TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 11
Number 2
June 2017

DOI: 10.12716/1001.11.02.03

Towards Automated Performance Assessment for Maritime Navigation

K.I. Øvergård, S. Nazir & A. S. Solberg
University College of Southeast Norway, Borre, Norway

ABSTRACT: This paper presents the outcome of a pre-project that resulted in an initial version (prototype) of an automated assessment algorithm for a specific maritime operation. The prototype is based on identified control requirements that human operators must meet to conduct safe navigation. Current assessment methods of navigation in simulators involve subject matter experts, whose evaluations unfortunately have some limitations related to reproducibility and consistency. Automated assessment algorithms may address these limitations. For a prototype, our algorithm had a large correlation with evaluations performed by subject matter experts in assessment of navigation routes. The results indicate that further research in automated assessment of maritime navigation has merit. The algorithm can be a stepping stone in developing a consistent, unbiased, and transparent assessment module for evaluating maritime navigation performance.

1 INTRODUCTION

The complexity of maritime navigation is continuously increasing. The growth in the world fleet size and the changing characteristics of vessels increases the need for skilled navigators. Assessment of navigation skills is therefore an ever more important issue. However, the methods for assessing the performance of human operators have mostly remained unchanged. Currently, the performance assessment during training in simulators is done as a subjective assessment by domain experts.

Research suggests limitations to the reliability of subjective assessments. In principle, all fair tests are designed to differentiate between those that have a trait (*e.g.* being competent) and those that lack the trait (*e.g.* those that are not competent). However, since a human subject matter expert (SME) is the assessment tool of the trainees' performance within the simulator, the assessment is affected by the biases that follows from such a subjective evaluation (Manca *et al.*, 2012; Nazir & Manca, 2015). A bias in assessment involves the tendency to systematically shift the evaluation away from a consistent score (Kahneman, 2011; Allen & Yen, 1979). The presence of biases lowers the reliability of the assessment (Cronbach *et al.*, 1972; Freedman, 2009). Biases can arise from the fact that humans are not perfectly rational decision-makers (Simon, 1979). Humans show non-optimal decision making and judgement even in situations where all necessary information is available to make an optimal decision (Kahneman, Slovic & Tversky, 1982). Also, human assessments do not have perfect test-retest reliability but can vary as a function of time (Fried & Feldman, 2008). Hence, identical performances at different times can lead to different assessments. Biases in assessment involving human judgment are a general phenomenon and are much researched in fields such as medicine (Higgins & Altman, 2008; Higgins *et al.*, 2011), and psychology (Kahneman *et al.*, 1982; Kahneman, 2011). These

limitations apply to all subjective assessments – including the evaluations done by subject matter experts (SMEs). Factors such as time of the day, fatigue, mood, and low blood sugar levels can also negatively affect the outcome of an expert evaluation (Danziger *et al.*, 2011).

We suggest the use of automated assessment algorithms to support the subjective assessment by SMEs. Unfortunately, there are numerous challenges to automating the assessment of maritime navigation. *First*, navigation is an open goal-oriented work task which is characterized by having multiple degrees of freedom. *Second*, navigators have the freedom to choose the sequence and timing of work tasks (*i.e.* there are few procedures that prescribe how a task shall be done in a specific situation). *Third*, navigation and manoeuvring must be context-sensitive and adaptive since vessels move around in a cluttered environment with multiple obstacles or objects (*e.g.* rocks, land, other ships or objects in the water). *Fourth*, the degrees of freedom are also exceptionally large because it often exists numerous *acceptable* ways of reaching a destination (*e.g.* sailing from A to B), meaning that there are several possible routes that all conform to the requirements for safety and efficiency (Bjørkli *et al*., 2007). *Fifth*, maritime navigation is related to a number of constraints related to physical laws, operational limits, societal laws/regulations, and organizational goals related to safety, economy, and the environment (Øvergård, 2012).

