University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

U.S. Navy Research | US Department of Defense

1-1-2011

The development of a prototype behavioral marker system for US Navy officers of the deck

Paul O'Connor
National University of Ireland, poc73@hotmail.com

W. Max Long
Naval Postgraduate School, wmlong@ymail.com

Follow this and additional works at: http://digitalcommons.unl.edu/usnavyresearch

O'Connor, Paul and Long, W. Max, "The development of a prototype behavioral marker system for US Navy officers of the deck" (2011). *U.S. Navy Research.* Paper 39.
http://digitalcommons.unl.edu/usnavyresearch/39

This Article is brought to you for free and open access by the US Department of Defense at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in U.S. Navy Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Safety Science 49 (2011) 1381–1387

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

The development of a prototype behavioral marker system for US Navy officers of the deck

Paul O'Connor [a,*], W. Max Long [b,1]

[a] Centre for Innovation and Structural Change, J.E. Cairnes School of Business & Economics, National University of Ireland, Galway, Ireland
[b] Operations Research Department, Naval Postgraduate School, 1411 Cunningham Rd, Monterey, CA 93943, USA

ARTICLE INFO

Article history:
Received 12 November 2010
Received in revised form 24 May 2011
Accepted 29 May 2011
Available online 25 June 2011

Keywords:
Human factors
Nontechnical skills
Behavioral markers
Maritime
Military

ABSTRACT

The officer of the deck (OOD) of a US Navy ship is in charge of the safe and proper operation of the ship, and accountable to the commanding officer for every event that occurs during his or her OOD watch. This paper discusses the development of a prototype behavioral marker system to evaluate, and provide feedback on, the nontechnical (cognitive, social, and personal resource) skills of OODs.

An initial set of 17 categories of nontechnical skills were identified from a literature review. A focus group with four qualified OODs used the skills identified from the literature review to develop an initial taxonomy of five categories, each with two or three corresponding behavioral elements. This taxonomy was then used to classify 149 statements concerned with the nontechnical skills of OODs collected from 16 critical incident interviews. After three iterations of adaptations to the taxonomy, two independent raters were able to reach acceptable levels of reliability in using the taxonomy to classify the statements. Although further development work is required, it is suggested that the prototype behavioral marker system has implications for improving safety and performance on military and civilian ships.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The officer of the deck (OOD) is a critical watch station on a US Navy ship. The OOD is accountable to the Captain for every event that occurs during his or her OOD watch. "*As the Captain's direct representative, the OOD is the only person on board who can make decisions that affect the safety of the ship and the lives of her crew*" (Stavridis and Girrier, 2007, p. 3). There are 13 separate watch stations (individuals in the ship's company responsible for carrying out particular tasks) that report directly to the OOD. When underway, the OOD is designated by the Captain to be in charge of the ship including its safe and proper operation (Chief of Naval Operations, 2005). "*Nowhere in military or civilian life is there a parallel to the range and degree of responsibility that is placed in the hands of the OOD. As direct representative of the Captain, he or she acts with all the authority of command and, next to the Captain and the Executive Officer (XO),* [the OOD] *is the most important person on the ship*" (Stavridis and Girrier, 2007, p. 1).

While technical skills and knowledge are obviously necessary for an effective OOD, they are not sufficient. The UK Marine Accident Investigation Branch (MAIB) stated that the majority of maritime accidents are attributed to human error (MAIB, 2000). Rothblum (2000) states that 75–96% of marine casualties are caused, at least in part, by human error. Similarly, Wagenaar and Groeneweg (1987) analyzed 100 Dutch shipping mishaps and found that in 96 of the mishaps those individuals involved could have prevented the mishap. In a recent study of mishaps involving Greek-flagged ships from 1995 to 2006, 57% of mishaps were attributed to human factors (Kokotos and Linardatos, 2011). Two examples of US Navy maritime mishaps in the public domain caused by poor nontechnical skills were the downing of an Iranian Airliner by the *USS Vincennes* in July 1988 (Klein, 1998), and the ramming of an anchored Spanish bulk carrier by the nuclear-powered aircraft carrier *USS Dwight D. Eisenhower* in August 1988 (National Transportation Safety Board, NTSB, 1990). It was determined that the probable causes of the ramming of the Spanish vessel were the delayed and insufficient action to correct the *USS Eisenhower*'s deviation from the intended track by the navigator and the OOD. The OOD did not adequately track the vessels movement (failure of situation awareness), and the commanding officer (CO) was not informed of a change of speed in a timely manner (failure of communication; NTSB, 1990).

