TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 17
Number 1
March 2023
DOI: 10.12716/1001.17.01.10

Seeking the Best Practices of Assessment in Maritime Simulator Training

H.M. Tusher[1], S. Nazir[1,2], S. Ghosh[3] & R. Rusli[4]
[1] *University of South-Eastern Norway, Borre, Norway*
[2] *Nord University, Bodo, Norway*
[3] *University of Tasmania, Launceston, Australia*
[4] *Universiti Teknologi Petronas, Seri Iskandar, Perak, Malaysia*

ABSTRACT: Simulator-based training has become an integral part of Maritime education, and its effectiveness hinges on the use of appropriate assessment protocols. Despite the existence of several subjective and objective assessment techniques, instructors face difficulties in selecting and implementing the best practices that fit different learning contexts. The contextualized utility of the available assessment techniques further complicates the contexts. This study adopts a systematic literature review approach to comprehensively analyse available assessment techniques employed in maritime simulator training and to elicit their relationship with the desired learning outcomes. The study also presents a nuanced understanding of the advantages and limitations of the identified assessment techniques. Further, the state-of-the-art of assessment methods is discussed along with a few proposals for the future considering both research and practical implications. The findings of this study are expected to provide valuable guidance to maritime instructors in selecting and implementing appropriate assessment techniques that align with desired learning outcomes in simulator training.

1 INTRODUCTION

A highly skilled workforce is essential in the seafaring industry to navigate through diverse operational scenarios. Successful operation depends on competent human operators where training plays a crucial part. Simulators are utilized for various seafarer training contexts where risk-free, repeated exercises are facilitated with considerably less time and cost than other traditional on-the-job training methods [1], [2]. Assessing trainee performance bears crucial significance in the overall success of maritime training whether it is classroom-based or simulator-based training [3], [4]. The conceptual coupling between specific learning outcomes and the assessment methods employed in training is also crucial from a theoretical point of view [5]. On the other hand, the employers require reliable indicators and evidences of seafarers' competence which forms the practical need of assessment [6]. Further, to justify the growing cost of training and its impact on workplace performance, standardized assessment scales have emerged as a means of evaluation [7]. In addition, the challenges related to attending the "workplace relevance" in maritime training and assessment only adds to the complexity of practical requirements [6].

1.1 *Conceptual foundations of assessment*

The definition of "educational assessment" has taken many forms over the years and is still under scrutiny. In the early 1990s, educational assessment simply meant to measure the outcome of learners' achievements relative to either their peers or based on their performance [8], [9]. There are two existing school of thoughts related to assessment of educational output namely the realist and the relativist approaches [10]. The realists tend to measure performance with regards to criterion-based standardized scales whereas the relativists rely on expert judgement in assessing learner's performance. Amidst the growing variation of assessment methods and confusion surrounding their application, Kraiger et al. (1993) suggested specific assessment methods to be employed to the corresponding learning outcomes, i.e., cognitive, skill-based, and affective outcomes (see Table 1).

In addition, authentic assessment is continually getting popular where assessment tasks are replicated according to the desired performance of real workplace, therefore, termed as "performance assessment" [8]. We also seek out "what constitutes a good assessment?". Gipps (1994) noted a few key elements of good assessment practices which include:

– a range of activities and wide opportunity to perform,
– carried out in a safe environment, i.e., a normal classroom setting with ample opportunity to interact with teacher,
– a range of media to perform except written exams.

The requirements of good assessment resonate with the literature related to maritime simulator training where trainees participate in a wide range of scenarios in a safe environment and in various modalities, i.e., full-mission, desktop-based, cloud and virtual reality (VR) simulators [2], [4], [12]–[15].

1.2 *Assessment in maritime simulator training*

The demand for diverse assessment methods in maritime training has arisen to meet complex learning objectives, such as ensuring pedagogical effectiveness [14] and developing practical job skills [6]. Consequently, a plethora of assessment methods have been developed and are currently utilized in maritime simulator training to address both theoretical (i.e., pedagogical) and practical (i.e., job relevant) aspects of assessment. Debriefing [1], dynamic assessment by analysing video data [14], psychophysiological evaluation [16] and differing computer-based tools [3], [17] are some of the sighted methods of assessment in the mentioned context.

Table 1. Prescribed assessment methods for differing learning outcomes (adapted from Kraiger et al., 1993)

Learning outcomes	Assessment methods	Measurement goals	Tools and method descriptions
Cognitive outcomes: a compilation of problem-solving strategies that include declarative and procedural knowledge [11]	Recognition and recall tests	Measuring declarative knowledge	Multiple-choice question, true-false, free recall exams
	Power tests	Measuring the accuracy of knowledge	The number of correct answers, without any time limit
	Speed tests	Measuring the rate at which an individual can access knowledge	The number of correct answers, or the reaction time to any single item
	Free sorts	Measuring knowledge structure, cognitive maps, or mental models	Observing if trainees can physically arrange elements according to expected clusters
	Structural assessment	(same as above)	Clustering or scoring algorithms
	Probed protocol analysis	Measuring cognitive strategies and task behaviour relative to goals	Questionnaire asking trainees specified probe questions at each task step
	Self-report	Measuring trainees' self-awareness of knowledge gained	Questionnaire related to the awareness of procedural level, additional learning requirements and awareness of mistakes
	Readiness for testing	Measuring the ability of learners to judge the future applicability of their knowledge	Questionnaire
Skill-based outcomes: skill-based outcomes related to technical and motor skills [5]	Targeted behavioural observation	Measuring the speed and fluidity of performance, error rates, overall theoretical conceptualization, and trainees' ability to generalize skills in different contexts	Observing frequency of desired behaviour, time, steps and sequencing requirements to complete a task, error counts etc.
	Hands-on testing		Observing if trainees can identify a series of correct steps and providing qualitative evaluation along the way
	Structured situational interviews		Questionnaire asking trainees about how they would perform a particular task
	Secondary task performance	Measuring available cognitive resources and the automaticity of performance	Observing stable trainee performance on a primary task and increasing performance in secondary task
	Interference problems		Observing the decrease in performance for inference tasks (i.e., normal tasks but with key information altered)
	Embedded measurement	(same as above)	Observing if trainees can perform a task without guidance and if they omit initial stages of a task
Affective outcomes: an internal state of behaviour which affect learners' attitude [11]	Self-report measures	Measuring trainee attitudes, mastery, perceived performance capability and goal setting	Questionnaire asking trainees about their capability and confidence in performing a task with varying difficulties
	Free recall measures	Measuring complexity of goal structures	Questionnaire, focused interviews or think aloud protocol for trainees
	Free sorts	Measuring goal commitment	Observing if trainees can physically arrange elements according to expected clusters