The sensitivity and complexity of making automated assessments of maritime navigation and manoeuvring has refrained many researchers from developing automated methods and procedures to assess navigation performance in real time. One exception is the Navigational Risk Detection and Assessment System (NARIDAS; Gauss, Rötting & Kersandt, 2007; Gauss & Kersandt, 2005; Hederström, Kersandt & Müller, 2012) – a system that combines multiple parameters to form a risk assessment of navigation. This system has focused on risk assessment of navigation, and not on the assessment of navigational performance as such.

2 CURRENT STATE OF AUTOMATED ASSESSMENT OF OPERATOR PERFORMANCE

A literature survey on automated procedures for operator performance assessment suggests more research is needed. The examples that exists comes from the aviation (Johannes *et al.*, 2007), the naval (McCormack, 2007; Bjørkli & Øvergård, 2012), and the surgery domain (Fried & Feldman, 2008). Johannes *et al.* (2007) validated an automated assessment method in a flight simulator by showing high correlations between the outcome of the assessment algorithm and expert trainer's rating of the operator's simulator performance. However, this approach requires a human expert trainer to visually identify the behaviour/actions made by the trainee - limiting the automaticity of the algorithm.

To date, there exist no fully functional automated assessment systems that are adapted for tasks with large degrees of freedom. A limited number of current objective assessment algorithms is employed on procedure-based work scenarios where the sequence and timings of actions can be pre-defined. Examples of assessment systems for procedure-based scenarios are the K-SIM ® Polaris – Ships Bridge simulator (Kongsberg Maritime, 2017) and systems for the automated assessment of operators' performance in a petrochemical process simulator (Manca *et al.*, 2012; Manca & Brambila, 2011; Nazir *et al*., 2013; Nazir *et al*., 2015). However, both systems focus on procedure-based work tasks, and are not designed to handle open goal-oriented dynamic work tasks such as coastal navigation.

It is becoming increasingly important to establish an unbiased evaluation system that can contribute consistent, unbiased, and transparent evaluation of operators' skills and competencies. To meet these challenges, we have quantified some of the *control requirements* (Petersen, 2004; Bjørkli *et al*., 2007; Øvergård *et al*., 2010) that a navigator must meet to conduct navigation in a safe manner. The research is part of the GruNT pre-project.

3 AIM OF THIS PAPER

The main aim of this paper is to present the first steps toward making an automated assessment algorithm for dynamic goal-oriented work tasks, such as maritime navigation. It presents the outcome of the GruNT pre-project, which includes the first validation study of a simple form of this algorithm.

4 METHOD

Identification of control requirements for the safety of navigation was done using open interviews with six SMEs who all held deck-officer certifications. Open interviews were chosen to allow the SMEs full freedom to talk about important parameters and requirements they believe are important in the assessment of navigation. Several control requirements were identified. Control requirements where compared to the information that was available in the log system of the K-Sim® simulator. We then selected the control requirements that had relevant parameters in the logging system. The control requirements selected for use in the pre-project are: 1) distance to land based on own ship length, 2) distance to moving objects (vessels) based on own ship length, 3) distance to floating objects based on ship length, 4) the deviation between ship heading and heading of dock (meaning that the ship should be parallel to the dock during the last part of docking), and 5) the minimum depth below the ship´s keel (the so-called 'safety depth').

Based upon the input from the SME´s, we defined hundred-point limits (HPL) and zero-point-limits (ZPL) for each of the parameters to fit the simulator model of the vessel "Thor Magni" (IMO 9679024). 'Thor Magni' is a 64.40 meters long offshore vessel. A draught of 5.70 meters was selected for the vessel in the scenarios. The HPL and ZPL values for the vessel are given in table 1.

Table 1. Performance Indicators and Control Requirements for "Thor Magni"

Performance Indicators	HPL	ZPL
Distance to land	>539m	<263m
Distance to small floating objects	>270m	<132m
Distance to moving objects (e.g. vessels)	>1058m	<539m
Safety depth (clearance under keel, aft)	HPL > 5m > ZPL	
Deviation between Heading of vessel and Heading of dock at 10 meters' distance (measured in degrees)	<3 deg.	>5 deg.