Nontechnical skills are the "*cognitive, social, and personal resource skills that complement technical skills, and contribute to safe and efficient task performance*" (Flin et al., 2008, p. 1). Stavridis and Girrier (2007) state that the nontechnical skills of forehandedness, vigilance, judgment, intuition, and leadership are just as important to OODs as technical skills and knowledge. However, despite the recognized importance of nontechnical skills, there has

* Corresponding author. Tel.: +353 (0) 91 492897; fax: +353 (0) 91 495524.
E-mail addresses: poc73@hotmail.com (P. O'Connor), wmlong@ymail.com (W. Max Long).
[1] Tel.: +1 757 450 7012.

0925-7535/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ssci.2011.05.009

This article is a U.S. government work, and is not subject to copyright in the United States.

not been any research conducted to identify the specific nontechnical skills required by the OOD. There are also no valid and reliable tools to evaluate and provide feedback on the nontechnical skills of OODs. The lack of a framework for evaluating nontechnical skills is not confined to the OOD watch station. There has been little human factors research conducted in the maritime industry more generally (Hetherington et al., 2006). The purpose of this paper is to carry out a task analysis to identify the nontechnical skills required by effective OODs, and develop a prototype behavioral marker system for evaluating these skills.

2. Behavioral markers

Behavioral markers are "*observable, nontechnical behaviors that contribute to superior or substandard performance within a work environment*" (Klampfer et al., 2001, p. 10). They are usually structured into a set of categories (e.g. co-operation, decision making, and situational awareness). Normally, these categories are then sub-divided into more specific nontechnical skills or elements.

The seminal research on behavioral markers comes from studies of civilian pilots carried out by Helmreich and colleagues at the University of Texas. In the late 1980s they developed a data collection form called the LINE/LOS Checklist (LLC) to gather information on flight crews' crew resource management performance (Helmreich et al., 1990). This checklist has been used as the basis of many airlines' behavioral marker systems (Flin and Martin, 2001).

Another example of behavioral marker system that is widely used in civil aviation is Nontechnical Skills (NOTECHS). It consists of four categories (co-operation, leadership and management skills, situation awareness, and decision making), with 15 associated elements (for example, the elements in the co-operation category are 'team building and maintaining', 'considering others', 'supporting others', and 'conflict solving'. Ratings from 1 (very poor) to 5 (very good) are given to the pilot and co-pilot on each element. The NOTECHS system was found to have acceptable levels of reliability when used by trained raters to evaluate the nontechnical skills of civilian pilots (Flin et al., 2008). Behavioral marker systems have also been developed for using by anesthesiologists (Anaesthetists' nontechnical skills, ANTS; Fletcher et al., 2004), surgeons (Nontechnical skills for surgeons, NOTSS; Yule et al., 2006), scrub nurses (scrub practitioners' list of intra-operative non technical skills, SPLINTS; Mitchell and Flin, 2009), and nuclear power control room teams (O'Connor et al., 2008).

Klampfer et al. (2001) identified five characteristics of a good behavioral marker: it should describe a specific and observable behavior instead of an attitude or personality trait; while it does not have to be present in every situation, the behavioral marker should have a causal relationship with the performance outcome; it needs to be described in domain specific language; it should employ simple phraseology; and it should describe a clear concept.

Flin et al. (2008) also identified three significant limitations to behavioral marker systems. First, they are never capable of capturing every possible aspect of performance. The variables that effect performance are simply too wide ranging to rate with one system. The second limitation is that it may not be possible to observe some behaviors. Important, but infrequent, capabilities such as conflict management may not occur often enough to rate on a consistent basis. The final limitation is the abilities of human observers. While human raters bring experience and skills that a fully automatic assessor cannot match, they also bring their own biases and perceptions.

2.1. Developing a behavioral marker system

The typical method for the initial development of behavioral marker systems is to carry out a literature review of previous domain specific research concerned with nontechnical skills, followed by interviews with subject matter experts designed to extract the nontechnical skills required to do their job effectively (e.g. Fletcher et al., 2004; Mitchell and Flin, 2009; Yule et al., 2006).

It might be expected that useful information could be obtained from accident and incident reports. However, a number of authors have commented that these types of reports contain little in the way of human factors information (e.g. Fletcher et al., 2004; O'Connor et al., 2008; Yule et al., 2008). This is also true of reports of US Navy surface warfare mishaps and incidents. Carter-Trahan (2009) found that in an examination of 111 major mishaps in the surface warfare community investigated by the Naval Safety Center from 1999 to 2009, only 23% of mishaps were attributed to human factors (the remaining were attributed to material causes, 12%; and unknown causes, 65%). The percentage of mishaps attributed to human factors in the surface warfare community is far below the 80% that are generally attributed to human error in other organizations, including US Naval aviation (Hollnagel, 1993; Reason, 1990; Wiegmann and Shappell, 1999).

