



A proposed quantitative methodology for the evaluation of the effectiveness of Human Element, Leadership and Management (HELM) training in the UK

F. Saeed¹ · A. Wall¹ · C. Roberts¹ · R. Riahi¹ ·
A. Bury¹

Received: 6 November 2014 / Accepted: 17 June 2016 / Published online: 12 July 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract In 2006, a review of maritime accidents found that non-technical skills (NTSs) are the single largest contributing factor towards such incidents. NTSs are composed of both interpersonal and cognitive elements. These include things such as situational awareness, teamwork, decision making, leadership, management and communication skills. In a crisis situation, good NTSs allow a deck officer to quickly recognise that a problem exists and then harness the resources that are at their disposal to safely and efficiently bring the situation back under control. This paper has two aims. The first is to develop a methodology which will enable educators to quantitatively assess the impact of Maritime and Coastguard Agency (MCA)-approved Human Element, Leadership and Management (HELM) training on deck officer's NTSs with a view to identifying further training requirements. The second is to determine whether the HELM training provided to develop the NTSs of trainee deck officers is fit for purpose. To achieve these aims, a three-phase approach was adopted. Initially, a taxonomy for deck officer's NTSs is established, behavioural markers are identified

✉ F. Saeed
F.Saeed@ljmu.ac.uk

A. Wall
A.D.Wall@ljmu.ac.uk

C. Roberts
C.C.Roberts@ljmu.ac.uk

R. Riahi
r.riahi@ljmu.ac.uk

A. Bury
A.S.Bury@ljmu.ac.uk

¹ School of Engineering, Technology and Maritime Operations, Liverpool John Moores University, James Parsons Building, Byrom Street, Liverpool L3 3AF, UK

and the relative importance of each attribute is calculated using the analytical hierarchy process (AHP). Subsequently, a set of scenarios were identified for the assessment of deck officer's NTSs in a ship bridge simulator environment. A random selection of students that have completed the Chief Mate (CM) programme was performed, and data regarding their NTS-related performance in the scenarios was collected. Finally, the collected data was fed into the evidential reasoning (ER) algorithm, utility values were produced and, having established these values, the effectiveness of the HELM training that the students have received was then evaluated.

Keywords Non-technical skills · Human element · Leadership · Management · Evidential reasoning · Analytical hierarchy process

1 Introduction

Non-technical skills (NTSs) are composed of both interpersonal and cognitive elements. These include situational awareness, teamwork, decision making, leadership, management and communication skills (Flin et al. 2003a, 2009). In a crisis situation, good NTSs allow a deck officer to quickly recognise that a problem exists and then harness the resources that are at their disposal to safely and efficiently bring the situation back under control. The ability to identify a deck officer's level of NTSs could enable maritime educators to assess the effectiveness of their Human Element, Leadership and Management (HELM) training programmes. As a result, the aim of this paper is twofold:

- To develop a methodology to enable educators to quantitatively assess the impact of Maritime and Coastguard Agency (MCA)-approved HELM training on deck officer's NTSs with a view to identifying further training requirements.
- To determine whether the HELM training provided to develop the NTSs of trainee deck officers is fit for purpose.

2 Background

A review of the maritime accident databases of the UK, the USA, Norway and Canada conducted by Barnett et al. (2006) found that human error is the main contributing factor in the majority of maritime accidents. The study illustrated that major maritime accidents are not caused by technical problems but rather by the failure of the crew to respond to situations appropriately. The main findings of the study were that (Barnett et al. 2006; Habberley et al. 2001):

- Whilst the total number of accidents is declining, human error continues to be the dominant factor in 80 to 85 % of maritime accidents.
- Human fatigue and task omission are closely related to failures in situation awareness.

Human error cannot possibly be eliminated altogether, but measures can be taken to reduce it. Analysis in a number of industrial sectors has indicated that up to 80 %

of accidents can be attributed to human factors (Gregory and Shanahan 2010; Flin et al. 2008). The Maritime Transportation Research Board, in the USA, also estimated that 80 % of accidents are due to *human error* (Goulielmos et al. 2012). Many court decisions about maritime accidents have also come to the same conclusion: “80% of all maritime accidents are due to human error” (Havold 2000). The Registro Italiano Navale (1996) suggests that the human factor is the root cause of between 60 and 80 % of serious maritime casualties. These casualties are often preceded by human error in the form of misconduct, omission, lack of competence, under-estimation of hazardous situations, lack of preparedness, lack of communication and responsibility (Goulielmos et al. 2012).

The concept of non-technical skills originated in the aviation industry when the National Transportation Safety Board of the USA investigated a number of airline accidents in the 1960s and 1970s (Smith 2001; Kern 2001). The first generation of the crew resource management (CRM) courses that were developed to address the causes of these accidents were delivered in 1981. These courses have evolved over time and have now reached their fifth generation (Smith 2001; Kern 2001). The fifth generation of CRM courses accepts that human error is inevitable. With this being the case, CRM courses are now seen as an inclusive part of the range of counter-measures that make up three lines of defence (Kern 2001):

- The avoidance of error
- The trapping of incipient errors before they are committed
- Mitigating the consequences of those errors that occur and are not trapped

Various other sectors of a safety critical nature, such as nuclear power (Crichton and Flin 2004), anaesthesia (Fletcher et al. 2003a) and surgery (Yule et al. 2006), have followed in the footsteps of the aviation industry. Each of these sectors has made their own efforts towards developing models of those non-technical skills that are deemed relevant within each of their respective domains. Within the shipping industry, the need for the training and assessment of non-technical skills was acknowledged in 2012 when the HELM training was made compulsory.

The road to this point began in 2010 when a comprehensive ongoing review of the 1978 Standards of Training, Certification and Watchkeeping for Seafarers (STCW) Convention culminated in a Conference of Parties to the STCW Convention being held in Manila. This conference adopted a significant number of amendments to the STCW Convention and the STCW Code. Amongst these amendments was the requirement to introduce mandatory training in resource management, leadership and teamwork at operational level and leadership and managerial skills at management level. These respective packages of training were to be referred to as HELM(O), the operational level course aimed at trainee officer of the watch-level students, and HELM(M), the management level course aimed at qualified officers studying for a chief mate’s licence.

