# The Validation of Non-Technical Behavioural Markers for Merchant Navy Officers

Devitt, K. R., Holford, S.D., Pantaleev, B and Sharma, D

Warsash Maritime Academy

Southampton Solent University

September 2012

# ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the following people and organisations, without whom this validation research would not have been possible.

Mike Blair and Peter Camp from Chevron

Krishan Unni from Exxon Mobil

The shipping division of Shell

The lecturers and technicians at Warsash Maritime Academy

# TABLE OF CONTENTS

1. Introduction........................................................................................................ 4

2. Literature Review ................................................................................................ 4

3. Methodology ........................................................................................................ 6

4. Data Collection ..................................................................................................... 7

5. Data Analysis ........................................................................................................ 8

Table 1: Extract of Excel document containing BM assessment ................. 8

Table 2: Frequency of observations for Decision Making .............................. 9

6. Summary ............................................................................................................... 10

7. Bibliography ......................................................................................................... 13

8. Appendix 1: Colour coded table of relative observability .................................. 16

9. Appendix 2: Table of ranked observed positive behaviours .............................. 17

# 1 INTRODUCTION

Resource management principles, first established in the aviation industry, have been adopted and adapted for the Merchant Navy and other safety critical areas where the importance of the blend of effective technical and non-technical skills is recognised.

The 2010 Manila amendments to the STCW Convention and Code include requirements for deck and engineering officers to show competence in the non-technical skills of resource management, leadership and managerial skills.

Research carried out by Devitt and Holford (2010) sought to identify whether the broad competence criteria described in the STCW amendments would be consistently interpreted by a range of maritime industry stakeholders, in the context of the behaviours that they would expect to see, hear or experience in demonstrating compliance with, and evidencing competence in, the criteria.

Building on the response of the research participants, the authors proposed behavioural markers (BMs) designed to assess effective and ineffective behaviours underpinning non-technical skills as identified in the Manila amendments. These markers were then used as a basis for a series of observations carried out on simulation exercises currently being run for experienced seafarers at Warsash Maritime Academy.

The intent was to establish the validity and usability of the BMs, in order to enable industry to assess the performance of the deck and engine room teams and to improve selection, training and promotion processes and procedures for merchant navy officers. This paper outlines the result of the follow-up validation of the proposed BMs.

## 2. LITERATURE REVIEW

Klampfer *et al (*2001) define BMs as "observable, non-technical behaviours that contribute to superior or substandard performance within a work environment" which can act as contributory factors in improving safety or in accidents and incidents in the aviation and other industries. They further define BMs as "observable behaviours of teams or individuals, usually structured into a set of categories". Those categories consist of sub-components that are considered differently in a range of behavioural marker systems (for example, the aviation industry's non-technical skills (NOTECHS): "markers" and "elements").

Klampfer *et al* (2001) also show that there is common agreement regarding the significance of interpersonal behaviours in technological environments, and the requirements for training these behaviours (sometimes called non-technical skills or behavioural markers) to complement technical training. Crew Resource Management (CRM) training was initiated to achieve this in aviation and is now recognised all over the world. BMs have been successfully implemented in aviation as valuable tools in preparation and evaluation of CRM methods, and in influencing training. BM systems continue to be developed for the purpose of performance evaluation in many industries such as aviation, nuclear power, railways and medicine. The rationale behind the BM systems differs, as do the current BM systems themselves, even if the main concepts are comparable. It appears that some uncertainty exists as to what precisely BMs could be used for, and how they are able to contribute to enhanced performance in high-tension situations (Klampfer *et al*, 2001).

In their research, Thomas *et al* (2004) suggested that BMs might contribute to the reduction and better management of errors. This view is shared by Aviation Teamwork (2011), a company specialising in assessment of non-technical skills (NOTECHS), proposing that BMs help the preparation of individual feedback to crews "and trainees on their non-technical skills development".

Evidence from recent maritime accident investigation reports has shown severe shortcomings in the competence of some Merchant Navy officers, who were unable to manage both resources and crises (ATSB, 2005; MAIB, 1999; MAIB, 2007; MAIB, 2011; NTSB, 2005). According to the American Bureau of Shipping (2005), human factors are the root cause in 70% of all shipping incidents. The UK P&I Club (UK P&I, 2010) considers that human error is directly accounted for in more than 58% of all major insurance claims. Gatfield (2004) argues that not all such incidents lead to crisis, but they do have the potential to do so.