Based upon a review of the processes used to develop behavioral marker systems in other domains, it was decided to use a three stage methodology consisting of a literature review, focus group interview, and critical incident interviews to develop a prototype behavioral marker system for OODs. Each of these stages is described in detail below.

3. Literature review

The purpose of the literature review was to create an exhaustive list of nontechnical skills that have been shown to be necessary for effective performance in high risk domains, with a particular focus on any research that had been carried out in the maritime industry. Several electronic databases (PsychINFO, ScienceDirect, Web of Science, Google Scholar, and the Defence Technical Information Center) were used to identify research articles on human factors in the maritime industry by using the following search terms: maritime, shipping, behavioral markers, human factors, situation awareness, decision making, communication, leadership, teamwork, safety, and shipping/maritime accidents.

An examination of the literature on nontechnical skills in the maritime industry found that little previous research had been carried out in this domain. Similar conclusions were drawn by Hetherington et al. (2006). In their review of human factors studies carried out in the maritime industry only 20 relevant articles were found. Schröder-Hinrichs (2010) also stated that it was noticeable that recently human factors had not been high on the international maritime agenda.

The only reference to behavioral markers in the maritime industry was by Gatfield (2005). He used simulator observations to develop a set of markers for assessing the competencies of merchant marine engineering officers during crisis management. However, the behavioral markers were not developed at the time of publication (nor are they discussed in any subsequent publications).

Given the dearth of research on nontechnical skills in the maritime industry, four behavioral marker systems that were available in the literature (LLC, NOTECHS, ANTS, and NOTSS) were examined. Also included were the nontechnical skills identified by Stavridis and Girrier (2007), the skills identified by Flin et al. (2008), and the relevant personnel issues identified by Hetherington et al. (2006). This literature review resulted in the identification of 17 categories of nontechnical skills. These are shown in Table 1. The purpose of identifying these categories of nontechnical skills was to develop a "start list" (Miles and Huberman, 1994) of domains derived from the literature, which would then be modified as

Table 1
Initial list of nontechnical skills.

• Situational awareness	• Judgment
• Decision making	• Intuition/experience
• Communication	• Energy
• Team working	• Co-operation
• Leadership	• Management skills
• Managing stress	• Task management
• Coping with fatigue	• Workload management
• Forehandedness	• Assertiveness
• Vigilance	

necessary in subsequent stages of the development of the behavioral marker system.

From Table 1 it can be seen that several of the skills were overlapping, lacked specificity, and may be unrelated to the OOD watch station. In the next stage of the study the list of nontechnical skills obtained from the literature was refined with input from subject matter experts (SMEs).

4. Focus group

A group of SMEs identified which of the nontechnical skills in Table 1 were appropriate for assessing OODs. Once this had been carried out, the goal was for the SMEs to then agree upon an initial taxonomy of skills sub-divided into specific elements (the same structure as NOTECHS).

4.1. Focus group participants

The focus group consisted of four US Navy junior officers led by an experienced OOD. Each participant was a Surface Warfare Officer (SWO) attending the Naval Postgraduate School. All of the participants had been qualified as an OOD for a minimum of a year. They had all stood the watch numerous times and spent time training others to do the same. They were also all Human Systems Integration Masters' students, and had received education in human factors, safety, individual differences, leadership, and team working. The rationale for choosing these particular participants was that they had all just completed a tour in which they stood the OOD watch on a US Navy surface ship. Therefore, they had recent experience of being an OOD to draw upon. Secondly, in addition to their operational expertise, their education in human factors aided them in considering the nontechnical skill required by the OOD from a more academic/theoretical perspective.

4.2. Focus group procedure

Prior to the focus group each participant was provided the list of nontechnical skills shown in Table 1, a definition of each skill, and examples of behavioral marker systems (NOTECHS and LLC). They were asked to familiarize themselves with the material prior to meeting. It was explained that the goal of the focus group was to identify which of the skills in Table 1 were applicable to the OOD watch station, and would make an effective behavioral marker.

The group began by eliminating nontechnical skills that they felt were either redundant or not applicable to the position of the OOD. For example, the participants removed 'judgment' because they felt it was included under the 'decision making' category. 'Co-operation' was eliminated because the focus group felt that is was not applicable since the study was addressing individual instead of team skills. The participants removed other skills such as 'forehandedness' and 'energy' as it was decided that the meanings were too ambiguous.

