York, NY. https://doi.org/10.1007/978-1-4419-8126-4_11

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., ... & Yu, L. F. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, *2*, 645153.

Appendices

Appendix 1. Full list of questionnaire questions used in this study.

Differentiated cognitive load questionnaire (Klepsch et al., 2017)

Item #	Type of Load	Question
1	ICL	For this task, many things needed to be kept in mind simultaneously.
2	ICL	This task was very complex.
3	GCL	I made an effort, not only to understand several details, but to understand the overall context.
4	GCL	My point while dealing with the task was to understand everything correctly.
5	GCL	The learning task consisted of elements supporting my comprehension of the task.
6	ECL	During this task, it was exhausting to find the important information.
7	ECL	The design of this task was very inconvenient for learning.
8	ECL	During this task, it was difficult to recognize and link the crucial information.

NASA TLX Questionnaire (Hart & Staveland, 1988)

Item #	Dimension	Question
1	Mental Demand	How mentally demanding was the task?
2	Physical Demand	How physically demanding was the task?
3	Temporal Demand	How hurried or rushed was the pace of the task?
4	Performance	How successful were you in accomplishing what you were asked to do?
5	Effort	How hard did you have to work to accomplish your level of performance?

6	Frustration	How insecure, discouraged, irritated, stressed, and annoyed were you?

Igroup Presense Questionnaire (IPQ) (Schubert et al., 2001)

G = General Immersion

SP = Spatial Presence

INV = Involvement

REAL = Realism

Item #	**Construct**	**Questions**
1	G1	In the learning environment, I had a sense of "being there"
2	SP1	Somehow I felt that the learning environment surrounded me.
3	SP2	I felt like I was just perceiving pictures.
4	SP3	I did not feel present in the learning environment.
5	SP4	I had a sense of acting in the learning environment, rather than operating something from outside.
6	SP5	I felt present in the learning environment.
7	INV1	How aware were you of the real world surrounding while navigating in the learning environment? (i.e. sounds, room temperature, other people, etc.)?
8	INV2	I was not aware of my real environment.
9	INV3	I still paid attention to the real environment.
10	INV4	I was completely captivated by the learning environment.
11	REAL1	How real did the learning environment seem to you?
12	REAL2	How much did your experience in the learning environment seem consistent with your real world experience?
13	REAL3	How real did the learning environment seem to you?
14	REAL4	The learning environment seemed more realistic than the real world.

Simulator Sickness Questionnaire (SSQ) (Kennedy et al., 1993)

Item #	**Symptoms**
1	General Discomfort

2	Fatigue
3	Headache
4	Eyestrain
5	Difficulty focusing
6	Increased salivation
7	Sweating
8	Nausea
9	Difficulty concentrating
10	Fullness of head
11	Blurred vision
12	Dizzy (eyes open)
13	Dizzy (eyes closed)
14	Vertigo
15	Stomach awareness
16	Burping

Appendix 2. Qualitative feedback on pre-training.

Please describe how it helped you to understand the task more effectively.	***How can the pre-training mock up activity be improved?***
The general layout of the map/track	*Possibly by having the same exact layout of checkpoints*
It remind me to slow down when making a turn	*Explain how many turns, what the fastest way of turning is*
Understand the parcour (course) I needed to sail, make slow turns	
Knowing more about what I'm going to face. Feel more confidence!	*The distances between each of the gates could be made more clear somehow, I was surprised by how close together they were.*
It gave me a basic understanding of what I was to expect	*maybe implementing a very simple PlayStation or Xbox game very close to the physical one proposed by the researcher, so that the person under examination would use both their hands like it happens in the VR simulation*
I knew I had go to through gates frontally while respecting a preceding order and possibly completing the route in the fastest way	*Pre training mockup should have a visual representation of the throttle and wheel*
Explanation and directions made me have a good idea of course outline	
i know try to be close to the inner side	

Appendix 3. Equal variance (Levene's test) and normal distribution (Shapiro-Wilk test) check scores.

TPI Mean scores:

Levene's test revealed that the variance of the mean TPI is approximately equal ($F = 3.18$, $p = .082$). Shapiro-Wilk test revealed that the mean TPI scores are considered approximately normally distributed ($W = .870$, $p = .616$). Based on the TPI mean scores can be interpreted using independent samples t-test.