Non-technical skills (NTS) are a crucial part of effectiveness in a maritime environment, but these skills are not addressed explicitly in standard training. Devitt and Holford (2010) proposed behavioural markers that are specific to the maritime industry and would enable resource management and leadership competencies to be evaluated against the framework of STCW.

It is not only in the maritime industry that this link between NTS and training is found to be deficient. Fletcher *et al* (2003) comment: "Non-technical skills are critical for good anaesthetic practice but are not addressed explicitly in normal training. Realization of the need to train and assess these skills is growing, but these activities must be based on properly developed skills frameworks and validated measurement tools."

## 3. METHODOLOGY

Devitt and Holford's (2010) proposed behavioural markers were developed and refined through the process of response coding, which were then put into prioritised markers. These markers were tested for validity and reliability in the simulated environments of both bridge and engine room. As suggested by Gatfield (2008) in his research on BMs for the assessment of competence in crisis management:

> "For a behavioural marker to be an effective assessment metric, it needs to be relevant to the competence being assessed. It should be easily evaluated as a demonstration of good or poor behaviour, easily observed and should occur quite frequently."

According to the original research carried out by Devitt and Holford (2010)

> "This will allow the markers to be refined into a user friendly matrix of positive and negative markers which can then be applied with some adaption to processes of selection, appraisal, training and promotion".

High and low performance markers were arranged in their respective categories and a marking system devised, using a system already tested by the Rail Safety and Standards Board, RSSB (2009), which commented:

"Once the draft set of behavioural markers had been developed, the Simulator Operators (SOs) were asked to think of particularly good and particularly poor assessments to help them in identifying relevant behaviours."

All elements that were analogous to other BM categories were removed.

A BM observation matrix was created in an A3 format for ease of observation, allowing an experienced observer to monitor three seafarers simultaneously. However, based on the experience of the research team, it is recommended that two observers work as a team, especially in the first months of their work with this set of BMs, otherwise the quality of the data may be compromised.

## 4. DATA COLLECTION

Over a period of 4 months, 60 simulator exercises were observed in the WMA simulators. These exercises formed part of Engine Room Resource Management (ERRM) and Bridge Team Management (BTM) courses. The deck officer sample was drawn from Shell, Exxon Mobil and Chevron; the engineer officers were drawn only from Chevron due to the suitability of the course set up for that specific company. There were thus more opportunities to observe deck officers (34 out of a total of 45 participants).

During the simulator exercises on the bridge, four participants are normally present. There are 8 simulations during the week and the deck officers are rotated between the following roles: Master, navigator, extra officer operating ARPA (Automatic Radar Plotting Aid) , and observer. As the observer is purely recording the performance of his colleagues, and offering suggestions for alternative actions, this role was not included in the observations. In total, the research team made 145 observations. Every seafarer was observed more than once because of the role rotation, and each participant during every exercise is counted as a separate observation. Thus, if one deck officer demonstrated an observable behaviour in 5 different exercises, five observations were made.

The engine room simulations have a similar role rotation but do not use the observer role as frequently as the bridge simulations.

Occasionally, the actions of the same individuals were validated more than once. For example, one participant carried out the role of Master, navigator and extra officer operating ARPA more than once during the course. Thus, three validations were recorded on one individual because these three roles are radically different.

The nationality and gender of participants were recorded, though analysis of these variables sits outwith the boundary of this research.

## 5. DATA ANALYSIS

For the purpose of data evaluation all the obtained data was transferred into an Excel Document. An extract dealing with BM assessment is shown in Table 1:

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Participant | Date | Personal Details | | | Relevent Experience | | | | |
| 2 | | | Dept | Gender | Nationality | Role in the Exe | Real Position at | | | |
| 3 | | | | | | | | (1) Recog | (2) Has p | (3) Optio |
| 4 | | | | | | | | | | |
| 5 | 001 M | C1-18/05/2011 | 1 | 1 | Polish | Master | Second Officer | N/A | N/A | N/A |
| 6 | 001 M Validation | C4-20/05/2011 | 1 | 1 | Polish | Master | Second Officer | N/A | 1 | 1 |
| 7 | 002 M | C2-18/05/2011 | 1 | 1 | Polish | Master | Master | 1 | 1 | 1 |
| 8 | 002 M Validation | C5-20/05/2011 | 1 | 1 | Polish | Master | Master | 1 | 1 | 1 |
| 9 | 003 M | C3-19/05/2011 | 1 | 1 | Croatian | Master | Second Officer | 0 | 0 | 1 |
| 10 | | | | | | | | | | |
| 11 | 004 M | S1-23/05/2011 | 1 | 1 | Croatian | Master | Master | 1 | 0 | 1 |
| 12 | 004 M Validation | S6-27/05/2011 | 1 | 1 | Croatian | Master | Master | 1 | 1 | 1 |
| 13 | 004 A | S2-24/05/2011 | 1 | 1 | Croatian | ARPA | Master | N/A | 1 | 1 |
| 14 | 004 A Validation | S3-25/05/2011 | 1 | 1 | Croatian | ARPA | Master | 0 | 2 | 1 |
| 15 | 005 M | S5-27/05/2011 | 1 | 1 | British | Master | Second Officer | 1 | 2 | 1 |
| 16 | 005 A | S1-23/05/2011 | 1 | 1 | British | ARPA | Second Officer | N/A | N/A | N/A |
| 17 | 005 A Validation | S4-26/05/2011 | 1 | 1 | British | ARPA | Second Officer | 0 | 0 | 1 |
| 18 | 006 M | S2-24/05/2011 | 1 | 1 | British | Master | 3rd Mate | 1 | 1 | 1 |
| 19 | 006 M Validation | S4-26/05/2011 | 1 | 1 | British | Master | 3rd Mate | 0 | 0 | 1 |
| 20 | 006 A | S5-27/05/2011 | 1 | 1 | British | ARPA | 3rd Mate | 1 | 2 | 1 |
| 21 | 007 M | S3-25/05/2011 | 1 | 1 | Croatian | Master | 3rd Mate | 0 | 2 | 1 |
| 22 | 007 A | S6-27/05/2011 | 1 | 1 | Croatian | ARPA | 3rd Mate | N/A | N/A | N/A |
| 23 | | | | | | | | | | |
| 24 | 008 M | C6-13/06/2011 | 1 | 1 | Italian | Master | Master | N/A | 1 | 1 |
| 25 | 008 A | C9-15/06/2011 | 1 | 1 | Italian | ARPA | Master | 1 | 1 | 1 |

**Table 1: Extract of Excel document containing BM assessment**

Marking of behaviours are as follows:

1 means that a positive behaviour has been observed

0 means that a negative behaviour has been observed

2 means that both a positive and a negative behaviour was observed

N/A means Not Applicable, and identifies a particular situation that does not require the suggested behaviour.

As Table 2 below indicates, each of the 65 elements could theoretically have been observed on 145 occasions. However, some of the pre-planned events of the simulator exercises were never encountered, and these were marked as N/A. The other frequencies show how many times the positive, the negative, or the combination of those two elements was recorded.

| | A | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Participant | | | | | | | | | | | | |
| 2 | | (1) DECISION MAKING | | | | | | | | | | | |
| 3 | | (1) Recog | (2) Has p | (3) Optio | (4) Range | (5) Regul | (6) Team | (7) Decisi | (8) Reviev | (9) Confic | (10) All rel | (11) Ensure | (1) D |
| 4 | | | | | | | | | | | | | |
| 218 | | | | | | | | | | | | | |
| 219 | Frequency of marker | 145 | 145 | 145 | 145 | 145 | 145 | 145 | 145 | 145 | 145 | 145 | 14 |
| 220 | Frequency of N/A | 31 | 13 | 4 | 17 | 64 | 4 | 6 | 13 | 8 | 4 | 18 | 2 |
| 221 | Frequency of Negative | 19 | 15 | 5 | 80 | 22 | 2 | 0 | 40 | 3 | 1 | 13 | 9 |
| 222 | Frequency of Positive | 91 | 104 | 136 | 47 | 55 | 139 | 139 | 92 | 123 | 140 | 110 | 11 |
| 223 | Frequency of Both | 4 | 13 | 0 | 1 | 4 | 0 | 0 | 0 | 11 | 0 | 4 | 1 |

**Table 2: Frequency of observations for Decision Making.**

A table was also created showing the relative observability of the BMs, - an example of observable positive behaviours in descending order of observability can be seen in Appendix 1. The different colours in that table indicate to which category of BM the respective elements belong.