3 Methodology

To achieve the aims of this paper, a three-phase approach was adopted (Fig. 1). In phase 1, a taxonomy for deck officer’s NTSs is established, behavioural markers are then identified and the relative importance of each attribute is calculated using the analytic

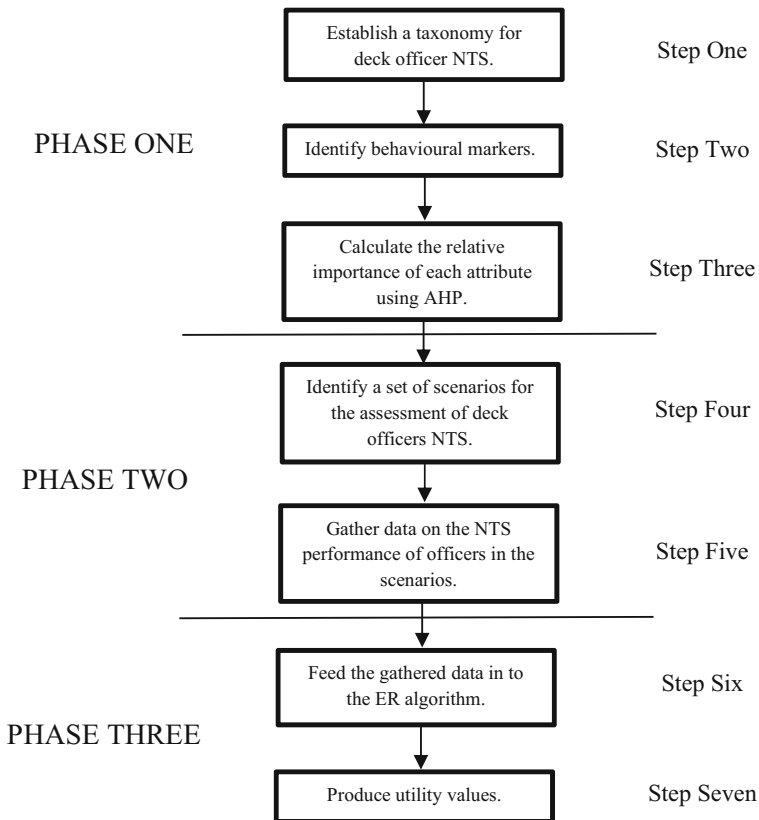


Fig. 1 A three-phase approach to assessing the effectiveness of MCA-approved HELM training (source: own)

hierarchy process (AHP). In phase 2, a set of scenarios are identified for the assessment of deck officer's NTSs in a simulated ship's bridge environment. Subsequently, a random selection of students that have completed the MCA-approved Chief Mate programme was performed and then they were observed performing their role as part of the bridge team during the identified scenarios. During this process, data is collected regarding their NTS-related performance. Finally, in phase 3, the data that was collected was fed into the evidential reasoning (ER) algorithm and utility values were subsequently produced that enable the effectiveness of the HELM training that the students have received to be evaluated.

3.1 Phase 1, step 1: establish a taxonomy for deck officer's NTSs

Based on a review of relevant literature (Flin et al. 2003a; International Association of Maritime Universities 2010; MNTB 2012) and input from experienced deck officers, a taxonomy for non-technical skills was developed. Twelve senior officers were asked to provide their input. To qualify for inclusion amongst these 12, the individuals had to have more than 10 years of experience at sea, hold an unlimited Masters licence and currently be studying towards postgraduate qualification. The relevant literature

established the boundaries within which the discussion was to take place and the responses of the senior deck officers as they were interviewed were then collated to construct a taxonomy for the issue. The taxonomy that was developed is shown in Table 1.

3.2 Phase 1, step 2: identify behavioural markers

Behavioural markers are used for the assessment of those undergoing training in a simulated environment. Such systems were first developed in the aviation industry (Helmreich et al. 1999). Later on, other safety critical sectors such as medicine and nuclear power developed their own behavioural marker systems. Any system of assessment that is based on a behavioural marker framework must, as far as possible, be designed to ensure that it is capable of capturing the fullest context of the environment in which the assessment is taking place (Gatfield 2008).

To determine which skills and behaviour are important for deck officers to have, interviews were conducted with 12 experienced senior deck officers, based in the UK and Sweden. To qualify for inclusion amongst these 12, the individuals had to have more than 10 years of experience at sea, hold an unlimited Masters licence and currently be studying towards postgraduate qualification. The same 12 officers used in step 1 were used again here in step 2. The purpose of this was to determine, from the perspective of experienced senior deck officers, the most significant elements of deck officer's non-technical skills. Based on the output of these interviews, a series of behavioural markers for assessing the teamwork, leadership, situational awareness and decision making of an officer in a ship bridge simulator were identified. These

Table 1 Taxonomy for non-technical skills amongst deck officers

Criteria	Sub-criteria
Teamwork	Team building and maintenance Considering others Supporting others Communication Information sharing
Leadership and management	Authority and assertiveness Providing and maintaining standards Planning and coordination Workload management Task prioritisation Task delegation Initial crisis management
Situational awareness	Awareness of bridge systems Awareness of external environment Awareness of time Assessment of situation
Decision making	Problem definition and diagnosis Option generation Risk assessment and option selection Outcome review

Source: own (amended from Flin et al. 2003a)

markers are demonstrated in Tables 2, 3, 4 and 5. There are five levels of performance against each of these behavioural markers ranging from very good practice to very poor practice. By using these behavioural markers, an assessor is able to rate a student's performance in a simulated ship's bridge environment.

3.2.1 *Teamwork*

The need for people to work together as a team and to achieve objectives which contribute to the overall aims of their organisation has become increasingly important as organisations have grown in size and become more complex (West 2012; Cohen and Bailey 1997). Organisations with team-based structures can respond quickly and effectively in the modern fast-changing environment (Cohen and Bailey 1997). Teamwork is important in many workplace settings but is especially important in higher-risk industries such as aviation, nuclear power, healthcare and maritime.

To be able to achieve its goals, a team must function effectively from the moment it is established. In the case of a ship's bridge team, team members must have a common understanding of how they will be expected to work together to manoeuvre their ship (Civil Aviation Authority (CAA) 2016). The effective operation of such a team is highly dependent on the team's ability to perform a range of skills. These include, but are not limited to, communication, coordination, cooperation and control (Stanton 1996). The selected teamwork sub-criteria along with their associated behavioural markers are shown in Table 2.

3.2.2 *Leadership and management*

Fiedler (1995) defines a team leader as "a person who is appointed, elected, or informally chosen to direct and co-ordinate the work of others in a group". Leadership is about encouraging team members to work together, assigning them tasks and assessing their performance, developing the knowledge base of the team as a whole, improving team members' skills and abilities, continuously motivating team members, planning and organising the execution of tasks and establishing a positive team atmosphere (Salas et al. 2004). The selected leadership and management sub-criteria along with their associated behavioural markers are shown in Table 3.

