Training, Education, and Learning Sciences, Vol. 109, 2023, 94–101
https://doi.org/10.54941/ahfe1003155

Exploring Alternative Performance Evaluation Method in Nautical Simulations

Tae-Eun Kim[1], Ziaul Haque Munim[2], and Hans-Joachim Schramm[3]

[1]Department of Technology and Safety, University of Tromsø (UiT)-The Arctic University of Norway, Norway
[2]Department of Maritime Operations, University of South-Eastern Norway, 3184 Horten, Norway
[3]Department of Global Business and Trade, Vienna University of Economics and Business, 1020 Vienna, Austria

ABSTRACT

Performance evaluation is fundamental for skill formation and competence development in nautical education and training. Accurate observation and evaluation of students' performance during simulation processes are essential for providing targeted feedback, identifying areas for improvement, and assessing whether the necessary knowledge and skills required to operate vessels safely have been acquired. Instructors are typically the primary assessors; differences in judgement could arise due to variations in their preferences and experience with the tasks, which can in turn influence how the performance is perceived and potentially result in variations for assessment results. This study reviews current methods for performance evaluation used in nautical training and put forth a technology-assisted method that use a combination of control inputs, audio, visual and motion data to classify behaviours and provide feedback for improvement. This can be further developed as an objective and automatic performance evaluation method that serve as a viable supplement to relying solely on human judgment.

Keywords: Performance assessment, Nautical competence, Human performance, Maritime education and training

INTRODUCTION

Simulation is an essential component in nautical skill development and training process, it enables students to practice and acquire hands-on experience in a safe, realistic, and immersive environment. Depending on the functionality of the simulators, a wide range of scenarios can be simulated, from routine operations (e.g., navigating in open waters, maneuvering to ports or narrow areas), to worst-case situations (e.g., ship grounding, collision). As such simulation training offers students a good learning opportunity to practice maneuvering skills, bridge resource management, crisis situation and emergency procedures to effectively respond to a variety of situations in a controlled environment without the costs and risks as in real-world exercises (Kim et al., 2021, Hanzu-Pazara et al., 2008).

To evaluate whether the students have met a desired level of knowledge and skills within the simulated training scenario, performance standard and assessment criteria need to be established to outline what is expected to achieve and how well the students are meeting those criteria. In the performance evaluation process as of present, instructors are typically the primary assessors, whose judgments are based on observing the control actions, how students communicate with others and how the responses were given to different information. Variations in judgements among the instructors are possible and could exist as each instructor may have own experience and opinions about how and when certain tasks should be performed that constitutes "good" or "bad" ship handling and maneuvering. This can influence how performance is perceived and potentially result in variations of assessments.

There are some ways to reduce subjectivity and increase reliability and validity in nautical performance evaluation, such as establishing clear and objective assessment criteria and consistently apply across all students, increasing the number of assessors to evaluate each student(s) performance, or using self-assessment and peer-assessment in addition to instructor-assessment to obtain multi-perspectives on performance to reduce the impact of individual biases and preferences. Besides the effort to improve performance assessment through human assessors, researchers have also looked into automated way of conducting performance assessment to provide a more objective and standardized way of evaluation that help to shape the way that educators approach assessment and evaluation practices in this digital age (Wong et al., 2014). Several studies have explored the use of simulator log data (Øvergård et al., 2017), eye tracking devices (Atik and Arslan, 2019), video recordings (Wei et al., 2009), EEG-based neuroimaging (Liu et al., 2017) for performance assessment as well as development of automated performance evaluation systems to increase objectivity while reducing time and cost associated with solely human performance assessment. The use of simulator outputs data, for instance, can provide a wealth of information on the student's control actions within a given training scenario. Eye tracking devices can be used to analyze gaze patterns and provide insights into their attentional processes and decision-making strategies, which could be used as a tool for measuring situation awareness. EEG-based neuroimaging can provide additional insights into brain activity and cognitive processes, which can be used to detect changes in stress, emotions and mental workload that reflect physiological responses of the student. This physiological data is not obtainable through direct observation and can offer additional inputs for more informed instructional decisions, such as adjusting the difficulty level of training scenarios or providing targeted support to address specific issues.