NOTE: The safety depth is the same as the minimum water depth below the vessel´s keel – corresponding to a water depth of 5.7 + 5 = 10.7 meters. HPL = Hundred-point limit, ZPL=Zero-point Limit, m = meters, deg. = degrees

If the score of one of the parameters was above the HPL the parameter was scored as 100 points. If the score was below the ZPL a score of 0 points was given. If the parameter was between the HPL and the ZPL a score equal to the linear interpolation between these scores was given. For example, if the distance to land was 401 meters a score of 50 points was given, indicating that the vessel did not have an optimal positing relative to land. The calculation is shown in table 2.

Table 2. Score calculation for distance to land

Situation	D > HPL	HPL ≥ D ≥ ZPL	ZPL > D
Score calculation	100	$\frac{D-ZPL}{HPL-ZPL}*100$	0

NOTE: D=Distance to land, HPL=Hundred-point limit, ZPL=Zero-point limit

4.1 *Creation of routes for assessment*

The third author created 20 different routes using the Kongsberg K-Sim® Navigation simulator version 2.2. All routes started just east of Mefjordbåen in the Oslo fjord and the destination was the deep-water dock in Horten on the western side of the Oslo fjord. The routes were intentionally made of different quality (from excellent to poor) – thereby creating variance that would allow us to measure the extent of covariance between the assessment made by the SMEs and the assessment algorithm.

Data from these simulator trials were logged once per second and saved in EXCEL-files. The parameters were then transformed into scores between 100 and 0 per the limits described in Table 1. The minimum score for each of the five parameters (during the whole session) was used as representative parameter scores for each of the 20 trials. The mean of the minimum scores were then used to calculate an overall score for each scenario.

4.2 *Validation of assessment algorithm*

Validation was conducted in the initial phase to reach consistent results. Two dedicated SMEs were requested to rank the 20 developed routes. In addition, another SME gave an individual ranking of the 20 routes. The rankings done by the SMEs were independent of each other. The SMEs were not informed of the output from the assessment algorithm.

The ranking of the 20 routes were done by showing printed images of the routes to the SMEs. Examples of the images are seen in Figures 1a-b. We also gave additional images showing the closest passage between a vessel and other vessels for each route – allowing the experts to assess whether the "Thor Magni" was too close to land, other vessels and floating objects.

We acknowledge that these pictures are not a suitable way to assess navigational performance during training or education. However, our intention was to see whether a human evaluation of a reduced set of information was similar as the output from our simple algorithm. Future research will of course involve more complex algorithms and assessment of real-time simulator-based navigation.

