



Barnett, Michael and Gatfield, David and Habberley, John. (2002). Shipboard Crisis Management: A Case Study. Proceedings of the Human Factors in Ship Design and Operation Conference., 2002,

Downloaded from <http://ssudl.solent.ac.uk/432/>

Usage Guidelines

Please refer to usage guidelines at <http://ssudl.solent.ac.uk/policies.html> or alternatively contact ir.admin@solent.ac.uk.

SHIPBOARD CRISIS MANAGEMENT: A CASE STUDY

Dr Michael Barnett, Warsash Maritime Centre, Southampton Institute, UK

David Gatfield, Warsash Maritime Centre, Southampton Institute, UK

Capt John Habberley, Warsash Maritime Centre, Southampton Institute, UK

SUMMARY

The loss of the “Green Lily” in 1997 is used as a case study to highlight the characteristics of escalating crises. As in similar safety critical industries, these situations are unpredictable events that may require co-ordinated but flexible and creative responses from individuals and teams working in stressful conditions. Fundamental skill requirements for crisis management are situational awareness and decision making. This paper reviews the naturalistic decision making (NDM) model for insights into the nature of these skills and considers the optimal training regimes to cultivate them. The paper concludes with a review of the issues regarding the assessment of crisis management skills and current research into the determination of behavioural markers for measuring competence.

AUTHORS BIOGRAPHY

Dr Mike Barnett is Head of the Maritime Research Centre in Southampton Institute. After a seafaring career to chief officer rank, Mike joined Warsash in 1985 as a lecturer in tanker safety, in which post he was involved in the development of the Centre’s liquid cargo operations simulator. He was awarded a PhD from the University of Wales, Cardiff in 1989 for his work on human error and the use of simulation in training for emergencies. He has been Head of Research at Warsash since 1991, during which time the Centre has developed its research capability into various aspects of maritime human factors. His current post encompasses responsibility for the management of research at Warsash and on the Southampton campus in the fields of marine environmental management and technology. Mike is a Fellow of the Nautical Institute and a current Vice-President of the Institute of Marine Engineering, Science and Technology (IMarEST).

David Gatfield is currently Senior Lecturer and Unit Manager of the Machinery Space Simulator at Warsash Maritime Centre. David served as a Marine Engineering Officer with Shell Tankers throughout the company's fleet from 1980 to 1987, before secondment as Computer Systems Trainer for the Shell fleet. In 1992 he was promoted to the post of Technical Superintendent and was responsible for the support of the instrumentation and control systems onboard all vessels in the Shell fleet. He joined Warsash Maritime Centre in 1996 and lectures in Engineering Systems Management, Risk Management, Petrochemical Tanker Engineering Systems and Gas Tanker Reliquefaction Systems. He is currently conducting research for a PhD in behavioural markers for the assessment of competence in crisis management. David is an Associate Member of the Institute of Marine Engineering, Science and Technology (IMarEST) and a Member of its Council.

Capt John Habberley is Head of Consultancy at Warsash Maritime Centre. John went to sea with P & O Lines and served in ranks up to Chief Officer. After obtaining an honours degree in Maritime Studies at the University of

Wales, Cardiff, John returned to sea as Master before joining the Warsash Maritime Centre in 1981. Since then he has taught ship handling and bridge team management using ship simulators. He was Research Manager for three years and specialised in research using ship’s bridge simulators to assess the collision avoidance behaviour of watchkeepers. He obtained his MPhil for research into the assessment of competence. From 1991 to 1995 he was Deputy Head of the Maritime Operations Centre at Warsash, and subsequently Head of Simulation for five years before his current post. John is a Fellow of both the Nautical Institute and the Royal Institute of Navigation.

1. INTRODUCTION

In the year 2000, the Maritime Coastguard Agency (MCA), following a recommendation of the Marine Accident Investigation Branch (MAIB) in response to the loss of the “Green Lily”, awarded a research project to Warsash Maritime Centre to investigate the potential use of simulators for training in the handling of escalating emergencies (MCA Project RP467). This project enabled the research team to review current concepts and models in the field of crisis management across a range of safety critical industries and to conduct a survey of expert opinion on the optimal training and assessment regimes for handling escalating emergencies. This paper describes some of the findings of that project. In particular, the paper addresses the following issues:

- What are the characteristics of crises and how do they differ from emergencies?
- What are the skill requirements for crisis management?
- What are the essential requirements of a training system to foster these skills?
- What forms of simulation offer the optimal training regimes for crisis management?
- What are the issues to be resolved in the assessment of competence in these skills?

The paper concludes with a description of current research being conducted at Warsash in the use of behavioural markers for assessing competence.

2. A CASE STUDY: THE “GREEN LILY”

On 18th November 1997, the 3,624 grt Bahamian registered vessel “Green Lily” sailed from Lerwick in the Shetland Islands with a cargo of frozen fish for the Ivory Coast. The weather on departure was bad with wind speeds increasing to severe gale force 9. The following morning, while hove to about 15 miles south-east of the island of Bressay in the Shetland Isles in storm force 10 winds, a sea water supply line fractured in the engine room. The engineers controlled the flooding and pumping out had begun when the main engine stopped. Unsuccessful attempts were made to restart the engine while the vessel drifted northwards towards Bressay. Shetland Coastguard was advised and three tugs, the Lerwick RNLI lifeboat and a coastguard helicopter prepared to proceed to the casualty.

