PERFORMANCE OF SEAFARERS DURING EXTENDED SIMULATION RUNS

A Kircher and M Lützhöft, Chalmers University of Technology, Sweden

SUMMARY

The term “good seamanship” is often used to describe the behaviour of deck officers who navigate according to a commonly accepted set of norms. Even though these norms offer a certain flexibility, most of them are covered by laws and regulations. The navigation performance of deck officers is also of interest, and it could be argued that there is a connection between the terms. For training and evaluation of deck officers an objective assessment of performance is also of great advantage. The scope of the paper is to analyse how different measures can be used to describe navigation performance, and the relationship between these in different collision situations. During an extended simulator study with 50 deck officers, navigation performance was assessed while participants spent one complete week in navigation simulators. Their task was a realistic 24/7 voyage through the English Channel, and each participant spent 64 or 90 hours in the simulator.

NOMENCLATURE

AIS  Automatic Identification System
COLREGs  Convention on the International Regulations for Preventing Collisions at Sea (Collision Regulations)
CPA  closest point of approach
Nm  nautical mile
rs  Spearman’s rho

1 INTRODUCTION

This paper describes part of the results of a project investigating fatigue-induced performance degradation of seafarers during an extended simulation run. Both navigation simulators and cargo handling simulators were employed in the study, although here only navigation simulator results are presented.

Excessive fatigue affects human performance negatively, as studies in different areas have shown [1]. While it is not the focus of the paper, it must be noted that fatigue is not the same as sleepiness. Fatigue is conceptualized as “a reduction in physical and/or mental capacity as the result of physical, mental, or emotional exertion which may impair nearly all physical abilities including: strength, speed, reaction time, coordination decision making or balance” [2]. Sleepiness, on the other hand, can be defined as a physiological drive to sleep [3]. So while fatigue is often considered to be a generic term, sleepiness is seen as one of the major sub-components. Here the term fatigue will be used. Fatigue is recognized as an important safety related factor in shipping, as several publications show (for an overview see [4]). The results of the project in relation to fatigue will be published in early 2012.

Performance itself is of main interest here, and several ways of assessing performance, as well as various sub-factors of performance, exist. Assessing human performance can be accomplished with a variety of methods, which are often subdivided into qualitative and quantitative methods. The first often refers to subjective methods where human expert judgement is involved. Quantitative methods usually imply some form of data collection and testing of hypotheses. It is argued that cognitive performance is central in wheelhouse work in commercial shipping, while physical performance plays a smaller role. Cognitive performance refers to “observable behaviour on cognitive tasks including intelligent behaviour, successful task taking and expert performance” [5].

One scope of the project was to produce and analyze measures of (navigation) performance that are less subjective and easy to quantify and score, in order to have scientifically sound results. Events which had to be mastered by the deck officers were scored with both subjective measures of performance, and such measures which one could consider being semi-objective and unbiased. These developed measures could also be used during future projects dealing with navigation performance, and even for training assessment.

In terms of impaired cognitive performance the following manifestations of fatigue [1] are of practical importance for shipping:

- Increasingly frequent lapses in performance
- General cognitive slowing, including a lowering of optimum performance capabilities
- Memory problems
- An increasing inability to maintain the vigilance required to perform the tasks required

Phillips [6] grouped fatigue manifestations as follows:

- Activation problems (attention failures, slips and lapses)
- Perception limitations (limiting visual and auditory sensations)
- Information processing problems (interpretation, encoding and correlational deficits)
- Aversion to effort (failure to act)
- Differing effort (failure to act properly)
It should be noted that the manifestations above are time-of-day related as well. In the study presented here critical situations or events that had a direct link to the factors in the list above were used.

A large number of measurement options for performance, sleepiness and fatigue in general exist (in fact several books on the subject can be found). However, the navigation performance (of a ship officer) is more complex to both define and assess.

From the point of view of a navigation course teacher or a bridge simulator instructor, the concept “good seamanship” would probably be preferred over “navigation performance”. In the maritime field, good seamanship has the advantage that professionals know what it implies. On the other hand, assessing which components make up good seamanship, and even finding ways to assess it in a scientific manner is challenging.

