A study of the performance of captains and crews in a full mission simulator

Andersen, H.B.; Sørensen, P.K.; Weber, Steen; Sørensen, C.

Publication date:
1996

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Andersen, H. B., Sørensen, P. K., Weber, S., & Sørensen, C. (1996). *A study of the performance of captains and crews in a full mission simulator*. Risø National Laboratory. Denmark. Forskningscenter Risoe. Risoe-R No. 916(EN)

Risø-R-916(EN)

A study of the performance of captains and crews in a full mission simulator

Henning Boje Andersen, Peter K. Sørensen, Steen Weber and Carsten Sørensen

Risø National Laboratory, Roskilde
December 1996

Abstract: This paper gives an overview of the main results of an analysis of the performance of captains and bridge crews in a full mission simulator during two types of voyages. The unit of analysis used in the paper is defined in terms of captains' performance during voyages and segments of these. A high correlation was found between the scores made by navigation instructor of captains' performance and objective incidents, viz., groundings and near-misses. It turned out that older captains obtained a lower performance score or "grade average", as judged by navigation instructor; yet, they had significantly fewer groundings or near-misses. At the same time, it was established that captains whose verbal communication contained a relatively greater percentage of utterances directed at the future had significantly fewer groundings or near-misses.

•Systems Analysis Department
Risø National Laboratory
P.O. Box 49
DK-4000 Roskilde, Denmark
Phone: (+45) 4677 5141
Fax: (+45) 4675 5170
Email: henning.b.andersen@risoe.dk
Web: www.risoe.dk

*Danish Maritime Institute
Hjortekærsvej 99
2800 Lyngby, Denmark
Phone: (+45) 4587 9325
Fax: (+45) 4587 9333

ISBN 87-550-2211-1
ISSN 0106-2840

Grafisk Service, Risø, 1996

Contents

1. Introduction 5

2. Description of data 5
2.1 Scorings made by navigation instructor 6
2.2 Classification of communication 6
2.3 Objective incidents - groundings and near-misses 6
2.4 Demographic data 7
2.5 Course data 7

3. Results and discussion 7
3.1 Correlations among navigation instructor’s scores 7
3.2 Age and incidents 8
3.3 Correlation between navigation instructor’s scores and incidents 9
3.4 Captains’ age and time in present position related to navigation instructor’s scores 10
3.5 Correlations between types of communication and incidents 13
3.6 The stability over time of scores by navigation instructor and psychologist 14

4. Conclusion 15

5. References 15

1. Introduction[1]

The study reported in this paper was carried out as a follow-up analysis of results of a Crew Resource Management course involving simulated voyages in a full mission simulator. The course was conducted at the Danish Maritime Institute in 1993-94 for bridge officers employed by a European ship operator.[2] Data from the course were complete for the purpose of statistical analysis for the majority of, but not all, captains and voyages involved, and the present analysis is based on data including video recordings and simulator logs covering 53 captains and 90 voyages.

The voyages consisted of two types: a relatively short departure scenario where the crew would depart a port and take their ship down an estuary (lasting on the average half an hour) and a much longer arrival scenario where the crew would go up the estuary to the port (lasting around two and a half hours). Half of the crews would start with the arrival scenario, proceeding with the departure scenario on one of the following days, while the other half of the crews would take the scenarios in the opposite order. The captains and the crews were put together on an ad hoc basis dependent on their shoreleave schedules; captains and mates therefore often did not know each other in advance or had not been working together on a bridge before. We have not been able to include the extent of acquaintance among officers as a variable in the analysis.

2. Description of data

The data that went into the analysis include the following

- Evaluation scores of each captain for segments of each voyage made by an experienced navigation instructor
- Classification of communication of captains and crews carried out by a psychologist into large linguistic groups
- Incidents (groundings and near-misses)
- Demographic data (age, time in present position)
- Course data (order of scenarios; course period in which a captain's simulated voyages had taken place).

We shall now briefly explain each of these groups of variables:

1 This report is an expanded and slightly modified version of a paper presented at the BIMCO/WMU Conference, 17-20 June 1996, Copenhagen, Denmark. The authors gratefully acknowledge helpful comments to earlier versions of this report by Claus Bornemann who has conducted most of the scoring of communication. We also thank our colleagues Jens Bay, Gil Guillermo, John Paulin Hansen and Jørgen Thau for suggesting interesting points to consider for the analysis.

2 The company kindly co-operated in the planning of the analysis and generously supported its execution.

2.1 Scorings made by navigation instructor

For each voyage an experienced navigation instructor would assess the performance of the captain involved for consecutive segments of the voyage. For the purpose of the analysis we have used the total, averaged score of the voyage as they were given for each of four separate categories:

- *Timing:* scores on this aspect of performance express whether course alterations and counter manoeuvres are carried out at the right time, handover of the watch is done at a proper time, information on the bridge is passed on at the right time etc. The scores also reflect whether the particular part of the voyage in question is carried out effectively, since a very slow (and apparently safe) passage may result in a low score on timing. It is important to note that a voyage should be both safe and efficient as the ships are working on fairly tight schedules.