Attempts were made by two of the tugs to secure a line and tow the “Green Lily” away from land but although initially successful, each line parted. The starboard anchor was released and the third tug attempted to snag the cable and pull her head to wind, but the cable parted. At this time, the lifeboat rescued five crewmen, including two injured, from the ship’s deck. The ten remaining crew members were rescued by the Coastguard helicopter but the winchman, who had remained on the deck of the ship, was swept into the sea and lost. The “Green Lily” went aground and started to break up. The investigation by the Marine Accident Investigation Branch (MAIB), published in June 1999, advised the cause of the grounding was:

“the lack of propulsion and failure to restart the main engine to arrest the drift of the vessel towards the shore in the prevailing environmental conditions. Contributory causes included flooding of the engine room, failure to reset the mechanical over-speed trip, inadequate knowledge of the cooling water system, failure of the towage attempts and inadequate teamwork” (MAIB, 1999 page 9)

The causal factors in this case suggest that this tragedy bears all the hallmarks of similar critical incidents across a range of safety critical industries:

- An initial technical failure has precipitated events and has been compounded by a hostile environment and further technical problems and failures. The situation was escalating in severity. An emergency was becoming a crisis, but the actors in this tragedy did not have the benefit of hindsight to read the script.

- The available emergency plans, which tend to be procedures based on single failures, were not applicable. The individuals involved were forced to fall back on their experience to cope with an increasingly complex and unpredictable set of circumstances.
- Initial diagnosis of the technical failure was incorrect and led to a faulty but persistent mental model of the situation. In this case, the chief and second engineers, together with the electrical engineer, failed to understand why the main engine stopped and were consequently unable to restart it. They believed that the main engine failure was due to the effect of the flooding, previously caused by the fracture of the sea suction pipe. The probable reason for the main engine stoppage was actually due to the mechanical over-speed trip either not being reset or reset incorrectly.
- Awareness of the overall situation by individuals was based on incomplete or inaccurate information. In this case, both the Master, based on his calculation of drift, and the engineers were over optimistic in their belief that a tow would be available before the ship ran aground. Meanwhile, the skippers of the rescue craft had unexpressed reservations about various aspects of the operation including the appropriateness of some of the towing gear, the weather conditions and sea room, and the ability of the ship’s crew to handle the towlines.
- Individuals and units were separated physically and several agencies were interacting through various forms of communication. In these circumstances, it was very difficult for the key players to communicate meaningfully and maintain a shared and agreed awareness of the rapidly changing situation.

These characteristics of critical incidents suggest that a distinction may be drawn between Emergencies and their management and Crises and their management. This distinction may be summarised as follows:

An Emergency can be defined as a situation outside normal operating parameters where corrective decisions and actions are based on documented procedures. In the maritime context, examples might be “Man overboard”, steering gear failure or a report of a fire in a cabin. Emergency procedures can be trained both at onshore training establishments and on board.

A Crisis differs from an emergency in that successful decisions and actions may not necessarily be based on documented procedures. Appropriate pre-defined responses may not exist, and even if they do, in practice they may have conflicting requirements. Those responsible for handling crises will have to think through the situation, and respond in creative and flexible ways.

This distinction between emergencies and crises has a significant impact on the training requirements for their management. Training in handling emergencies may simply be training in following pre-prescribed procedures and drills. Training in crisis management is likely to require a much more demanding approach to practise the skills required in these situations.

So what skills are required to handle crises? There is now considerable evidence from both military and civilian sources that the main requirements are for the high-level cognitive skills of problem solving and decision making and this is discussed in more detail in Section 5. Crichton and Flin (2002) suggest that, at its most simplified, there are two fundamental and inter-related skill requirements:

- Situation assessment – “what’s the problem”
- Decision making – “what shall I do”.

The following section reviews current concepts in decision making in critical situations.

3. DECISION MAKING UNDER STRESS

Modern concepts for understanding decision-making have progressed from classic rational choice models to ones that try to reflect the way decisions are actually made in the real world. The most influential of these models is called the naturalistic decision-making (NDM) model and has been defined as follows:

“The study of NDM asks how experienced people, working as individuals or groups in dynamic, uncertain, and often fast-paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and the larger organisation in which they operate.” (Pruitt et al, 1997)

This definition reveals a number of characteristics of the situations in which NDM takes place. These characteristics suggest that NDM is an appropriate model for the understanding of decision making under stress in escalating marine emergencies. These characteristics are:

- The situations in which decisions are made are uncertain, unpredictable and dangerous.
- Knowledge of the situation is incomplete, and constantly changing.
- The consequences of decisions and actions based on poor situational awareness are potentially catastrophic.
- Experienced people, not novices, generally conduct decision making in such situations.

Another important feature of NDM, which reflects its value in understanding real world decision making, is that, unlike classical models of decision making, where

the objective is to provide optimal decisions, the objective for real world decision makers is to arrive at actions based on decisions that will satisfy the immediate concerns of the situation, without those decisions necessarily having to be the best ones. There are a number of different models within an NDM approach to describe the process by which decisions are made. The dominant model is known as the Recognition-Primed Decision (RPD) model. Orasanu (1997) provides a comprehensive description of the process:

“Its basic principle is that experts use their knowledge to recognise a problem situation as an instance of a type, and then retrieve from their store of patterns in memory an appropriate response associated with that particular problem type. The response is evaluated for adequacy in the present context, and if it passes, it is adopted. If it is found wanting, either another interpretation of the situation is sought or a second level response is retrieved and evaluated.”