5 RESULTS

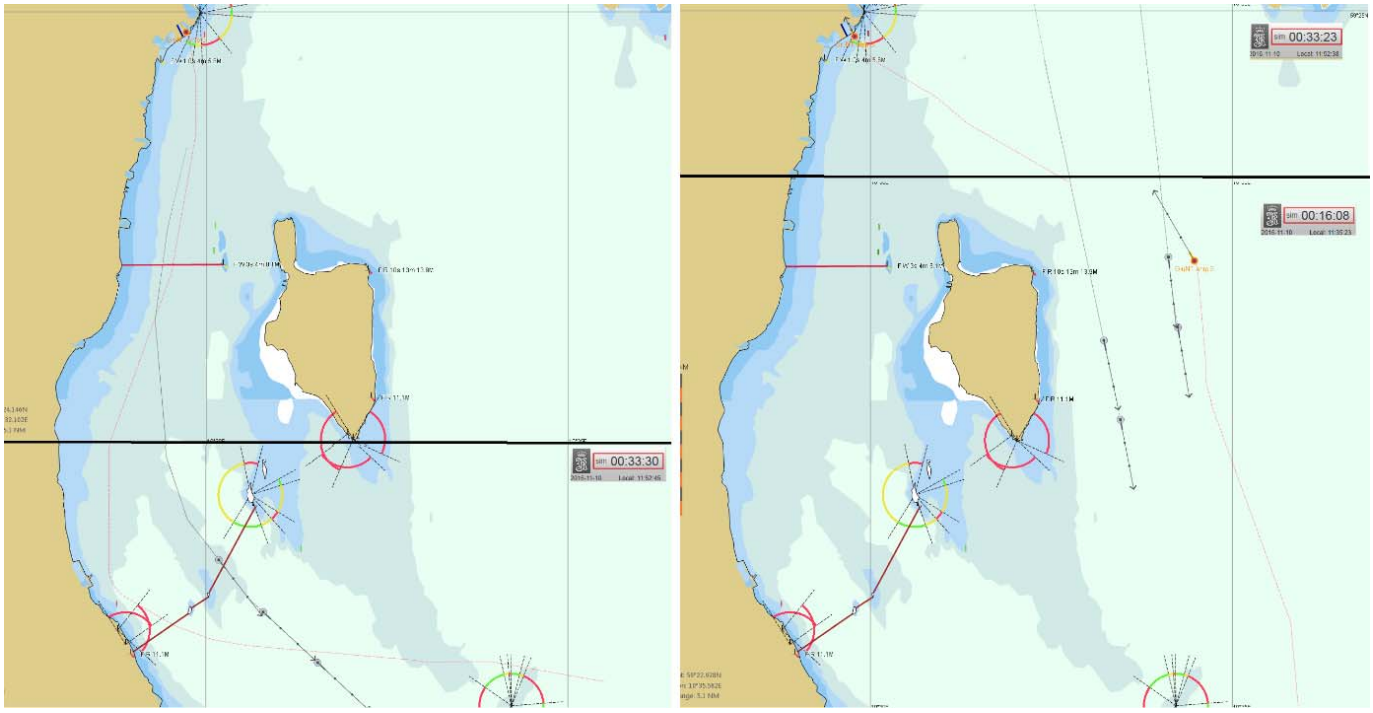
We first took the mean scores from the assessment algorithms and ranked them using mean ranks. If the same scores were obtained for multiple routes each of these were assigned the mean rank (for example, three routes scored 100 points, these where all given the rank of 1+2+3 = 6/3 = 2).

Rating of the effect sizes as 'small', 'medium' and 'large' are done in accordance with Cohen´s (1988) classification of effect sizes. We compared the association between the algorithm´s ranking of the routes with the two independent rankings from the SMEs. Spearman´s Rank Correlation Coefficient (r_s) between the algorithm and SME 1&2 was large (r_s = 0.61, 95% CI [0.177, 0.885]).

To give a visual representation of the correlations we have printed scatterplots showing the relationship between the algorithm and SME 1&2 (see Figure 2).

The correlation algorithm and SME 3 was also large (r_s = 0.551, 95% CI [0.117, 0.859]). Figure 3 shows a graphical representation of the relationship between SME3 and the algorithm.

The correlation between the SME 1 & 2 and SME 3 was also large (r_s = 0.815, 95% CI [0.542, 0.942]), indicating that the SMEs agreed to a larger extent with each other than they did with the algorithm.



Figures 1a-b: In the middle of the images you see Bastøy island. The starting point of each route is just east of Mefjordbåen which is the beacon in the image´s lower end. The routes went to Horten harbor which is seen in the upper part of Figure 1a-b. The figures are created out of multiple screen shots from the K-SIM® Navigation Simulator.