- *Safety:* this dimension of performance concerns the *safety* with which the sailing of the vessel during this segment is carried out - does the vessel ground or does it have a near-miss, is the speed appropriate under the circumstances, is the planned track followed, is traffic monitored properly, is there a high level of shared situational awareness, is the passage carried out with proper considerations of factors determined by the environment?

- *Resource management:* this covers the use of both technical and human resources - are resources used in an optimal way, are manoeuvres overdone, is there a clear and well-defined task distribution, is there a satisfactory level of synergy on the bridge, is the captain exercising an appropriate level of authority, does teamwork ensure that nobody becomes overloaded?

- C*ommunication*: scores on this performance aspect concern the quality and efficiency of communication on the bridge both within the team and between the team and other ships or shore (Vessel Traffic Service, agent, terminal, etc.). Does the captain leave room for the other officers to speak up (exercise assertiveness) if they discover anomalies, does the captain inform the others of his intentions and actions. Is communication loud and distinct and does it follow the standard maritime vocabulary, are orders acknowledged?

The navigation instructor's scores were meant to reflect his assessment of the quality of a captain's bridge behaviour in terms of these four different dimensions of performance. This assessment was expressed in terms of a numerical score or "grade" on a 9-point scale from 0 to 8. For the purpose of analysis we have transformed scores to a 1 to 9 scale.

2.2 Classification of communication

While the navigation instructor would assess communication in terms of its quality and efficiency, another type of rating was performed on communication as well. This rating was not directed at the content or efficiency of communication but was meant to record the forms or types of utterances made on the bridge. Thus, during the simulation sessions a psychologist would perform an on-line classification (entered into a time-stepped spread sheet) of the types of communication among the crew. For each utterance, the psychologist would record its originator, whether the utterance was a command, an observation, a question or a reply (or inchoate) and whether it concerned the *past*, the *future* or the *current situation*. (The distinctions into past and future were, in fact, further divided into 'distant' and 'near' past or future; but since it turned out that there were very few utterances about distant future or distant past, no meaningful conclusions about possible correlations with other variables can be made on these two categories.)

2.3 Objective incidents - groundings and near-misses

Based on the navigation instructor's scorings in the category *Safety* an identification was made of voyage segments that contained a grounding or a near-miss (numerical score 0 and 1, respectively). The occurrence and nature of these incidents have been verified by subsequent analysis of video and trackplot. Altogether there were 22 voyage segments containing a grounding or a near-miss in the 90 voyages covered by our analysis. (For brevity, we sometimes refer in subsequent figures to this variable as just "*grounding*").

2.4 Demographic data

Based on data supplied by the company in which captains and crews were employed, the following variables were selected for each captain: (a) age and (b) total time in present position. We should add that "time in present position" is a slight misnomer, since it covers only the time in which a captain has been in his present position in his present company. Some captains received their promotion before they were employed by their present company (we were not able to get data for total time in position of captain for our subjects). Therefore, we should expect "age" to be better correlated with experience than "time in present position". We will return to this distinction below when results are discussed.

2.5 Course data

The 90 voyages included in the data analysis were conducted in the DMI full mission simulator over a period of almost one and a half years. Each captain and crew would normally conduct their two scenarios on two consecutive days or with at most two or three days between the voyages. It should be added that a captain would have the same crew for both voyages - or, in some rare instances, nearly the same crew, in case a crew member was replaced because of illness etc.

The period of the course in which a captain performed his voyages was of potential interest, since this could reveal if the navigation instructor's scores as well as the psychologist's scores into utterance categories were uniform over the course period. This information was therefore included as a variable in the analysis and was given by yearly quarters.

Voyages were of two kinds, as mentioned above, and it has been included as a variable in the analysis whether a captain started with the relatively short departure scenario and then went on to the longer arrival scenario or took the scenarios in the opposite order. However, the courses changed midway in a few minor ways. Before the change, crews were allowed to observe by audio and radar how another crew performed before they themselves began the voyage. After the change, this "advantage" was no longer allowed. The data on sequence on voyages, therefore, are very scarce, since it is not possible to compare the two periods.

3. Results and discussion

3.1 Correlations among navigation instructor's scores

The statistical analysis established that there was a significant correlation among scores given on *Timing, Safety, Resource Management* and *Communication*: If a captain was rated at the higher (or lower) end of the scale on one of these, he would most likely be rated at the higher (or lower) end on the other three. Yet, the correlation between *Communication* and each of the other three categories was, while impressive, still a bit weaker. In the following figure we show as an example the correlation between scores on *Safety* and scores on *Timing*.