A number of conclusions may be drawn from this description of the RPD process, which are relevant to the handling of emergencies in stressful circumstances:

- Normal control of operations is managed through a series of approximating or “satisficing” decisions, i.e. a situation is recognised as being typical and a number of decisions and actions taken. The situation is then monitored and further decisions and actions taken to refine the original response.
- The more experienced the decision maker, the richer the store of experiences to draw from and the more extensive their repertoire of actions. Orasanu (1997) records a number of research studies that confirm that this is one of the most consistent differences between novices and experts.
- The most critical aspect of the decision making process is awareness of the situation, not the generation of options. Orasanu (1997) also records that in most aviation accidents, crews have exhibited poor situation assessment rather than faulty selection of a course of action.

RPD works well when the situation can be recognised, i.e. in normal situations. The paradox is that in emergency situations, just when the expert needs to draw on a reliable repertoire, the situation is unpredictable and atypical, so no repertoire can be called upon. The emergency handler has to revert to a creative response i.e. they have to think their way through the novel situation.

It is this requirement to think through the situation that creates stress, which in turn may affect the quality of decision making. If decisions need to be made quickly, then time pressure also becomes an additional stressor. Orasanu (1997) describes the situation thus:

“In situations that support perception-based memory retrieval, stress effects should be minimal. These situations tend to be those that deal with familiar and unambiguous problems for which rule-based responses are available. Situations that require attentional scanning and impose demands on working memory are ones that are either unfamiliar or for which cues indicating a problem are ambiguous, thus requiring information search or diagnostic strategies.”

Given the nature of decision making under stress, the challenge is how best can we prepare individuals and groups for the unpredictable or as Orasanu (1997) puts it – how might we strengthen the weak links? She provides three possible answers: through training; through better procedures; and through the design of better decision making aids.

This paper is concerned with shipboard crisis management and the training that could evolve from that. So what are the implications of NDM and RPD theory for training?

NDM and RPD theories raise a number of issues that suggest ways in which training may be enhanced to prepare people for handling escalating emergencies under stress (Orasanu 1997):

- Decision-making is a skill. Like all skills it may be learned through practice. By reducing the cognitive load through practice, experts will be less stressed than novices in threatening situations.
- In addition to specific skills, there is a set of more general decision making skills, known as metacognitive skills. The direct development of such generalised situation awareness skills might counteract the consequences of stress.
- Educating people about stress, and providing them with techniques and strategies to cope with stress may be helpful.
- Stress effects of decision making may be reduced by sharing the decision making process within the members of a team.

The primary justification for the direct training for crisis management is based on the belief that by exposing individuals or teams to a variety of potential crisis scenarios, their mental models of situations will be enriched, thus enhancing their situational awareness techniques and their repertoires of decision making. A secondary justification is that by exposing people to such situations it provides them with the self-confidence that they can handle future unknown situations.

The key to this approach is in the “richness” of the mental models developed by the individual or team, but

paradoxically, the problem is that if the training scenarios are too prescriptive, then the learned repertoires may be inappropriate to the real emergency encountered. Crego and Spinks (1997) express this dilemma in their description of the “Minerva” simulation, which is the command and control training system for senior police officers:

“It is vital that any patterns presented to the learner should not be reduced to individual cues...at various periods during the unfolding event. This behaviourist approach may well be successful in engendering recognition of a particular pattern set, but this recognition may not be transferable when the learner is faced with similar patterns in different contexts. What is needed is a more fluid, flexible simulation that is responsive to decisions made and, as a result, is very much driven by the actions (or indeed the inactions) of the participants. At the same time, unfolding incidents must be multi-threaded and at times parallel, if the impression that the learner is being led through prescribed problems towards pre-planned outcomes and solutions, is to be avoided. Only through such patterning and sequencing, combined with effective team driven communication, will the command team gain a real sense of ‘situational awareness’ and fidelity with the real world of command be sustained.”

Clearly, some form of simulation offers one of the most beneficial training mediums for the direct learning and practice of situational awareness and decision making skills in a safe environment. In the short extract above, Crego and Spinks also highlight the major issues to be addressed in the use of simulation for the training of emergency handling:

- To what extent will RPD type skills, learned in a simulated environment, **transfer** to the real emergency?
- What level of context or **fidelity** has to be provided in a simulated environment to provide the sort of fluid, responsive simulation that is required?

These issues of transfer and fidelity are addressed in the next section.

4 TRAINING

Before considering the sophistication and type of simulation required to address the training requirement, it is necessary to consider the theoretical aspects of training transfer and fidelity:

Classic definitions of transfer (MSA Project 340,1995) refer to the degree to which learning to perform one task is made easier by the prior learning of another task. Simulator based training involves the systematic development in a simulated environment of the knowledge, skills and attitudes required to perform a task

in the real world. In this context, transfer refers to the degree to which learning in the real environment is made easier by prior learning in the simulated situation. If the learning situation facilitates performance in the second situation, positive transfer has occurred. If the learning situation impairs subsequent performance, negative transfer has taken place.