TPI 1st Trial scores:

Levene's test revealed that the variance of the first trial TPI scores is approximately equal ($F = 1.794$, $p = .202$). Shapiro-Wilk test revealed that the TPI mean scores are considered approximately normally distributed ($W = .935$, $p = .289$). Based on these findings, the first trial TPI scores can be interpreted using an independent samples t-test.

NASA TLX dimensions:

For equal variance, physical demand ($F = 5.497$, $p = .034$) and effort ($F = 6.512$, $p = .023$) violated the assumption of both equal variances and distributed normally. For normality, temporal demand ($W = .758$, $p = <.001$) and effort ($W = .869$, $p = .026$) didn't meet the normal distribution assumptions.

NASA TLX Dimensions	Levene's Test (Equal Variances)	Shapiro-Wilk Test (Normality)
Mental Demand	.802	.080
Physical Demand	.034*	.263
Temporal Demand	.377	<.001*
Performance	.113	.688
Effort	.023*	.026*
Frustration	.945	.106

Note: The significance level is .050.

DCL:

Upon testing the assumption of equal variance and normality for the DCL measure, the Shapiro-Wilk test showed that the data is normally distributed across all three cognitive load types, with ICL (W = 0.941, p = 0.364), GCL (W = 0.948, p = 0.455), and ECL (W = 0.944, p = 0.401) respectively. Also, Levene's test indicated that all variables have equal variances (ICL: p = 0.612; GCL: p = 0.852; ECL: p = 1.000)., meaning a t-test can be used for the DCL measurement.

Appendix 4. Full pre-training script.

<u>Story/Setting (Individual Components):</u>

- **Purpose:** The purpose of this mock-up pre-training is to help you visualize the VR environment and get familiarized with the VR simulation task components and the interaction between them before the actual VR task.
- **Setting:** Imagine you are the captain of a RHIB (the fast boat) and you are tasked to operate the RHIB through a slalom course. The RHIB is represented by the black and white dice. The slalom course contains multiple buoy obstacles, they are represented by the coloured dice.
- **Task:** In this task, you have to navigate through this course from start point to finish point <u>as fast as possible, take the shortest route possible, and miss as few obstacles as possible</u>. The course is marked by buoys that indicate a zigzag path that you must follow.
- There are 5 obstacles in total in this mock-up, but the VR task contains 14 in total. One center in the beginning, then right, left, right, and then the last one (14th) is also at the center.
- **Rule:**
    - Hitting a buoy incurs a time penalty – you have to restart from the beginning.
    - There will be 5 trials, the first 2 are for practicing, last 3 will be timed.
    - Your ship should be touching the map the whole time, don't lift it up
    - You cannot pass through the obstacle from the back/from behind. That incurs an error.

<u>Tips (to highlight element interactivity and indicate causal relationship between elements):</u>

Here are some tips (structure: tip, then reasoning):

- Be mindful of your speed and direction – Higher speeds increase the risk of losing stability, especially during turns. Slowing down minimizes the chances of capsizing or losing control.
- Slow down the boat before the obstacles to improve the precision of the turn – it gives you more time to adjust the course accurately. Reducing speed also provides you with more time to assess the surroundings and obstacles, improving situational awareness.
- Look ahead and anticipate upcoming turns or obstacles. Planning your maneuvers in advance allows for smoother adjustments and reduces the chance of sharp and sudden turns, which may lead to overshoot and loss of control.

- Avoid accelerating or decelerating sharply during a turn, and make smooth and controlled steering movements, that help to maintain the stability and balance of the boat.
- Strive to maintain a steady and consistent boat speed throughout the course. – This may help to reduce motion sickness, as well as cybersickness induced by the VR.
- If you miss a turn, the best way is to try again by backing up instead of going a big circle. This is the best way to save time and length.
- In the actual simulation, waves could have an impact on the speed. Try to avoid hitting the waves head-on. Avoiding hitting waves head-on can help you maintain your speed, provide control, and reduce motion sickness and cybersickness induced by the VR.

Appendix 5. Data tables.

Table 1. Descriptive Statistics for between-subject factorial analysis of variance (ANOVA) of TPI over three trials.

