Scenario design, data measurement approaches, and analysis methods in maritime simulator training: A systematic review

Ziaul Haque Munim[1], Helene Luise Sonna Krabbel[1], Per Haavardtun[1], Tae-Eun Kim[2], Morten Bustgaard[1], Haakon Thorvaldsen[1]

[1] Faculty of Technology, Natural, and Maritime Sciences, University of South-Eastern Norway, 3184 Horten, Norway.
[2] Department of Technology and Safety, University of Tromsø (UiT) -The Arctic University of Norway, 9019 Tromsø, Norway
Corresponding author: ziaul.h.munim@usn.no

Abstract. Developing an objective assessment approach in maritime simulator training can be a highly challenging task due to the complexity of simulating realistic scenarios, capturing relevant performance indicators and establishing good assessment protocols. This study provides a synthesis of simulation scenario contexts, data collection approaches, and data analysis methods in published simulator training studies. A systematic literature review (SLR) approach was adopted for identifying the relevant studies and analysing their content in-depth. The findings reveal that the reviewed studies focused on full-mission simulator-based assessment using collected data from various tools including surveys, eye-tracking, ECG, video or voice recording etc. The findings hold relevance in the development of learning analytics for facilitating objective assessment in maritime simulator training.

Keywords: Navigation simulator, Navigation competence, Training assessment, Learning analytics.

1 Introduction

Objective assessment approaches in maritime simulator training and education have been consistently advocated for implementation [1]. While rapid technological advancements in simulation modalities have diversified the maritime learning approaches and delivery methods, the assessment aspect still mainly relies on human involvement. Instructors play an essential role in evaluating the learning performance and determining the competence and readiness of the learners. While the Standards of Training, Certification and Watchkeeping for Seafarers (STCW convention and codes) provides competence requirements and outlines assessment criteria, the specific assessment methods, scoring criteria and detailed performance standards may vary depending on the instructors and training institutions.

Recently, scholars have been investigating alternative assessment approaches in maritime simulator training and seek to explore innovative ways to evaluate seafarers' performance and competence. For example, an automatic assessment approach was explored by Juszkiewicz and Żukowska [2]. The assessment approaches or tools need to be perceived as reliable and useful by the instructors and students. Munim and Kim [3] proposed a learning analytics dashboard (LAD) outline based on underlying pattern detection from simulator data. To develop a functioning LAD, data from multiple sources can be collected and analysed. The focus of assessment, data collection, and analysis approaches have to be aligned. Hence, the following research questions are addressed in this study:

(1) What are the characteristics of the simulation scenarios in terms of context, duration, and level of complexity?
(2) What are the data collection approaches used in published maritime simulation training studies?
(3) What are the common data analysis methods and tools in the assessment of maritime simulator training?

2 Methodology

Data for this review study was collected using a systematic approach. The scientific records database Scopus was used to identify relevant articles. Scopus is one of the largest databases and widely used in literature review studies, see for example Macke and Genari [4]. The literature search was conducted using Boolean expressions to identify relevant articles on various assessment approaches of maritime simulator training in March 2023. The search was limited to titles, abstracts, and keywords, and only journal articles written in English, which reverted to 348 potential records. The Boolean expression is reported in Table 1.

After the initial search, articles published prior to 2010 were not considered (86), leaving 262 articles for manual scrutiny. The abstracts of the records were screened to identify the studies that are highly relevant to the research questions of this study. Most of the studies (141) did not focus on Maritime Education and Training (MET) at all or were focused on general simulator set-up without involving assessment approaches. Two papers were duplicate but not recognized as such initially due to minor inconsistencies in the title; 17 focused on the role of the instructor in MET, and not on the assessment of navigational competency; 18 on the engine simulation; 36 were related to autonomous shipping, and 9 were on ports. After screening the abstracts, 36 studies were left, which were read in detail. Out of these, 13 were considered of high relevance. One article was added manually at this point. Hence, a total of 14 articles were transcribed in a literature review matrix to compose the results presented in the next section.

Table 1. Literature search terms

Search string 1: ("maritime" AND "simulat*" AND ("training" OR "education"))
OR
Search string 2: ("maritime" OR "marine" OR "sea") AND ("pilotage" OR "navigation*" OR "bridge" OR "seafarer") AND ("education" OR "training") AND ("simulator*" OR "simulation")

3 Results

A qualitative content analysis approach was adopted to answer the research questions. The 14 relevant studies were coded for content using a structured literature review matrix in Microsoft Excel. The matrix included coding of simulator use information, description of simulation scenario, exercise duration, key performance indicators (KPIs) for assessment, variables for assessment, data collection approaches, and analysis methods and software. A summary of the reviewed studies is presented in Table 2.