Bibliometric Analysis on Technology-Assisted Performance Evaluation

In recent years, research in technology-assisted performance evaluation has gained popularity as a way to increase the efficiency, objectivity, and reliability of performance assessment across various domains, including education, transportation, healthcare, and sports (Stevens-Adams et al., 2010). The capability to automatically assess the performance is often achieved through a combination of sensors, cameras, software applications and Machine Learning (ML) algorithms to process the collected data and recognize patterns or anomalies that are indicative of performance status. Such a system is not only complex in its setup but also in the process of ensuring its effectiveness and reliability. One of the main challenges of developing it is ensuring that the data being collected is accurate and reliable. This requires careful calibration of sensors and cameras used to capture performance data, as well as repetitive testing to ensure that the datasets being collected are consistent and accurate. Another challenge is related to ML algorithms development. As a large amount of training data is needed to train these algorithms, ongoing refinement is required to tune these underlying ML algorithms to ensure that they are providing accurate analysis and useful feedback.

A bibliometric analysis of literatures on technology-assisted performance evaluation published in last 30 years between 1992-01-01 and 2023-04-11 were conducted based on the data obtained from Web of Science (WoS) collection database. Total 9,805 journal articles were obtained using keywords i.e., “automated performance assessment”, “computerized performance assessment”, “educational data mining”, “human performance evaluation” and other similar terms. Only journal articles written in English were selected during the process of retrieval. Keywords occurred for more than 50 times in WoS core collection were analyzed. Of 37,383 keywords, 96 met the threshold. “Performance assessment” appeared as the most frequent keywords with total link strength of 2,425, which had strong links to model, design, optimization, and reliability. The research field under study has revealed six clusters and lead by several influential researchers (Figure 1): The first cluster, represented by the blue color, centers on the composition of performance assessment system and methodologies, which involves investigating structure and design of performance assessment tools. The orange cluster is concerned with validity and reliability of performance assessment system through simulation and assessment. The black cluster delves into the algorithms, behaviors, and characteristics of performance assessment systems. The purple cluster explores the impacts and efficiency of using performance assessment tools on improving learning outcomes and focuses on the effects of performance assessment. The red cluster is centered around performance monitoring and diagnosis, whereas the green cluster focuses on utilizing data mining techniques and analytics to derive insights for informed decisions within educational settings.

Since 2018, a noticeable rise in research publications occurred (see Figure 2) and a large proportion of these articles were published in Elsevier’s journals like engineering structures, reliability engineering and system safety, which have been cited 3,795 and 2,732 times respectively. Most recent studies (green cluster) are primarily focused on learning analytics, ML models and algorithm optimizations, indicating a growing interest on leveraging data-driven approaches and ML techniques to enhance educational outcomes.

Total 8,419 organizations have research outputs in this area, of which 131 have over 20 published articles. The top ten most active countries and organizations are listed in Table 1. Among these, University of Alberta counts