However, research indicates that these assessment methods may not guarantee the acquisition of competency and the development of a competent workforce [4], [6], [18]. The lack of authenticity in current assessment practices during simulator training is considered a hindrance to the development of essential competencies and preparation for workplace challenges. Consequently, the need for "authentic assessment" has been highlighted in the literature [6], [19]–[21].

In addition, instructors' subjectivity, bias, and uncertainty around assessment methods pose significant challenges in operationalizing efficient assessment during simulator training [3], [4]. For example, instructors may develop scenarios catering to a particular group of trainees or test students on simulators based on their own experience (e.g., cargo ship experience), which may not be suitable for other trainees on a different context (e.g., emergency on a passenger ship).

Maritime simulators provide a simulated virtual environment for education, making them ideal technology-based learning environments [22]. Thus, the same assessment challenges that exist in other technology-based learning environments also apply to maritime simulator training. Meeting the challenges of maritime simulator training requires addressing differences in learner characteristics, technical capabilities, and pedagogical design. These differences can lead to inconsistent learning outcomes. For example, variations in learners' characteristics and simulator fidelity can result in mixed learning outcomes [23]. In addition, the misalignment of simulation practices with pedagogical theories is a recognized issue [24].

Assessing trainee performance in maritime training is a complex task with varying opinions on its objectives, especially during simulator training. Some view assessment as a means of objectively categorizing maritime trainees based on competence [3], [25], while others see it as an assistive learning instrument [15], [26], [27]. Objective assessment aims to ensure the validity and reliability of evaluation measures, but it is a fallacy to believe that learning can be accurately and reliably assessed [8]. The need for professional intersubjectivity of instructors undermines the requirements of validity and objectivity of assessment measures [15], [28]. Sadler (2005) proposed that the ideal performance assessment should focus on "standard performance" rather than "criterion-based objective assessment" focusing on validity only. There is a lack of consensus if assessment of maritime simulator trainees should purely aim for objective measures or depend on expert evaluation of instructors. Ideal assessment frameworks can be conceptualized as a method where trainee performance can be reliably assessed without compromising the expert-in-the-loop feature.

Therefore, in-depth knowledge on the operationalization of existing assessment methods in maritime simulator training is crucial to ensure that theoretical and practical aspects of evaluation are satisfied. It is also important to identify which evaluation methods are best suited for specific simulator training scenarios and their specific advantages and limitations. Such analysis would pave the way for assessment best-practices that can help maritime instructors to administer appropriate assessment methods for specific training needs, while also being aware of their benefits and pitfalls. Awareness of the assessment best-practices can also assist instructors in developing valid and reliable assessment scenarios and adapting their training approaches to achieve desired learning outcomes. Furthermore, educators would acquire enhanced comprehension regarding the optimal timing and extent of expert intervention in the evaluation process, enabling them to make informed decisions on when and how much to intervene.

1.3 *The goal of this study*

In this study, we examined the empirical studies that employed various assessment methods within the context of maritime simulator training. The goal was to systematically assemble differing assessment methods including their objectives, their specific advantages and limitations coupled with the in-depth analysis of their suitability with differing learning outcomes. Thus, we also shed light on the state-of-the-art of assessment methods identifying their gaps and discuss the future requirements in maritime simulator training context.

The following research questions have been formed:

– RQ1: How differing assessment methods are operationalized in maritime simulator training?
– RQ2: What are the advantages and limitations of operationalized assessment methods in maritime simulator training?

2 METHODS

To address the research questions, a systemic literature review method was adopted in this study. The following keywords were utilized for searching documents in two different databases (i.e., Web of Science and Scopus): (“maritime” or “shipping” or “seafarer*”) AND “simulator*” and “training” and “assessment”. A set of inclusion and exclusion criteria is used during the database search (see Table 2).

Table 2: Inclusion and exclusion criteria for database search

Inclusion criteria	Exclusion criteria
Peer-reviewed	Not related to simulator training
English articles	No intervention used during the study or experiment
Seafarer training related empirical studies	Conceptual and non-empirical studies
Studies operationalizing specific assessment method	White paper or technical reports
Studies with clear learning/ training outcome (Cognitive, Skill-based or Affective outcomes)	

The initial number of search output from two (02) separate databases resulted in a total of 147 studies including one (01) document added through snowballing. The overall literature review process followed a systemic approach as depicted in Figure 1 aligning with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

[30]. Notably, all conference articles had to be removed due to the lack of information of differing peer-review processes. A total of 18 studies remained to be screened for final analysis.

Subsequently, these studies were qualitatively synthesized to elicit the goals of the assessment, type of simulator used, assessment measures and methods along with the associated learning outcomes. In addition, the advantages and limitations associated with each assessment method as mentioned in the studies are also included in the analysis. Another co-author separately screened the excerpts of the analysis in excel format for inter-rater reliability (see Table 3).