Once the focus group had identified a list of nontechnical skills, they were asked to use these to formulate a taxonomy of categories and associated elements. The participants were encouraged to draw from their experience as an OOD, and their human factors education, to identify specific behavioral markers skills. Following the discussion, the group reached a consensus on five categories, each with two or three corresponding elements. The taxonomy that resulted from the focus group is shown in Table 2.

Table 2
Initial OOD nontechnical skills taxonomy.

Category	Element
Leadership	Establishing authority Managing workload Maintaining the standards of the watch
Decision making	Defining the problem Generating possible solutions Implementing best solution
Situational awareness	Actively gathering information Responding to changes in information Anticipating future events
Communication	Selecting correct medium Sending information clearly and concisely Effectively receiving information
Managing stress	Maintaining concentration Coping with stressors

5. Critical incident interviews

The purpose of this stage of the research was to evaluate whether the taxonomy developed by the focus group (shown in Table 2) could be used to reliably classify the nontechnical skills used by OODs in actual scenarios. It was anticipated that this process would be iterative, and the raters would need to make changed to the taxonomy in order to achieve acceptable reliability. The scenarios were developed using the critical incident technique (CIT) to generate interview data for analysis. The data was then analyzed using an adaption of the consensual qualitative methodology (Hill et al., 1997, 2005). The combination of these two techniques was used by O'Connor et al. (2007) to identify the causal factors of US Navy diving mishaps, and by O'Connor et al. (2008) to identify the team skills require by nuclear operations personnel.

5.1. Interview sample

A total of eight US Naval junior officers were interviewed. The participants were all qualified OODs and had spent a minimum of a year standing the watch on a US Navy ship. Each of the eight participants related between one and three challenging situations they had experienced while standing watch as an OOD. A total of 16 separate scenarios were collected. The interviews lasted between 25 and 60 min with an average of 45 min.

5.2. Interview procedure

The CIT interview is a task analysis method used for evaluating behavior in work environments. CIT has been widely used in studies of human error and safety (Kirwan and Ainsworth, 1992). It enables the researcher to identify the (often tacit) knowledge of skills and expertise possessed by respondents by asking them to describe a challenging incident. It goes beyond procedural knowledge by probing the behavioral aspects of experience (O'Connor et al., 2008). The CIT interviews were conducted in four sweeps by a qualified OOD knowledgeable about human factors.

Sweep 1 – Prompting the interviewee to identify a relevant incident: Each participant was asked to select and describe an event

that occurred when they were standing watch as the OOD which they found to be particularly challenging. They were asked to describe the event from their own perspective and to provide details, stage by stage as it developed.

Sweep 2 – Filling in gaps in the incident: The interviewer repeated the reported incident back to the respondent, in order to check understanding. The respondent was told they should correct any mistakes in the account or add any information that was omitted during the recounting. This sweep helps to pinpoint gaps, both in time and events, and typically aids in recall of the missing portions.

Sweep 3 – Expanding on the incident to look for cues and factors affecting teamwork: The interviewer went through the event again, this time probing at various points and asking for more detailed description of the nontechnical aspects of the situation. This sweep involved questioning the reasoning process and looking for cues and rationale for the actions taken by team members.

Sweep 4 – "What if" queries: The interviewer asks questions about the participant's perceptions, thoughts, judgments, and actions, and what would have happened if aspects of the scenario had been different. Each question is designed to extract more information that is applicable and gain a better understanding of the story as a whole.

Interview transcription. Rather than transcribe the whole interview, the recording was used to develop a single full report of the incident. The events were put into chronological order and repetitions omitted. The interviews were 'edited' into a standard format that was concise, clear and comparable across the 16 interviews. A total of 149 statements concerned with the nontechnical skills of the OOD were identified from the interview transcripts.

These statements were selected based upon the opinions of two US Naval officers who read through the interviews together, and unanimously decided whether or not a statement was concerned with the behavior of the OOD. One of the coder was a qualified OOD and a Human Systems Integration Masters student at the Naval Postgraduate School, Monterey, California. The other coder had a Ph.D in psychology with a background in human factors and the coding of interview transcripts. Statements concerned with attitudes (e.g., "the commanding officer was difficult to work for"), knowledge about the team (e.g., "everyone did not carry out the task in exactly the same way"), or technical knowledge or information (e.g., "the sailing boat was heading on a bearing of 300") were not selected.