In law, good seamanship is synonymous with “ordinary practice of seamanship” (an American court decision on a ship collision in 1876 formally gave the same meaning to the two terms). Rules following are defined in the COLREGs [7], where proper look-out, speed and foresight are major elements. However, in addition to the COLREGs, seafarers have to comply with broader obligations, often unwritten rules such as recognized customs, where some may even location-dependent. Even a departure from the COLREGs may be necessary in special circumstances ([7], rule 2b). Here, a discussion of the legal implications of good seamanship is not attempted, instead the focus is measurable data to create a link between good seamanship and navigation performance.

For example communication is explicitly named in the COLREGs as being important for collision prevention. The IMO requires certain communication standards for navigation purposes [8], which are in fact a requisite for professional certification. The IMO standard contains a detailed guide on language, terminology, spelling, etc. for precise and effective communication. To assess if (radio) communication has the appropriate standard the single factors named in the IMO documents can be used.

1.1 PERFORMANCE SCORING PROCEDURE

Nine navigation instructors, some having been active seafarers, others still being active, were asked to name aspects that they consider important when assessing students and professionals seafarers. This was done with open interviews using possible collision situation as example. Of special interest were single observable or measurable elements, which they felt would build up an overall assessment of how well a situation was handled.

Simulator events, mainly related to COLREGs situations, were designed keeping in mind what could be expected from the literature (mainly the factors described by Dinges [1] and Phillips [6]).

The obtained scoring points were then adapted to a few four scale questions (see Table 1), or yes/no questions, since it was important that real time assessment would be possible, limiting the amount of different items that an observer would be able to score more or less simultaneously. The obtained scoring procedure is presented in the results section, followed by the results from data analysis.

1.2 DELIMITATION

This paper will not discuss possible performance alteration due to increasing fatigue and different watch patterns. The interested reader will find reports on these issues on the project webpage (www.project-horizon.eu) and in relevant scientific media.

2. METHOD

The experiment was conducted during four full weeks in 2010 at Chalmers University of Technology, Department of Shipping and Marine Technology. Each of the 50 participants spent one week (seven days) in one of the five navigation simulators (see Figure 2 and Figure 3). Before the experiment, two full-scale pilot studies were carried out to test scenarios and logistics.

The 50 participants were active desk officers recruited by a marine manning agency. Most participants were male (48 of 50), the mean age was 31.6 ± 9.8 years.

Inclusion criteria for participating in the experiments were: no sleep problems, body mass index less than 30, history of good health, no anxiety or depression problems, no heavy smokers, at least 12 months professional experience at sea and experience in liquid cargo handling (this because a part of the experiment included cargo handling operations). Participants received compensation equivalent to one week of work aboard a ship for their participation. Approval from the ethical committee was granted for the study. Information about the study was sent to participants well before starting the experiments, more detailed information (such as the Master Standing Orders) were handed out one day before the first experiment day. The entire first day was dedicated to familiarization with the simulators, the procedures, and medical examination to ascertain that all participants were fit to take part in the study. A nurse would always be available during the experiment weeks in case it was necessary.

Participants were randomly assigned to either a 3-watch (4 on 8 off) watch system, or a 2-watch (6 on 6 off) watch system. The complete experiment duration was seven days for each participant (in addition one day was spent on training). Since five identical simulators were available, ten participants per week worked in the 2-
watch system, and 15 participants in the 3-watch system. Each participant spent around 56 hours in the navigation simulator for the 4 on 8 off watch system, and 84 hours in the 6 on 6 off watch system. Attention was put on creating as much realism as possible, without losing on experimental control: participants had their accommodation on a vessel moored in proximity to the university, food was served onboard the vessel, and all free time had to be spent onboard. Communication in the simulator (to master or lookout) was via VHF radio. Navigation equipment in the simulators was equivalent to what is found on real vessels, including ARPA radar, paper charts, etc.

A real master was not present for each bridge, instead a navigation instructor played the role of the master/lookout.

The area sailed according to the voyage plan stretched from the Dutch coast passing through the English Channel to approximately Longitude 2°W, and back. Each participant sailed the same route. Participants would not see each other on the radar, since each simulation was running “in its own world”. Traffic was modelled to a great extent according to real AIS traces of vessel traffic in the actual area. A map of the sailed area is shown in Figure 1.