A distinction must be made between non-observable behaviours, and behaviours which were not observed. The former would indicate that behaviours could never be observed; the latter suggests that during the period of observation, the behaviours were not demonstrated. The latter case may have been due to a number of factors. The use of the simulation exercises to observe behaviours was opportunistic, that is, the exercises were not designed for that purpose, but were a convenient context in which to validate the markers. A significant number of BMs were observed within the simulators, though it was recognised that a limitation of this context might be that certain behaviours might not be observed. However, this did not necessarily mean that those BMs would be of no use – work on board ship differs from pre-planned exercises, and therefore non-observed BMs might still be relevant in that context. An example would be a BM which came near to the bottom of the list of observed positive behaviours at Appendix 2 - "member uses good conflict management strategies". It would be invidious to suggest that such a positive behaviour would not be observed at sea, as would its negative opposite - "bullying".

Another possible contributory factor to non-observed markers could might be the awareness of the students in the simulator exercises that they were being watched – and therefore on their "best behaviour" – the Hawthorne effect (Landsberger, 1958).

Where behaviours could be observed, it became clear that certain behaviours were demonstrating ineffective performance. In more than the half of the observed simulator exercises, navigation and technical problems were addressed by using the first solution option generated by the team. It was rare to observe contingency plans being considered or implemented. On board a vessel, such a one sided approach would have left the crew badly exposed if their initial idea failed and there was no Plan B.

It would be incautious to generalise to the wider industry, however, as 45 participants - albeit from three oil majors - are insufficient statistically. However, it does indicate that further research would be useful to identify whether this lack of contingency planning is more widespread.

## 6. SUMMARY

The specific objectives of this project were to:

1 Identify whether the behavioural markers proposed by Devitt and Holford (2010) were appropriate for use within the industry to measure effective competence in resource management skills laid out in the STCW Manila amendments of 2010.

2 Compare the selected categories and elements and check for overlapping.

3 Identify the appropriate number of behavioural markers that could be practically observed in each category

After observing a statistically valid sample of seafarers, the overwhelming majority of the BMs appear to be both relevant and observable. There are caveats regarding some BMs which were not observed in the simulators, but which are highly relevant in a real time situation at sea. An example of this relates to handovers between team members, which was hardly ever observed in the simulators due to the constraints of the exercises, but would clearly happen regularly at sea. It would be appropriate to carry out further research in a seafaring environment to test this hypothesis.

There is a limit to the number of BMs that any observer can reasonably be expected to observe, even after training. These can be determined under each category, based on the findings of observability frequency. However, should training organisations or companies wish to spotlight in depth behaviours within a particular category, they could utilise the whole span of behaviours pertinent to that category.

Structuring the behavioural markers in terms of frequency of positive and negative behaviours allows shipping companies and training establishments to pinpoint areas where training may usefully be focused. It speaks highly for the three oil majors supporting this project that there is not a single BM where more than 10% of negative results were observed.

Strengths of both individual seafarers and entire bridge and engine room teams can be assessed, when observations are repeatedly carried out. Further, these BMs could be used for specific selection purposes. For example, one of the most positively observed BMs was "Communication acknowledges cultural diversity". This suggests that that the seafarers participating in this research should be well suited to work as a part of a multinational crew.

The BM matrix can be also used for evaluation of the effectiveness of BRM and ERM courses. It was noted that over a period of a week, some behaviours showed significant improvement. This was the case for the BM "Clearly and unambiguously briefs key personnel on roles and duties and main priorities of the task and solicits feedback"

As a result of this validation programme, the researchers suggest that the BM matrix proposed contains valid BMs that are usable within the maritime industry to assess effectiveness in non-technical skills. The matrix could be used in selection, appointment and promotion, and after suitable training, assessors can utilise this tool both at sea and shoreside.