Table 2. Summary of the reviewed studies

Study	Data collection	Data frequency	Data analysis	Software
[5]	Survey	10 minutes interval	Descriptive statistics, Spearman Rank Difference mean	Excel
[6]	Eye-tracking	Continuously, 100 Hz	Descriptive statistics, Kruskal Wallis test, Mann-Whitney U Test, Heat-map	Tobii Pro Lab Analyzer; SPSS
[7]	ECDIS	Every 10 seconds	Deviation calculation	Not reported, most likely probably Excel
[2]	SEA tool in the simulator, observation	NA	SEA and review by experts	SEA program (automatic evaluation) and evaluation by

				qualified persons
[8]	EEG; Videos to label the events and match them	Continuously	SVM; EEG data processing; Evaluating the brain states using ML	C++ on Visual Studio; HTML and JavaScript
[9]	1. EEG 2. ECG 3. Eye-tracking 4.1 NASA TLX 4.2. Likert scale	1-3 continuously; 4.1 after each exercise; 4.2. every 2mins	Likert scale transformed into a continuous numerical function	Kubios HRV
[10]	Data from the simulation	NA	ANOVA; Perason correlation	IBM SPSS
[11]	Cameras, microphones, eye tracking, simulator data storage	Continuously	Tobi Pro Glasses 2	Tobii Pro software development kit (SDK)
[12]	Revised NEO Personality Inventory, SARS questionnaires; ECG; Simulator data	5-min interval for ECG; The revised NEO Personality Inventory one week prior, SARS after the simulation	Hierarchical multiple regression; Pearson correlation; ANOVA: Turkey's post hoc test.	NA
[13]	Video recording	Continuously	Cataloguing to Excel, analytical search	Excel

[14]	SA: 15 questions; Decision making: questionnaire; Navigational performance: Simulator data	Simulator data: Continuously;	Mann–Whitney U test; Wilcoxon signed-rank test	NA
[15]	Observations from the instructors and participants	Continuously	Safety compliance	NA
[16]	Empatica E4 band (wrist)	Body temp: 4 Hz; in °C; EDA: 4 Hz, in μS; BVP: 64 Hz; PPG; all signal are downsampled to 1 Hz for data analysis	Convolutional neural network	NA
[17]	fNIRS data; Ship distance data from simulator	Continuously	ANOVA, Artificial Neural Networks	NA

Simulation scenario characteristics

All the studies in the sample used full-mission simulators. Four studies reported the use of Kongsberg simulators [2, 5, 7, 16], while others did not report the simulator provider's name. The simulation scenarios varied in terms of context, duration, and level of complexity. The majority of the simulation scenarios included anti-collision scenarios [2, 5, 6, 13, 14, 17]. Some focused on sub-sea gas compression installations [11], manoeuvring [8, 10], docking [7], berthing [9] and offloading operations [15]. Overall, simulation exercises included tasks related to navigation (situational awareness, Radar, ECDIS), watchkeeping, manoeuvring, docking/berthing/offloading (dynamic positioning), collision avoidance (COLREG regulations), radio reporting, and decision making.

Some of the sample studies described the simulation scenario context in detail. Simulation scenarios covered real voyages in the Strait of Gibraltar [5], the Oslofjord [7], the coast of Western Norway [11], and the Dover Strait in the English Channel [13]. Others focused more on the tasks, required activities, and/or phases of operations. For instance, the simulation training experiment in Orlandi et al., [10] included two phases: (1) developing four manoeuvring plans, and (2) executing the plans during voyage. Likewise, the task requirement in Juszkiewicz & Żukowska [2] included turning on and adjusting the ARPA to weather conditions, acquiring targets, assessing the situation, planning and executing an anticollision maneuver (if necessary), and also planning and executing a returning maneuver (if necessary). Studies that attempted to assess higher levels of navigation competency, usually included anti-collision, decision making [6, 17], and berthing operations [9, 10].

The length of simulation scenarios varied among the sample studies to a great extent. The shortest reported is 13 minutes [6] and longest ones lasted up to 02 hours [9, 15]. The scenario length was about 60 minutes in several studies [8, 14, 16]. Most of the simulation scenarios can be divided into three to five phases. Typical phases are familiarization or baseline (5-15 minutes), operational tasks (5 to 10 minutes), duties or watchkeeping (15-30 minutes), crisis/emergency management or decision making (5-10 minutes), and feedback or debriefing (15-30 minutes). It is recommended to include these phases in future maritime simulation scenario design.