Table 3. Performance assessment in maritime simulator training

SL	Source	Assessment goals	Simulator types	Assessment measures	Corresponding method	Learning outcome	Advantages	Limitation
1	[31]	Measuring situational awareness of navigational officers in the event of autopilot failure	Bridge simulator	Reaction time to failure	Speed test; secondary task performance	Cognitive, Skill-based	(Not listed in the study)	Overall time limit of the exercises. It can be argued that participants could correct the issues if given enough time.
2	[32]	Assessing electronic navigation competency	Bridge simulator	Eye tracking data for fixation duration and fixation counts	Targeted behavioural observation	Skill-based	Eye-tracking provides novel data, e.g., focus of attention, facilitates objective observation of trainees	High cost of eye tracking equipment
3	[3]	Assessing pilotage operation	Bridge simulator	Computer-aided Performance Assessment tool (CAPA) utilizing Analytical Hierarchy Process (AHP) and Bayesian Network for binary assessment of checklisted items	Targeted behavioural observation	Cognitive, Skill-based	Inter-rater reliability was fair for CAPA-tool compared to conventional methods in terms of teamwork performance	Methodological challenges related to utilizing AHP tools for weighting performance score, CAPA tool relies heavily on human observation or interpretation, thus human bias cannot be fully eliminated, Criterion validity of the tool could not be established
4	[33]	Assessing situational awareness of bridge watchkeepers	Bridge simulator	Eye tracking data for fixation duration, heatmap along with debriefing	Targeted behavioural observation; hands-on testing	Skill-based	Opportunity for objective and continuous monitoring of students by eye tracking	High cost of eye tracking equipment. Limitation of analysis software to analyse dynamic scenarios as ships' motion is unstable. The bigger size and layout of bridge simulators is not optimal for the analysis of large amount of eye-tracking data
5	[34]	Assessing visual attention during heavy lifting operation	Heavy lift simulator	Eye-tracking, briefing-debriefing, questionnaire	Targeted behavioural observation; hands-on testing; recognition & recall test	Skill-based, Cognitive	(Not listed in the study)	Time consuming; inclusion of subjective factors in the assessment such as defining optimal measures (e.g., trigger time)
6	[16]	Psychophysiological (cognitive workload, stress) evaluation of seafarers	Bridge simulator	EEG for measuring heartrate variability	Targated behavioural observation (physiological data)	Skill-based	Psychophysiological evaluation complements the current simulator-aided assessment. High reliability of the proposed system.	(Not listed in the study)
7	[12]	Assessing the effect of introducing complexity in different levels of simulator training	Bridge simulator	ECDIS data to calculate cross-track error	Targeted behavioural observation	Skill-based	(Not listed in the study)	Smaller sample size and lack of variation in the vessel types being used during assessment
8	[35]	Evaluating individual risk perception with a focus on situational awareness	Offshore crane simulator	Multi-layer and multi-sensor fusion to analyse bio-metric data	Targeted behavioural observation; hands-on testing	Skill-based	Provides new insights into novel SA assessment methodologies	Lack of robustness and validity concerns with method
9	[36]	Measuring SA during an engine-supervisory task environment in ships	Engine plant simulator	NASA-TLX for perceived SA, SA sensitivity	Self-report; structured situational interviews	Cognitive	Subjective measurement is sensitive to task complexity but not to participants experience while the objective measurement is sensitive to both. The objective measurement also provides freedom to the evaluator to develop scenario relevant to the purpose of the study.	Objective measure: the simulation required to be frozen several times to administer the questionnaire; subjective measure: limited in depicting the effect of different level of participants expertise during measurement; familiarity effect was higher for subsequent measurements; sensitive to different workload levels; participants' actions during the scenario was passive; the mental

								model of participants action or communication instances were not evaluated.
10	[15]	Evaluating navigational competence	Bridge simulator	Observation with assessment sheet indicating rate of turn, turn rate and speed in manoeuvring along with interviews	Hands-on testing; probed protocol analysis; embedded measurement	Skill-based, Cognitive	Keeps instructor in the loop during assessment	There is scope for assessors' subjectivity and individual bias
11	[37]	Evaluating lifeboat launching operation	Lifeboat simulator	Questionnaire including accuracy of recall, order of actions	Power test; hands-on testing	Cognitive, Skill-based	Verbal administration was possible without lowering the lifeboats	The type of questionnaire favoured one type of simulator participants more than the other
12	[38]	Assessing navigational competency	Bridge simulator	Observation, interviews, video recording for COLREG deviations	Hands-on testing; targeted behavioural observation	Skill-based	Possibility of instructor-in-the-loop during training, i.e., instructors intervene during training to correct mistakes and gives inputs to fulfil the learning objectives	As instructors provide selective correction to a few students but not all, the fairness of assessment is questioned
13	[39]	Human error evaluation	Engine plant simulator	Questionnaire on different operational tasks on mimic panels	Targeted behavioural observation; hands-on testing	Skill-based	Possible to measure both operational task competency and situational awareness, and can be used to select best candidates for certain operation	(Not listed in the study)
14	[40]	Evaluate and rank students based on their actions with regards to different error producing conditions	Engine plant simulator	Shipboard Operation Human Reliability Analysis (SOHRA) for overall trainee performance	Targeted behavioural observation; hands-on testing	Skill-based	This method can be utilized to select best candidate for a specific operation	(Not listed in the study)
15	[41]	Assessing deck officers NTS	Bridge simulator	AHP, Evidential Reasoning (ER) algorithm for identified behavioural markers	Targeted behavioural observation	Skill-based	The effectiveness of any training methodology can be determined	Inadequate data for conclusive results
16	[26]	Evaluating Manoeuvring performance in differing situations	Bridge simulator	Computer Based Evaluation (CBE) including course, speed, overshoot angle, rudder angle, track deviation and time	Targeted behavioural observation; hands-on testing	Skill-based	CBE provides additional support for simulator instructors and provides opportunity for increasing the objectivity in evaluation	Lack of clear evaluation criteria, communication and individual situation-awareness aspects are difficult to monitor or measure in CBE
17	[27]	Proposing an assessment method for evaluating the required time for training and the level of navigational competency	Bridge simulator	Observation, checklist including Rules of the Road, communication, vessel positioning, lookout, manoeuvring	Hands-on testing; targeted behavioural observation	Skill-based	This method can determine the seafarers' learning process, the impact of training, and the necessary number of assessments for achieving a satisfactory level of competency	The assessment must be conducted concurrently with the progression of the training scenario to ensure objectivity; otherwise, the final evaluation at the conclusion of the training scenario may become subjective as the assessors may not have a complete recollection of all events.
18	[42]	Assessing perceived situational awareness (SA), learning outcome and perceived realism	Bridge simulator	Self-reported Situational Awareness Rating Scale (SARS) questionnaire, ECG for SA, workload	Self-report	Cognitive, Affective	Provides new insights into how SA affects learning outcome during simulator training	Experience of seafarers and their perceived realism of simulator training supress the effect of measurement