Three complementary models of transfer are to be found in the literature. (Rolfe, 1991). One of the earlier explanations of transfer argued that positive transfer would occur to a second task if that task contained component activities that were present in the first task. Transfer is thus dependent on identical elements within both tasks.

A later model argues that transfer is dependent on the extent to which there is similarity between the representation of a stimulus and the response demands of the learning and actual performance situations. The importance of this model is that it introduces the idea that transfer can be obtained with simulations that are not replicas of the real situation.

More recent approaches to transfer stress the importance of the trainee in the learning process. Therefore, one of the pre-requisites for positive transfer is the motivation of the trainee to acquire new skills.

Total transfer is rare; i.e. some further learning in the operational setting is usually necessary. Transfer is not uniform; some skills will transfer more readily than others. Some aspects of the learning situation may transfer inadvertently. By deliberately omitting some tasks, trainees might infer incorrectly that these tasks are not considered important, and then neglect them in the real situation. Equally, beneficial aspects may transfer even though they were not specific training objectives. Training in teams may also influence the transfer of learning of individuals, for better or worse. For example, if one team member dominates, others may not learn so much. Equally, collaboration may enhance individual learning and transfer.

The importance of transfer in simulator training is that it is the key measure of the effectiveness of that training (Barnett, 1996). However, there are few recorded transfer experiments in the marine simulator literature. (Muirhead, 1991)

The most abundant source of transfer study experiments are in relation to flight simulators, and even here, results from studies are surprisingly inconclusive in providing hard evidence of positive transfer. (Billings et al, 1975; Rolfe, 1991)

Findings from these studies stress that transfer experiments need to be carefully designed. Transfer of training experiments are notoriously difficult to control (Caird, 1996). Problems involve the lack of experimental

control, insufficient sample sizes, insufficient time in the simulator, insufficient time for evaluating transfer in the operational setting, and insensitive measures (Waag, 1991).

One implication of the classic models of transfer is that to be cost-effective, any simulator should be designed so that it simulates the operational situation *only to the extent necessary* to provide transfer of the skills required by the initial Training Needs Analysis (TNA). As Jackson (1993) puts it:

“It is desirable that simulation fidelity and capability is sufficient to ensure the required transfer of training, but not to grossly exceed it since this would generally increase system cost with no return.

The result of the training needs analysis will be a specification of the cues required and their minimum fidelity in order to achieve the required transfer of training.”

This concept establishes the linkage between transfer and the level of fidelity required in order to provide that transfer.

In the simulation literature, there are a confusing number of definitions of fidelity in terms of realism, equipment fidelity, environmental fidelity, behavioural fidelity, psychological fidelity, physical similarity and total context fidelity. (Allerton and Ross, 1991)

Hays and Singer (1989) reduced these multiple definitions to two main dimensions:

“Simulation fidelity is the degree of similarity between the training situation and the operational situation which is being simulated. It is a two dimensional measurement of this similarity in terms of: (1) the physical characteristics...and (2) the functional characteristics.”

Caird (1996) develops this concept to produce the following definitions of physical and psychological fidelity:

“Physical fidelity has been defined as the degree that the physical simulation resembles the operational environment, whereas psychological fidelity is the degree that a simulation produces the sensory and cognitive processes within the trainee as they might occur in operational theatres.”

Early simulator design and training development progressed in the belief that by producing the highest level of physical fidelity possible, such realism alone would lead to effective transfer. As Caird (1996) puts it:

“For decades, the naïve but persistent theory of fidelity has guided the fit of simulation systems to training.”

Consequently, more modern thinking recognises that the level of fidelity required depends on the nature of the skills being trained. Both Jackson (1993) and Caird (1996) point out that greater degrees of physical fidelity are needed where physical or manual tasks are required, whereas functional or operational fidelity is required for cognitive tasks.

Furthermore, and this is the crucially important point, the level of fidelity required depends on whether physical or cognitive tasks are being developed, to the extent that in some cases, the lack or distortion of realism may provide increased training benefit:

“For cognitive and procedural training there need be less emphasis on realism, indeed it may be advantageous from a training transfer aspect to remove distracting ‘realistic’ features. It may even be advantageous to actively distort the reality in order to better demonstrate the subject matter” (non Euclidean environments i.e. outside of the normal laws of geometry). (Jackson, 1993)

“...there is some evidence from flight simulation that higher levels of fidelity have little or no effect on skill transfer and reductions in fidelity actually improve training. Reductions of complexity may aid working memory and attention as skills and knowledge are initially acquired. Perhaps errors on the side of more fidelity reflect failed attempts to completely understand the underlying physical to cognitive mappings. (Caird, 1996)

In order to ascertain the optimal types of simulation to provide training for Crisis Management situations, the Warsash research team used a panel of 15 experts and a modified “Delphi” technique to investigate the issue. The expert panel was drawn from marine simulation resources as well as researchers and practitioners from other similar safety critical industries. Within this project, the Delphi Method was used in a particular way termed Policy Delphi (Turoff 1970). The Policy Delphi process is a form of policy analysis that provides a decision maker with the strongest arguments on each side of the issue. A range of future implementation scenarios were proposed as training policies that could meet the perceived training requirements relating to the handling of escalating emergencies. These policies were presented to the panel of experts. A subsequent workshop involving some of the panel experts was also used to confirm and develop their responses.