![Figure 1. Map of the area sailed by the participants. Dover port was approached twice.](image)

Random radio communication from the area was pre-recorded, and played during the navigation runs.

The ship sailed was a small coaster with an overall length of 50 metres and a displacement of 1,200 m³. This size of vessel was chosen because there is a certain probability that an officer would be alone on the bridge on board similar ships (although according to regulations this is not allowed).

The five navigation simulators were equipped with standard equipment found on merchant vessels. Each simulator housed one watch team, but participants were alone on the bridge. The simulators were developed by Kongsberg Maritime Division (“Polaris” stations). The actual software version was 5.5.1. Four of the simulators were identical, with three large plasma screens for the visualization (120° FOV, see figure 3). The fifth simulator had a more advanced projection system and larger bridge space (see Figure 2), but the same instrumentation as the other four simulators and the same software.

![Figure 2. Large navigation simulator bridge.](image)

![Figure 3. Small navigation simulator bridge.](image)

A variety of data was collected, here mainly data related to navigation performance will be discussed. As on a real 4 on 8 off or 6 on 6 off watch system, each participant worked two shifts per day. During each shift a number of unexpected, sometimes critical events were present. These events, developed as described in Chapter 1.1, were:

- S1: Fishing boats on the ship’s track. Action is necessary to avoid collision. Event occurs once.
- S2: First time buoy adrift (North cardinal mark) adrift, located close to track. The second time there will be a life raft instead of a buoy. Event occurs twice.
- S3: Two relevant radio messages (among several irrelevant messages) inform about an
area to avoid, because there is a seismic survey vessel (first time) and a submarine exercise (second time). This area lies in the own ship’s track, and is reported with four latitude/longitude positions. Event occurs twice.

- S4: Collision course with ship that is supposed to give way. Event occurs four times.
- S5: High speed ferry from astern, overtaking own ship, and crossing her track. Event occurs once.

During these events measures listed in Table 1 were scored. All simulator instructors were trained on the events beforehand in order to minimize inter-rater variability.

Various additional data was collected, both related to cognitive performance (psychomotor vigilance task), subjective and objective sleepiness (Karolinska Sleepiness scale and electroencephalogram), and status of the participants (extensive diaries filled in by participants, actigraphy to assess sleep duration and quality). These will not be discussed here.

## 3. RESULTS

The performance scoring form is reported in the results section, as it is considered an outcome of the project. As mentioned before, scoring was performed in real time while the event occurred, the experiment leaders (which were expert instructors) had very limited time, thus the scoring had to allow for quick completion, furthermore the language needed to be simple.

### Table 1. Scoring criteria during events in the simulation. A subjective 10-point score was present for each task as well.

<table>
<thead>
<tr>
<th>Task (short name)</th>
<th>Criteria monitored</th>
<th>Ratings</th>
<th>Relevance for event</th>
</tr>
</thead>
</table>
| Use all available means to determine vessel position (“position taking”) | Use of GPS + radar to determine position and frequency of positions plotted | 1- Few positions plotted, only by one mean, 0–1 per watch  
2- 2–3 positions plotted but only by one mean  
3- 3–4 positions plotted mainly by one mean (at least 1 by 2 means)  
4- At least 4 positions plotted and regularly double-checked by other means | S2 |
| Compliance with COLREGs (“COLREGs”) | How well COLREGs are followed (distinct action in time/distance) | 1- COLREGs ignored  
2- Followed COLREGs but action taken very late (distance less than 1 nm)  
3- Followed COLREGs but action taken late (distance 1–1.5 nm) or not distinctively  
4- Strictly applying COLREGs taking action in time and distinctively | S1  
S3  
S4  
S5 |
| Detection range of targets (“detection range”) | Plotting of radar targets and range | 1- Never plotting relevant targets  
2- Rarely plotting targets and only when distance around 3 nm or less  
3- Plotting all relevant targets but at rather short range (approx. 3 nm)  
4- Plotting all relevant targets at ranges of at least 4 nm | S1  
S2  
S3  
S4  
S5 |
| Keeping a safe CPA (“safe CPA”) | Master standing orders (open sea: 2 nm, confined: 1 nm) | 1- CPA’s of less than 0.3 nm  
2- CPA’s of 0.4  
3- CPA’s of 0.5  
4- CPA’s of min. 0.7 | S1  
S3  
S4  
S5 |
| Communication and attention (“Communication”) | Correct reaction to radio message | NO: no action taken  
YES: proper action taken | S3 |
| Position report (radio communication) (“position report”) | Correct radio communication according to standards | 1- No position report  
2- Late reporting (definition in CALDOVREP)  
3- In time reporting, but incomplete report or wrong facts  
4- Timely and complete report | S2 |
| Spot object adrift (“spot object”) | Correct report and handover | 1- No handover report and no marking of object  
2- Mentions object during handover, but no marking and no report to coastguard, or only marked on chart  
3- Identified and mentioned object during handover and marks position on chart  
4- Identifies object, correct radio call to authorities and mentions during handover and mark on chart | S2 |
The scoring criteria are shown in Table 1. Besides these scores, each of the single events was scored on a subjective 10-point scale ("How well did the participant handle the event?"), ranging from very poorly to perfectly. This score was used as predicted value in the subsequent regression analysis.