5.3. Categorization of statements

The same two coders that identified the statements from the interview transcripts also independently categorized the statements identified from the interviews for each of the three iterations of categorization.

Iteration 1. A total of 25 randomly selected statements were independently categorized by the raters using the taxonomy developed by the focus group (see Table 2). Cohen's kappa was used to calculate the inter-rater reliability between the raters. An inter-rater reliability of $\kappa = 0.43$ resulted (moderate agreement; Landis and Koch, 1977). From the complete lack of agreement between the raters for the statements categorized by at least one of the raters using one of the 'leadership' or 'managing stress' it was concluded that these elements were too ambiguous and required clarification. There was also a lack of acceptable levels of agreement as to when to utilize the 'decision making' or 'communication' categories.

The two coders discussed the areas of agreement and disagreement, and were also sensitive to the need to ensure that system was as parsimonious as possible, with the minimal overlap between elements (Klampfer et al., 2001). On the basis of the discussion, the following changes to the taxonomy in an attempt to improve the inter-rater reliability.

- The 'managing stress' category was removed, and added as an element in the leadership category. The rationale was that 'maintaining concentration' was not a skill that could be readily observed. Also, it was decided that the ability to manage stress in oneself, and other team members, is a leadership skill.
- The leadership elements 'establishing authority' and 'maintaining the standards of the watch' were removed. It was decided that these two elements did not need to be specifically delineated, as they overlap with other elements such as 'managing workload', 'managing stress', and 'following orders and procedures.'
- The elements of the decision making category were changed to 'considering options', 'balancing risks', and 'implementing and reviewing decisions.' This terminology is closer to the decision making training that is taught to all US Naval personnel as part of Operational Risk Management (see Bellenkes, 2010 for more details).
- The elements of the communication category were changed to 'issuing orders', 'following orders and procedures', and 'providing information.' It was clear from the interview transcripts that issuing orders, providing information, and responding to direct orders are key skills of the OOD that were not addressed in the initial OOD nontechnical skills taxonomy. It was also concluded that 'effectively receiving information' was more a function of the watch stander providing the information to the OOD, and 'selecting the correct medium' was largely redundant.

Iteration 2. A random set of 25 statements (that were not coded in the first iteration) were independently analyzed using the revised taxonomy. An inter-rater reliability of $\kappa = 0.65$ resulted (substantial agreement; Landis and Koch, 1977). There was perfect agreement for the elements in the 'leadership' category, and there was acceptable agreement for the 'communication' elements. Unfortunately, there was still disagreement between the raters for the 'decision making' and 'situational awareness' elements.

After a discussion, the raters decided that more broadly defined 'decision making' elements would likely improve the inter-rater reliability. It was decided to use the three naturalistic decision making styles (analytical, rule-based, and intuitive; Klein, 2008) as the basis of the decision making elements. However, rule-based decision making was changed to 'following orders and procedures' to better fit the OOD environment. Since this element had previously been an element under communication, it was removed from the communication category, and replaced with 'receiving information'.

Iteration 3. The raters took 25 previously un-coded statements and tested the reliability of the second revision to the taxonomy. An inter-rater reliability of $\kappa = 0.70$ (substantial agreement) resulted. The raters were able to agree on the use of the 'leadership' and 'decision making' elements. However, there was some confusion between the elements from the 'communication' category and elements from the 'situational awareness' category. Specifically, it was difficult to decide whether the OOD was 'gathering information' or 'receiving information'. In the example of a lookout making a report, the OOD was certainly receiving information, but was also improving his/her situational awareness. To address this problem it was decided to eliminate the 'receiving information' element from the 'communication' category. The author also reworded two of the situational awareness elements to provide further clarification. 'Actively gathering information' was changed to 'gathering awareness', and 'responding to changes in information'

Table 3
Prototype nontechnical skills for officers of the deck (NTSOD) taxonomy with percentages of classifications attributed to each skill and element.