3.2.3 *Situational awareness*

Endsley (1995) defines situational awareness as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future". It has been widely established and accepted that situational awareness is a contributory factor to many accidents and incidents in safety critical industries (Grech et al. 2002). However, situational awareness only began to receive attention in the late 1980s when related research started in the aviation industry (Salmon et al. 2009). The selected situation awareness sub-criteria along with their associated behavioural markers are shown in Table 4.

Table 2 Behavioural markers of teamwork sub-criteria

Sub-criteria	Very good practice	Good practice	Acceptable practice	Poor practice	Very poor practice
Team building	Fully encourages input and feedback from others	Significantly encourages input and feedback from others	Moderately encourages input and feedback from others	Little encouragement of input and feedback from others	No encouragement of input and feedback from others
Considering others	Takes full notice of the suggestions of other team members	Takes significant notice of the suggestions of other team members	Takes moderate notice of the suggestions of other team members	Takes little notice of the suggestions of other team members	Takes no notice of the suggestions of other team members
Supporting others	Shows full consideration for the condition of other team members	Shows significant consideration for the condition of other team members	Shows moderate consideration for the condition of other team members	Shows little consideration for the condition of other team members	Shows no consideration for the condition of other team members
Communication	Provides detailed personal feedback to other team members	Provides significant personal feedback to other team members	Provides moderate personal feedback to other team members	Provides little personal feedback to other team members	Provides no personal feedback to other team members
Information sharing	Provides full help to other team members in demanding situations	Provides significant help to other team members in demanding situations	Provides moderate help to other team members in demanding situations	Provides little help to other team members in demanding situations	Provides no help to other team members in demanding situations
	Establishes an atmosphere that allows completely open communication	Establishes an atmosphere that allows significantly open communication	Establishes an atmosphere that allows moderately open communication	Establishes an atmosphere that allows limited communication	Establishes an atmosphere that allows no communication
	Communicates with exceptional effectiveness as an individual	Communicates with significant effectiveness as an individual	Communicates with moderate effectiveness as an individual	Communicates with limited effectiveness as an individual	Communicates with no effectiveness as an individual
	Full sharing of information between team members	Significant sharing of information between team members	Moderate sharing of information between team members	Little sharing of information between team members	No sharing of information between team members

Source: own (amended from Flin et al. 2003a)

Table 3 Behavioural markers of the leadership sub-criteria

Sub-criteria	Very good practice	Good practice	Acceptable practice	Poor practice	Very poor practice
Authority and assertiveness	Full initiative is used to ensure crew involvement and task completion Willing to take full control if the situation requires it	Significant initiative is used to ensure crew involvement and task completion Willing to take significant control if the situation requires it	Moderate initiative is used to ensure crew involvement and task completion Willing to take moderate control if the situation requires it	Little initiative is used to ensure crew involvement and task completion Willing to take limited control if the situation requires it	No initiative is used to ensure crew involvement and task completion Not willing to take control, even if the situation requires it
Providing and maintaining standards	Always considers the suggestions of team members Completely committed to achieving top performance	Frequently considers the suggestions of team members Significant commitment to achieving top performance	Infrequently considers the suggestions of team members Moderate commitment to achieving top performance	Considers the suggestions of team members on very limited occasions Limited commitment to achieving top performance	Ignores the suggestions of team members Not at all committed to achieving top performance
Planning and coordination	Encourages complete crew participation in planning and task completion Plan is exceptionally clearly stated and confirmed with team members	Encourages significant crew participation in planning and task completion Plan is clearly stated and confirmed with team members	Encourages moderate crew participation in planning and task completion Plan is moderately clearly stated and confirmed with team members	Encourages limited crew participation in planning and task completion Plan is not clearly stated but is confirmed with team members	Does not encourage crew participation in planning and task completion Plan is not clearly stated or confirmed with team members
Workload management	States the goals and boundaries for task completion with exceptional clarity Shows complete notice of the signs of stress and fatigue Allows sufficient time for task completion with a sufficient amount for contingencies	Clearly states the goals and boundaries for task completion Shows significant notice of the signs of stress and fatigue Allows sufficient time for task completion with an insufficient amount for contingencies	Adequately states the goals and boundaries for task completion Shows moderate notice of the signs of stress and fatigue Allows sufficient time for task completion	States the goals and boundaries for task completion but in a limited fashion Shows limited notice of the signs of stress and fatigue Allows marginally insufficient time for task completion	Does not state the goals and boundaries for task completion Does not notice signs of stress or fatigue Allows obviously insufficient time for task completion

Table 3 (continued)

Sub-criteria	Very good practice	Good practice	Acceptable practice	Poor practice	Very poor practice
Task prioritisation	Demonstrates complete prioritisation of tasks	Demonstrate significant prioritisation of tasks	Demonstrates moderate prioritisation of tasks	Demonstrate limited prioritisation of tasks	Demonstrates no prioritisation of tasks
Task delegation	Delegates all necessary tasks in a good manner	Delegates all necessary tasks in an acceptable manner	Delegates a limited number of tasks but in an acceptable manner	Delegates a limited number of tasks but in a poor manner	Does not delegate tasks
Initial crisis management	Identifies crisis situation very quickly and responds accordingly	Identifies crisis situation quickly and responds accordingly	Identifies crisis situation after some time and responds accordingly	Identifies crisis situation after some time but responds inappropriately	Does not identify the crisis situation and responds inappropriately

Source: own (amended from Flin et al. 2003a)

Table 4 Behavioural markers of the situational awareness elements

Sub-criteria	Very good practice	Good practice	Acceptable practice	Poor Practice	Very poor practice
Awareness of bridge systems	Fully monitors and reports all changes in the systems status	Significantly monitors the system and reports the majority of changes in the systems status	Moderately monitors the system and reports some changes in the systems status	Monitors the system in a limited way and reports few changes in the systems status	Does not monitor the system and does not report any changes in the systems status
Awareness of external environment	Collects a full range of information about the environment (e.g. own ship's position, traffic and weather)	Collects a significant range of information about the environment (e.g. own ship's position, traffic and weather)	Collects a moderate range of information about the environment (e.g. own ship's position, traffic and weather)	Collects a limited range of information about the environment (e.g. own ship's position, traffic and weather)	Does not collect any information about the environment (e.g. own ship's position, traffic and weather)
Awareness of time	Shares all key information about the environment with team members	Shares significant sufficient amounts of key information about the environment with team members	Shares a moderate amount of key information about the environment with team members	Shares a limited amount of key information about the environment with team members	Shares no key information about the environment with team members
Assessment of situation	Has a full discussion about time constraints with other team members	Has a significant discussion about time constraints with other team members	Has a moderate discussion about time constraints with other team members	Has a brief discussion about time constraints with other team members	Does not discuss time constraints with other team members
	Makes a complete attempt to assess the changing situation	Makes a significant attempt to assess the changing situation	Makes a moderate attempt to assess the changing situation	Makes a limited attempt to assess the changing situation	Makes no attempt to assess the changing situation

