Behavioural Markers for the Assessment of Competence in Crisis Management

DAVID GATFIELD

A thesis submitted in partial fulfilment of the
requirements of The Nottingham Trent University
and Southampton Solent University
for the degree of Doctor of Philosophy

October 2008

ABSTRACT

Keywords: **behavioural markers, crisis management, assessment of competence, marine engineering officers.**

A lack of competence in crisis management has been shown to be a causal factor in a number of recent maritime accidents. In safety critical industries other than commercial shipping, such as civil aviation, nuclear and petrochemical, research is being undertaken to identify behavioural markers that can be used to assess competence in crisis management. Although there is now a general acceptance of the core concepts for the non-technical or resource management skills required for competence in crisis management, there is also an acceptance that the behaviours associated with these skills are context specific. This research programme improves the understanding of how a behavioural marker system can be used to assess the competence in crisis management of merchant marine engineering officers within the context of a merchant vessel engine room control room.

This research reviews the current practice in using behavioural markers for the assessment of competence in crisis management within safety critical industries and the military. The differences between the assessment frameworks and environments in which behavioural markers are currently being used for this assessment of competence are discussed. The influences of these differences on the use of behavioural markers for the assessment of competence in crisis management within the context of a merchant vessel engine room control room are investigated.

Through the use of ethnographic study, the research presents a set of behavioural markers that can be used to assess competence in crisis management within the context of a simulated merchant vessel's engine room control room. The research concludes that these behavioural markers can be used as a valid objective assessment framework for the assessment of competence in crisis management of merchant navy engineering officers.

ACKNOWLEDGEMENTS

I would like to acknowledge my Director of Studies, Professor Mike Barnett for his sound advice and confidence in my ability to complete this research.

I would like to acknowledge the role of my Second Supervisors, Professor Edward Borodzicz and Professor Malcolm Cook for their lateral input to my thinking.

I would also like to acknowledge all of the people who volunteered to participate in the various experiments that were necessary to complete this research.

I would finally like to acknowledge my wife, Fiona, for her support and understanding.