Condition	Trial Sequence	*M*	*SD*	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
No Pre-training	1	.315	.180	.189	.442
	2	.415	.241	.289	.541
	3	.566	.225	.420	.712
Pre-training	1	.282	.124	.156	.409
	2	.450	.106	.324	.577
	3	.483	.150	.337	.629

Note: n = 16. The significance level is .050.

Table 2. Descriptive Statistics NASA TLX questionnaire.

	Condition (1 = NPT; 2 = PT)	*M*	*SD*	Independent-Samples Mann-Whitney (*U*)	Effect Size (*r*)
Mental Demand	1	67.50	7.071		
	2	67.50	8.864	.878	0.042
Physical Demand	1	41.88	17.513		
	2	41.25	32.814	.721	-0.106
Temporal Demand	1	61.25	22.160		
	2	65.63	17.410	.878	0.040
Performance	1	45.00	16.903		
	2	41.25	27.484	.798	-0.079
Effort	1	64.38	9.425		
	2	76.25	15.980	.105	0.425*
Frustration	1	61.25	19.226		
	2	50.00	17.113	.328	-0.255

Note: n = 16, significance level for independent-samples Mann-Whitney *U* is .050.

Table 3. Descriptive statistics for DCL scores (Klepsch et al., 2017).

	Group (1=ctrl, 2=exp)	*M*	*SD*
ICL	1	4.000	1.195
	2	5.063	1.016
	Combined	4.531	1.204
GCL	1	5.125	0.755
	2	5.250	0.922
	Combined	5.188	0.816
ECL	1	3.000	1.141
	2	3.000	1.155
	Combined	3.000	1.109

Appendix 6: Detailed hypotheses.

To elaborate on hypotheses from *This Study* section, the following are the detailed sub-questions and hypotheses that correspond to the main research question:

Task performance

- ***How does pre-training affect the Task Performance Index (TPI) mean score?***
    - Null Hypothesis (H0): pre-training does not lead to an increase in the TPI mean score.
    - Alternative Hypothesis (H1): pre-training leads to an increase in the TPI mean score.
- ***How does pre-training affect the TPI of 1st trial?***
    - H0: pre-training does not lead to an increase in the TPI of the 1st trial.
    - H1: pre-training leads to an increase in the TPI of the 1st trial.
- ***How does pre-training affect the TPI over the three trials?***
    - H0: Pre-training does not moderate an increase in the TPI over three trials.
    - H1: Alternative hypothesis: Pre-training moderates an increase in the TPI over three trials.

Cognitive Workload:

- ***How does pre-training affect each of the differentiated cognitive load types individually – intrinsic, extraneous, and germane load?***
    - H0: ICL – Participants who receive pre-training experience the same level of ICL compared to participants who do not receive pre-training.
    - H1: ICL – Participants who receive pre-training experience lower levels of ICL compared to participants who do not receive pre-training.
    - H0: ECL and GCL – There is no significant difference in the levels of ECL and GCL between participants who receive pre-training and those who do not.
    - H1: ECL and GCL – There is a significant difference in the levels of ECL and GCL between participants who receive pre-training and those who do not.

- ***How does pre-training affect each of the load dimensions in NASA TLX individually?***

- H0: Mental load dimension – Participants who receive pre-training experience have the same level of mental demand on the NASA TLX compared to participants who do not receive pre-training.
- H1: Mental load dimension – Participants who receive pre-training experience a lower level of mental demand on the NASA TLX compared to participants who do not receive pre-training.
- H0: Other 5 dimensions – There is no significant difference in the levels of physical demand, temporal demand, performance, effort, and frustration between participants who receive pre-training and those who do not.
- H1: Other 5 dimensions – There is a significant difference in the levels of physical demand, temporal demand, performance, effort, and frustration between participants who receive pre-training and those who do not.

Other Hypotheses:

- ***How does SOP differ between low-fidelity (LF) mock-up pre-training and high-fidelity (HF) VR simulation?***
  - H0: There is no difference between the SOP level in the LF mock-up and the HF VR simulation.
  - H1: The LF mock-up pre-training offers a low level of SOP, and the HF VR simulation offers a high level of SOP.
- ***How does cyber-sickness affect task performance?***
  - H0: Cyber-sickness does not affect the TPI of the last trial.
  - H1: Cyber-sickness affects the TPI of the last trial.