## Bibliography

American Bureau of Shipping (2005). *Accident database review of human element concerns: what do the results mean for classification?* [online] [viewed 14 January 2011]. Available from: http://www.eagle.org/eagleExternalPortalWEB/ShowProperty/BEA%20Repository/References/Technical%20Papers/2005/AccidentDatabaseReview%20of%20HumanElementConcerns

Australian Transport Safety Bureau (2005). *Independent investigation into equipment failure aboard the Hong Kong flag container vessel Maersk Tacoma* [online] [viewed 04 January 2011]. Available from: http://www.atsb.gov.au/media/24966/mair171_001.pdf

Aviation Teamwork (2011). *NOTECHS - Behavioural Markers*. [online] [viewed 04 January 2011]. Available from: http://www.aviationteamwork.com/safetymanagement/markers_atticus.asp

Devitt, K., and Holford, S. (2010). *Development of resource management and leadership behavioural markers for the merchant navy*. In: Maritime Human Resource Solutions Conference, 28-30 September, 2010, St John's, Newfoundland. Canada

Fletcher, G., Flin, R., Mcgeorge, P., Glavin, R., Maran, N. and Patey, R (2003). Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *British Journal of Anaesthesia* 90(5), 580-8

Gatfield, D., (2004). Behavioural Markers: A framework for the objective assessment of competence in crisis management. *Proceeding of 6th International Conference on Engine Room Simulators*. Wuhan, China: Wuhan University of Technology

Gatfield, D., (2008). *Behavioural Markers for the Assessment of Competence in Crisis Management*. PhD thesis

Goldsmith, T.E., and Johnson, P.J. (2002). Assessing and improving evaluation of aircrew performance. *International Journal of Aviation Psychology*, 12, 223-240

James, L.R., Demaree, R.G., and Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology*, 69, 85-98

James, L.R., Demaree, R.G., and Wolf, G. (1993). An assessment of within-group interrater agreement. *Journal of Applied Psychology*, 78, 306-309

Jartel Consortium, (2002*). Joint Aviation Requirement: Translation and Elaboration of Legislation.* Final Report, Contract n AI-97-SC.2037. [online] [viewed 08 December 2010]. Available from: http://www.transportresearch.info/Upload/Documents/200310/jartelrep.pdf

Johnson, P.J., Goldsmith, T.E., (1998). *The importance of quality data in evaluating aircrew performance*. US Federal Aviation Authority Technical Report. [online] [viewed 24 January 2011]. Available from: www.faa.gov/avr/afs/aqphome

Klampfer, B., Flin R., Helmreich, R.L., Haeusler, R., Sexton, B., Fletcher, G., Staender, S., Lauche, K., Dieckmann, P. & A. Amacher, (2001*). Enhancing Performance In High Risk Environments: Recommendations for the use of Behavioural Markers*. [online] [viewed 24 January 2011]. Available from: http://homepage.psy.utexas.edu/homepage/group/helmreichlab/publications/pubfiles/pub262.pdf

Landsberger, H. A, (1958) *Hawthorne Revisited*, Ithaca

MAIB, (2007). *Report on the investigation of the machinery breakdown and subsequent fire on board Maersk Doha*. [online] [viewed 28 March 2011]. Available from: http://www.maib.gov.uk/cms_resources.cfm?file=/Maersk%20Doha-published.pdf

MAIB, (2011). *Report on the investigation of the machinery space fire on board Oscar Wilde* [online] [viewed 28 April 2011]. Available from: http://www.maib.gov.uk/cms_resources.cfm?file=/Oscar_Wilde_Report.pdf

National Transportation Safety Board, (2005). *Grounding of Malaysian-flag Bulk Carrier M/V Selendang Ayu* [online] [viewed 04 January 2011]. Available from: http://www.ntsb.gov/publictn/2006/MAB0601.htm

RSSB. (2009) *A model for the measurement of non-technical skills and the management of errors in the simulator*. Rail Safety and Standard Board, London.

Thomas, E., Sexton, J. & R. Helmreich, (2004). Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *Quality and Safety in Health Care*. 2004 October, 13 (Suppl1): 57-6

UK P&I CLUB, (2010). *The Human Element* [online] [viewed 04 January 2011]. Available from: http://www.ukpandi.com/loss-prevention/risk-management-advice/the-human-element/