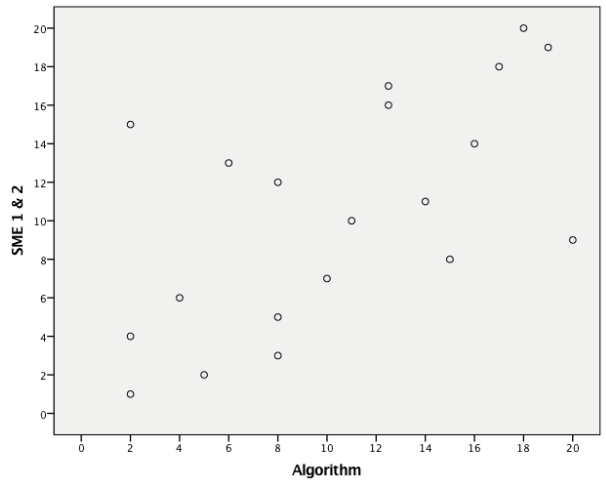


Figure 2. SME 1&2 vs. Algorithm

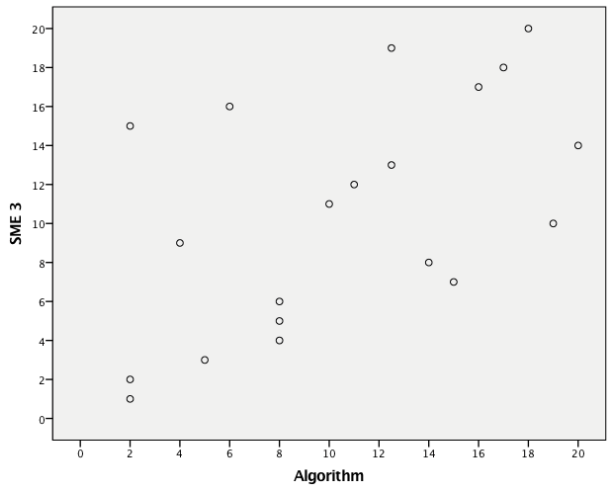


Figure 3: SME 3 vs. Algorithm

6 DISCUSSION

This paper describes the work done in the GruNT pre-project. The project aimed to do a proof-of-concept test of an automated assessment algorithm for the assessment of a simple scenario involving maritime navigation and manoeuvring. A total of 20 sailing routes between Mefjordbåen and Horten in the Oslo fjord were evaluated and ranked on two different occasions by different SMEs. The SME´s rankings were then compared with the assessment algorithm´s ranking of the same 20 sailing routes.

The results showed large rank-based correlations ('large' is >.50 per Cohen´s (1992) classification of Effect Sizes) between the SMEs and the assessment algorithm (r_s = 0.515 and 0.610). The correlations indicate that there is a quite good fit between the algorithm and the SMEs when evaluating a reduced and very simple case of coastal navigation.

The agreement between SMEs and the algorithm demonstrates the algorithm´s *criterion validity* by showing that there exists covariance between multiple different measurements of the same phenomenon (Fried & Feldman, 2008). The control requirements also have *face validity*, as distance to objects, land, and other vessels are important factors with respect to collisions and grounding. The deviation of the vessel´s heading to the heading of the dock is admittedly only relevant for a sub-set of the available docking procedures. Therefore, the algorithm needs to be further improved so it becomes more complex (more parameters and multi-dimensional parameters) and that it allows for more general navigation scenarios. This is also supported by the fact that correlation between the SME´s rankings were higher (r_s = 0.813) than between SMEs and the algorithm (r_s = 0.61 and 0.515), indicating that the assessment algorithm lacks some criteria that the human SMEs

are using and/or that the limits (HPL and ZPL) needs to be altered. A notable case of mismatch between SMEs and algorithm is the route which have been ranked as number 15 by both SME 1&2 and SME3 but which is ranked second best (2) by the algorithm, see Figure 2 and 3).

The model at this stage is simplified to assess only a few aspects of navigation. For example, environmental and efficiency aspects are not assessed at all, and there is a need for many more parameters to assess safety of navigation and manoeuvring. In contrast, SMEs tend to evaluate the simulations as real navigation exercises. However, the correlation achieved between the model and the SMEs indicate that the proof-of-concept assessment algorithm has merit – despite its apparent simplicity.

7 LIMITATIONS

The limitations towards assessing real navigation are apparent, and the authors´ acknowledge this. The limited nature of the assessed scenario, the limited number of parameters in the assessment algorithm, and the way that the SMEs assessed and ranked the 20 sailing routes are all points of improvement. Hence, we do not attempt to generalize our findings beyond saying that we think that our conceptual idea for automated assessment of maritime navigation has merit.

Related to the scenario we used, we are also faced with the fundamental difference between coastal navigation and harbour manoeuvring. During harbour manoeuvring the focus is on controlling the forces between the ship and the waters to ensure desired movement. During coastal navigation, less focus is needed on force controlling since the vessel generally moves in one direction with less sharp turns. In such circumstances a focus on safety distances, and displaying intentions to other ships may be more important.