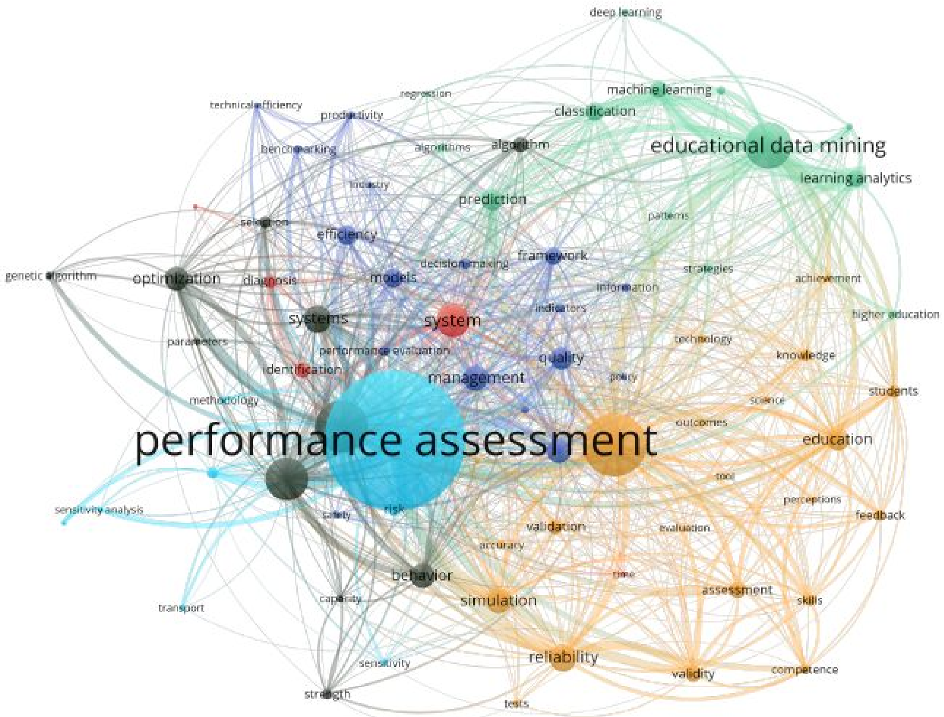

Figure 1: Co-occurrence of keywords in technology assisted performance evaluation literature.

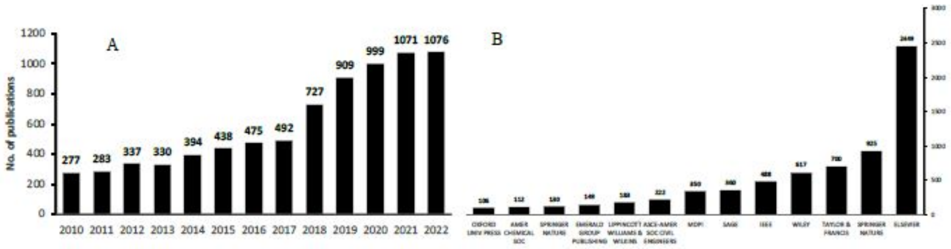

Figure 2: (A) Number of articles published by years (B) top ten journals with most published articles in the field of technology assisted performance evaluation.

the most publications (105) with a total of 2,505 citations. About 26.6% (2,609) of 9,805 articles were from US and have been cited 65,252 times with a total link strength of 228,964 (bibliographic coupling links with other countries).

The development of technology assisted performance evaluation system for objective assessment during simulation requires consideration of various data sources and analysis techniques. Such a system would need to be able to capture, process and analyse data and to provide instructors and students with immediate or summative feedback during or after the simulation exercises. There are significant challenges associated with its development and validation due to the architectural complexities and analytical challenges. Despite these challenges, there is growing recognition of the potential

Table 1. Top ten most active countries and organizations.

Subject	No. of publications	Total No. of citations	Total link strength
Country			
United States	2,609	65,252	228,964
China	1247	17,161	158,819
United Kingdom	728	19,522	100,445
Italy	612	13,106	70,670
Canada	605	14,783	105,667
India	555	5,393	66,903
Germany	450	12,523	68,234
Spain	417	10,741	70,225
Australia	393	8,478	66,223
Iran	383	4,068	70,610
Organization			
University of Alberta	105	2,505	18,622
Islamic Azad University	90	1,158	9,497
Politecnico di Milano	80	1,951	4,546
Sandia National Laboratories	78	1,516	13,534
University of Michigan	77	2,073	6,511
University of British Columbia	72	2,347	7,332
Stanford University	68	1,971	6,096
University of Toronto	68	1,517	6,528
University of Tehran	67	1,028	8311
Chinese Academy of Sciences	66	1,600	2,528

benefits of technology-assisted performance evaluation in professional training, particularly in fields where hands-on experience and practical skills are essential for competence development. In healthcare, technology-assisted performance evaluation has shown to improve the effectiveness of training and reduce the risk of medical errors (Murali et al., 2020). Similarly, a study in the aviation context has noted that the instructor group using Enhancing Performance with Improved Coordination (EPIC) tool had more differentiated and accurate assessment of team performance compared to the control group (Deaton et al., 2007).