Skill	Element	Definition	Example	Skill %	Element % within skill	Overall element %
Leadership	Managing watch team	Effectively setting and maintaining the standards of the watch team	*The OOD utilized the dead time in the schedule to review the emergency procedures with the helmsman*	12.1	44.4	5.4
	Coping with stress	Retaining a calm demeanor when under pressure and demonstrating to the watch that one is under control	*Despite the added pressure from the XO, the OOD managed the stress and performed proficiently*		55.6	6.7
Communication	Providing information	Passing information along to other watch stations throughout the ship, as well as other assets in the area	*The OOD called the other ships in formation to inform them that the passing oiler was dimly lit and difficult to see*	28.2	33.3	9.4
	Issuing orders	Effectively giving orders to other members of the watch team and other individuals as required	*The OOD ordered the Engineering Officer of the Watch to start another engine*		66.7	18.8
Situational awareness	Gathering awareness	Actively gathering information to keep up with the changing situation	*At two nautical miles out, the OOD visually inspected the contact through binoculars*	39.6	66.1	26.2
	Understanding awareness	Achieving an understanding of what the available information means	*The OOD identified the contact as a fishing vessel by analyzing the lighting configuration*		25.4	10.1
	Anticipating future events	Forward planning in order to anticipate possible future problems	*The OOD had the Conning Officer drive slightly right of the intended track because he knew the wind and current would push the ship to the left*		8.5	3.4
Decision making	Analytical decision making	Generating and comparing multiple courses of actions to come up with the optimal solution	*Once the oiler was located, the OOD decided to start driving towards her early to ensure that there was plenty of time to set up later*	20.1	53.3	10.7
	Following orders & procedures	Following documented procedures or direct orders from superior officers	*The OOD used the wind envelope guide to make sure that the winds were sufficient to conduct flight operations*		40.0	8.1
	Intuitive decision making	Making quick decisions based upon prior experience and intuition	*When the mysterious light finally materialized as a sailboat 300 yards off the port bow, the OOD immediately ordered "Hard Right Rudder."*		6.7	1.3

was changed to 'understanding awareness' (the taxonomy is shown in Table 3).

All 149 statements were then independently classified by the raters, and a kappa of 0.91 resulted. As the level of inter-rater reliability was higher than a kappa of 0.8 (perfect agreement as defined by Landis and Koch, 1977), no further changes were made to the taxonomy. Table 3 shows the distribution of the statements across the skills and elements. For the nine statements for which the raters did not agree, the ratings of the coder who was a qualified OOD were used as he had a better understanding of the operational environment.

6. Discussion

The prototype Nontechnical Skills for Officers of the Deck (NTSOD) taxonomy, shown in Table 3, provides a structure for evaluating the nontechnical skills of OODs. However, further developmental work is required to ensure that the system can be reliably used by evaluators. Once a prototype behavioral marker system has been developed, the next stage is to get the anticipated end users to rate standardized video enactments of scenarios similar to those encountered in the actual environment (Fletcher et al., 2004; O'Connor et al., 2002; Yule et al., 2008). The purpose of this evaluation is to assess the sensitivity, inter-rater reliability, and internal reliability of the system. The scenarios are typically filmed in a simulator with subject matter experts acting the main roles.

Should an acceptable level of reliability and sensitivity be achieved, the usability of the behavioral marker system is then examined by using it to evaluate the behaviors of trainees in the actual environment (i.e. flight deck, operating theater, ship's bridge, etc.). Completed rating forms are examined to assess whether all the elements and categories are being used, and information is obtained from the raters on the usability of the behavioral marker system.

The NTSOD taxonomy not only has implications for the evaluation of the nontechnical skills of OODs, but could also be used to provide a research foundation for the content of the US Navy's bridge resource management (BRM) program. As stated by Hetherington et al. (2006) "*a review of the literature reveals that there appears to be no empirical foundation for this type of course beyond research that was originally conducted in the formation of aviation CRM* [crew resource management] *courses*" (p. 407). The US Navy's BRM program is no exception. O'Connor (in press) compared the attitudes towards, and knowledge of, the human factors that are causal to mishaps of two groups of US Naval officers – Surface Warfare Officers (SWOs, those officers that serve aboard ships) and aviators. It was found that Navy's BRM training did not have a significant effect on the attitudes and knowledge of SWOs. Moreover, naval aviators were significantly more knowledgeable, and generally held attitudes that were significantly more positive than SWOs who had attended BRM training and those that had not. O'Connor (in press) concluded that there was a need for a systematic research effort to identify the particular human factors issues that should be included as part of the Navy's BRM program.

The first stage in the development of any training program should be a needs assessment. In the context of nontechnical skills training, such as BRM/CRM, a training needs assessment is necessary to identify the skills to be trained (Flin et al., 2008). To illustrate, extensive background research was carried out to identify the seven skills that form the basis of US Naval aviation CRM training (decision making, assertiveness, mission analysis, communication, leadership, adaptability/flexibility, and situational awareness; Prince and Salas, 1993). The needs assessment ensured that the training was based upon a sound research foundation, and that limited time and resources for training were being effectively utilized.

