(article starts on next page)

Certifying Navigational Skills: A Video-based Study on Assessments in Simulated Environments

C. Sellberg, O. Lindmark & M. Lundin
Chalmers University of Technology, Gothenburg, Sweden
University of Gothenburg, Gothenburg, Sweden

ABSTRACT: In Maritime Education and Training (MET) where students are trained for professions with high standards of safety, the use of simulators is taken to provide opportunities for safe and cost-effective training. Although the use of simulators for training and certifying technical proficiency and so-called non-technical skills is well established and regulated by international standards, previous research suggests that simulator-based assessment has been poorly implemented in the MET system. Now the challenge is to contribute with knowledge about how to conduct consistent, unbiased, and transparent assessments of navigational skills and competencies. However, in current research it is not evident how training of non-technical skills in simulated environments should be assessed. The aim of this study is to explore the pedagogical challenges instructors face when assessing students' navigational skills and competencies in a simulated environment. The study is based on video-recorded data from the certification part in a navigation course for second year master mariner students. A situated approach to cognition and learning is employed to analyze the co-construction of assessment in the simulated exercises by means of instructors' questions and students' answers. Results reveal an assessment practice where the students are still developing their navigational skills with instructional support from examiners whilst being certified on using Radar equipment in accordance to COLREG.

1 INTRODUCTION

In this study we explore the instructional challenges examiners in Maritime Education and Training (MET) face when assessing students' navigational competencies and skills in a simulated environment. In MET the international convention Standards of Training, Certification and Watchkeeping for Seafarers (STCW) provides regulations for performance-based competency tests, highlighting that such tests should be performed in a simulator (Regulation I/12). In navigation courses, the STCW code emphasizes both technical proficiency and so-called non-technical skills as minimum standard for certification as master mariners. The term non-technical skills have been defined by Flin, O'Connor and Crichton (2008) as cognitive, social and personal skills that contribute to safe and efficient task performance in safety critical domains. The learning objectives concerning non-technical skills to be assessed through simulation-based tests are: content knowledge, application and intent of the regulations for preventing collisions at sea (COLREG), in addition to skills in team- and resource management for contributing to a safe watch. Such skills include for example, the abilities to understand orders and to use the appropriate internal communication and alarm systems, i.e. clear and concise communication at all times and orders "in a seamanlike manner".

However, in a recent overview of the field, Sellberg (2017) found that several aspects of the

current training and assessment system stand out as alarming. For example, Emad and Roth (2008) conclude that MET fails to achieve its learning objectives. Rather, MET has actually changed the learning objectives to help students pass competence tests in accordance with STCW. Ghosh, Bowles, Ranmuthugala and Brooks (2014) argue that this change in learning objectives led to the use of assessment methods that are failing to develop professional skills that enable seafarer students to put to use their competence from MET to workplace contexts on board ships. Moreover, simulator-based competence tests in MET are claimed to be lacking in validity, reliability and security (Gekara, Bloor & Sampson, 2011; Sampson, Gekara & Bloor, 2011). The argument is that the current MET system fails to train students' so-called higher cognitive skills (e.g. comprehension, application, analysis, synthesis and evaluation), which is highlighted in STCW. What Gekara et al. (2011) identify thus a focus on aspects such as maintaining vessel course and speed, safe distance from other vessels and the required draft. Furthermore, the scenarios in the competence test are argued to be very similar to the scenarios carried out during training. Based on such results, Gekara et al. (2011) conclude that the current MET system favors "examination coaching" and "rote learning" rather than high-quality training or the effective evaluation of "essential knowledge and skills" (p. 98). Now the challenge is, as argued by Øvergård, Nazir and Solberg (2017), to contribute with knowledge about how to conduct consistent, unbiased and transparent assessments of navigational skills and competencies. However, in current research it is not yet evident how training of non-technical skills in simulated environments could nor should be assessed (Conceição et al., 2017).