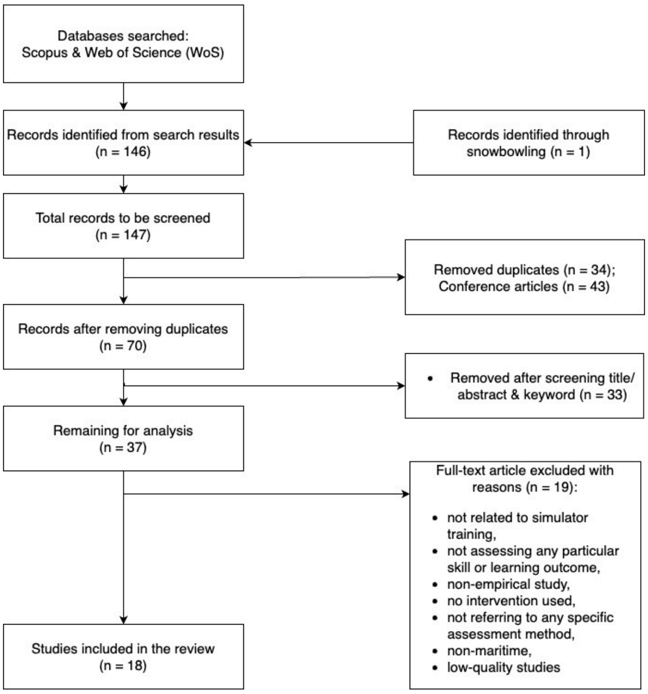

Figure 1. Systematic literature review process

3 RESULTS

3.1 *RQ1: How differing assessment methods are operationalized in maritime simulator training?*

To explain how the different assessment methods are operationalized in simulator training, the findings were categorized in three (03) segments:

– assessment goals,
– assessment tools, and
– assessment methods

3.1.1 *Assessment goals*

Apart from the differences of theoretical concepts, assessment in maritime simulator training has a wide variance of application in practice as revealed during the analysis of 18 selected studies. The goals of assessment in maritime simulators as identified based on their frequency of appearance in these studies are calculated. The analysis reveals that assessing navigational competence (33.3%) and situational awareness (28%) are the most focused assessment goals, followed by assessing human error (11%), non-technical skills (5.6%), lifeboat launching skills (5.6%), cognitive workload (5.6%), visual attention (5.6%) and others (5.6%). Bridge simulators (66.7%) are most used types of simulators in assessment contexts followed by engine plant simulator (16.7%), crane simulator (11.1%) and lifeboat simulators (5.6%) (see Figure 2).

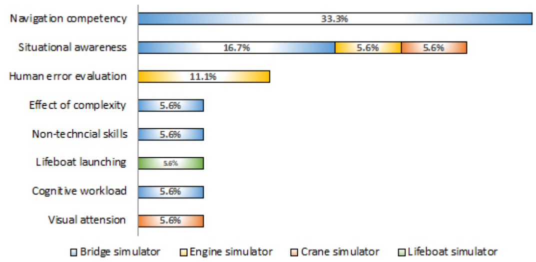

Figure 2. Assessment goals in different types of simulators

3.1.2 *Assessment tools*

The results pointed out the predominance of questionnaire (32%) and observation techniques (28%) as assessment tools, where evaluators' subjectivity and expertise played a crucial role in the overall judgement. Alternatively, heart rate variability and workload analysis from ECG (electrocardiogram) and EEG (electroencephalogram) signals, eye fixation duration and counts from eye tracking data, and algorithm-based analysis were sighted as quantitative measurement techniques (see Figure 3). However, irrespective of the tool or parameter used, the evaluators were involved either in the scale development or in determining what constituted good or bad performance, resulting in a relativist process of assessment [10]. This involvement nullified the potential objectivity of the seemingly realist quantitative techniques i.e., use of a criterion-based assessment.

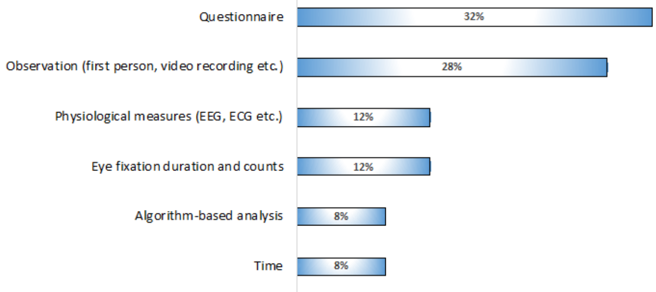

Figure 3. Assessment measures utilized in simulators

3.1.3 *Assessment methods*

The analysis of the selected literature suggested that targeted behavioural observation and hands-on testing are the most used assessment methods depicting the prevalence of skill-based skill training in maritime simulator training. Other methods such as embedded measurement, structured situational interviews, secondary task performance and a few novel methods (i.e., EEG for measuring heartrate variability) are operationalized for evaluating skill-based learning outcomes. Measuring cognitive learning outcomes include self-report, power test, probed protocol analysis, speed test, recognition and recall methods of assessment. On the other hand, self-report measures are the solely used method for measuring affective learning outcomes in maritime simulator training contexts (see Figure 4).