Data collection approaches

Based on the coding of information in the structured literature review matrix, a framework guiding the adoption of data collection approaches for assessing various navigation competency KPIs is proposed in Figure 1. In the sample of reviewed studies, five types of navigation competency KPIs are assessed: (1) navigation performance mainly through deviation from the planned route or course, (2) psychological and psychophysical factors, (3) situational awareness, (4) navigation rules and regulations, and (5) crisis management and decision making.

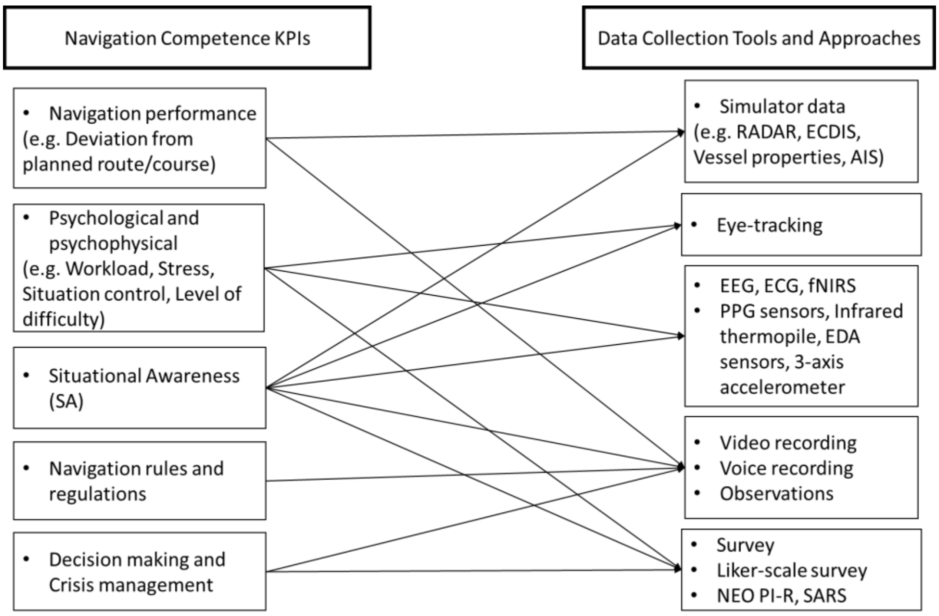
The tools and approaches to data collection varied depending on the type of navigation competency KPIs. To assess *(Type 1)* navigation performance, data were collected mainly from the simulator log data [2, 7, 10, 14]. Simulator data included ECDIS, RADAR/ARPA, AIS, and vessel properties data. Additionally, observation by the instructor was reported [2].

For assessing *(Type 2)* psychological and psychophysical factors such as workload, stress level, situation control, and level of difficulty, data were collected using eye-tracking, Electroencephalogram (EEG), Electrocardiogram (ECG), self-reported likert scale surveys [8, 9], and functional near-infrared spectroscopy (fNIRS) [17]. While EEG and fNIRS are used for measuring brain activity, ECGs are used to record the electrical activity of the heart.

Several studies focused on the assessment of *(Type 3)* situational awareness (SA), which has been assessed using data from diverse tools and approaches. Atik [6] has used only eye-tracking data for SA assessment. In addition to eye-tracking and simulator data, Sanfilippo [11] used data from video recording cameras and voice recordings from microphones for SA assessment. Türkistanli and Kuleyin [14] used survey questions multiple times at different phases of the simulation exercise. SA was assessed through variation in psychological and psychophysical measures as well using cardiovascular responses (CVR), heart rate variability (HRV, can be also derived from blood volume pulse), photoplethysmograph (PPG) sensor (measures blood volume pulse), infrared thermopile (reads peripheral skin temp), EDA sensor (measures the fluctuating changes in certain electrical properties of the skin), and 3-axis accelerometer (to capture motion-based activity) [12, 16]. The Empatica E4 (wrist) band was used for data collection by Xue et al., [16]. Saus et al., [12] used data from multiple sources such as the Situational Awareness Rating Scale (SARS), Revised NEO Personality Inventory (NEO PI-R), Ambulatory Monitoring System using 1 cm Ag/AgCl ECG electrodes, and simulator data.