Against this background, the aim of this study is to explore the pedagogical challenges instructors face when assessing master mariner students' navigational competencies and skills in a simulated environment. Drawing on a situated approach to cognition and learning, which implies analyzing interactional details by means of video data, gives us the opportunities to explore how assessment practices unfold during simulations (Heath, Hindmarsh & Luff, 2010). Specifically, by scrutinizing competency tests in the simulator environment in their own right, avoiding theoretical distinctions of technical versus non-technical skills in the analytical process, the aim is to deliver an adequate explication of existing assessment practices in simulator environments. The study is based on video data collected in a navigation course for second year master mariner students, and captures 30 students being certified on using Radar and ARPA (Automatic Radar Plotting Aid) equipment in a bridge operation simulator.

2 LITERATURE BACKGROUND

In safety critical domains, such as aviation, medicine and the offshore industry, different models to rate non-technical skills in simulated environments have been developed to make sure that each team member receives as "fair and objective" assessment as possible (Flin et al., 2003, p. 109). The basic premises of such models are twofold: 1) only observable behavior can be assessed, and 2) assessment system should have a rating system that indicates acceptable and unacceptable behavior. In the maritime domain, which is in focus in this study, a similar model has been developed, focusing on rating non-technical skills of naval cadets.

Table 1. Behavioral marker system for rating non-technical skills developed by Conceição et al. (2017).

Skill	Behavioral marker
Leadership	Takes the initiative
	Sets intentions and goals
	Establishes and control standards
Situation awareness	Monitors and reports changes of situations
	Collects external information
	Identifies potential danger or problems
Communication	Shares information
	Keeps a continuous, clear and effective flow of information
	Promotes a constructive environment for communications
Team work	Considers all the elements of the team
	Coordinates the tasks of the team
	Assesses the capabilities and corrects procedures
Decision making	Establishes alternative lines of action
	Assesses and verifies consequences of decisions and actions
	Considers and shares with others, risks of different lines of action

Models in these domains traditionally draw on a classic cognitivist view where technical and non-technical skills are seen as different and separable set of competencies (Flin et al., 2008). However, in the assessment model by Conceição et al. (2017) in Table 1above technical skills are not included, as technical skills are "already objectively evaluated through the compliance of procedures and the effectiveness of the decisions and actions" (p. 257). In contrast, studies adopting a situated approach to cognition and learning technical and non-technical skills are viewed as inherently intertwined and difficult, if not impossible, to separate in the navigational work trained and assessed in simulators (see e.g. Hontvedt, 2015; Sellberg & Lundin, 2018). For example, Sellberg and Lundin (2018) found that episodes which at first glance seemed to consist of quite technical instructions on the use of radar technologies and of keeping a safe distance to other vessels, when analyzed in detail the episodes entailed instructions on so-called non-technical skills such as situation awareness and decision-making. The analytical focus on interactional details is characteristic to the situated approach, drawing on the method of systematic video analysis of naturally occurring learning practices to render knowledge and competencies observable in work practices (Stahl, 2005). Shifting the perspective from measuring the development of various skills as is prime focus of cognitive perspectives, to analyzing the details of training for work practice, puts emphasis on simulator-based training as an interactional achievement, highlighting the role of the instructors' continuous assessments and instructional strategies in the simulator environment to support student learning towards a profession (Hontvedt 2015; Sellberg & Lundin, 2017).

In research on simulator-based assessments in aviation, assessment models with behavioral markers for rating non-technical skills have proven to be difficult to use for examiners. Several studies have revealed large disagreement and discrepancies in assessment outcomes between flight examiners (e.g. Roth & Mavin, 2013; Weber et. al., 2013; Roth, 2015). For example, Weber et. al. (2013) showed that examiners apply the same or similar reasons to arrive at different assessments or use different reasons to arrive at the same assessment. Hence, interrater reliability between examiners tends to be moderate to low. When interviewing flight examiners on how they conduct assessment, Mavin and Roth (2014) found that cockpit performance was discussed as holistic events rather than separable skills. Moreover, through observation of flight examinations, interviews and the use of think-aloud protocols during examinations, Roth (2015) found that flight examiners based their assessments of non-technical skills, such as situation awareness and decision-making, on a large number of observations put together into a coherent storyline even when using rating scales. Hence, before uncritically putting assessment models into use in MET, there is a need for empirical studies of current assessment practices to identify how assessments are interactionally achieved and thus how they can be developed.