Examining the NTSOD skills and elements that were most commonly used to categorize the interview statements provides

evidence as to the nontechnical skills that should be emphasized during OOD and BRM training. From Table 3 it can be seen that the most frequently used elements were 'issuing orders' and 'gathering information'. These two elements were used to categorize almost half of the interview statements. Therefore, assertiveness and the giving of direct and explicit orders would appear to be an important nontechnical skill that should be emphasized in the training of junior OODs. This finding is not unexpected given that the OOD is the leader of the bridge team, and must provide direction to the 13 different watch stations that report to him or her.

Gathering information was the most commonly used NTSOD element in this study – it was used to classify 26.2% of the interview statements. The failure of US Navy sailors to gather information was a key causal factor in the ramming of the Spanish freighter by the *USS Eisenhower* (NTSB, 1990). In Endsley's (1995) three stage model of situational awareness, the first stage of gathering information is where most errors occur. Jones and Endsley (1996) examined situational awareness errors in civil aviation, and Sneddon et al. (2006) examined them in offshore oil drilling. It was found that the most situational awareness errors occurred while gather information (76.3% and 67.0% respectively), as opposed to understanding the meaning of the information (20.3% and 20.0% of errors respectively), or anticipating future states (3.4% and 13.0% of errors respectively). Therefore, given the importance to the OOD of the skill of gathering information, in combination with it being where most errors occur, the OOD must be given training that emphasizes and provides practice in effectively gathering information (e.g. maintaining a good scan; see Flin et al., 2008, for more details).

Intuitive decision making was rarely used to classify the interview statements (it was only used to classify 1.3% of the statements). Part of the explanation for this may be that the operations of a US Navy ship are highly proceduralized. However, although it may be rare, the ability to make intuitive decisions is crucial for the OOD in risky fast moving non-normal situations for which there is no procedure.

7. Conclusion

Hetherington et al. (2006) stated that there are many gaps in the maritime human factors literature. We believe that the NTSOD taxonomy is a step towards filling these gaps. Although further research is required to establish the reliability and validity of the NTSOD taxonomy, it is one of the first research based behavioral markers systems developed for use in a maritime environment. It is suggested that the NTSOD taxonomy has implications for the teaching and evaluation of nontechnical skills beyond the OOD. Other military and civilian ships have individuals who stand a watch similar in scope to the US Navy OOD. Therefore, the NTSOD taxonomy provides a useful starting point to identify the nontechnical skills required by other watch standers on the bridge of military and civilian ships.

Human error can never be eliminated. However, identifying, training, and giving feedback on the nontechnical skills that are required for safe and effective performance will ensure that civilian and military sailors have the appropriate skills for minimizing, catching, and mitigating error before it leads to an mishap.

Acknowledgment

All opinions stated in this paper are those of the authors and do not necessarily represent the opinion or position of the US Navy, the Naval Postgraduate School, or the National University of Ireland, Galway.