To assess (*Type 4*) navigation rules and regulations competency, video recording and observation were used [2, 13]. Finally, (*Type 5*) crisis management and decision-making were assessed using data collected through observations from the instructors and participants [15], and questionnaire surveys [14].

Figure 1. A framework for data collection in the assessment of maritime simulator training



Abbreviations: Situational Awareness Rating Scale (SARS), Revised NEO Personality Inventory (NEO PI-R)

Data analysis methods and tools

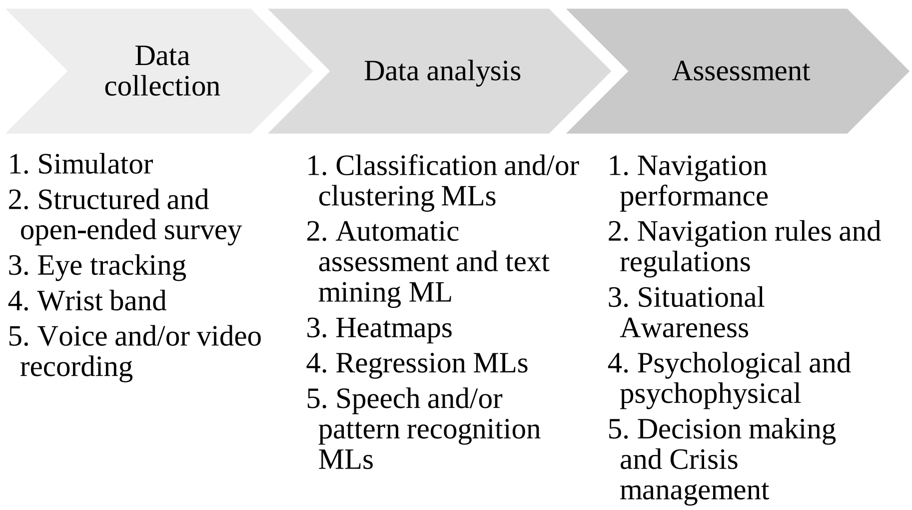
Data collected through various tools and approaches were mostly analysed using simpler statistical methods and widely used software. Descriptive analysis and mean comparison tests such as ANOVA, Kruskal-Wallis, Spearman Rank, and Mann-Whitney tests were evident in experimental studies [5, 6, 10, 14, 17]. Use of correlation and multiple regressions were evident, too [12]. In addition to statistical analysis, heat map analysis was used for eye-tracking [6]. For navigation performance assessment using simulator data, deviations were calculated [2, 7]. For EEG, ECG, and fNIRS data analysis, machine learning approaches have been used, for instance, artificial neural networks (ANN) [17], convolutional neural network (CNN) [16], and support vector machine (SVM) [8].

In terms of software used, Tobii Pro was used when eye-tracking data was analysed [6, 11]. SPSS and Excel were mainly used for statistical analysis and deviation calculations [5, 10]. For machine learning applications, the software was not explicitly reported, but now-a-days ANN, CNN, and SVM can be used though many software packages including Matlab, Stata, R, python etc.

Proposal for a comprehensive assessment approach

Since an inventory of data collection approaches, analysis methods, and how they are used in maritime simulator training have identified, a comprehensive assessment approach is proposed utilizing them in Figure 2. For instance, to assess navigation performance, simulator data can be used which can be analysed using classification and/or clustering ML algorithms. Similarly, eye tracking can be used to assess situational awareness by heatmap analysis. This approach can be adopted to develop a LAD for students and instructor in MET. For assessment of five dimensions of MET, at least five different data sources are recommended, and should be analysed through at least five separate models or algorithms, but the outcomes need to be visualized in one dashboard. The most important information, depending on the user’s (student or instructor) preference, should stand out in the dashboard.

Figure 2. Proposed assessment approach integrating emerging technologies



4 Conclusions and future research

This study provides a synthesis of simulation design characteristics, data collection approaches, and analysis tools in maritime simulation training. A systematic literature review approach was adopted to identify relevant published studies on the topic and analyse their contents. The findings reveal that the simulation scenarios varied in terms of context, duration, and level of difficulty. Data collection approaches include simulator data (including ECDIS and Radar data), eye-tracking, questionnaire surveys, observations, EEG, ECG, fNIRS, wristbands, video recording, and voice recording. Statistical analysis utilizing descriptive statistics and mean comparison tests were the most common analysis methods, while SPSS and Excel were the main analysis software tools. Based on the findings, a comprehensive assessment approach in maritime simulator training has been proposed.