While simulation-based learning is widely implemented in the field of maritime education and training (MET), very few studies only focused on technology-assisted performance evaluation system for objective evaluation of learning outcomes in maritime context (Wu et al., 2017, Ernstsen and Nazir, 2020, Øvergård et al., 2017). Among the 9,805 articles being reviewed, only 6 studies were focused on technology assisted performance assessment system in MET, which presented an opportunity for researchers to further investigate this area and contribute to the advancement of objective performance evaluation in nautical training programs.

Architecture Evaluation of Semi-Automated Performance Evaluation System (SAPES) in Nautical Simulations

In the following Figure 3, a semi-automated performance evaluation system (SAPES) is outlined as a technology-assisted method of performance evaluation that utilizes a combination of simulator multiple data inputs to classify trainee behaviors and provide feedback for improvement.

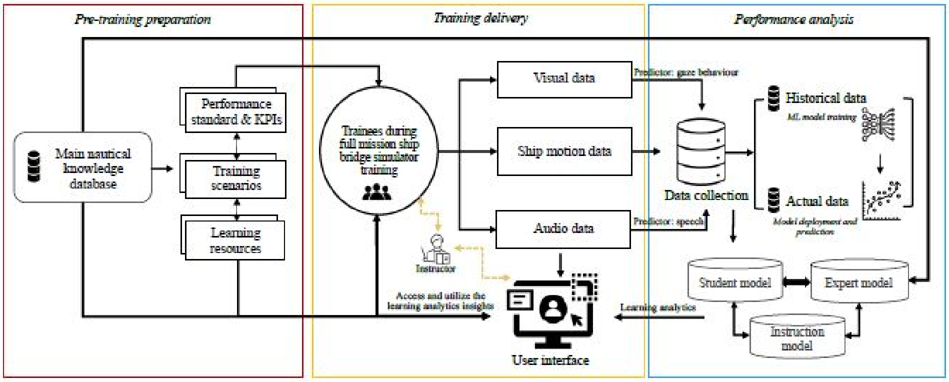

Figure 3: Overview of performance evaluation system in nautical simulations.

The SAPES here consists of a combination of sensors, software algorithms, and a graphical user interface as performance dashboard. The data outputs collected from eye-tracking glass, cameras, audio recorders and ship simulators will serve as the data sources and inputs to develop and train the algorithms to classify behaviors related to navigation, communication, decision-making, and situational awareness. There are several datasets that would be needed for this process, which include visual data, ship motion data and audio data, all of which could be collected during training delivery. The gaze behaviors generated through the eye tracker could be made visible, and audio data complements the performance evaluation by providing additional information and context. The system transcribes and analyzes the spoken content through integrating a "Speech to Text" API. Gaze behavior can provide insights into a trainee's attention and cognitive processes. If a trainee has frequent and prolonged fixations on a particular value or screen, it could suggests that they are paying close attention to that particular content, while frequent shifts in gaze with short fixations may indicate distraction, confusion or difficulty maintaining focus. The simulator log data could provide information regarding the ship motion, and reflect on the accuracy of control orders, response times to changing ship motion, adherence to predefined protocols or procedures, and overall proficiency in ship handling. These datasets need to be preprocessed to ensure their quality and compatibility for further analysis and modelling, this could include normalizing the data, remove irrelevant segments, align the timestamps across different data sources to ensure synchronization between eye tracker data, audio data, and simulator log data. After the features are extracted, ML algorithms can be developed and employed to analyze the datasets and uncover patterns, correlations or predictive relationships, this would form the basis for adaptive learning in the system. The user interface could serve as a performance dashboard that provides visual presentation of the students' performance, allowing the instructor to monitor learning progress and identify areas for its improvement.