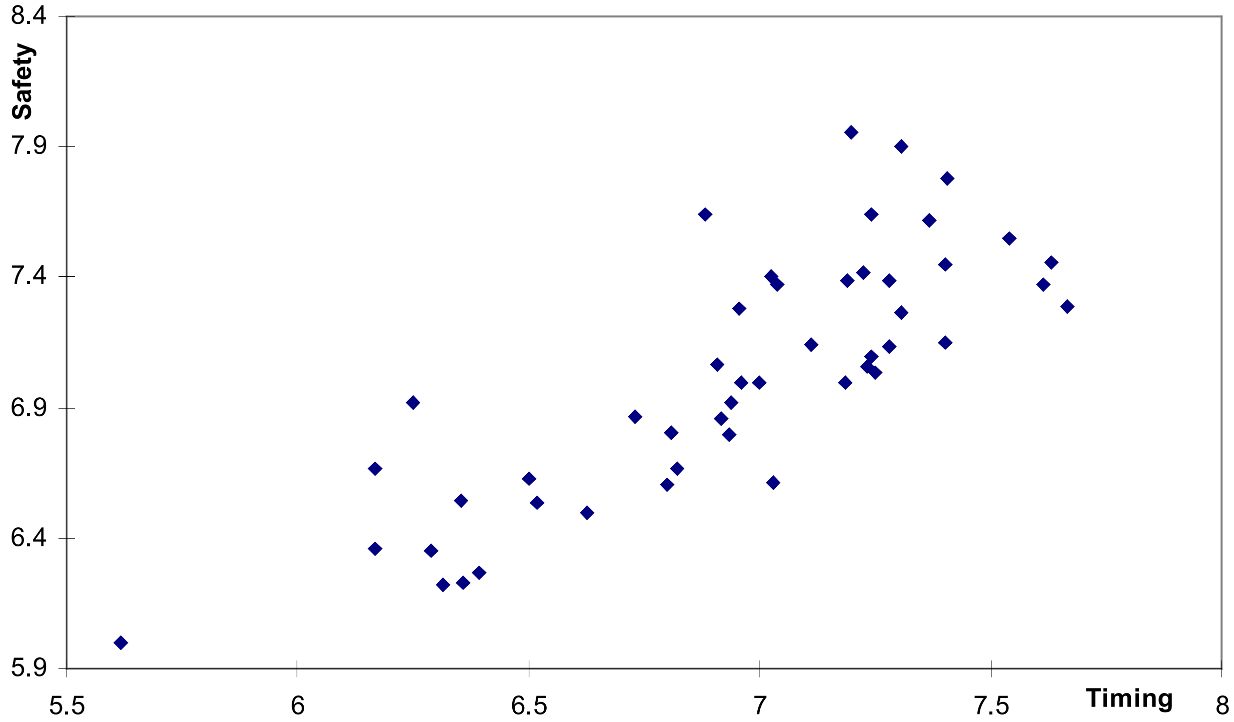

Fig. 1: Scatter plot showing correlation between scores on Safety and Timing. This is illustrative of the correlations that obtained among navigation instructor scores when any pair of these four performance categories is computed.

	Safety	Resource Management	Timing	Communication
Safety		p<0.01% R^2=72.2%	p<0.01% R^2=66.6%	p<0.01% R^2=34.4%
Resource Management	p<0.01% R^2=72.2%		p<0.01% R^2=69.7%	p<0.01% R^2=52.3%
Timing	p<0.01% R^2=72.2%	p<0.01% R^2=69.7%		p<0.01% R^2=38.7%
Communication	p<0.01% R^2=34.4%	p<0.01% R^2=52.3%	p<0.01% R^2=38.7%	

Table 1: Correlations among navigation instructor scores for each of the four performance categories. It can be seen that "Communication" tends to be less well correlated with the rest.

3.2 Age and incidents

It turned out that a captain's age was correlated with incidents: older captains had a significantly *smaller* risk of having had a grounding or near-miss during their simulated voyages. Below in fig. 2 we depict the age profile on the Y-axis and on the X-axis the two groups who had no incident or at least one incident, respectively

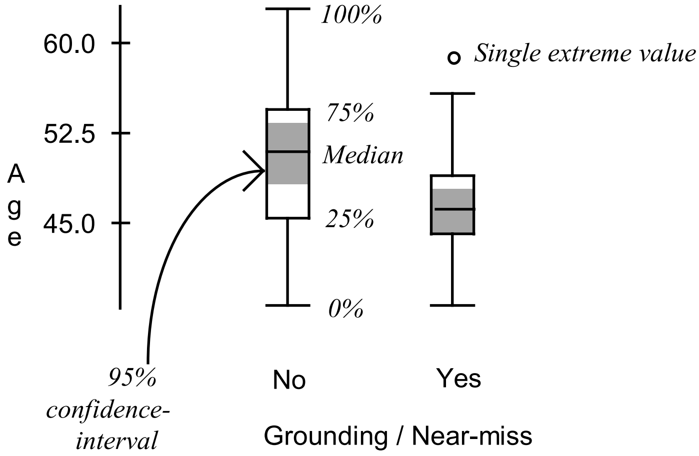

Fig. 2: Box plot showing the relation between captains' age and the "grounding-yes" and the "grounding-no" groups.[3]