Source: own (amended from Flin et al. 2003a)

Table 5 Behavioural markers of the decision making elements

Sub-criteria	Very good practice	Good practice	Acceptable practice	Poor practice	Very poor practice
Problem definition and diagnosis	Gathers all information to identify the problem Full discussion of probable causal factors with team members	Gathers a significant amount of information to identify the problem Significant discussion of probable causal factors with team members	Gathers a moderate amount of information to identify the problem Moderate discussion of probable causal factors with team members	Gathers a limited amount of information to identify the problem Limited discussion of probable causal factors with team members	Gathers no information to identify the problem No discussion of probable causal factors with team members
Option generation	States all alternative options Asks team members for a wide range of options	States a significant number of alternative options Asks crew members for a significant number of options	States a moderate number of alternative options Asks crew members for a moderate number of options	States a limited number of alternative options Asks crew members for a limited number of options	Does not state any alternative options Does not ask team members for alternatives
Risk assessment and option selection	In-depth discussion of limiting factors with team members States and confirms all selected options	Significant discussion of limiting factors with team members States and confirms a significant number of selected options	Moderate discussion of limiting factors with team members States and confirms a moderate number of selected options	Inadequate discussion of limiting factors with team members States and confirms a limited number of selected options	No discussion of limiting factors with team members Does not state and/or confirm any selected options
Outcome review	Complete checking of the plan outcome against the plan	Significant checking of the plan outcome against the plan	Moderate checking of the plan outcome against the plan	Limited checking of the plan outcome against the plan	Does not check the selected outcome against the plan

Source: own (amended from Flin et al. 2003a)

3.2.4 Decision making

Good decision making is an essential skill for the completion of operations in any high-risk organisation. In aviation, decision making is defined as “The process of reaching a judgement or choosing an option” (Flin et al. 2003b). Although this definition is labelled as aeronautical decision making, it may also be taken as a universal definition for all high-risk industries. Like an aeroplane pilot, a ship’s master also has to make different types of decisions for different situations. The selected decision making sub-criteria along with their associated behavioural markers are shown in Table 5.

3.3 Phase 1, step 3: calculate the relative importance of each attribute using AHP

The behavioural markers identified from interviewing 12 experienced senior deck officers, based in the UK and Sweden, allowed a non-technical skills assessment framework for deck officers to be established (Fig. 2). The next stage of the assessment process is to assign a weight to each criterion in the framework by applying a mathematical decision making method known as analytic hierarchy process (AHP).

3.3.1 Analytic hierarchy process

The AHP was pioneered by Saaty (1980). It is a popular method that is widely used in decision making tasks. Although it has been used in military decision making, it is not restricted to military problems (Coyle 2004). It is a multi-criteria decision making (MCDM) method that helps the decision maker deliver an informed decision when dealing with a complex situation (Ishizaka and Labib 2009). To make a comparison between the available alternatives, numbers (i.e. intensity of importance) are assigned by experts. The intensity of importance indicates how many times more important one element is over another element. The scale for comparison is based on Saaty (1990, 2008) and is shown in Table 6. When considering a number of attributes for evaluation, the main objective of the technique is to provide judgements on the relative importance of each of these attributes to each other. Once this is accomplished, it is then necessary to ensure that the judgements are quantified to an extent that permits their quantitative interpretation (Pillay and Wang 2003).

Riahi et al. (2012) state that the quantified judgements between pairs of attributes (A_i and A_j) can be represented by an n -by- n matrix (D). The entries a_{ij} are defined by the following entry rules:

Rule 1: If $a_{ij} = \alpha$, then $a_{ji} = 1 / \alpha$, $\alpha \neq 0$.

Rule 2: If A_i is judged to be of equal relative importance as A_j , then $a_{ij} = a_{ji} = 1$.

$$D = \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{bmatrix} \quad (1)$$

Fig. 2 Assessment framework for deck officer's non-technical skills (source: own)

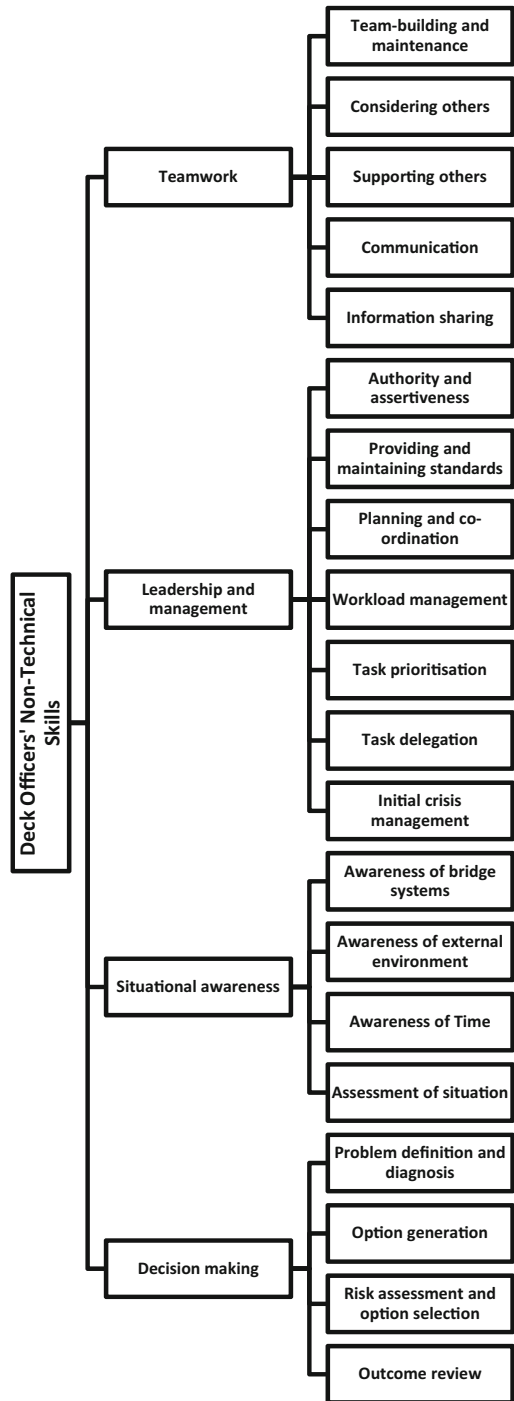


Table 6 Pairwise comparison scale

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
2	Weak or slight	
3	Moderate importance	Experience and judgement slightly favour one activity over another
4	Moderate plus	
5	strong importance	Experience and judgement strongly favour one activity over another
6	Strong plus	
7	Very strong	Experience and judgement very strongly favours one activity over another; its dominance being demonstrated in practice
8	Very, very strong	
9	Extreme importance	Experience and judgement favour one activity over another in a way that affirms its dominance in the highest possible order

Source: Riahi et al. (2012)

In this formula, $i, j = 1, 2, 3, \dots, n$ and each a_{ij} is the relative importance of attribute A_i to attribute A_j .