This algorithm-based performance evaluation setup is beneficial for several reasons. Firstly, it enforces standardized evaluation criteria and metrics, which could provide evaluations consistently and impartially compared to instructor-based evaluations, as they rely on data-driven analysis rather than personal judgment. Secondly, it allows for automated and efficient analysis of a large amount of data, which could save time and effort from instructors to review and assess individual cases. Instructors can also bring their expertise and qualitative insights to the evaluation process, which algorithms may not capture fully. A conceptual evaluation of the SAPES will be performed as the next research step to test it in a simulated nautical environment to determine its accuracy and effectiveness in assessing training performance. The evaluation involves comparing the performance measurements generated by the SAPES with those generated by human assessors only. This comparison can provide information related to a level of agreement between the SAPES and human assessors and can help identify any discrepancies or biases in the performance assessment process.

CONCLUSION

Simulator-based learning is common in professional training process due to its ability to provide a safe and controlled practical learning environment. However, learning effectiveness depends on the quality of the simulation curriculum and the ability of assessors to objectively evaluate students against the learning outcomes. The SAPES outlined in this paper has the potential to serve as a viable alternative or supplement in nautical education and training than relying solely on human judgment. Its development and implementation, however, require careful consideration of data sources, analysis techniques and architectural complexities.

ACKNOWLEDGMENT

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

REFERENCES

Atik, O. & Arslan, O. 2019. Use of eye tracking for assessment of electronic navigation competency in maritime training. *Journal of eye movement research,* 12.

Deaton, J. E., Bell, B., Fowlkes, J., Bowers, C., Jentsch, F. & Bell, M. A. 2007. Enhancing team training and performance with automated performance assessment tools. *International Journal of Aviation Psychology,* 17, 317–331.

Ernstsen, J. & Nazir, S. 2020. Performance assessment in full-scale simulators - A case of maritime pilotage operations. *Safety Science,* 129, 14.

Hanzu-Pazara, R., Barsan, E., Arsenie, P., Chiotoroiu, L. & Raicu, G. 2008. Reducing of maritime accidents caused by human factors using simulators in training process. *Journal of Maritime Research,* 5, 3–18.

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, 135–150.

Liu, Y., Sourina, O., Liew, H. P., Salem, H. S. & Ang, E. 2017. Human factors evaluation in maritime virtual simulators using mobile EEG-based neuroimaging. *Transdisciplinary Engineering: A Paradigm Shift,* 5, 261–268.

Murali, B., Belvroy, V. M., Pandey, S., Byrne, M. D., Bismuth, J. & O'malley, M. K. 2020 Towards Automated Performance Assessment using Velocity-based Motion Quality Metrics. 2020. International Symposium on Medical Robotics (ISMR), 18–20 Nov. 2020. 36–42.

Øvergård, K. I., Nazir, S. & Solberg, A. S. 2017. Towards automated performance assessment for maritime navigation. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation,* 11, 229–234.

Stevens-Adams, S. M., Basilico, J. D., Abbott, R. G., Gieseler, C. J. & Forsythe, C. 2010. Using after-action review based on automated performance assessment to enhance training effectiveness. 2010. SAGE Publications Sage CA: Los Angeles, CA, 2309–2313.

Wei, H., Nguyen, H., Ramu, P., Raju, C., Liu, X.& Yadegar, J. Automated intelligent video surveillance system for ships, Proc. SPIE 7306, Optics and Photonics in Global Homeland Security V and Biometric Technology for Human Identification VI, 73061N. https://doi.org/10.1117/12.819051

Wong, J. H., Ogren, L., Peters, S., Bratt, E. O. & Newport, R. Developing and evaluating an intelligent tutoring system for advanced shiphandling. Proceedings of the Interservice/Industry, Training, Simulation, and Education Conference (I/ITSEC) Annual Meeting, 2014.

Wu, Y., Miwa, T. & Uchida, M. 2017. Advantages and obstacles of applying physiological computing in real world: Lessons learned from simulator based maritime training. *International Journal of Maritime Engineering,* 159, A149–A157.