3.3 Correlation between navigation instructor's scores and incidents

A high correlation was found between incidents (groundings and near-misses) and the navigation instructor's scores on each of the three categories *Timing, Safety* and *Resource Management* (in the following termed the *TSR categories).* Conversely, there was no correlation between incidents and the navigation instructor's rating on *Communication.* Thus, if a captain has received a rating at the higher end on any of the *TSR categories*, there is a high probability that none of his voyages contained a grounding or near-miss; and similarly, if a captain has had a grounding or near-miss, he will probably have received a rating at the lower end of the scale on the *TSR categories*. The correlation is especially strong for *Safety* and *Resource Management* and somewhat weaker, yet marked, for *Timing.* On the other hand, the fact that a captain experienced a grounding or near-miss cannot be used to predict at all his score on the category *Communication*.

---

[3] In the figures we use box plots to illustrate relations between ordinal and nominal variables. Ordinal variables will have values distributed along a linear scale, e.g., age or time as navigator. Nominal variables divide data into discrete categories, e.g., "grounding-yes" and "grounding-no". A box plot consists of a box with two "wings", a median (denoting the value at which half of the observations are above and half are below) and a shaded area that represents the 95% confidence interval. This interval refers to certainty with which we may fix the position of the median. As a rough but useful rule of thumb one may say that if the shaded 95% confidence intervals in a box plot do not overlap, then it is likely that there is a difference between the medians of the ordinary variable ("Age") when we divide our observations into the groups of the nominal variable ("grounding-yes" and "grounding-no"). In the example given in Fig. 2, where the ordinal variable Age is plotted against the nominal value Grounding-Yes/No, it is revealed that there is probably a relation between a captain's belonging to one the latter two groups and his age. The bottom horizontal line indicates 0% of the observations and the top horizontal line 100%, and the bottom of the unshaded box indicates 25% and its top 75% of the observations. These percentages are approximate, since singe outliers may end up outside.

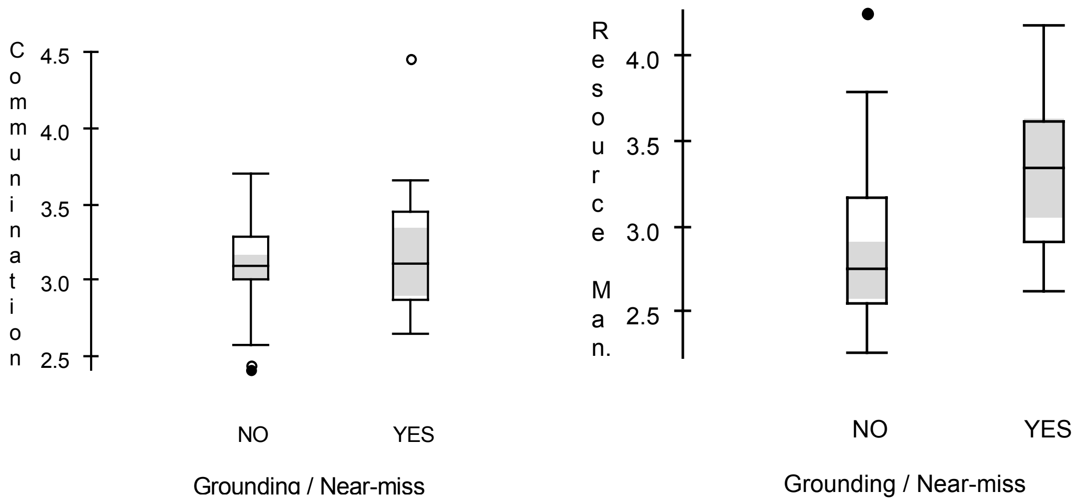

Fig. 3 The left hand box plot shows that there is no difference between the captains who had and those who did not have a grounding or near miss when we plot this against the scores they received on Communication. The right hand box plot, in contrast, shows that there is a marked difference in scores on the category Resource Management when we divide captains into these two groups