Having recorded the quantified judgments of comparison on pair (A_i, A_j) as the numerical entry a_{ij} in the matrix D , what is left is assigned to the n contingencies (A_1, A_2, \dots, A_n) , a set of numerical weights (w_1, w_2, \dots, w_n) that should reflect the recorded judgements. Generally weights w_1, w_2, \dots, w_n can be calculated by using the following equation, where a_{ij} represents the entry of row i and column j in a comparison matrix of order n :

$$\omega_k = \frac{1}{n} \sum_{j=1}^n \frac{a_{kj}}{\sum_{i=1}^n a_{ij}} \quad (k = 1, 2, 3, \dots, n) \quad (2)$$

The weight vector of the comparison matrix will provide the priority order, but it cannot confirm the consistency of the pairwise judgement. The AHP provides a measure of the consistency of the pairwise comparisons by computing a consistency ratio (CR) (Aull-Hyde et al. 2006). The CR is devised in such a way that a value of less than 0.10 is deemed consistent in the pairwise judgement. A decision maker should review the pairwise judgements if the resultant value is more than 0.10. The CR value is calculated according to the following equations where CI is the consistency index, RI is the average random index (Table 7), n is the matrix order and λ_{\max} is the maximum weight value of the n -by- n comparison matrix D :

$$CR = \frac{CI}{RI} \quad (3)$$

Table 7 Value of RI versus matrix order (Endsley 1995)

Number	RI
1	0
2	0
3	0.58
4	0.9
5	1.12
6	1.24
7	1.32
8	1.41
9	1.45
10	1.49

Source: Riahi et al. (2012)

$$CI = \frac{\lambda_{\max} - n}{n - 1} \tag{4}$$

$$\lambda_{\max} = \frac{\sum_{j=1}^n \left[\left(\sum_{k=1}^n w_k a_{jk} \right) / w_j \right]}{n} \tag{5}$$

3.3.2 Geometric mean method

Expert judgements were aggregated using the geometric mean method. The formula below [28] demonstrates this method. In this formula, e_{kij} is the k th expert judgement on pair of attributes A_i and A_j :

$$\text{Geometric mean}_{ij} = [e_{1ij}, e_{2ij}, e_{3ij}, \dots, e_{kij}]^{\frac{1}{k}} \tag{6}$$

3.4 Phase 2, step 4: identify a set of scenarios for the assessment of deck officer’s NTS

It was necessary to produce scenarios that would give participants the opportunity to demonstrate the recognised elements (teamwork, situational awareness, leadership and decision making) of their NTSs in a ship’s bridge simulator. With this as the aim, the following three scenarios were generated:

Scenario 1 The team’s vessel (own vessel) is alongside a jetty in Southampton. Sufficient time will be given prior to the commencement of the exercise for those involved to be able to check all the equipment and familiarise themselves with the pre-prepared passage plan. The bridge equipment needs to be tested and the relevant

checklists need to be completed prior to the exercise commencing. The team will pilot their own vessel and maintain all the records as required. Each team will independently manoeuvre their own vessel without the use of tug boats. However, the bow thruster can be used if the team feels that it is necessary. During the exercise, the team's own vessel will encounter a number of other vessels that are either inbound to or outbound from Southampton. There will be a grounded vessel in the vicinity of the Nab Tower with a salvage operation underway that requests a wide berth. Whilst passing the Fawley Terminal, gyro number 1 will begin to drift at a rate of 1° per s. Based on the position of the vessels at the time of passing, there will also be the possibility of interaction with a large inbound containership.

Scenario 2 This exercise is set in the approaches to the Bosphorus, Turkey. Sufficient time will be given prior to the commencement of the exercise for those involved to be able to check all the equipment and familiarise themselves with the pre-prepared passage plan. The Master begins the exercise in the debriefing room and is available to be called to the bridge as required. The bridge equipment needs to be tested, and the relevant checklists need to be completed prior to the exercise commencing. The vessel is to proceed to an anchorage to begin bunkering operations. There will be a number of vessels in the area concerned. These will range from vessels at anchor, vessels approaching the team's vessel from different directions, overtaking and numerous ferries crossing. A strong tide is present that makes it more difficult to steer. As time progresses, the vessel will go on to proceed through the Bosphorus towards the Black Sea. At this stage, there will be a number of south-bound vessels, strong cross currents and ferry operations. All of these will require strict adherence to the passage plan as well as manoeuvring to avoid collision with other ships.

Scenario 3 This exercise starts at the handover of a watch. Sufficient time will be given prior to the commencement of the exercise for those involved to be able to check all the equipment and familiarise themselves with the pre-prepared passage plan. The bridge equipment needs to be tested and the relevant checklists need to be completed prior to the exercise commencing. The Third Officer (3/O) commences the exercise as the Officer of the Watch (OOW). The instructor will act as the lookout and is available on a walkie-talkie/telephone. An individual from each of the three bridges will be playing the role of the Second Officer and will proceed to the bridge to take over the watch. The handover will take place using the appropriate procedure and checklists. When the watch has been handed over, the relieved OOW will return to the debrief room and will take up the role of Chief Mate (CM). Bridge 1 will be situated in such a way that in the initial 20 min of running the exercise, it will be just in a visual range of a target showing a strobe light (normally fitted on a life raft). This target will have a very poor radar return and will only be detected if the radar controls are set appropriately. Another target to the north (a vessel in distress) will provide a weak radar return but will not be within visual range. Bridges 2 and 3 will be within very high frequency (VHF) range of bridge 1. The scenario is set in Global Maritime Distress and Safety System (GMDSS) sea area A2 (medium frequency range). Each bridge will have a Master and CM available to assist the OOW if requested. The exercise will progress in the anticipation that the OOW on bridge 1 will identify the life raft, summon the master and instigate a search and rescue plan. In the event that the OOW does not take the appropriate action,

the virtual lookout (the instructor) will call the bridge and report the sighting. The exercise will then be conducted in line with the delegates' responses. One of the bridges will be tasked as being the On Scene Commander (OSC), and there will be a minimum intervention from the Maritime Rescue Coordination Centre (MRCC). There will be other vessels in the area. These consist of the following:

- A warship to the west that is in medium frequency (MF) range and has an operational helicopter.
- A fishing vessel to the north that will offer assistance and has the benefit of a low freeboard. Her position is such that the Emergency Position Indicating Radio Beacon (EPIRB) position of the causality can be reached if she is utilised immediately.