The limits described in Table 1 are based upon input from only six SMEs. Hence, we do not know if these limits are something that the large population of navigators would agree to. This will be the focus for further research.

Also, our assessment algorithm is at present only designed for assessing technical skills such as handling and navigation of the vessel. It cannot (currently) assess aspects related to the interaction between humans during navigation. This includes features like non-technical skills (Flin et al., 2008), situation awareness (Endsley, 1995), or team communication (Øvergård et al., 2015). To assess these 'soft' skills we must combine other assessment methodologies with our assessment algorithm. This is of course one of the future research challenges.

The model does not yet describe a real-world navigation situation, but rather a simplified model. At this stage of research, the results presented herein is encouraging, and indicates that further research into automated performance assessment in the maritime domain is warranted. The need for further research is also supported by the fact that the industry has shown interests in systems that can assess the performance of truly autonomous vessels. The current model, albeit limited. illustrates some of the methodological challenges and give an indication of the feasibility of automated algorithm-based assessment in the maritime domain.

8 FUTURE RESEARCH

Future research will focus on identifying more control requirements for safe, efficient and 'green' navigation. The aim will be to identify the limits for these control requirements by investigating a large multi-national sample of experienced navigators.

Another research challenge is how to combine and to create consistent weights for the assessment scores of safe, efficient and 'green' navigation. Methods such as Analytic Hierarchy Process (AHP; Saaty, 1980) exist, but this approach cannot solve problems where the weights change dynamically (Saaty, 2007). Hence, we will research ways to ensure a consistent set of weights between different parameters that will allow us to assess navigation and manoeuvring in both open and confined waters.

Also, we aim to extract information from AIS-data about historical sailing routes to determine where vessels normally navigate. Based upon this, we hope to supplement the data we get from talking to experienced navigators by also identifying the statistical distributions of acceptable distances between vessels, land, floating objects as a function of characteristics of the vessel and the situation.

The research presented may also have relevance for the automated assessment of the performance of truly autonomous vessels. Further research into automated assessment is likely to consider the development in the field of autonomous vessels.

9 CONCLUSION

We have developed a simple version of an automated assessment module based upon the quantification of control requirements for safe navigation. There is a large degree of agreement between SMEs and our assessment algorithm, indicating that our simple proof-of-concept model may have merit. We believe the model presented in this paper may be a stepping stone into larger research efforts.

ACKNOWLEDGEMENTS

The GruNT pre-project was funded by Kongsberg Digital and the Oslofjord Regional Research Fund in Norway (project number 258894).

REFERENCES

Allen, M. J. & Yen, W. M. (1979). *Introduction to Measurement Theory*. Belmont, CA: Wadsworth.

Bjørkli, C. A., Øvergård, K. I., Røed, B. K., & Hoff, T. (2007). Control Situations in High-Speed Craft Operation. *Cognition, Technology, and Work*, 9, 67-80. doi: 10.1007/s10111-006-0042-z

Bjørkli, C. A. & Øvergård, K. I. (2012). *Automated assessment of docking maneuvers: When do we know when an operator performs well?* Presentation at Scandinavian Maritime Conference 2012, 28-29 November at Vestfold University College, Horten, Norway.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements*. London, England: John Wiley.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences*, *108*(17), 6889-6892. doi: 10.1073/pnas.1018033108

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32-64. doi: 10.1518/001872095779049543

Flin, R. H., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end: a guide to non-technical skills*. Aldershot, England: Ashgate.

Freedman, D. A. (2009). *Statistical Models: Theory and Practice, rev. ed.* Cambridge, England: Cambridge University Press.

Fried, G. M., & Feldman, L. S. (2008). Objective assessment of technical performance. *World Journal of Surgery*, *32*, 156-160. doi: 10.1007/s00268-007-9143-y

Gauss, B., & Kersandt, D. (2005). NARIDAS-Navigational Risk Detection and Assessment System for the Ship's Bridge. In Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, 2005 (Vol. 2, pp. 612-617). IEEE.