3 METHOD AND DATA

The methodological approach in this study is guided by Heath et al. (2011) principles for using video in research. Following these principles, our aim is to explore human-technology interactions "in the wild," in this case naturally occurring examinations in the simulator. Moreover, these principles put emphasis on the relationship between temporal, material and social aspects in activities. This makes video recorded data an important source for analysis, since video creates stable records of the verbal, visual and material practices under study, enabling detailed and collaborative analysis (reference).

Video recorded data of (training and) examinations in one navigation course for second year master mariners at a Swedish university was collected. While the training sessions on the bridge operation simulator have been analyzed in prior studies (e.g. Sellberg & Lundin, 2017; Sellberg & Lundin 2018), this study draws on video data from simulator-based competence tests at the end of the course. During competence tests students are being certified on using Radar and ARPA equipment in different traffic and weather conditions, which is one of the learning objectives. Another learning objective outlined in the syllabus is the ability to interpret and apply COLREG in various situations. In the empirical data, the students are performing a simulated crossing of the Dover Strait Traffic Separation Scheme (TSS). In TSS-lanes, the crossing should be done with as close to a right angle as practicable according to COLREG. There are two main reasons for this: First, to show the intention to cross the lane to surrounding vessels in the strait. Second, in order to minimize the amount of time spent crossing a TSS-lane. A safe and effective crossing means to adjust the ship's speed

and/or course in accordance to the situation. How to make a safe and effective crossing is this not straightforward but a judgement to be carried out by the student. For example, it can be considered effective to slow down if this is done in order to avoid making adjustments to one's own ship's course. In another situation, a slight change of course can be preferred over a change in speed. Rather, in accordance to the practices of "good seamanship" (Taylor, 1998), the overall goal is to maintain the traffic flow in the TSS.

During competence tests in the simulator, one or two examiners, who were also instructors during the course, usually monitor the students from an adjacent instructor's room. In the instructor's room, several computer screens display different aspects of the activities on the five different bridges: instrument settings, video surveillance of students' work on the bridge, as well as how the students' view the marine environment through visual look-out. The examiners also have an overall view of the scenario through a screen showing the actions of each vessel from a birds-eye perspective. Performance during the test is assessed based on several instances: 1) observable actions taken during the scenario 2) interviews with students on the bridge, 3) observations. For this purpose, two different assessment sheets are used, one for assessing the setting of different instrument, and one for assessing the students' understanding of the use of instruments and the traffic situation at hand. At the end of the course, the students are also examined by means of a theoretical, written examination on COLREG. However, the theoretical tests have been left out of this study.

Table 2. An overview of the video-recorded examinations.

Day	Student group	Examiner(s)	Nᵒ of students	Nᵒ of pass	Nᵒ of fail
Day 1	Group 1	n=1	n=5	n=4	n=1
Day 1	Group 2	n=1	n=5	n=5	n=0
Day 2	Group 3	n=1	n=5	n=5	n=0
Day 2	Group 4	n=1	n=5	n=3	n=2
Day 3	Group 5	n=2	n=5	n=4	n=1
Day 3	Group 6	n=2	n=5	n=1	n=4

Three different sessions of competence tests in the simulator were recorded, in all 30 students (see Table 2). In order to capture the activities that take place during competence tests, wall-mounted go-pro cameras were used on each of the five bridges in the simulator, and a fixed camera was placed in the instructor's room to capture the examiners' use of the monitoring technologies.