3.5 Phase 2, step 5: gather data on the NTS performance of officers in the scenarios

Participants were randomly selected from a cohort of students that had completed the CM training programme. They were then divided into two groups. The first group had not yet received non-technical skills training and the second group had received non-technical skills training through the MCA-approved HELM course. Within these groups, they were then divided in sub-groups which would make up their bridge teams. Based on the bridge simulator scenarios detailed in step 4, the qualitative characteristics (teamwork, leadership and management, situational awareness, decision making) of each participant were assessed using the identified behavioural markers (Tables 2, 3, 4 and 5). Data was gathered on their performance in each of the scenarios. The data gathered was then aggregated using the ER algorithm.

3.6 Phase 3, step 6: feed the gathered data into the ER algorithm

The theory of evidence was first generated by Dempster (1967) and further developed by Shafer (1976). After the further development, the theory came to be referred to as the Dempster-Shafer (D-S) theory of evidence. The D-S theory was originally used as an approximation tool for information aggregation in expert systems. Subsequently, it also came to be used in decision making under uncertainty. Through the ongoing evolution of the theory of evidence the ER algorithm was developed. The ER algorithm can be explained as follows (Riahi et al. 2012):

Let R represent a set with five linguistic terms (very poor, poor, average, good and very good). The associated belief degrees (β) can be synthesised by two subsets R_1 and R_2 from two different assessments. Then, for example, R , R_1 and R_2 can separately be expressed by

$$\begin{aligned} R &= \{\beta^1 \text{ very poor}, \beta^2 \text{ poor}, \beta^3 \text{ average}, \beta^4 \text{ good}, \beta^5 \text{ very good}\} \\ R_1 &= \{\beta_1^1 \text{ very poor}, \beta_1^2 \text{ poor}, \beta_1^3 \text{ average}, \beta_1^4 \text{ good}, \beta_1^5 \text{ very good}\} \\ R_2 &= \{\beta_2^1 \text{ very poor}, \beta_2^2 \text{ poor}, \beta_2^3 \text{ average}, \beta_2^4 \text{ good}, \beta_2^5 \text{ very good}\} \end{aligned} \quad (7)$$

Suppose that the normalised relative weights of two assessments in the evaluation process are given as w_1 and w_2 ($w_1 + w_2 = 1$). Then, w_1 and w_2 can be estimated by using the AHP. Suppose that M_1^m and M_2^m ($m = 1, 2, 3, 4$ or 5) are individual degrees to which the subsets R_1 and R_2 support the hypothesis that the evaluation is confirmed to the five linguistic terms. Then, M_1^m and M_2^m are obtained as

$$\begin{aligned} M_1^m &= w_1 \beta_1^m \\ M_2^m &= w_2 \beta_2^m \end{aligned} \tag{8}$$

Suppose that H_1 and H_2 are the individual remaining belief values unassigned for M_1^m and M_2^m ($m = 1, 2, 3, 4$ or 5). Then, H_1 and H_2 are expressed as

$$\begin{aligned} H_1 &= H_1 + \tilde{H}_1 \\ H_2 &= H_2 + \tilde{H}_2 \end{aligned} \tag{9}$$

where \bar{H}_n ($n = 1$ or 2) represents the degree to which the other assessors can play a role in the assessment and H_n ($n = 1$ or 2) is caused by the possible incompleteness in the subsets R_1 and R_2 . \bar{H}_n ($n = 1$ or 2) and H_n ($n = 1$ or 2) are described as

$$\begin{aligned} H_1 &= 1 - w_1 = w_2 \\ H_2 &= 1 - w_2 = w_1 \\ \tilde{H}_1 &= w_1 \left(1 - \sum_{m=1}^5 \beta_1^m \right) \\ \tilde{H}_2 &= w_2 \left(1 - \sum_{m=1}^5 \beta_2^m \right) \end{aligned} \tag{10}$$

Suppose that $\beta^{m'}$ ($m = 1, 2, 3, 4$ or 5) represents the non-normalised degree to which the reliability evaluation is confirmed to each of the five linguistic terms as a result of the synthesis of the judgements produced by assessors 1 and 2. Suppose that H'_U represents the non-normalised remaining belief unassigned after the commitment of belief to the five linguistic terms because of the synthesis of the judgements produced by assessors 1 and 2. The evidential reasoning algorithm is stated as

$$\begin{aligned} \beta^{m'} &= K(M_1^m M_2^m + M_1^m H_2 + M_2^m H_1) \\ H'_U &= K(H_1 H_2) \\ \tilde{H}'_U &= K(\tilde{H}_1 \tilde{H}_2 + \tilde{H}_1 H_2 + \tilde{H}_2 H_1) \\ K &= \left(1 - \sum_{T=1}^5 \sum_{\substack{R=1 \\ R \neq T}}^5 M_1^T M_2^R \right)^{-1} \end{aligned} \tag{11}$$

After aggregation, the combined degrees of belief are generated by assigning values of \overline{H}'_U to the five linguistic terms using the normalisation process. Within this formula, H_U is the unassigned degree of belief representing the extent of incompleteness in the overall assessment

$$\beta^m = \frac{\beta^{m'}}{1-H'_U} \quad (m = 1, 2, 3, 4 \text{ or } 5)$$

$$H_U = \frac{\widetilde{H}'_U}{1-H'_U} \tag{12}$$

The above gives the process of combining two subsets. If three subsets are required to be combined, the result obtained from the combination of any two subsets can be further synthesised with the third subset using the above algorithm. In a similar way, the judgements of multiple assessors of lower-level criteria in the chain system (i.e. components or subsystems) can be combined.

3.7 Phase 3, step 7: produce utility values

The main aim of using a utility approach is to obtain a single crisp number for the final output result or goal in order to rank them. Let the utility of an evaluation grade H_n be denoted by $u(H_n)$ and $u(H_{n+1}) > u(H_n)$ if H_{n+1} is preferred to H_n ; $u(H_n)$ can be estimated using the decision marker’s preferences. If no preference information is available, it could be assumed that the utilities of evaluation grades are equidistantly distributed in a normalised utility space. The utilities of evaluation grades that are equidistantly distributed in a normalised utility space are calculated as follows (Salmon et al. 2009):

$$\mu(H_n) = \frac{V_n - V_{\min}}{V_{\max} - V_{\min}} \tag{13}$$

where V_n is the ranking value of the linguistic term H_n that has been considered, V_{\max} is the ranking value of the most-preferred linguistic term H_N and V_{\min} is the ranking value of the least-preferred linguistic term H_1 .