Since the reviewed studies only focused on full-mission simulation training, future research should focus more on other simulator modalities such as desktop, cloud and virtual reality. Future development of navigation simulator training might further benefit from the adaptation of the metaverse. Despite the use of wearable tools in data collection in the reviewed studies such as eye-tracker, fNIRS, EEG and ECG, their feasibility of use during regular training sessions needs to be assessed. Further, in terms of data analysis, machine learning applications were limited in the sample of reviewed studies, which should be further explored in future research.

Funding

This research is funded by the European Union’s Horizon Europe research and innovation programme under grant agreement No 101060107.

References

[1] H. Kobayashi, “Use of simulators in assessment, learning and teaching of mariners,” *WMU Journal of maritime affairs*, vol. 4, pp. 57–75, 2005.

[2] W. Juszkiewicz and A. Żukowska, “The Use of the K-Sim Polaris Simulator in the Process of Automatic Assessment of Navigator Competence in the Aspect of Anticollision Activities,” *Applied Sciences*, vol. 13, no. 2, p. 915, 2023.

[3] Z. H. Munim and T.-E. Kim, “A Review of Learning Analytics Dashboard and a Novel Application in Maritime Simulator Training,” in *Proceedings of the AHFE Conference*, 2023.

[4] J. Macke and D. Genari, “Systematic literature review on sustainable human resource management,” *Journal of cleaner production*, vol. 208, pp. 806–815, 2019.

[5] J. I. A. Jiménez, “Modelling the relationship between performance and ship-handling simulator,” *Journal of Maritime Research*, vol. 17, no. 3, pp. 68–73, 2020.

[6] O. Atik, “Eye tracking for assessment of situational awareness in bridge resource management training,” *Journal of eye movement research*, vol. 12, no. 3, 2019.

[7] K. Hjelmervik, S. Nazir, and A. Myhrvold, “Simulator training for maritime complex tasks: an experimental study,” *WMU Journal of Maritime Affairs*, vol. 17, no. 1, pp. 17–30, 2018.

[8] Y. Liu *et al.*, “Psychophysiological evaluation of seafarers to improve training in maritime virtual simulator,” *Advanced Engineering Informatics*, vol. 44, p. 101048, 2020.

[9] L. Orlandi and B. Brooks, “Measuring mental workload and physiological reactions in marine pilots: Building bridges towards redlines of performance,” *Applied ergonomics*, vol. 69, pp. 74–92, 2018.

[10] L. Orlandi, B. Brooks, and M. Bowles, “A comparison of marine pilots’ planning and manoeuvring skills: uncovering mental models to assess shiphandling and explore expertise,” *The Journal of Navigation*, vol. 68, no. 5, pp. 897–914, 2015.

[11] F. Sanfilippo, “A multi-sensor fusion framework for improving situational awareness in demanding maritime training,” *Reliability Engineering & System Safety*, vol. 161, pp. 12–24, 2017.

[12] E.-R. Saus, B. H. Johnsen, J. Eid, and J. F. Thayer, “Who benefits from simulator training: Personality and heart rate variability in relation to situation awareness during navigation training,” *Computers in Human Behavior*, vol. 28, no. 4, pp. 1262–1268, 2012.

[13] C. Sellberg, M. Lundin, and R. Säljö, “Assessment in the zone of proximal development: simulator-based competence tests and the dynamic evaluation of knowledge-in-action,” *Classroom Discourse*, vol. 13, no. 4, pp. 365–385, 2022.

[14] T. T. Türkistanli and B. Kuleyin, “Game-based learning for better decision-making: A collision prevention training for maritime transportation engineering students,” *Computer Applications in Engineering Education*, vol. 30, no. 3, pp. 917–933, 2022.

[15] A. Wahl, T. Kongsvik, and S. Antonsen, “Balancing Safety I and Safety II: Learning to manage performance variability at sea using simulator-based training,” *Reliability Engineering & System Safety*, vol. 195, p. 106698, 2020.

[16] H. Xue, B.-M. Batalden, P. Sharma, J. A. Johansen, and D. K. Prasad, “Biosignal-based driving skill classification using machine learning: A case study of maritime navigation,” *Applied Sciences*, vol. 11, no. 20, p. 9765, 2021.

[17] S. Fan and Z. Yang, “Towards objective human performance measurement for maritime safety: A new psychophysiological data-driven machine learning method,” *Reliability Engineering & System Safety*, p. 109103, 2023.