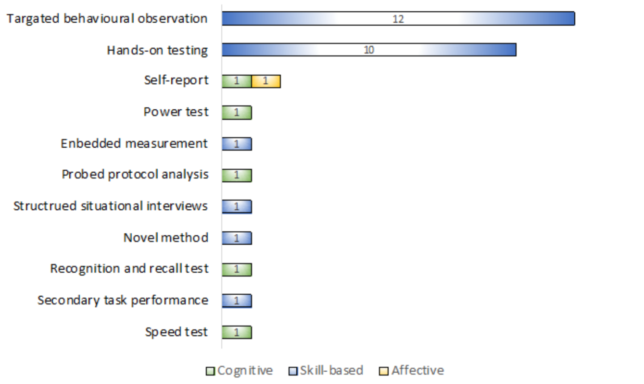

Figure 4. Assessment methods and associated learning outcomes

3.2 *RQ2: What are the advantages and limitations of operationalized assessment methods in maritime simulator training?*

The review revealed context-specific advantages and limitations of differing assessment methods. For example, targeted behavioural observation is predominantly used for subjective evaluation; however, it facilitates objective evaluation if coupled with eye tracking data or other computer-based assessment measures. Similarly, hands-on testing facilitates instructors' crucial input in different stages of learning which may be considered a disadvantage if the overall goal is to provide more objective assessment. Henceforth, a succinct overview of the pros and cons of every assessment approach, as implemented in differing maritime simulator training context, is presented:

3.2.1 *Advantages*

State-of-the-art tools such as eye-tracking, EEG exhibits novel promises in training evaluation facilitating objective and continuous observation of trainees [16], [33]. Computer-aided assessment tools (e.g., CAPA) demonstrate inter-rater reliability potentially reducing the involvement of instructors and subsequent bias in the assessment process [3] while other computer-based methods assure instructional support for the users [26], both aiming for objective evaluation of trainees. Other targeted behavioural observation and hands-on testing instances allows for continuous monitoring of trainees [33] as well as providing novel insights especially while using sensor fusion and biometric data [35] in measuring situational awareness (SA). On the other hand, the methodological characteristics of several assessment protocols (e.g., hands-on testing, probed protocol analysis) allows for more instructor involvement facilitating expert-in-the-loop and efficient subjective evaluation during training [15], [37], [38]. In addition, subjective methods (e.g., self-report) are found particularly useful to administer with a wider population of trainees especially during SA measurements [36].

3.2.2 *Limitations*

The identified limitations as described in respective studies utilizing various assessment methods can be categorized broadly in two classes namely hardware-based limitations and methodical limitations. The former encompasses difficulties arising from the resource intensive processes such as expensive eye-trackers as well as limitation of analysis software in dynamic assessment situations [32], [33]. On the other hand, methodical limitations encompass a range of issues such as insufficient assessment time [31] or being too time consuming [34], lack of variation in scenarios [12]; unintentional favouritism or bias [37], [38], familiarity effects [36], immeasurable behavioural constructs [26], [36], unclear evaluation criteria [26], ambiguity among instructors about assessment tools and procedures [27], subjectivity of these tools [3], [27] and related validity concerns [3], [35].

4 DISCUSSIONS

The focus of this study is to explore different assessment methods with regards to their objectives, suitability to the learning outcomes (i.e., cognitive, skill-based, and affective outcomes) and context-specific advantages and limitations. Based on the results of this study, we discuss the state-of-the-art maritime simulator training and assessment and propose a few future extents where the emerging assessment methods should focus on.

4.1 *State-of-the-art maritime training and assessment*

The concept of simulator training stems from the notion of competency-based training facilitating seafarers' knowledge and skill acquisition required for their professional work [2], [43]. However, this review highlights the ubiquitous goal of simulator training being navigational competency training and situational awareness assessment in bridge simulators (see Figure 2) while other forms of competency training (e.g., engine room operators' training, emergency procedural tasks, non-technical skill training etc.) are less prevalent in maritime institutes. Thus, understating specific competencies during simulator training ultimately defeats the purpose of all-round competency development for seafarers, while also contradicting the goal of authentic training in simulators, where the actual work environment may require seafarers to be competent in a diverse set of skills.

The results also suggest that a handful of assessment methods are frequently used while other prescribed methods such as the ones suggested by Kraiger et al. (1993) are underutilized. For example, targeted behavioural observation and hands-on testing are most prevalent while other types of assessment methods are barely utilized in maritime simulator training context. This could be due to the disproportionately large focus on measuring cognitive and skill-based learning outcomes in maritime simulator training while emphasizing less on affective learning outcomes. Nevertheless, from a human factor's perspective, affective learning outcomes are found to be correlated with real-world performance. Maritime research highlights the importance of affective components such as emotion [44], self-efficacy, motivation [23] and attitude in maritime training [45], [46].

In addition, instructors find themselves using traditional tools such as questionnaire and observation techniques prevalently in maritime simulator training despite the growing evidence related to the utility of physiological measures in performance assessment such as EEG [16], eye-tracking [32], [33] and functional Near-Infrared Spectroscopy (fNIRS) technology [47] in measuring cognitive resources such as workload, attention, stress etc. In addition, novel assessment protocols utilizing deep learning algorithms (e.g., artificial neural network) are envisaged as promising in categorizing trainee performance during maritime simulator training [48]. The instructors' lack of interest in implementing new assessment methods could be attributed to either their unfamiliarity with such methods or concerns over their validity.

4.2 *The future of assessment in maritime simulator training*

This review has identified a significant gap in the assessment of maritime trainees' affective learning outcomes (e.g., motivation, self-efficacy etc.), indicating a need for further research to explore innovative approaches to measuring such outcomes in the context of simulator training. This concept aligns with the conclusions drawn from other scholarly investigations as well [38], [49]. Therefore, future studies should prioritize investigating new and effective methods for assessing affective learning outcomes in this domain.