The following is a summary of the responses received from the panel of experts in reply to 19 questions sent to them in order to further clarify the main arguments for and against the proposed training policies.

Training Policy 1: Using a Full Mission Simulator with Team Based Exercises

There was general agreement that training and assessment should only ever be undertaken separately.

There were a number of concerns expressed about how any form of assessment would be undertaken in order to ensure objectivity.

Strengths of this policy option were seen to be the ability to undertake team-based activities and the greater fidelity of the training environment.

The main weaknesses of this policy option were seen to be the high cost of full mission simulators and the difficulties in carrying out assessments of individuals undertaking team-based activities.

There was general agreement that the tutor should never also be the assessor within the same time-frame.

Training Policy 2: Full Mission Simulator with Single Trainee Exercises

There was agreement that this policy option was not generally beneficial, but could be useful in special circumstances such as remedial and pre-team training.

Training Policy 3: Virtual Environments

Although there was still a very positive response to this policy option, little empirical evidence was cited to support the opinions given.

There was general agreement that the communications systems used within this policy option could be embedded, as long as they allowed actual voice communications, and this could be used in a similar way to real communication systems.

Most responses indicated that the co-workers within virtual reality training environments should be real and not simulated in order to facilitate effective team training. However, the possibility was raised that simulated co-workers could be used to afford a greater variety of training opportunities for team members.

There was general agreement that a high level of fidelity was required for certain elements of the virtual environment, but there was a wide diversity of opinion as to what these elements were. The elements discussed were all part of the functional representation of the real environment, both physical and procedural. One response stated that virtual environment did not have to have a high degree of fidelity as long as it allowed for the replication of the skills inherent in the task being trained.

Training Policy 4: Desktop Computer Simulation

There was agreement that this policy option required a certain level of interactivity to be effective and that an increase in interactivity could improve effectiveness and

efficiency up to a point, beyond which the trainee may start to feel confused.

A number of ways of improving interactivity were proposed including the:

- creation of multiple training paths
- provision of training scenarios with more than one acceptable outcome
- use of a facilitator to guide the trainee.

If this policy option could be team-based there was general agreement that this would be more beneficial, because it would allow trainees to discuss alternative solutions. However, one response indicated that if the simulation were more team-based it would become more difficult to control and it would be more difficult to carry out assessments.

It was suggested that, within this policy option, team-based activity could be achieved by having a facilitator guide each training scenario.

There was general agreement that it would be difficult to use this policy option to undertake assessments at a distance, as this type of assessment would only be based upon the training outcome and not the process leading to it. The issue of ensuring the authenticity of a candidate was seen to be a further difficulty if undertaking assessment at a distance.

Training Policy 5: Table-top simulation

All participants agreed that this policy option could be used for training. However, there were arguments made both for and against the use of this policy option for undertaking assessment.

The argument against was based on the lack of fidelity provided by this type of simulation and the difficulty in observing relevant competent behaviour in a context that is very different from the actual workplace.

The argument for was based on assessment being undertaken against those relevant behavioural markers that could be observed within the context of the simulation.

There was general agreement that it is important to create stress during the training and assessment for handling escalating emergencies.

A number of ways of creating stress were proposed, the most common of these being to introduce time constraints into the simulation exercises. Other methods discussed for introducing stress were increasing information flows, increasing exercise complexity, introducing unexpected failures and malfunctions, increasing noise levels and the use of role playing facilitators.

There was a strong difference of opinion as to whether the level of fidelity correlates with the level of stress induced within a simulation. A number of responses suggested that, unless the simulation had a high level of fidelity, stress would not be induced. However, there were other responses that strongly opposed this view, suggesting that stress is psychologically induced and is therefore more dependent upon the exercise participant's perception of presence within the simulation scenario, than the level of fidelity of the simulation itself.

Training Policy 6: Class Room Based Workshops

There was general agreement that this policy option is best suited to training only.

The following strengths were associated with this policy option:

- cost beneficial
- flexible
- gives the opportunity to discuss operational / emergency problems with others
- tutor guided

The following weaknesses were associated with this policy option:

- there is no environment to manage
- not suitable for the assessment of competence

One response suggested that any weaknesses associated with this policy option could be overcome by providing a good tutor and ensuring interactivity. There was a wide spread of opinion regarding which other methods of training this policy option could be usefully used in conjunction with. The overall range of opinion covered all of the remaining five policy options. One response suggested that classroom-based workshops followed by practice in context would allow increased transfer.

The following were proposed as being suitable to be trained using this policy option:

- appreciation of technical risks
- knowledge of systems
- knowledge of procedures
- theoretical knowledge
- planning
- risk management
- problem solving

The workshop concluded that the inclusion of full mission simulation was the only viable assessment option. This method is used extensively by the nuclear and aviation industries. The argument is that it is the only safe method that guarantees that the majority of the cues that seem important are present and that the perceived required skills may be demonstrated.