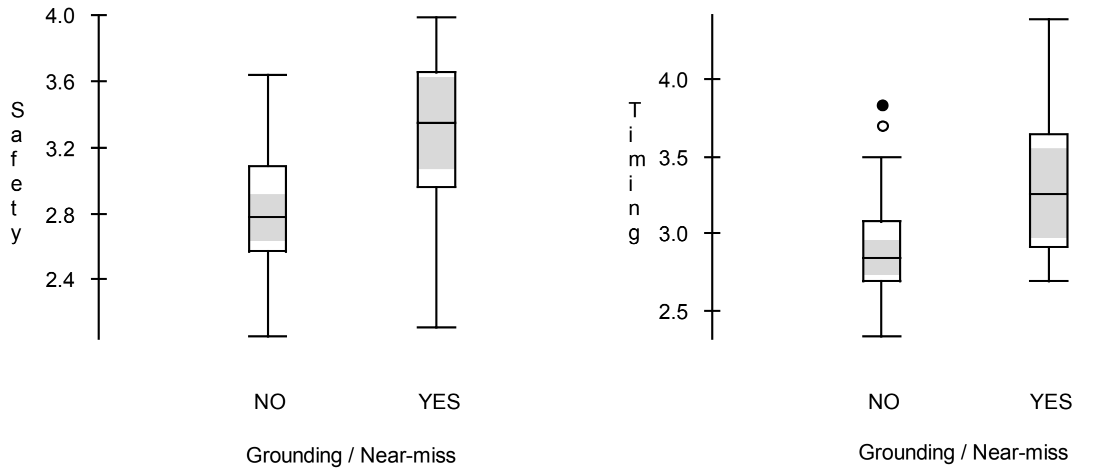

Fig. 4: As in the case of Resource Management, we find a clear difference between the scores on Safety (the left hand box plot) as well as Timing (the right hand box plot) when we divide the captains into the group who did and the group who did not experience a grounding or near miss

So far, we have seen that high scores on the navigation instructor categories (except *Communication)* can be used to predict a “no-grounding/near-miss” with a reasonably good probability; and that the older captains have a smaller risk of belonging to the group of “grounding and near-miss”. A reader may be inclined to expect now that the older captains would tend to obtain higher scores on the navigation instructor’s assessment for TSR categories. But as we shall see now, the data revealed quite a different picture.

3.4 Captains’ age and time in present position related to navigation instructor’s scores

No significant relations were found between captains’ time in present position and the scores received on any of the four navigation instructor categories. As was noted above, the variable “time in present position” means “time in present position in present company” and, therefore, cannot be expected to reflect a captain’s experience in his position. But while there was, as expected, a high correlation between *Age* and *Time in present position*, there was in fact also a significant correlations between *Age* and navigation instructor scores - but not in the direction one might have expected:

Thus, it turned out that the older captains had obtained a lower rating on the categories Safety, Resource Management and Communication, but not on Timing. (Resource Management: p=2,39% R2=10.8% - i.e., the probability that this relation is obtained by chance alone is 2.39% and Age can explain 10.8% of the distribution. Communication: p=0.05% and R2=16.3%. Safety: p=1.05% R2=13.7%). Confer figures 5a-d.

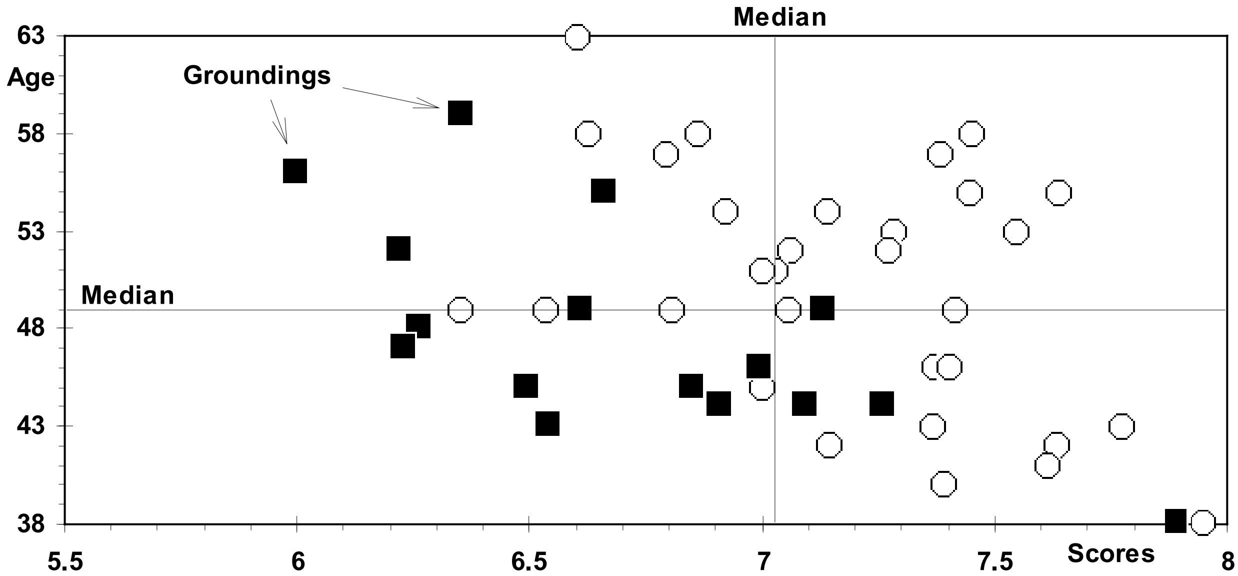

*Fig 5a: Scores on **Safety** plotted against age. Squares represent captains with at least one grounding, circles those with no groundings*