TABLE OF CONTENTS

TITLE PAGE……… i

ABSTRACT……… ii

ACKNOWLEDGEMENTS……… iii

CHAPTER 1. INTRODUCTION

1.1 BACKGROUND AND RATIONALE FOR THE RESEARCH……… 6

1.2 AIMS AND OBJECTIVES OF THE RESEARCH………11

CHAPTER 2. LITERATURE REVIEW

2.1 THEMES OF A CRISIS………13

2.2 THE USE OF OVERT BEHAVIOURAL MARKERS FOR ASSESSMENT………17

2.3 THE USE OF VERBAL PROTOCOLS FOR ASSESSMENT………22

2.4 THE USE OF COGNITIVE ANALYSIS FOR ASSESSMENT………26

2.4.1 COGNITIVE PROCESSES………26

2.4.2 COGNITIVE SKILLS………28

2.4.3 COGNITIVE GOALS………32

2.4.4 ASSESSING COGNITIVE SKILLS………35

2.4.5 COGNITIVE ANALYSIS FOR ASSESSMENT – CONCLUSIONS………39

2.5 THE ATTRIBUTES OF A CRISIS MANAGER………40

2.6 THE ATTRIBUTES OF A CRISIS MANAGEMENT TEAM………43

2.7 THE ATTRIBUTES OF A CRISIS MANAGEMENT TEAM MEMBER………47

2.8 BEHAVIOURAL BIAS AND OPTION GENERATION………50

2.9 THE DECISION-MAKING RESEARCH ENVIRONMENT………51

2.10 FACTORS TYPIFYING DECISION-MAKING IN NATURALISTIC ENVIRONMENTS..53

2.11 CONTROL OF COMPLEX DYNAMIC TASKS………55

2.12 EXPERIMENTAL METHODOLOGIES FOR RESEARCH INTO DECISION-MAKING...57

2.13 NON-TECHNICAL SKILLS (NOTECHS)………61

2.14 CAUSAL CHAINS………64

2.15 PROCESS VERSUS OUTCOME………65

2.16 BEHAVIOURAL MARKERS AND THE ASSESSMENT OF CRISIS MANAGEMENT…68

2.17 SUMMARY OF RELEVANCE OF LITERATURE REVIEW TO RESEARCH…..………79

CHAPTER 3. REVIEW OF EXISTING ASSESSMENT PRACTICE

3.1 MILITARY……………………………..…………………………………………………….84

3.2 CIVIL AVIATION………………………………...……………………….…….……...88

3.3 OFFSHORE INDUSTRY……………………………………..……………………..91

3.4 FIRE BRIGADE………...……………………………….……..………………………...…93

3.5 ANAESTHESIOLOGY…………………………………………………...…………..…95

3.6 COMMERCIAL SHIPPING……………………………………………...…………….97

3.7 SUMMARY OF RELEVANCE OF ASSESSMENT PRACTICE IN OTHER DOMAINS.100

CHAPTER 4. RESEARCH METHODOLOGY (PHASE ONE)

4.1 BACKGROUND AND HYPOTHESIS........................………….................................…....104

4.2 ETHNOGRAPHIC ANALYSIS……………………………………....….............…......105

4.3 ETHICAL CONSIDERATIONS...…………...………......…....108

4.3.1 INFORMED CONSENT……………………………………........................……109

4.3.2 CONFIDENTIALITY……………………………….............................…………109

4.3.3 RIGHT TO REFUSE TO PARTICIPATE OR WITHDRAW……………….…………..…109

4.3.4 RISK ASSESSMENT…………………………………………………………………..109

4.3.5 COMPLAINTS PROCEDURE…………………………………..……………...….110

4.4 EXPERIMENT PARTICIPANTS...…………......….......110

4.5 EXPERIMENTAL PROTOCOL (PHASE ONE) ....……......... 112

4.5.1 INITIAL PLANT CONDITION……………………………………………………….114

4.5.2 INITIAL SITUATION……………………………………………………………....115

4.5.3 PERSONNEL ROLES……………………………………………………………....115

4.5.4 CRISIS SCENARIO INTERVENTIONS……………………………………..……….116

4.5.5 RECORDING OF NEW OBSERVED BEHAVIOURAL MARKERS……………………..117

4.5.6 RATING OF OBSERVED BEHAVIOURAL MARKERS…………………………………118

4.5.7 DETERMINATION OF EXPERIMENTAL ASSESSMENT FRAMEWORK……….……119

4.5.8 EXPERT EVALUATION OF EXPERIMENTAL ASSESSMENT FRAMEWORK………..119

CHAPTER 5. RESULTS AND ANALYSIS (PHASE ONE)

5.1 BEHAVIOURAL MARKERS ARISING OUT OF EXPERIMENTS……….....................121

5.2 EVALUATION OF BEHAVIOURAL MARKERS…………………………………….……122

5.3 CHARACTERISATION OF BEHAVIOURAL MARKERS……………………..………132

5.4 EVALUATION BY EXPERT CRISIS MANAGEMENT ASSESSORS…………………….136

CHAPTER 6. RESEARCH METHODOLOGY (PHASE TWO)

6.1 INTRODUCTION………..……………………………………..……………………..152

6.2 RESEARCH HYPOTHESES…………………………………....……………….……152

6.3 EXPERIMENTAL PROTOCOL (PHASE TWO)…………………..…….……………..……153

6.4 METRICS …………………...…155

6.4.1 INTERRATER RELIABILITY AND INTERRATER AGREEMENT……….….…………155

6.4.2 RANK ORDER……………………………………………………..….........….……160

6.4.3 OBSERVABILITY……………….…...……161

CHAPTER 7. RESULTS AND ANALYSIS (PHASE TWO)……………………164

7.1 INTERRATER AGREEMENT ………..………………………………………….……..165

7.2 RANK ORDER ……………………………………………………………………..……169

7.3 OBSERVABILITY…………………………………………………………………..………174

7.4 SUMMARY………………………………………………………………………………….176

CHAPTER 8. CONCLUSIONS ……………………………………………………….178

CHAPTER 9. REFLECTIONS AND RECOMMENDATIONS………………….……186

REFERENCES...……………………………………..................…188

APPENDICES…………………………………………….…………………….……….…..202

Appendix 1: Research Participant Consent Form

Appendix 2: Final Assessment Framework used by Group 2 Assessors

Appendix 3: Phase 2 Assessment Data

Appendix 4: Publications

LIST OF TABLES:

Table 1: Communications Behaviours.
Table 2: NOTECHS framework.
Table 3: Behavioural markers potentially related to competence in crisis management.
Table 4: Examples of Military Command and Control assessment criteria.
Table 5. Military Command and Control assessment rating scale.
Table 6: Examples of Civil Aviation Non-Technical Skills assessment criteria.
Table 7. Examples of Line Operations Safety Audits (LOSA) assessment criteria.
Table 8. Line Operations Safety Audits (LOSA) assessment rating scale.
Table 9. Examples of Offshore Industry assessment criteria.
Table 10. Examples of Fire Brigade assessment criteria.
Table 11. Fire Brigade Incident Command assessment rating scale.
Table 12. Examples of Anaesthesiology assessment criteria.
Table 13. Anaesthesiology assessment rating scale.
Table 14. Merchant Navy assessment criteria from the *Specification of minimum standard of competence in crisis management and human behaviour*.
Table 15. Comparison of the Assessment Frameworks for Assessing Competence in Crisis Management within different Safety Critical Organisations
Table 16. Research Participant Simulated Engine Room Environment Familiarisation Schedule.
Table 17. Behavioural markers arising out of experiments.
Table 18. Evaluation of Behavioural Markers.
Table 19. Number of assessment criteria in different assessment frameworks.
Table 20. Behavioural markers with the highest index.
Table 21. Characterisation of Behavioural Markers
Table 22. Evaluation by expert crisis management assessors.
Table 23. Behavioural marker indices from expert assessor evaluation.
Table 24. Comparison of behavioural marker indices.
Table 25. Comparison of all evaluation ratings and indices.
Table 26. Final Experimental Assessment Framework.
Table 27. Within-Group Interrater Agreement for uniform null distribution r_{wg_un}
Table 28. Rank Order of Chief Engineers as assessed by Group 2.
Table 29. Rank Order of Chief Engineers as assessed by Group 1.
Table 30. Rank Order of Chief Engineers as assessed by Group 3.
Table 31. Comparison of the Mean Rank Order of Chief Engineers by each group of assessors.
Table 32. Percentages of Behavioural Markers being Observed.

LIST OF FIGURES:

Figure 1: Communication Direction Diagram.

Figure 2: The 'wife/mother-in-law' reversible figure.

Figure 3: Cognitive processing module relationships with working storage and knowledge bases.

Figure 4: Relationship between behavioural marker characteristics and interrater agreement.

Figure 5: Assessment Rating Scale

LIST OF GRAPHS:

Graph 1. Mean Within-Group Interrater Agreement for Chief Engineers

Graph 2. Mean Within-Group Interrater Agreement for Behavioural Markers

Chapter 1. INTRODUCTION

1.1 BACKGROUND AND RATIONALE FOR THE RESEARCH

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.

On 19th November 1997, the 3,624 grt Bahamian registered vessel "Green Lily" grounded on the island of Bressay in the Shetland Isles in Force 10 winds and subsequently broke up. All crew members were rescued by a Coastguard helicopter but the helicopter winchman, who remained on the deck of the ship, was swept into the sea and lost. One of the reasons for the vessel grounding, determined by the UK Marine Accident Investigation Branch (MAIB) when investigating this incident, was that there had been *"inadequate teamwork"* by the vessel's crew.

The MAIB report, recalling previous investigations and noting the "Green Lily" investigation, advised that many of the accidents investigated showed that team cohesion failed when non standard emergency or crisis situations occurred, leading to rising levels of personal stress. Under these stressful conditions:

> *"engineer officers often show a lack of diagnostic skills, while deck officers fail to operate as an effective bridge team"* (MAIB, 1999)

One of the recommendations within the MAIB report into the "Green Lily" incident was directed towards the UK Maritime and Coastguard Agency (MCA), and it was:

> *"to commission a research study into how bridge and engine room simulators can best be used for bridge and engine room resource management training that includes escalating emergencies and increasing levels of stress. The results should be used to develop effective training for handling emergencies at sea"* (MAIB, 1999)

The UK MCA responded to this recommendation by commissioning a research project to study 'simulator training for handling escalating emergencies' (Habberley et al., 2001). One of this research project's recommendations 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."*

It was this research outcome of the Habberley et al (2001) project that was the catalyst for this new research into the field of using behavioural markers for assessing competence in crisis management for merchant navy engineer officers.