The search for a single cost-effective training option to deliver the required standard of competence may be misplaced. The principle enshrined in STCW95 and National Vocational Qualifications (NVQs) is that once the standard of competence has been defined, how an individual reaches that standard is irrelevant. Among a number of variables, it is the motivation of the learner and the ingenuity of the trainer that will determine the most cost-effective training option. In an ideal world, the trainer would select the most appropriate method from his/her training "toolbox" to suit the individual trainee, their learning style, and stage of development identified through continuous assessment.

Recent research by Crichton and Rattray (2002) describes the potential of Tactical Decision Games (TDGs) for crisis management training. TDGs are a low-cost, low fidelity classroom based simulation that focuses on improved decision making and heightened situational awareness. Evaluation of their effectiveness and their validity and reliability as a competence assessment tool is currently underway.

In summary, the most cost-effective training option will be determined by a number of "local" factors, including the ingenuity of the instructor. At present, however, the assessment of competence, particularly for marine certification purposes, through the use of currently available Full Mission Simulations represents the most viable option.

5. ASSESSMENT

All safety critical organisations consider how they would manage a crisis situation and undertake some form of preparedness training. This training concentrates mostly on how to deal with an emergency, where a laid down procedure can be put into action. Few of these organisations take their training into the realms of a crisis situation, where there is no procedure to call upon, and where lateral thinking and rapid decision making are required of their managers. Even fewer organisations try to assess their personnel's competence in managing a crisis.

One of the recommendations of the United Kingdom Maritime and Coastguard Agency Project RP467 (Habberley et al., 2001) was that:

"Crisis management standards of competence are ill defined and consequently so are their 'behavioural markers' by which the standard may be assessed. More research is needed in this area, particularly in assessing the team working competencies."

So how do safety critical organisations assess the competence of their crisis managers? How do they do this objectively, and what are the assessment criteria they use?

Of all the safety critical organisations, the military have taken crisis management training and assessment the furthest. This is done for a very good reason, as all combat situations are, by their very nature, crises. Confirming the experience of researchers in other domains, Tollcott (1992) states that the two primary components of military decision making are:

- situation assessment (what is happening); and
- action selection (what to do about it).

The first of these components requires crisis managers to generate hypotheses to account for the information that is being received. The second of these components requires the generation and evaluation of alternate actions. During a crisis these tasks have to be performed within a highly demanding decision environment.

In certain circumstances this demanding decision environment may become too demanding for the crisis manager, and they may find themselves unable to cope. This is described by Salas et al. (1996) as a situation when:

"environmental demands evoke an appraisal process in which perceived demand exceeds resources and results in undesirable physiological, psychological behavioural or social outcomes."

So it is important within any safety critical organisation to try and determine whether the personnel placed in the role of potential crisis manager will be able to cope when a crisis arises.

Following their participation in a major US military research project, 'Tactical Decision Making Under Stress', Cannon-Bowers and Salas (1998) proposed a set of knowledge, skill and attitude requirements for teams to work effectively during crisis situations:

Team Knowledge Requirements:

*cue strategy association
knowledge of team-mate characteristics
shared task models
knowledge of team interaction patterns
task sequencing*

Team Skill Requirements:

*adaptability
shared situational awareness
mutual performance monitoring
communication
decision making
interpersonal skills
team leadership*

assertiveness
conflict resolution

Team Attitude Requirements:

Collective efficacy
shared vision
team cohesion
mutual trust
collective orientation
importance of teamwork

If indeed these are the requirements for an effective crisis management team, the assessment of competence in crisis management based upon these requirements is a daunting task. If, as discussed in Section 4 of this paper, assessment should be undertaken in an environment that closely resembles the real world situation, the capture of data to evaluate against assessment criteria relating to all of these requirements is a truly enormous task.

Through their use of war games, the military attempt this task. They use large numbers of assessors, dispersed throughout the war gaming environment during an assessment exercise. After the assessment exercise, the assessors meet to discuss their observations during the exercise, and to evaluate the actions of the team against set assessment criteria. Examples of these criteria are:

'was there a good flow of information into the control position at all times'

and

'was the incident picture well kept'

These criteria are assessed as having been either 'met' or 'not met'. A discussion is then held between assessors to give an overall assessment of how the team performed. Due to the severe time restraints imposed on the assessment process, because of the operational requirements of the military, and the sheer complexity of the war gaming environment, subjective assessments are inevitable. However, because of the large number of assessors used, effective assessments can be achieved through moderation.

The civil aviation industry has recently been undertaking research into the possibility of assessing the non-technical skills of aircrew. Non-technical skills are defined as those skills, in addition to technical skills, required for competence in crisis management. Through the Joint Aviation Requirements Translation and Elaboration of Legislation research project (JAR TEL Consortium, 2001), a methodology for assessing the non-technical skills of aircrew, by observing individual overt behaviours, has been proposed.

The cockpit environment is very different to that of a war gaming environment, but the non-technical skills of co-

operation, leadership and management, situational awareness and decision making, as metrics for assessing competence in crisis management, are common to both. A major difference between the assessment of competence in crisis management within the military context and the civil aviation context is that within the military context a team is assessed, whereas within the civil aviation context it is the assessment of an individual working within a team that is undertaken.

The JAR TEL non-technical skills or 'NOTECHS' assessment framework provides definitions of the non-technical skills to be assessed and gives the assessor examples of overt behaviours that indicate good or poor practice of these skills.

An example skill element under the category of 'Co-operation' is *"team building and maintaining"*.