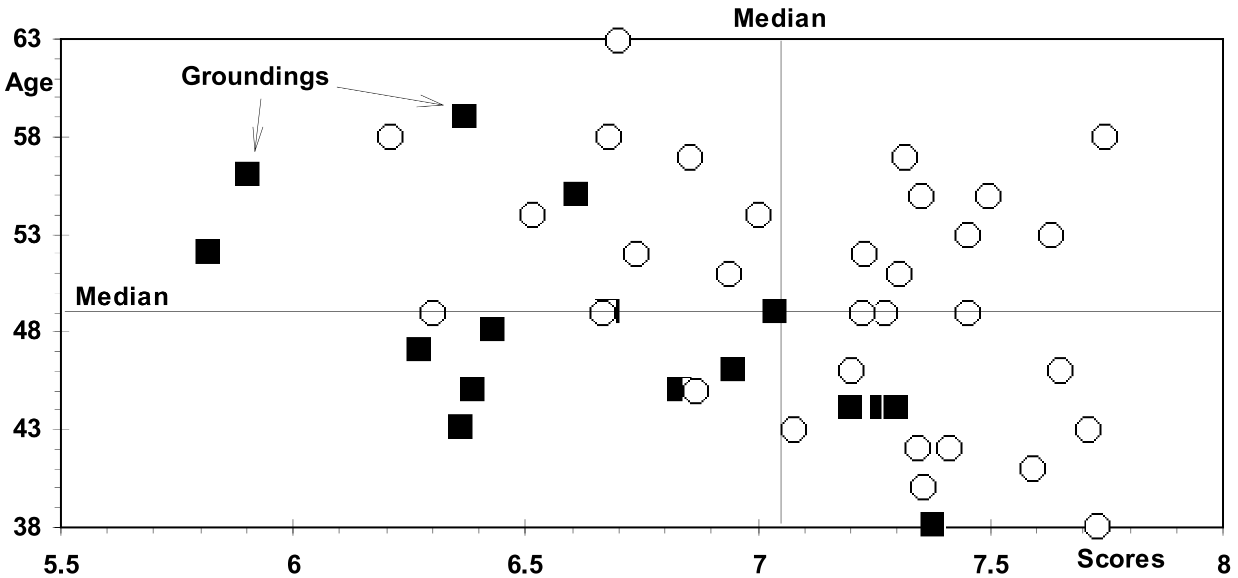

*Fig 5b: Scores on **Resource Management** plotted against age. Squares represent captains with at least one grounding, circles those with no groundings*

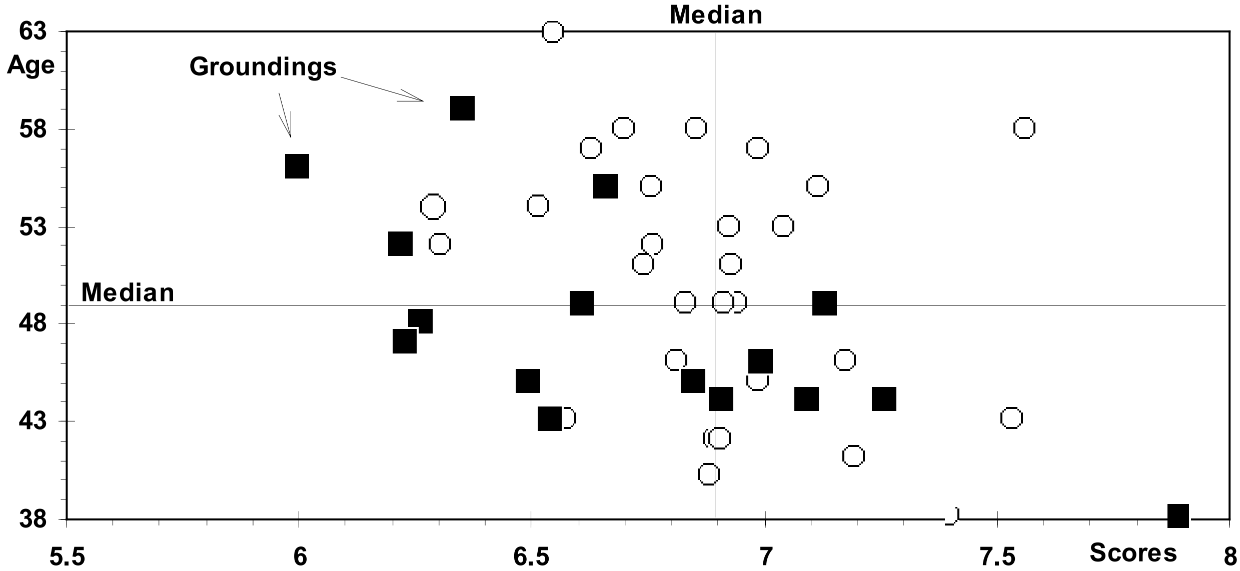

*Fig 5c: Scores on **Communication** plotted against age. Squares represent captains with at least one grounding, circles those with no groundings*