Analyses were carried out on SPSS 19; confidence level chosen: 95%, p-values are only reported as above or below 0.05 (if not stated otherwise). CATREG Version 3 by Data Theory Scaling System Group (DTSS), Leiden University, The Netherlands, used for categorical regression (note that no cross-validation on a separate data set was performed).

All but one participant completed the experiment (data from this participant are excluded from the analyses). Participants were of different nationalities, with the majority being Croatian, Latvian, and Romanian (all about 28%). A number of debriefing questions were asked after the experiments, these had a free text option, and were encoded afterwards. 75% of the participants reported that their performance in the simulator was approximately equal to when at sea, and 22.9% reported a worse performance in the simulator than at sea. The simulation was considered realistic or somewhat realistic by 75% of the participants, while 25% did not find the simulation realistic. When asked about what was considered a safe CPA in different waters, participants reported the following mean values (standard deviation reported as well): in confined waters 0.51 ± 0.26 nm, in open sea 1.29 ± 0.66 nm and in the Dover Straight 0.5 ± 0.26 nm. For both confined waters and Dover Straight, minimal CPA of down to 0.1 nm was considered safe! Having observed colleagues falling asleep while working on the bridge in real life was reported by 67.3% of the participants.

The scores in Table 1 were analysed in order to find associations between scoring criteria and subjective rating (note that for each event one additional subjective expert rating was present). Parametric correlation and categorical regression are applied.

Event S1 (fishing boats): Correlation between the relevant scoring criteria (table 1) and the overall subjective scale were all significant at the .01 level. “COLREGs” had the highest correlation (r=.83). Categorical regression confirmed that the COLREGs score was the strongest predictor for the subjective score (Beta=.54, p<.05, overall model: R^2=.73,6.40, p<.05).

Event S2 (buoy adrift): during the first occurrence, correlation between the subjective score and the relevant scoring criteria was significant (sig.<.005) for “position taking” and “spot object”. No participant reported the buoy adrift. Categorical regression showed position taking as most relevant (Beta=.63, p<.05, overall model: R^2=.78,7.19, p<.05). The second time a life raft appeared as object adrift, the correlation was significant for position report and spot object. Only five of the 49 participants reported the life raft adrift. Regression was not carried out because of missing data.

Event S3 (acting upon radio message): during the first occurrence COLREGs, safe CPA and communication were significantly correlated with the subjective score. Correct reaction if receiving the radio message had the highest impact on the subjective score. No analysis was performed the second time the event occurred, because of a high number of missing data. No categorical regression was performed because of missing data.

Event S4 (collision course): during the four times the event occurred, all relevant scoring criteria correlated significantly with the overall score, and COLREGs had the highest correlation (ranging from .69 to .85). Categorical regression for the first time the event occurred pointed out COLREGs as most important factor for the subjective scale (Beta=.51, p<.05, overall model R^2=.8,9(4,2), p<.05). For the remaining three times the event was scored, COLREGs was again the most important factor.

Event S5 (high speed ferry): all relevant scoring criteria correlated significantly with the overall score. Both COLREGs and detection range had a correlation of r=.93. Categorical regression pointed out COLREGs as most important factor (Beta=.59, p<.05, overall model R^2=.98,6.45 p<.05).