The goal of simulator training is to address the issues related to the lack of authentic learning and assessment contexts [50] while the assessment aims to bridge the gap between theory and practice [15], [22]. However, the lack of "workplace-relevance", i.e., authentic assessment protocols is a well-known issue in maritime training context [6], [24]. Also, the effect of learning depends on the level of authenticity in simulator training [50]. Future studies should focus on establishing authentic assessment protocols satisfying the required criteria such as authentic training context [51], having real-world relevance and facilitating opportunity for collaboration [52] along with its seamless integration with the training activity [51].

The review highlights the subjective nature of traditional assessment tools such as questionnaires and observations, and the challenges associated with implementing them, including high costs (e.g., expensive eye-tracking equipment) and time-consuming processes (e.g., long briefing and debriefing sessions). Moreover, the instructors may lack experience or clarity on evaluation criteria [26], exacerbating these challenges. To address these issues, there is a growing trend towards using more objective and standardized methods for assessing maritime trainees in simulated environments [3], [16], [47]. Such methods are deemed beneficial in reducing subjectivity and bias while providing deeper insights into trainees' performance. Research has also identified that students correlate fairness of assessment with their own engagement in the process [53] which is often ignored in the current assessment practices in maritime simulator training. Future studies should focus on the investigating novel learner-centred methods focusing more on instructor-trainee collaboration, validating emerging tools and methods through empirical research to increase instructors' confidence in using them.

The discussions above enunciate that future assessment methods should focus on all learning outcomes, be authentic and integrated into realistic training contexts, and be practical in terms of cost and time efficiency. These methods should also be easy to understand and to administer by the instructors. Additionally, it is important to incorporate teacher-student collaboration in assessment to reduce bias while still retaining expert input from instructors (see Figure 5).

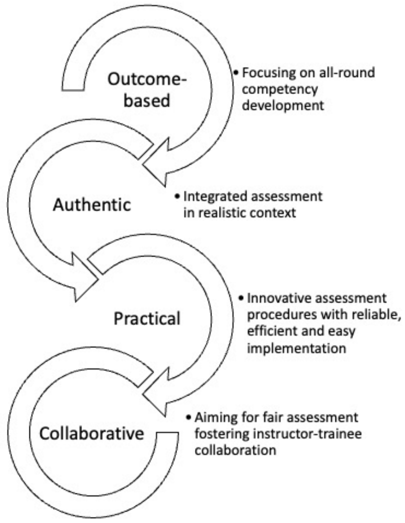

Figure 5. Characteristics of future assessment methods as envisaged

5 CONCLUSIONS

This review focuses on exploring how differing assessment methods are operationalized in current maritime simulator training practices. Also, the suitability of various assessment methods in differing contexts, taking into consideration the learning outcomes in terms of cognitive, skill-based, and affective competencies are investigated. Our findings demonstrate that while some assessment methods align well with these learning outcomes, there is a lack of methods to measure affective competencies. Additionally, there is an overemphasis on navigation training at the expense of other competencies, which could hinder the all-round competency development of seafarers. Furthermore, there are existing challenges in operationalizing various assessment methods. Based on our review, a detailed analysis of current assessment methods is presented to propose an envisaged best-practice for the future, considering their specific advantages and limitations and identifying areas for improvement. This analysis will enable maritime simulator instructors to select appropriate assessment methods, design assessment episodes to capitalize on their advantages while avoiding potential drawbacks and adapt their training to meet the needs of their students. Overall, it is essential to prioritize outcome-based, authentic, practical, and collaborative assessment methods to enhance the effectiveness of maritime simulator training.

ACKNOWLEDGEMENT

The authors acknowledge the support Centre of Excellence in Maritime Simulator Training and Assessment (COAST) in Norway, funded by the Directorate for Higher Education and Competence (HK-dir). The 1st, 2nd, and 4th author appreciate the support of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823904 (Project: ENHANCing Human Performance in Complex Socio-Technical SystEms, ENHANCE).