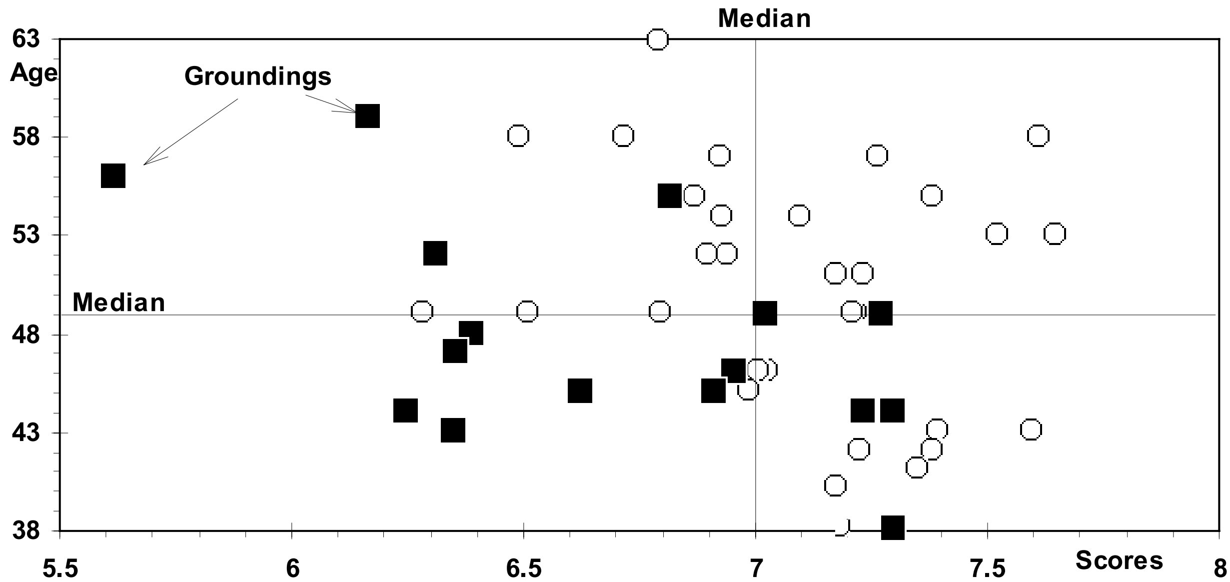

*Fig 5d: Scores on **timing** plotted against age. Squares represent captains with at least one grounding, circles those with no groundings*

It is natural to speculate about possible explanations of the following chain of correlations:

- instructor scores on the TSR-categories predict membership of the incident group quite well: high TSR scores indicate less risk of incidents
- older captains obtained a lower score on *Safety, Resource Management* and *Communication* (but not on *Timing*);

and yet:

- older captains had significantly **fewer** incidents.

At a first glance, this seems surprising. Lower-end scores correlate roughly with greater risk of incidents; older subjects receive relatively low scores - hence, one might expect older subjects to have more incidents. But it is exactly opposite: they have a smaller risk of having incidents.

We suggest that a partial explanation for this chain of correlations may be given along the following lines. First, although we have no data on the length of experience as captains our subjects had, we are strongly inclined to hold that "age" is tied to experience for our sample of captains. Second, it is generally recognised that experience is a significant factor for avoiding incidents. But, third, it is seemingly more difficult for the relatively older captains to adapt to bridge behaviour oriented towards more "modern" styles of leadership. As was mentioned above, immediately before the officers started on their full mission simulator course, they had participated in a 3-day Crew Resource Management (CRM) course in which participants received a broad range of all those non-technical aspects that influence safety. Courses like these are based, first, on the observation that about two-thirds of all accidents at sea are caused by human errors and, second, that human errors on the bridge can be reduced by, inter alia, training that highlights how inadequate communication and use of available human resources on the bridge can jeopardise safety. So, such a course will tend to change participants' behaviour towards a greater recognition of the importance of "modern" styles of leadership and communication. On the other hand, judged on informal observations by several experts, older captains are less liable to adapt to "new" styles of leadership.

The results concerning the variables involved in instructor scores, incidents and age, can be summarised broadly as saying that a captain has a diminished risk of having an incident in the realistic environment provided by a full mission simulator if he either receives relatively high scores by an experienced navigation instructor or if he belongs to the elder (and presumably more experienced) group of captains. Of course, fulfilling both criteria is even better.

3.5 Correlations between types of communication and incidents

As the simulated voyages were conducted, communication on the bridge was classified on-line, as mentioned above, by a psychologist. Since this classification was done in real time, it was inevitable that utterances would be lost or missed; however, by comparing results on when both of the psychologist raters (either of whom served as rater during the courses) were scoring independently of each other, it turned out that there was a high degree of congruence in their scores. These raters, recall, would score bridge communication into commands, observations, questions and replies and by originator and would note whether utterances concerned the past, the future or the current situation.

No correlations among these scores and the navigation instructor scores were found - a result which by itself came as a surprise to some of the training staff and analysts. Equally, there was no correlation between the psychologists' scores of utterances and age, nor was there any correlation between incidents and *absolute* numbers of *utterance types* (commands, observations, etc. or their *temporal aspect*). However, a significant correlation was found between incidents and the percentage of communication directed at the future vs. the present. Captains who were using a relatively large percentage of their communication on future matters had a significantly greater chance of belonging to the "grounding-no" group; and equally, captains who devoted a relatively larger percentage of their communication to present events were more likely to belong to the "grounding-yes" group.