An example of an overt behaviour indicating poor practice of this skill element is:

"Keeps barriers between crew members."

An example of an overt behaviour indicating good practice of this skill element is:

"Encourages inputs and feedback from others (lowers the barriers)."

Although the 'NOTECHS' framework has moved the assessment of competence in crisis management, within the context of civil aviation, towards a more objective foundation, the experimental results of inter-rater reliability trials showed that in the more complex assessment scenarios there were significantly divergent assessments.

The JAR TEL report states that there are some strongly held reservations, by some members of the aviation fraternity, about the very concept of the assessment of non-technical skills. One of the prime reservations being that:

"it is felt that the criteria on which assessment is based are largely subjective and thus cannot easily be monitored for fairness and accuracy"

Through the STCW Code Table A-V/2 (IMO, 1995), the International Maritime Organisation (IMO) has provided the competence specification of a minimum standard of competence in crisis management and human behaviour for those officers who have responsibilities for passengers. As within the civil aviation industry, these competencies relate to individuals working within a team. The required underpinning knowledge, understanding and proficiency, are stated for each competence, along with methods for demonstrating competence and criteria for evaluating competence.

IMO does not differentiate between crises and emergencies, and the Table A-V/2 relates primarily to the management of emergencies, citing the use of procedures and actions in accordance with established plans as a criterion for evaluating competence.

The assessment criteria given in Table A-V/2 of STCW 95 are also highly subjective, an example being:

“Information given to individuals, emergency response teams and passengers is accurate, relevant and timely.”

From the examples above it can be seen that safety critical organisations undertake the assessment of competence in crisis management in very different ways. These have been summarised in Table 1 at the end of this paper.

Experience within the military context has shown that the crisis management assessment framework used has been both fair and effective. However, this has been achieved through the use of a huge amount of resources, both within the assessment environment and the assessor team.

Within the context of civil aviation the use of overt behavioural markers as criteria for assessing competence in crisis management skills has been attempted, and has been shown to be successful when used to assess personnel within simple, non-crisis, scenarios. However, the current assessment framework has been shown to be unreliable when used to assess personnel within complex, crisis, scenarios.

Within the merchant marine context, the assessment framework for crisis management and human behaviour is too open to interpretation to be effective.

Any framework for the assessment of competence in crisis management within the context of the merchant marine would not have the resources available to it that the military has. The civil aviation assessment framework for non-technical skills, although feasible to apply within the merchant marine context, has not yet been shown to be reliable in assessing competence in crisis management.

In order to provide the international maritime community with an understanding of how a behavioural marker system could be applied for the assessment of competence in crisis management of merchant marine officers, research is currently being undertaken at Warsash Maritime Centre.

The aims of this research are:

- To understand how behavioural markers can be used to objectively assess competence in crisis management of merchant marine officers.

- To understand the methods by which these behavioural markers can be elicited and assessed.

Data is being collected and analysed using ethnographic study techniques during simulated crisis scenarios within a high-fidelity ships engine control room environment.

It is hoped that this research will lead to the development of an assessment framework that can be applied within the merchant marine context for the fair and effective assessment of competence in crisis management.

6. CONCLUSIONS

The characteristics of crisis situations in which decisions have to be made under stress by often experienced individuals are uncertainty, unpredictability and danger. Situational awareness is incomplete, and constantly changing. However, the consequences of decisions and actions based on poor situational awareness are potentially catastrophic. The key skills therefore of successful crisis management are situation awareness and decision making.

Training in crisis management is likely to require a much more demanding approach to practise the skills required in these situations than procedurally based emergency training.

The direct training of crisis management skills is based on the belief that by exposing individuals or teams to a variety of simulated crisis scenarios, their mental models of situations will be enriched, thus enhancing their situational awareness techniques and their repertoires of decision making. The key to this approach is in the “richness” of the mental models developed by the individual or team.

The results of a survey of marine and other experts in simulator training suggest that a variety of simulation based options have different strengths and weaknesses for crisis management training. The most cost-effective training option, therefore, is likely to be determined by a number of factors. However, at present, the only really viable option for the assessment of competence, for marine certification purposes, is in the use of full mission simulation.

A comparison of assessment methods used in other safety critical organisations reveals that the assessment of competence in crisis management is largely open to subjective interpretation, even where criteria exist. This situation is moderated, in some cases such as the military, by the number of assessors used. Another interesting difference between civilian and military practice is that the military are often assessing the competence of the team whereas, certainly in the maritime context, it is the performance of the individual

within the team that is of primary concern for the purposes of certification.

Research is currently being conducted at Warsash Maritime centre in order to provide the international maritime community with an understanding of how a behavioural marker system could be applied for the assessment of competence in crisis management of merchant marine officers.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the professional help and assistance provided by other members of the Warsash research team and the experts from the maritime and other industries who contributed to the project. The authors would also like to express their thanks to Phil White, the MCA project officer, for permission to use extracts from the MCA report.

8. REFERENCES

Allerton, D.J. and Ross, M.J. (1991) "Evaluation of a Part-Task Trainer for Ab Initio Pilot Training." Proc. Training Transfer Conf. The Royal Aeronautical Society. November, 1991.