4. DISCUSSION AND CONCLUSIONS

Assuming a common agreement of experts about how well a (COLREGs) situation is mastered, the factors that build up this agreement were of main interest. Developing the semi-objective scores for the different scenarios was exactly this: identifying the essence of a certain expert score for a specific situation.

The data analyses pointed out that the criteria rating from Table 1 related well to the overall subjective scores for each event. COLREGs compliance, which had the extreme scoring points total ignorance of rules and perfect compliance, and as intermediate points partial compliance depending on how well in time action was taken, was the best predictor of the subjective score. Categorical regression gave high modelling power, but it should be pointed out that no data set for validation was used. But even replacing the subjective score by using only COLREGs compliance would not be sufficient for reproducing the subjective score adequately.

The experience during the interviews with navigation instructors shows that they consider not only the actual actions of a participant at the time when an encounter happens, but include how the behaviour of the participant was well before the situation commenced. Here some factors were identified:
- Selected range of the radar, depending on area, traffic density, route characteristics
- Tracking of vessels that may choose a route that intersects the own route or approaches the own ship
- Logbook keeping manner
- Posture of participants on the bridge
- Frequency of use of radio communication (here in fact communication that is limited to few occurrences of major importance seems to be preferred to more frequent communication)
- Correctness, completeness and language of radio communication

The factors above are as well very subjective, and different experts (or Masters) may not share the same opinion in what is the best choice.

A number of episodes were strongly commented by the experts, for example that very few participants actually reported the objects adrift.

Inter-rater variability would have been better accounted for with statistical methods, instead of only training experts with events to be presented to participants. It was observed that although the differences in opinion between the experts were not large, often the "best way" to solve a situation differed based on the background of the expert. One implication is clearly the professional training of seafarers: if experts or instructors do not share a common view of COLREGs situations, how are students expected to learn or interpret them and act correctly when confronted with such situations in real life?

A second observation could be: are the COLREGs not clear enough for more complex ship to ship encounters, where decisions have to be made quickly? In fact, comments by the experts during the four weeks of simulator runs point out this as a challenge.

The article has attempted to illustrate how different measures of navigation performance are interlinked. Difficulties in both the definition of what performance is, as well as inter-rater variability of expert scorings, confirm that more research is needed.

Scoring of navigation performance in COLREGs situations where action is required to avoid accidents may be partly possible with semi-objective measures, as attempted in the simulator study. However, the absence of an accident, even the absence of critical encounter is far from being a proof of good (navigation) performance. Talking about fatigue, an officer alone on the bridge may fall asleep for several minutes without anything critical occurring (or anyone noticing). In this aspect shipping may appear more forgiving than for example driving a lorry, where a few seconds of sleep are likely to cause accidents.

Assessment of radio communication according to standards was limited to a four-grade scale, as time did not allow using a more extensive scoring forms based on IMO standards.

5.1 LIMITATIONS OF THE STUDY

There are obvious limitations of simulator studies, one is the difficulty to characterize long-term effects, another is the "syntheticity" of presented critical situations. It cannot be assessed how much effort participants put into "behaving as they would in reality" (which was part of the instructions).

Participants were not homogenous in certain aspects; nationality, education and training differences of seafarers may have had an impact on the results of the study. Inter-rater variability was not assessed by scientific means, instead it was attempted to minimize it by training instructors on how to score events previous to the experiments.

6. ACKNOWLEDGMENTS

The project was financed by the EU 7 framework programme. All participants of the study are gratefully acknowledged.

7. REFERENCES


8 AUTHORS’ BIOGRAPHIES

Albert Kircher currently holds the position of research engineer at Chalmers University of Technology, department of Shipping and Marine Technology, maritime human factors group. He conducts human factors related research and participates in different EU funded projects. His previous experience comprises traffic safety research with focus on human behaviour and medical informatics. He has a Master in Electronics and Biomedical Technology.

Margareta Lützhöft a Master Mariner and sailed for 13 years on Swedish ships. After leaving the sea, she studied for a Bachelor’s degree in Cognitive Science and a Master’s in Computer Science. In December 2004 she received a PhD in Human-Machine Interaction.

She presently works as an Associate Professor at Chalmers University of Technology, leading the Maritime Human Factors research group at the Department of Shipping and Marine Technology, within the Lighthouse competence centre.