4 ANALYTIC RESULTS

In all, the video data analyzed in this study consists of 85 episodes of examiner-student interactions and 3 instances of student-student interactions during competence tests. While most examiner-student interactions in the data corpus were routinely performed in order to fill out the assessment sheets, there were also instances of examiner-student interactions where the examiner visited some students

on their bridges and intervened in midst of the test (n=8). These occasions were not part of the observation and interview part of the competence test. In previous studies on simulator-based maritime training, the instructors' monitoring of students' on-going activities makes it possible to attend to specific details of the students' conduct, which lays the ground for making assessments (Sellberg & Lundin, 2017; Sellberg & Lundin, 2018). Such assessments represent a continuous and on-going process that is grounded in the instructors' abilities to recognize the fit or gap between the learning objectives and the students' activities in the simulator. In this way, these assessments lay the ground for supporting each student with corrections, clarifications or approvals on their performance. The instructional work is argued to be fostering the students into the maritime work practice and discourses of what constitutes "good seamanship", and seen as essential for training the next generation of mariners (Sellberg & Lundin, 2017; Sellberg & Lundin, 2018). In this study, we found that this kind of instructive work, i.e. intervention on the bridge based on observations made in the instructor's room, also takes place during examinations, which calls for further analysis. The eight episodes identified in the data corpus were transcribed and analyzed with attention to the sequential organization of talk and bodily conduct to gain knowledge on how these interventions or visits enter into the examination practice in the simulators (cf. Heath et al. 2011). In the following text we will analyze two of the eight episodes where the instructors pay the student a visit on the bridge in the midst of competence tests.

In the first episode, the examiner enters one of the bridges and confronts the student on slow speed approaching the point of beginning crossing the TSS-lane:

```
Excerpt 1.
01. EXAMINER: Why do you drive so slowly?
    There is no reason to
02. STUDENT:  Nooo
03. EXAMINER:   When you are about to cross
    th-then… you're trailing for a speed you
04. shouldn't keep
05. ((3 seconds pause))
06. Right?
07. STUDENT: Yeah but I've trailed on the
    speed I'm going right now but…
08. EXAMINER: Yes you have but should you
    cross so slowly?
09. STUDENT:  Yes I don't really need to
10. EXAMINER: Need to? You shouldn't! You
    should pass as quickly as possible…
11. Well yes
```

In the first turn, the examiner topicalizes the issue for intervening in the midst of examination which is speed, or the lack thereof, by asking the student "why do you drive so slowly?". The instructor directly accounts for the relevancy of such a question: "There is no reason to" (line 01). The student responds with some uncertainty: "nooo". The student's response can be seen as inadequate, either as revealing a problem with the examiners' formulation of the question, or the student's lack of understanding of the "too slow" situation. The examiner tries to clarify the student's

problem: "you're trailing for a speed you shouldn't keep" (lines 03-04), making explicit a wrong behavior in the situation. A 3 second pause follows, the instructor by a "right?" (line 06), almost rushes the student to answer the question or make some indication of understanding this incorrect behavior. In his response, the student accounts for using the trail-function in relation to the current speed (line 07). This response is evidently taken as insufficient as the examiner continues by again asking the student if he *should* go so slow (line 08). The examiner turns the student's account of the current (slow) speed into a question of what *could* and what *should* be done. In this manner the instructor identifies a lack in the student's understanding of the situation. The student seems thus to display an understanding of such a distinction: "yeah I don't really need to" (line 09), which is then treated as a correctable matter by the examiner in line 10. The sharp remark "you shouldn't!" and the following account of the reason why is delivered in a higher tone of voice, highlighting that the student's slow speed is inappropriate, as well as unsafe, in this particular situation. This episode ends as the instructor rushes out and the student is left alone again. After the examiners' intervention, the student increases his speed and perform a crossing that receive a passing grade.