REFERENCES

[1] M. Hontvedt and H. C. Arnseth, 'On the bridge to learn: Analysing the social organization of nautical instruction in a ship simulator', Int. J. Comput.-Support. Collab. Learn., vol. 8, no. 1, pp. 89–112, Mar. 2013, doi: 10.1007/s11412-013-9166-3.
[2] T. Kim et al., 'The continuum of simulator-based maritime training and education', WMU J. Marit. Aff., vol. 20, no. 2, pp. 135–150, Jun. 2021, doi: 10.1007/s13437-021-00242-2.
[3] J. Ernstsen and S. Nazir, 'Performance assessment in full-scale simulators–A case of maritime pilotage operations', Saf. Sci., vol. 129, p. 104775, 2020, doi: https://doi.org/10.1016/j.ssci.2020.104775.
[4] C. Sellberg, 'Simulators in bridge operations training and assessment: a systematic review and qualitative synthesis', WMU J. Marit. Aff., vol. 16, no. 2, pp. 247–263, May 2017, doi: 10.1007/s13437-016-0114-8.
[5] K. Kraiger, J. K. Ford, and E. Salas, 'Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation.', J. Appl. Psychol., vol. 78, no. 2, pp. 311–328, Apr. 1993, doi: 10.1037/0021-9010.78.2.311.
[6] S. Ghosh, M. Bowles, D. Ranmuthugala, and B. Brooks, 'Reviewing seafarer assessment methods to determine the need for authentic assessment', Aust. J. Marit. Ocean Aff., vol. 6, no. 1, pp. 49–63, Jan. 2014, doi: 10.1080/18366503.2014.888133.
[7] B. S. Bell, S. I. Tannenbaum, J. K. Ford, R. A. Noe, and K. Kraiger, '100 years of training and development research: What we know and where we should go.', J. Appl. Psychol., vol. 102, no. 3, pp. 305–323, 2017, doi: 10.1037/apl0000142.
[8] C. Gipps, 'Developments in Educational Assessment: what makes a good test?', Assess. Educ. Princ. Policy Pract., vol. 1, no. 3, pp. 283–292, Jan. 1994, doi: 10.1080/0969594940010304.
[9] W. L. Sanders and S. P. Horn, 'Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes', 1995.
[10] M. Yorke, 'Summative assessment: dealing with the "measurement fallacy"', Stud. High. Educ., vol. 36, no. 3, pp. 251–273, 2011.
[11] R. M. Gagne, 'Learning outcomes and their effects: Useful categories of human performance.', Am. Psychol., vol. 39, no. 4, p. 377, 1984.
[12] K. Hjelmervik, S. Nazir, and A. Myhrvold, 'Simulator training for maritime complex tasks: an experimental study', WMU J. Marit. Aff., vol. 17, no. 1, pp. 17–30, Mar. 2018, doi: 10.1007/s13437-017-0133-0.
[13] K. I. ØvergÁrd, S. Nazir, and A. Solberg, 'Towards Automated Performance Assessment for Maritime Navigation', TransNav Int. J. Mar. Navig. Saf. Sea Transp., vol. 11, no. 2, pp. 43–48, 2017, doi: 10.12716/1001.11.02.03.
[14] C. Sellberg, 'Pedagogical dilemmas in dynamic assessment situations: perspectives on video data from simulator-based competence tests', WMU J. Marit. Aff., vol. 19, no. 4, pp. 493–508, Dec. 2020, doi: 10.1007/s13437-020-00210-2.
[15] 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', Classr. Discourse, pp. 1–21, Nov. 2021, doi: 10.1080/19463014.2021.1981957.
[16] Y. Liu et al., 'Psychophysiological evaluation of seafarers to improve training in maritime virtual simulator', Adv. Eng. Inform., vol. 44, p. 101048, Apr. 2020, doi: 10.1016/j.aei.2020.101048.
[17] L. Orlandi, B. Brooks, and M. Bowles, 'The development of a shiphandling assessment tool (SAT): A methodology and an integrated approach to assess manoeuvring expertise in a full mission bridge simulator', in 15th Annual General Assembly of the International Association of Maritime Universities, IAMU AGA 2014-Looking Ahead: Innovation in Maritime Education, Training and Research, 2014, pp. 131–140.
[18] V. O. Gekara, M. Bloor, and H. Sampson, 'Computer-based assessment in safety-critical industries: the case of shipping', J. Vocat. Educ. Train., vol. 63, no. 1, pp. 87–100, 2011.
[19] S. Ghosh, 'Can authentic assessment find its place in seafarer education and training?', Aust. J. Marit. Ocean Aff., vol. 9, no. 4, pp. 213–226, Oct. 2017, doi: 10.1080/18366503.2017.1320828.
[20] S. Ghosh and M. Bowles, 'Challenges and implications in achieving content validity of an authentic assessment task designed to assess seafarer's leadership and managerial skills', WMU J. Marit. Aff., vol. 19, no. 3, pp. 373–391, Sep. 2020, doi: 10.1007/s13437-020-00209-9.
[21] C. Sellberg, A. C. Wiig, and R. Säljö, 'Mastering the artful practice of navigation: The situated endorsement of professional competence in post-simulation evaluations', Stud. Educ. Eval., vol. 72, p. 101111, Mar. 2022, doi: 10.1016/j.stueduc.2021.101111.
[22] M. G. Jamil and Z. Bhuiyan, 'Deep learning elements in maritime simulation programmes: a pedagogical exploration of learner experiences', Int. J. Educ. Technol. High. Educ., vol. 18, no. 1, p. 18, Dec. 2021, doi: 10.1186/s41239-021-00255-0.
[23] S. K. Renganayagalu, S. Mallam, S. Nazir, J. Ernstsen, and P. Haavardtun, 'Impact of Simulation Fidelity on Student Self-efficacy and Perceived Skill Development in Maritime Training', TransNav Int. J. Mar. Navig. Saf. Sea Transp., vol. 13, no. 3, pp. 663–669, 2019, doi: 10.12716/1001.13.03.25.
[24] C. Sellberg, 'From briefing, through scenario, to debriefing: the maritime instructor's work during simulator-based training', Cogn. Technol. Work, vol. 20, no. 1, pp. 49–62, Feb. 2018, doi: 10.1007/s10111-017-0446-y.
[25] 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', Comput. Hum. Behav., vol. 28, no. 4, pp. 1262–1268, Jul. 2012, doi: 10.1016/j.chb.2012.02.009.