Gauss, B., Rötting, M., & Kersandt, D. (2007). NARIDAS–evaluation of a risk assessment system for the ship's bridge. In Human Factors in Ship Design, Safety and Operation. RINA-The Royal Institution of Naval Architects. International Conference.

Hederström, H., Kersandt, D., & Müller, B. (2012). Task-oriented structure of the navigation process and quality control of its properties by a nautical task management monitor (ntmm). European Journal of Navigation, 10(3).

Higgins, J. P. T. & Altman, D. G. (2008). Assessing risk of bias in included study. In J. P. T. Higgins and S. Green (eds.). *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 187-242). West Sussex, England: John Wiley & Sons.

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savovic, *et al.* (2011). The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 343*(7829), d5928. doi: 10.1136/bmj.d5928

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, England: Cambridge University Press.

Kongsberg Maritime (2017) K-Sim Navigation – Kongsberg. Web site Kongsberg Maritime [Available at] https://www.kongsberg.com/en/kongsberg-digital/maritime%20simulation/k-sim%20navigation%20-page/

Manca, D., Nazir, S., Colombo, S., & Kluge, A. (2014). Procedure for automated assessment of industrial operators. *Chemical Engineering Transactions*, *36*, 391-396. doi: 10.3303/CET1436066

Manca, D., & Brambilla, S. (2011). A methodology based on the Analytic Hierarchy Process for the quantitative assessment of emergency preparedness and response in road tunnels. *Transport Policy*, *18*(5), 657-664. doi: 10.1016/j.tranpol.2010.12.003

Manca, D., Nazir, S., Lucernoni, F., & Colombo, S. (2012). Performance indicators for the assessment of industrial operator. *Computer Aided Chemical Engineering*, 30, 1422-1426. Doi:10.1016/B978-0-444-59520-1.50143-3.

McCormack, W. (2007). Automated Operator and System Performance Assessment. In T. Bastiaens & S. Carliner (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* 2007 (pp. 7252-7259). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Nazir, S., Colombo, S., & Manca, D. (2013). Minimizing the risk in the process industry by using a plant simulator: a novel approach. *Chemical Engineering Transactions*, *32*, 109-114. doi: 10.3303/ACOS1311028

Nazir, S., & Manca, D. (2015). How a plant simulator can improve industrial safety. Process Safety Progress, 34(3), 237-243. doi:10.1002/prs.11714

Nazir, S., Sorensen, L. J., Øvergård, K. I. & Manca, D. (2015). Impact of training methods on distributed situation awareness of industrial operators. *Safety Science*, *73*, 136-145. doi: 10.1016/j.ssci.2014.11.015

Petersen, J. (2004). Control situations in supervisory control. *Cognition, Technology, and Work, 6*, 266-274. doi: 10.1007/s10111-004-0164-0

Saaty, T. L. (1980). *The analytic hierarchy process: planning, priority setting, resources allocation*. New York, NY: McGraw-Hill.

Saaty, T. L. (2007). Time dependent decision-making; dynamic priorities in the AHP/ANP: Generalizing from points to functions and from real to complex variables. *Mathematical and Computer Modelling*, *46*(7), 860-891.

Øvergård, K. I. (2012). *Absolute constraints, situation awareness and modelling of socio-technical systems.* Presentation at the Scandinavian Maritime Conference 2012, 28-29 November at Vestfold University College, Horten, Norway.

Øvergård, K. I., Bjørkli, C. A., Røed, B. K. & Hoff, T. (2010). Control strategies used by experienced marine navigators: observations of verbal conversations during navigation training. *Cognition, Technology, and Work, 12(3)*, 163-179. doi: 10.1007/s10111-009-0132-9

Øvergård, K. I., Nielsen, A. R., Nazir, S., & Sorensen, L. J. (2015). Assessing navigational teamwork through the situational correctness and relevance of communication. *Procedia Manufacturing*, *3*, 2589-2596. doi: 10.1016/j.promfg.2015.07.579

.