In the second episode, the examiner enters the student's bridge at the same time as the student is frenetically pushing the steering gear to avoid a close-quarter situation with another vessel in the strait. In this episode the student provides an account directly as the examiner enters the bridge, showing an understanding of the situation and the intervention of the examiner:

```
Excerpt 2.
12. STUDENT: Yeah… I didn't see that
bastard
    comin'
13. EXAMINER: No! Cause you don't look
ahead
14. What are you doing here? ((points to the
    right side of the radar
15. display))
16. You should be there… ((points towards
the
    center of the radar
17. display))
18. ((2 second pause))
19. … in the center
20. STUDENT: ((sighs)) yes I really messed
up
    there
21. I'm tryin' to salvage the situation
    right now
22. EXAMINER: yeah that's that
23. STUDENT: ((gasping))
```

The student anticipates the critique to be delivered by the examiner, saying that he "didn't see that bastard comin'" (line 12), and reveals his understanding of the problem the examiner is about to address. The evaluation delivered by the examiner suggests a reason why the student "didn't see that bastard comin'", which is failing to look ahead (line 13). Hence, the negative evaluation

concerns an issue closely tied to gaining and maintaining situation awareness (Conceição et al., 2017). The examiner then are using starboard radar display to point out the incorrect position, followed by the preferred position of the students' vessel (lines 14-19). The student responds to this with a deep sigh, and comments that he "messed up" and is "tryin' to salvage the situation" (lines 20-21), and thus are agreeing with the examiner's problem definition "not thinking ahead" in line 13 and suggested correct alternative in lines 16-19. The examiner closes the intervention (line 22) and as he leaves the room, the student gasps, revealing his frustration with the current situation (line 23). The student is however able to correct the situation and continues the passing according to protocol and passes the exam.

However, it is worth noting that the students in both episodes passed the examination. Of the eight students that were being corrected on their performance after the examiner identified some kind of trouble with their TSS crossing during the competence test, six passed the examination and two failed.

5 CONCLUSION AND DISCUSSION

In this study we have explored authentic instances of examiner-student interactions during simulator-based competence tests. While the instructors' continuous and on-going process of monitoring, assessing and correcting students during training is argued to be essential for fostering students into the maritime work practice in previous research, findings from the current study show that this kind of instructive work also takes place during individual certifications in the simulator environment. The examiners interventions during competence tests in the simulator found in the data corpus were organized as brief corrections with clear directives for improvement. In that sense, they differ from instructions during training, which are oriented towards developing students' professional reasoning (e.g. Sellberg & Lundin, 2017; Sellberg & Lundin, 2018). However, the students' needing and receiving instructional support during competence tests in the simulator suggest that students are still developing their professional competence at this point in training. Hence, there are reasons to consider the practice of assessing students that are only halfway through their education for the purpose of receiving professional certificates.

Although this analysis is based on a small sample of video recorded data, the data corpus offers a complex and interesting starting point for analyzing the existing assessment practices in MET. In this data, preliminary findings show that corrections during competence tests are regularly made, but not all students are provided with this kind of support. In regards to this, findings in the empirical data raises critical and important questions in regards to what it means to produce as "fair and objective" assessment as possible (Flin et al., 2003, p. 109). While the examiners work systematically as instructors to support students' learning throughout the course, the overall goal of the competence test is to conduct consistent, unbiased, and transparent assessments (Øvergård et. al., 2017). Development of assessment tools that support examiners work is one possible part of the solution, as proposed by Øvergård et al. (2017). Another part of the solution is to develop the examiners' knowledge on how to conduct valid and reliable assessments of performance in the simulator. In regards to this challenge, there are reasons to be careful before putting different assessment models for rating non-technical skills to use in MET (cf. Conceição et al., 2017). As pointed out in the background, results from aviation reveal a number of problems when using these models as ground for making assessments (e.g. Mavin & Roth, 2014). Hence, there is need for future studies that analyze the current assessment practices to identify areas of improvement, and develop a practice where simulator-based assessments of competence ensure the validity and reliability of MET certificates.