Barnett, ML "The role of simulators and the qualifications of instructors and assessors under the STCW Convention" Marine Simulation and Ship Manoeuvrability (Marsim 1996) Copenhagen September 1996.

Billings, C.E., Gerke, R.J and Wick Jr., R.L (1975) "Comparisons of pilot performance in simulated and actual flight." *Aviation Space Environmental Medicine*, Vol.46, No.3, pp.304-308

Canon-Bowers, J.A. & Salas, E. (1998). Individual and Team Decision Making Under Stress. In: J.A. Cannon-Bowers & E. Salas (Eds) *Making Decisions Under Stress: Implications for individual and team training*. Washington D.C.: American Psychological Society.

Caird, J.K. (1996) "Persistent issues in the application of virtual environment systems to training." Proc. HICS'96: 'Third Annual Symposium on Human Interaction with Complex Systems'. Los Alamitos, CA: IEEE Computer Society Press, pp.124-132

Crego J, and Spinks T. (1997) Critical incident management simulation in 'Decision making under stress: emerging themes and applications' (eds: R Flin, E Salas, M Strub and L Martin) Ashgate.

Crichton, M and Flin R. (2002) "Command Decision Making" In 'Incident Command: Tales from the Hot Seat.' (eds R Flin and K Arbuthnot) Ashgate.

Crichton, M and Rattray, W (2002) "Tactical Decision Games: What are they and how can they help?" Proc. Critical incident Management Conference, Aberdeen, June 2002.

Habberley, J., Barnett, M., Gatfield, D., Musselwhite, C. & McNeil, G. (2001). Simulator Training for Handling Escalating Emergencies. *United Kingdom Maritime and Coastguard Agency Research Project 467*. Southampton: The Maritime and Coastguard Agency

Hays, R.T. and Singer, M.J. (1989) "Simulation Fidelity in Training System Design: Bridging the Gap between Reality and Training." New York: Springer-Verlag.

International Maritime Organization. (1995). *Seafarer's Training, Certification and Watchkeeping Code (STCW Code)*. London: IMO.

Jackson, P. (1993) "Applications of Virtual Reality in Training Simulation" In: K. Warwick, J. Gray and D.Roberts, eds. 'Virtual Reality in Engineering'. London: The Institution of Electrical Engineers.

JAR TEL Consortium. (2001). *Joint Aviation Requirements: Translation and Elaboration of Legislation - Results of Experiment* Workpackage 3 Final Report. Reference JARTEL/BA/WP3/D5_20. European Commission DG TREN.

Marine Accident Investigation Branch (MAIB) Marine Accident Report 5/99 "Report of the Inspector's Inquiry into the loss of mv Green Lily on 19 November 1997 off the East Coast of Bressay, Shetland Islands".

Maritime Safety Agency (1995) "The effectiveness and practical application of simulators as tools for training and examining seafarers. Project 340 report. M L Barnett, Warsash Maritime Centre.

Muirhead, P.M. (1991) "Transfer of Training to the Real World." Simulator workshop: Transfer Training. IMSF Meeting Venice July 1991.

Orasanu, J.M (1997). Stress and naturalistic decision making : Strengthening the weak links in 'Decision making under stress: emerging

themes and applications' (eds: R Flin, E Salas, M Strub and L Martin) Ashgate.

Pruitt, J S, Cannon-Bowers, J A, and Salas E. (1997) In search of naturalistic decisions in 'Decision making under stress: emerging themes and applications' (eds: R Flin, E Salas, M Strub and L Martin) Ashgate.

Rolfe, J.M. (1991) "Transfer of Training." Proc. Training Transfer Conf. The Royal Aeronautical Society. November, 1991.

Salas, E., Driskell, J., & Hughes, S. (1996) Introduction: the study of stress and human performance. In: J. Driskell & E. Salas (Eds) *Stress and human performance*. Mahwah, N.J.: Erlbaum, 1-45

Tollcott, M.A. (1992). Understanding and aiding military decisions. In: S.C. Collyer (Ed) *The 27th International Applied Military Psychology Symposium: A focus on decision making research*. London: Office of Naval Research, ONR Europe Report 92-4-W, 33-48

Turoff, M, (1970) "The Policy Delphi," *Journal of Technological Forecasting and Social Change*, (2:2), 1970.

Waag, W.L. (1991) "The Value of Air Combat Simulation – Strong Opinions but Little Evidence." Proc. Training Transfer Conf. The Royal Aeronautical Society. November, 1991.

TABLE 1: COMPARISON OF THE ASSESSMENT FRAMEWORKS FOR ASSESSING COMPETENCE IN CRISIS MANAGEMENT WITHIN DIFFERENT SAFETY CRITICAL ORGANISATIONS

Context	Assessment Environment	Assessor(s)	Assessed	Assessment Criteria	Remarks
Military	High-fidelity War Game within real environment	Multiple Distributed throughout assessment environment	Team	Prescriptive and detailed, but complexity of assessment environment leads to subjective interpretation	Fairness achieved through moderation
Civil Aviation	High-fidelity simulator	Single	Individual working within team	Overt behavioural markers with examples given of good and poor practice	Assessment framework difficult to use in complex scenarios leading to divergence of assessment
Merchant Marine	Simulations onboard and table top	Single	Individual working within team	Prescriptive, but very open to subjective interpretation	Assessment framework too open to interpretation by assessing authority