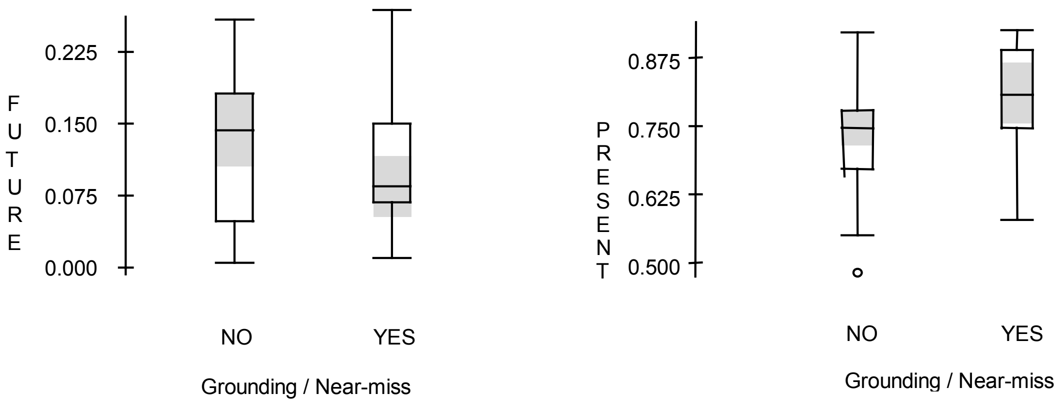

Fig. 6: The left hand box plot shows the percentage of communication directed at future events against "grounding-yes/no"; the right hand box plot shows the percentage of communication directed at current events

Since groundings and near misses may well evoke communication about the present it might be conjectured that this alone can explain why the correlation obtained. That is, perhaps it was the groundings / near-misses that prompted the captains to speak so relatively much more about the present. However, this conjecture about why there is a relation ship between temporal aspects of communication and incidents fails, for it must be stressed that the segments of the voyage in which an incident would occur were quite short in comparison with the total voyages (30 min and 150 min.). As was pointed out, there was no relationship between the absolute number of utterance types and incidents - only between the percentual distribution into future / present and incidents. We suggest, first, that the total amount of utterances represents an individual characteristic which, if our data are valid, bears no relation to the likelihood of an incident; and second, that the distribution between utterances about the future and the 'here-and-now' may be taken as a *measure of the attention* of the captain and his crew. The result says, therefore, that crews whose attention has a relatively greater focus on the current situation and a relatively minor focus on things ahead will run a greater risk of experiencing an incident. Corresponding to this hypothesised explanation of why the relationship obtains, one may also say that it reflects the familiar fact that those captains and crews who have the requisite professional resources (knowledge, skills, well prepared voyage) will tend to *both* avoid having incidents *and* be concerned with planning and with priming their attention and actions to things ahead. So, the underlying skills and knowledge are causally responsible for likelihood of incidents *and* for the ability of a captain to spend a relatively greater amount of his communication (and attention) on things to come.

3.6 The stability over time of scores by navigation instructor and psychologist

An analysis was made in order to detect possible changes and movements in scores and scores over the roughly six quarters during which the 90 voyages of the course took place. An outstanding result of this is that there is hardly any variation across the one-and-a-half years the course took place and none which may not be explained by chance. In total, both the navigation instructor's scores as well as the psychologists' classification of utterances showed a surprisingly robust stability across yearly quarters.

4. Conclusion

The findings reported above will be compared with results of analysis of data which are scheduled to be collected from future simulator training courses at the DMI simulator facilities. In particular, we wish to investigate further the relations between the categories of age and experience, instructor scores and objective incidents, and we aim at examining the influence of CRM-type training on navigator's performance in a full mission simulator. Studies of the relationships between age and performance (Pélegrin et al., 1995; Salthouse 1990; Smith 1990; Davies et al. 1992) reveal uniform degradation of most cognitive functions at a gradual and very modest rate until after the 60s, but there is little evidence in the literature of how age interacts with the learning of new social skills. This issue seems to deserve further study considering the almost universal recognition that management of resources is the single most important single factor behind accidents in areas such as aviation and navigation.

5. References

Pélegrin, C., Maho, V. & Amalberti, R.: "Pilot age and training performance". In N. Johnston, R. Fuller & N. McDonald (eds.): *Aviation Psychology: Training and Selection*, *Vol 2*, Avebury Aviation, Aldershot, 1995

Salthouse, T.: "Influence of experience on age differences in cognitive functioning". *Human Factors*, 32, 551-69

Smith, D.V.D.: "Human factors and ageing: An overview research needs and application opportunities". *Human Factors*, 32, 509-26

Davies, D., Taylor, A. & Dorn, L.: "Aging and human performance". In A.P Smith & D.M. Jones (eds.): *Handbook of Human Performance, Vol 3,* Academic Press, London 1992