REFERENCES

Conceição, V. P., Basso, J. C., Lopes, C. F., & Dahlman, J. (2017). Development of a behavioural marker system for rating cadet's non-technical skills. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*. Doi:10.12716/1001.11.02.07

Emad, G., & Roth, W. M. (2008). Contradictions in the practices of training for and assessment of competency: A case study from the maritime domain. *Education+ Training*. Doi: https://doi.org/10.1108/00400910810874026

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

Flin, R., Martin, L., Goeters, K-M., Hörmann, H-J., Amalberti, R., Valot, C. & Nijhuis., H. (2003). Development of the NOTECHS (non-tecnical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, 3(2), 97-119.

Gekara, V. O., Bloor, M. & Sampson, H. (2011). Computer-based assessment in safety-critical industries: The case of shipping. *Journal of Vocational Education & Training*. Doi: https://doi.org/10.1080/13636820.2010.536850

Ghosh, S, Bowles, M, Ranmuthugala, D. & Brooks, B. (2014). Reviewing seafarer assessment methods to determine the need for authentic assessment. *Australian Journal of Maritime & Ocean Affairs*. Doi: 10.1080/18366503.2014.888133

Heath, C., Hindmarsh, J. & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. SAGE Publications Ltd, London.

Hontvedt, M. (2015). Professional vision in simulated environments—Examining professional maritime pilots' performance of work tasks in a full-mission ship simulator. *Learning, Culture and Social Interaction*. Doi: https://doi.org/10.1016/j.lcsi.2015.07.003

Mavin, T., & Roth, W-M. (2014). A holistic view of cockpit performance: An analysis of the assessment discourse of flight examiners. *International Journal of Aviation Psychology*. Doi: https://doi.org/10.1080/10508414.2014.918434

Roth, W-M. (2015). Flight Examiners' Methods of Ascertaining Pilot Proficiency. *The International Journal of Aviation Psychology*. Doi: 10.1080/10508414.2015.1162642

Roth, W-M. & Mavin, T. (2015). Peer Assessment of Aviation Performance: Inconsistent for Good Reasons. *Cognitive Science*. Doi: 10.1111/cogs.12152

Sampson, H., Gekara, V. & Bloor, M. (2011). Water-tight or sinking? A consideration of the standards of the contemporary assessment practices underpinning seafarer licence examinations and their implications for employers. *Maritime Policy Management*. Doi: https://doi-org.ezproxy.ub.gu.se/10.1080/03088839.2010.533713

Sellberg, C. (2017). Simulators in bridge operation training and assessment: A systematic review and qualitative synthesis. *WMU Journal of Maritime Affairs*. Doi: https://doi.org/10.1007/s13437-016-0114-8

Sellberg, C. & Lundin, M. (2017). Sellberg, C., & Lundin, M. (2017). Demonstrating professional intersubjectivity: The instructor's work in simulator-based learning environments. *Learning, culture and social interaction*. Doi: https://doi.org/10.1016/j.lcsi.2017.02.003

Sellberg, C., & Lundin, M. (2018). Tasks and instructions on the simulated bridge: Discourses of temporality in maritime training. *Discourse Studies*. Doi: https://doi.org/10.1177/1461445617734956

Stahl G (2005) Group cognition in computer-assisted collaborative learning. *Journal of Computer Assisted Learning*, Doi: https://doi.org/10.1111/j.1365-2729.2005.00115.x

Taylor, D. H. (1998). Rules and regulations in maritime collision avoidance: New directions for bridge team training. *Journal of Navigation*, 51(1), 67–72.

Weber, D., Roth, W-M., Mavin, T. & S. Dekker. (2013). Should we pursue interrater reliability or diversity? An empirical study of pilot performance assessment. *Aviation in Focus*, 4(2), 34-58.

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*. Doi:10.12716/1001.11.02.03