The utility of the top level or general criterion $S(E)$ is denoted by $u(S(E))$. If $\beta_H \neq 0$ (i.e. the assessment is incomplete, $\beta_H = 1 - \sum_{n=1}^N \beta_n$), there is a belief interval $[\beta_n, (\beta_n + \beta_H)]$, which provides the likelihood that $S(E)$ is assessed to H_n . Without loss of generality, suppose that the least-preferred linguistic term having the lowest utility is denoted by $u(H_l)$ and the most-preferred linguistic term having the highest utility is denoted by $u(H_N)$. Then, the minimum, maximum and average utilities are defined as follows (Riahi et al. 2012):

$$u_{\min}(S(E)) = \sum_{n=2}^{N-1} \beta_n u(H_n) + (\beta_l + \beta_H) u(H_l)$$

$$u_{\max}(S(E)) = \sum_{n=1}^{N-1} \beta_n u(H_n) + (\beta_N + \beta_H) u(H_N)$$

$$u_{\text{average}}(S(E)) = \frac{u_{\min}(S(E)) + u_{\max}(S(E))}{2} \tag{14}$$

If all the assessments are complete, then $\beta_H = 0$ and the maximum, minimum and average utilities of $S(E)$ will be the same. Therefore, $u(S(E))$ can be calculated as

$$u(S(E)) = \sum_{n=1}^N \beta_n u(H_n) \quad (15)$$

It is perhaps worth mentioning that the above utilities are used only for characterising an assessment and not for criteria aggregation (Riahi et al. 2012).

4 Results

A three-phase approach was adopted to address the issue of whether the existing MCA-approved HELM training is fit for purpose. Each phase delivered its own set of results that were fed into the subsequent phase. The results generated in each of these phases were as follows:

4.1 Phase 1: analytic hierarchy process

Data was collected from 12 experienced management level seafarers. To qualify for inclusion amongst these 12, the individuals had to have more than 10 years of experience at sea, hold an unlimited Masters licence and currently be studying towards a postgraduate qualification. The results obtained from four of these experts were considered to be inconsistent (CR was greater than 0.1), so they were not included in the subsequent calculations. The consistent results from the remaining eight experts were fully utilised in the AHP. The resultant weights of the criteria and sub-criteria can be seen in Table 8.

4.2 Phase 2: scenarios

Based on the scenarios that were developed and the established behavioural marker assessment framework, observations were made of the performance of those involved in each scenario. After conducting extensive simulator observations, enough data was collected to be fed into the ER algorithm.

4.3 Phase 3: evidential reasoning

Through utilising the evidential reasoning algorithm and the utility approach, the data collected from observations made during the scenarios was used to make a comparison between the average performance of groups with the HELM training and the average performance of groups without the HELM training. To do this, the utility values were calculated and the mean for each group was identified to allow it to be compared against the average group performance. Table 9 shows the utility values of the groups with the HELM training and those without. Table 10 shows the rankwise sequencing of all the groups. Table 11 compares the average utility value of those groups with HELM training against those groups without.

Table 8 Weights of the criteria and sub-criteria

Criteria	Sub-criteria
Teamwork (0.1913)	Team-building and maintenance (0.2066)
	Considering others (0.1860)
	Supporting others (0.1831)
	Communication (0.2436)
	Information sharing (0.1807)
Leadership and management (0.2878)	Authority and assertiveness (0.1579)
	Providing and maintaining standards (0.0857)
	Planning and coordination (0.1437)
	Workload management (0.1280)
	Task prioritisation (0.1255)
	Task delegation (0.1316)
	Initial crisis management (0.2276)
Situational awareness (0.2863)	Awareness of bridge systems (0.2433)
	Awareness of external environment (0.2375)
	Awareness of time (0.1860)
	Assessment of situation (0.3332)
Decision making (0.2346)	Problem definition and diagnosis (0.2447)
	Option generation (0.2069)
	Risk assessment and option selection (0.2426)
	Outcome review (0.3058)

Table 10 shows that there is no relationship between the position of a group within the order of ranking and whether that group has received HELM training. This is best

Table 9 Utility values of the groups with HELM training and those without

Group number	Utility values (%)
Without HELM training	
1	0.2459 (24.59)
2	0.2724 (27.24)
3	0.3917 (39.17)
4	0.1459 (14.59)
5	0.3487 (34.87)
6	0.5409 (54.09)
With HELM training	
7	0.4751 (47.51)
8	0.1501 (15.01)
9	0.2797 (27.97)
10	0.3888 (38.88)
11	0.4423 (44.23)
12	0.2576 (25.76)

Table 10 Rankwise sequencing of all the groups

Rank	Group number	Training	Utility value (%)
1	6	Without HELM	0.5409 (54.09)
2	7	With HELM	0.4751 (47.51)
3	11	With HELM	0.4423 (44.23)
4	3	Without HELM	0.3917 (39.17)
5	10	With HELM	0.3888 (38.88)
6	5	Without HELM	0.3487 (34.87)
7	9	With HELM	0.2797 (27.97)
8	2	Without HELM	0.2724 (27.24)
9	12	With HELM	0.2576 (25.76)
10	1	Without HELM	0.2459 (24.59)
11	8	With HELM	0.1501 (15.01)
12	4	Without HELM	0.1459 (14.59)

exemplified by group number 6, a group which has not received HELM training, being ranked as the group with the greatest utility value.

Table 11 shows that the average utility value of groups with the HELM training is only 0.8 % greater than the average utility value of groups without the HELM training. It was evident during the observations that the students with the HELM training did not apply the non-technical skills that were taught during the course. Generally, students were found to be especially weak in situational awareness and decision making. Lack of anticipation resulted in poor decisions. In some instances, task delegation was not clear and this resulted in task omissions.

5 Conclusions

This paper had two aims. The first was to develop a methodology to enable educators to quantitatively assess the impact of MCA-approved HELM training on deck officer's NTSs with a view to identifying further training requirements. To achieve this aim, a three-phase approach was adopted where initially a taxonomy for deck officer's NTSs was established and then each attribute was allocated a weight through the AHP. Subsequently, a set of scenarios were identified for the assessment of deck officer's NTSs in a simulated ship's bridge environment. Finally, the data collected through these scenarios were fed into the ER algorithm and the utility values were produced. In this regard, a methodology has been devised and tested to demonstrate its ability to assess the NTSs of individuals operating as part of a bridge team in a simulated bridge

Table 11 A comparison of the average utility value of the groups with HELM training against those without

	Utility value
Without HELM training	32.4
With HELM training	33.2

environment. As a result, it can be considered that the first objective has been successfully met.