References

Bellenkes, A., 2010. Military aviation safety policy and managing the culture of risk. In: O'Connor, P., Cohn, J. (Eds.), Human Performance Enhancements in High-Risk Environments: Insights Developments, and Future Directions from Military Research. ABC-Clio, Santa Barbara, CA, pp. 150–168.
Carter-Trahan, A., 2009. An Examination of the Human Factors Attitudes and Knowledge of Surface Warfare Officers. Masters Thesis. Naval Postgraduate School, Monterey, CA.
Chief of Naval Operations, 2005. Standard Organization and Regulations of the US Navy. OPNAVINST 3120.32C Washington, DC: Author.
Endsley, M., 1995. Toward a theory of situation awareness in dynamic systems. Human Factors 37, 32–64.
Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., Patey, R., 2004. Rating nontechnical skills: developing a behavioral marker system for use in anaesthesia. Cognition, Technology, and Work 6, 165–171.
Flin, R., Martin, L., 2001. Behavioral markers for CRM: a review of current practice. International Journal of Aviation Psychology 11, 95–118.
Flin, R., O'Connor, P., Crichton, M., 2008. Safety at the Sharp End: Training Nontechnical Skills. Ashgate Publishing Ltd., Aldershot, England.
Gatfield, D., 2005. Using simulation to determine a framework for the objective assessment of competence in maritime crisis management. Paper Presented at the Society for the Advancement of Games and Simulations in Education and Training Annual Conference, Portsmouth, England.
Helmreich, R., Wilhelm, J., Kello, J., Taggart, E., Butler, R., 1990. Reinforcing and Evaluating Crew Resource Management: Evaluator/LOS Instructor Manual. Austin: NASA/UT/FAA Aerospace Group.
Hetherington, C., Flin, R., Mearns, K., 2006. Safety in shipping: the human elements. Journal of Safety Research 37, 401–411.
Hill, C.E., Knox, S., Thompson, B.J., Williams, E.N., Hess, S.A., Landany, N., 2005. Consensual qualitative research: an update. Journal of Counselling Psychology 52 (2), 196–205.
Hill, C.E., Thompson, B.J., Williams, E.N., 1997. A guide to conducting consensual qualitative research. The Counseling Psychologist 25, 207–217.
Hollnagel, E., 1993. Human Reliability Analysis: Context and Control. Harcourt Brace, London, UK.
Jones, D., Endsley, M., 1996. Sources of situation awareness errors in aviation. Aviation Space, and Environmental Medicine 67, 507–512.
Kirwan, B., Ainsworth, L.K., 1992. A Guide to Task Analysis. Taylor & Francis, London.
Klampfer, B., Flin, R., Helmreich, R., Häusler, R., Sexton, B., Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, P., Amacher, A., 2001. Enhancing Performance in High Risk Environments: Recommendations for the use of Behavioral Markers. Report from the Behavioral Markers Workshop, Zürich, June. Berlin: Damler Benz Foundation.
Klein, G., 2008. Naturalistic decision making. Human Factors 50, 456–460.
Klein, G., 1998. Sources of Power: How People Make Decisions. MIT Press, Cambridge.
Kokotos, D.X., Linardatos, D.S., 2011. An application of data mining tools for the study of shipping safety in restricted waters. Safety Science 49, 192–197.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33 (1), 159–174.
MAIB, 2000. Annual Report 1999. Department of the Environment Transport and Regions, London.
Miles, M.B., Huberman, A.M., 1994. Qualitative Data Analysis. Sage, Thousand Oaks, CA.
Mitchell, L., Flin, R., 2009. Scrub practitioners' list of intra-operative nontechnical skills-SPLINTS. In: Flin, R., Mitchell, L. (Eds.), Safer Surgery. Ashgate Publishing Ltd., Aldershot, England, pp. 67–82.
National Transportation Safety Board, 1990. Ramming of the Spanish Bulk Carrier Urduliz by the USS Dwight D. Eisenhower (CVN 69). National Transportation Safety Board, Washington, DC.
O'Connor, P., in press. An evaluation of the effectiveness of bridge resource management training. International Journal of Aviation Psychology.
O'Connor, P., Hörmann, H.-J., Flin, R., Lodge, M., Goeters, K.M.the JARTEL group., 2002. Developing a method for evaluating CRM skills: a European perspective. International Journal of Aviation Psychology 12 (3), 263–286.
O'Connor, P., O'Dea, A., Flin, R., 2008. Identifying the team skills required by nuclear operations personnel. International Journal of Industrial Ergonomics 38, 1028–1037.
O'Connor, P., O'Dea, A., Melton, J., 2007. A methodology for identifying human error in US Navy diving accidents. Human Factors 49 (2), 214–226.
Prince, C., Salas, E., 1993. Training and research for teamwork in the military aircrew. In: Wiener, E., Kanki, B., Helmreich, R. (Eds.), Cockpit Resource Management. Academic Press, San Diego, pp. 337–366.
Reason, J., 1990. Human Error. Cambridge University Press, New York.
Rothblum, A.R., 2000. Human error and marine safety. Paper presented at the National Safety Council Congress and Expo, Orlando, FL.
Schröder-Hinrichs, J.-U., 2010. Human and organizational factors in the maritime world – are we keeping up to speed? WMU Journal of Maritime Affairs 9 (1), 1–3.
Sneddon, A., Mearns, K., Flin, R., 2006. Situation awareness and safety in offshore drill crews. Cognition, Technology, and Work 8, 255–267.
Stavridis, J., Girrier, R., 2007. Watch Officer's Guide. Naval Institute Press, Annapolis, MD.

Wagenaar, W.A., Groeneweg, J., 1987. Accidents at sea: multiple causes and impossible consequences. International Journal of Man–Machine Studies 27, 587–598.

Wiegmann, D.A., Shappell, S.A., 1999. Human error and crew resource management failures in Naval safety center data, 1990–1996. Aviation, Space and Environmental Medicine 70, 1147–1151.

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G.G., Paterson-Brown, S., 2008. Surgeons' nontechnical skills in the operating room: reliability testing of the NOTSS behaviour rating system. World Journal of Surgery 32, 548–556.

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., Rowley, D., 2006. Development of a rating system for surgeons' nontechnical skills. Medical Education 50, 1098–1104.