[26] K. Benedict, M. Baldauf, C. Felsenstein, and M. Kirchhoff, 'Computer-based support for the evaluation of ship handling exercise results', WMU J. Marit. Aff., vol. 5, no. 1, pp. 17–35, Apr. 2006, doi: 10.1007/BF03195079.
[27] H. Kobayashi, 'Use of simulators in assessment, learning and teaching of mariners', WMU J. Marit. Aff., vol. 4, no. 1, pp. 57–75, Apr. 2005, doi: 10.1007/BF03195064.
[28] C. Sellberg and M. Lundin, 'Demonstrating professional intersubjectivity: The instructor's work in simulator-based learning environments', Learn. Cult. Soc. Interact., vol. 13, pp. 60–74, Jun. 2017, doi: 10.1016/j.lcsi.2017.02.003.
[29] D. R. Sadler, 'Interpretations of criteria-based assessment and grading in higher education', Assess. Eval. High. Educ., vol. 30, no. 2, pp. 175–194, 2005.
[30] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and The PRISMA Group, 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement', PLoS Med., vol. 6, no. 7, p. e1000097, Jul. 2009, doi: 10.1371/journal.pmed.1000097.
[31] J. P. Chan, R. Norman, K. Pazouki, and D. Golightly, 'Autonomous maritime operations and the influence of situational awareness within maritime navigation', WMU J. Marit. Aff., vol. 21, no. 2, pp. 121–140, Jun. 2022, doi: 10.1007/s13437-022-00264-4.
[32] O. Atik and O. Arslan, 'Use of eye tracking for assessment of electronic navigation competency in maritime training', J. Eye Mov. Res., vol. 12, no. 3, Jul. 2019, doi: 10.16910/jemr.12.3.2.
[33] O. Atik, 'Eye tracking for assessment of situational awareness in bridge resource management training', J. Eye Mov. Res., vol. 12, no. 3, Apr. 2020, doi: 10.16910/jemr.12.3.7.
[34] G. Li, R. Mao, H. P. Hildre, and H. Zhang, 'Visual Attention Assessment for Expert-in-the-Loop Training in a Maritime Operation Simulator', IEEE Trans. Ind. Inform., vol. 16, no. 1, pp. 522–531, Jan. 2020, doi: 10.1109/TII.2019.2945361.
[35] F. Sanfilippo, 'A multi-sensor fusion framework for improving situational awareness in demanding maritime training', Reliab. Eng. Syst. Saf., vol. 161, pp. 12–24, May 2017, doi: 10.1016/j.ress.2016.12.015.
[36] A. M. Nizar, T. Miwa, and M. Uchida, 'Measurement of situation awareness in engine control room: approach for non-technical skill assessment in engine resource management', WMU J. Marit. Aff., vol. 21, no. 3, pp. 401–419, Sep. 2022, doi: 10.1007/s13437-022-00270-6.
[37] J. Jung and Y. J. Ahn, 'Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study: A comparative assessment interfaces in virtual training', Comput. Animat. Virtual Worlds, vol. 29, no. 3–4, p. e1812, May 2018, doi: 10.1002/cav.1812.
[38] C. Sellberg, O. Lindmark, and M. Lundin, 'Certifying Navigational Skills: A Video-based Study on Assessments in Simulated Environments', TransNav Int. J. Mar. Navig. Saf. Sea Transp., vol. 13, no. 4, pp. 881–886, 2019, doi: 10.12716/1001.13.04.23.
[39] C. Kandemir, O. Soner, and M. Celik, 'Proposing a practical training assessment technique to adopt simulators into marine engineering education', WMU J. Marit. Aff., vol. 17, no. 1, pp. 1–15, Mar. 2018, doi: 10.1007/s13437-018-0137-4.
[40] C. Kandemir, and M. Celik, 'A Human Reliability Assessment of Marine Engineering Students through Engine Room Simulator Technology', Simul. Gaming, vol. 52, no. 5, pp. 635–649, Oct. 2021, doi: 10.1177/10468781211013851.
[41] F. Saeed, A. Wall, C. Roberts, R. Riahi, and A. Bury, 'A proposed quantitative methodology for the evaluation of the effectiveness of Human Element, Leadership and Management (HELM) training in the UK', WMU J. Marit. Aff., vol. 16, no. 1, pp. 115–138, Jan. 2017, doi: 10.1007/s13437-016-0107-7.
[42] E.-R. Saus, B. H. Johnsen, J. E.-R. Saus, and J. Eid, 'Perceived learning outcome: The relationship between experience, realism and situation awareness during simulator training', Int. Marit. Health, vol. 62, no. 4, pp. 258–264, 2010.
[43] G. Emad and W. M. Roth, 'Contradictions in the practices of training for and assessment of competency: A case study from the maritime domain', Educ. Train., vol. 50, no. 3, pp. 260–272, Apr. 2008, doi: 10.1108/00400910810874026.
[44] S. Fan, J. Zhang, E. Blanco-Davis, Z. Yang, J. Wang, and X. Yan, 'Effects of seafarers' emotion on human performance using bridge simulation', Ocean Eng., vol. 170, pp. 111–119, Dec. 2018, doi: 10.1016/j.oceaneng.2018.10.021.
[45] S. Jensen, M. Lutzen, L. L. Mikkelsen, H. B. Rasmussen, P. V. Pedersen, and P. Schamby, 'Energy-efficient operational training in a ship bridge simulator', J. Clean. Prod., vol. 171, pp. 175–183, Jan. 2018, doi: 10.1016/j.jclepro.2017.10.026.
[46] T. Kim, A. K. Sydnes, and B.-M. Batalden, 'Development and validation of a safety leadership Self-Efficacy Scale (SLSES) in maritime context', Saf. Sci., vol. 134, p. 105031, Feb. 2021, doi: 10.1016/j.ssci.2020.105031.
[47] S. Fan and Z. Yang, 'Towards objective human performance measurement for maritime safety: A new psychophysiological data-driven machine learning method', Reliab. Eng. Syst. Saf., vol. 233, p. 109103, May 2023, doi: 10.1016/j.ress.2023.109103.
[48] H. M. Tusher, S. Nazir, S. Mallam, and Z. H. Munim, 'Artificial Neural Network (ANN) for Performance Assessment in Virtual Reality (VR) Simulators: From Surgical to Maritime Training', in 2022 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Kuala Lumpur, Malaysia, Dec. 2022, pp. 0334–0338. doi: 10.1109/IEEM55944.2022.9989816.
[49] A. M. Wahl and T. Kongsvik, 'Crew resource management training in the maritime industry: a literature review', WMU J. Marit. Aff., vol. 17, no. 3, pp. 377–396, Sep. 2018, doi: 10.1007/s13437-018-0150-7.
[50] O. Chernikova, N. Heitzmann, M. Stadler, D. Holzberger, T. Seidel, and F. Fischer, 'Simulation-Based Learning in Higher Education: A Meta-Analysis', Rev. Educ. Res., vol. 90, no. 4, pp. 499–541, Aug. 2020, doi: 10.3102/0034654320933544.
[51] J. Herrington and L. Kervin, 'Authentic Learning Supported by Technology: Ten suggestions and cases of integration in classrooms', Educ. Media Int., vol. 44, no. 3, pp. 219–236, Sep. 2007, doi: 10.1080/09523980701491666.
[52] T. C. Reeves, J. Herrington, and R. Oliver, 'A development research agenda for online collaborative learning', Educ. Technol. Res. Dev., vol. 52, no. 4, pp. 53–65, 2004.
[53] M. A. Flores, A. M. Veiga Simão, A. Barros, and D. Pereira, 'Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education', Stud. High. Educ., vol. 40, no. 9, pp. 1523–1534, 2015.