The second aim of this paper was to establish whether the HELM training provided to deck officers is fit for purpose. It can clearly be seen that the results derived from the methodology established in this paper show that the average utility value of the groups with HELM training is higher than the average utility value of the groups without HELM training. However, with only a 0.8 % difference between the values, this is such a small margin that it is misleading to state that HELM training clearly benefits all those that participate in it. With this being the case, it suggests that the current HELM training course is an ineffective method for improving all deck officer's NTS. However, a larger data set would be required to enable a conclusive result to be arrived at.

The conclusion that the present HELM course is ineffective is based on simulator scenarios performed by CM students only. It is possible that this may not be a representative population to draw such a conclusion from.

Acknowledgments The authors would like to acknowledge the participation of those who undertook the simulated scenarios and the support of the International Association of Maritime Universities (IAMU) and The Nippon Foundation in Japan.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Aull-Hyde R, Erdogan S, Duke JM (2006) An experiment on the consistency of aggregated comparison matrices in AHP. *Eur J Oper Res* 171(1):290–295
- Barnett, M., Gatfield, D., Peckan, C. (2006) "Non-technical skills: the vital ingredient in world maritime technology", London; Institute of Marine Engineering, Science and Technology (IMarEST)
- Civil Aviation Authority (CAA) "Crew Resource Management Training" <https://www.caa.co.uk/Safety-initiatives-and-resources/Working-with-industry/Human-factors/Human-factors/> [Accessed 3 Feb. 2016]
- Cohen SG, Bailey DE (1997) What makes teams work, group effectiveness research from the shop floor to the executive suite. *J Manag* 23(3):239–290
- Coyle, G. (2004) "The analytic hierarchy process (AHP)", practical strategy, Open Access Material AHP, Pearson Education Limited
- Crichton MT, Flin R (2004) Identifying and training non-technical skills of nuclear emergency response teams. *Ann Nucl Energy* 31:1317–1330
- Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat* 38(2): 325–339
- Endsley MR (1995) Towards a theory of situation awareness in dynamic systems. *Hum Factors* 37(1):32–64
- Fiedler FE (1995) Cognitive resources and leadership performance. *Appl Psychol* 44(1):5–28
- Fletcher, G., Flin, R and McGeorge, P. (2003a) "Review of human factors research in anaesthesia", University of Aberdeen SCPMDE Project: RDNES/991/C
- Flin R, Martin L, Geosters K, Hoermann J, Amalberti R, Valot C, Nijhuis H (2003a) Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Hum Factors Aerosp Saf* 3(2): 95–117
- Flin R, Martin L, Geosters K, Hoermann J, Amalberti R, Valot C, Nijhuis H (2003b) Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Hum Factors Aerosp Saf* 3(2): 95–117
- Flin R, O'Connor P, Crichton M (2008) *Safety at the sharp end*. Ashgate Publishing Limited, Abingdon

- Flin, R, Winter, J, Sarac, C and Raduma, M (2009) Human factors in patient safety: review of topics and tools. Report for methods and measures working group of WHO patient safety: World Health Organization
- Gatfield D (2008) "Behavioural markers for the assessment of competence in crisis management". In: PhD Thesis. Southampton Solent University,, United Kingdom
- Goulielmos AM, Lathouraki G, Giziakis C (2012) "The quest of marine accidents due to human error, 1998–2011". *Int J Emerg Serv* 1(1):39–70
- Grech MR, Horberry T, Smith A (2002) Human error in maritime operations: analysis of accident reports using the Leximancer tool. Proceedings of the Human factors and Ergonomics society 46th annual meeting, Baltimore September 30–October 4, 2002
- Gregory D, Shanahan P (2010) The human element guide to human behaviour in the shipping industry. United Kingdom Maritime and Coastguard Agency (MCA), Southampton
- Habberley J, Barnett M, Gatfield D, Musselwhite C, Mcneil G (2001) "Simulator training for handling escalating emergencies", United Kingdom Maritime and Coastguard Agency (MCA), Research Project 467
- Havold JI (2000) "Culture in maritime safety". *J Marit Policy Manag* 27(1):79–88
- Helmreich RL, Merritt AC, Wilhelm JA (1999) The evolution of crew resource management training in commercial aviation. *Int J Aviat Psychol* 9(1):19–32
- International Association of Maritime Universities (2010) "Review of manila amendments", presented at the IAMU annual general assembly 14–18th October 2010, Pusan, Korea.
- Ishizaka A, Labib A (2009) Analytic hierarchy process and expert choice: benefits and limitations. *OR Insights* 22(4):201–220
- Kern AT (2001) "Controlling pilot error: culture, environment and crew resource management", controlling pilot error series. McGraw-Hill Professional, New York
- MNTB (2012) Human element leadership and management—operational and management levels. 1st edition, Merchant Navy Training Board
- Pillay A, Wang J (2003) A risk ranking approach incorporating fuzzy set theory and grey theory. *Reliab Eng Syst Saf* 79(1):61–67
- Registro Italiano Navale (1996) Class booklet. Registro Italiano Navale (RINA) Classification Society, Geneva
- Riahi R, Bonsall S, Jenkinson I, Wang J (2012) "A seafarer's reliability assessment incorporating subjective judgements", proceedings of the institution of mechanical engineers. Part M: *J Eng Marit Environ* 226(4): 313–334
- Saaty TL (1980) The analytic hierarchy process. McGraw-Hill, New York
- Saaty TL (1990) How to make decisions: the analytic hierarchy process. *J Oper Res* 48(1):9–26
- Saaty TL (2008) Decision making with the analytic hierarchy process. *Int J Serv Sci* 1(1):83–98
- Salas E, Burke CS, Stagl KC (2004) Developing teams and team leaders: strategies and principles. In: Day DV, Zaccaro SJ, Halpin SM (eds) *Leader development for transforming organizations: growing leaders for tomorrow*, series in applied psychology. Lawrence Erlbaum Associates, Mahwah, pp. 325–355
- Salmon PM, Stanton NA, Walker GH, Jenkins DP (2009) Distributed situation awareness: theory, measurement and application to teamwork. Ashgate Publishing Limited, Abingdon
- Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
- Smith RD (2001) "controlling pilot error: controlled flight into terrain (CFIT/CFITT)", controlling pilot error series. McGraw-Hill Professional, New York
- Stanton N (1996) Team performance: communication, co-ordination, co-operation and control. In: Stanton N (ed) *Human factors in nuclear safety*. Taylor and Francis, Philadelphia, pp. 197–215
- West MA (2012) *Effective teamwork, practical lessons from organizational research*. BPS Blackwell, Chichester
- Yule S, Flin R, Peterson-Brown S, Maran N, Rowley D (2006) Development of a rating system for surgeons' non-technical skills. *Med Educ* 40:1098–1104