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.

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-making environment.

In certain circumstances, this demanding decision-making 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."*

Therefore, it is important within any safety critical organisation to try to 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 associations; 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 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 (JAR TEL) 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. However, 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 interrater 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 Seafarer’s Training, Certification and Watchkeeping (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 different safety critical organisations undertake the assessment of competence in crisis management in very different ways.

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.

1.2 AIMS AND OBJECTIVES OF THE RESEARCH

This research has been undertaken 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.

The aims of this research were:

- to understand how behavioural markers can be used to objectively assess competence in crisis management of merchant navy engineer officers.
- to understand the methods by which these behavioural markers can be elicited and assessed.

Data were collected and analysed using ethnographic study techniques during simulated crisis scenarios within a high-fidelity ship’s engine control room environment.

It should be stated that this high-fidelity ship's engine control room environment was supplied by the use of an engine room simulator. This simulator provided a realistic working environment and all of the environmental cues that would be present in a real ship's engine room control room. Without the use of such a simulator it would be impossible to undertake this research as crisis scenarios cannot be presented for research purposes in the real world, as this presents too great a risk to personnel, material assets and the environment.

This research has led to the development of an assessment framework that can be applied within the merchant navy engine room control room context for the fair and effective assessment of competence in crisis management.

Chapter 2. LITERATURE REVIEW

2.1 THEMES OF A CRISIS

In order to review the literature relating to competence in crisis management, the term 'crisis' should first be defined, and a differentiation should be made between a crisis and an emergency.

A crisis can be defined as:

> *"A crucial stage or turning point in the course of something, especially a sequence of events" (Makins, 1994).*

The implication from this definition is that a crisis situation will demand a decision be made; doing nothing is not an option.

An emergency can be defined as:

> An unexpected situation that requires the implementation of a previously planned procedure in order to prevent its escalation leading to possible harm to property, the environment and life.

The implication from this definition is that an emergency also demands that decisions be made, the difference being that in an emergency these decisions can be based upon a predefined response plan.

A differentiation can be made between emergencies and crises, which can be used to clarify what is meant by the term 'crisis management'. Habberley et al (2001) highlight the following issues that can lead an emergency to become a crisis:

- ➢ there is no emergency procedure for the particular situation presented
- ➢ the emergency procedure for the particular situation is not known to those involved
- ➢ the relevant emergency procedure is incorrectly followed due to a lack of emergency preparedness training

- there are multiple emergencies occurring simultaneously which have conflicting resource requirements

- there are insufficient resources to carry out the relevant emergency procedure

- the personnel involved have insufficient knowledge and / or experience to recognise the emergency situation

- the personnel involved have insufficient knowledge and / or experience to carry out the relevant emergency procedure.

Habberley et al (2001) also argue that if there is an overall lack of knowledge and experience within a team faced with an emergency situation, aspects of group polarisation and groupthink can play an important part in the escalation of an emergency to a crisis. Group polarisation being the observable fact that teams tend to be more extreme in their decision-making than individuals. Groupthink being the term used to describe the problem in decision-making where highly cohesive teams tend to strive for the unanimity of decisions, rather than appraising different courses of action (Janis, 1982).

A crisis could therefore be further defined as a situation that has developed from the escalation of an emergency, due to an inability to respond effectively to that emergency. This inability to respond effectively may be due to a lack of knowledge, experience, or resources, or any combination of these. In the American Bureau of Shipping review and analysis of maritime accident databases (American Bureau of Shipping, 2004) it is shown that in up to 28% of maritime accidents, the primary causal factor is the crewmembers' lack of awareness of the situation they are in. If there is a lack of awareness of the emergency situation facing a team, it will much more readily develop into a crisis. This is because the emergency may well have escalated outside the limits of any applicable emergency procedure, or the abilities of the team to deal with the emergency, before the team becomes aware of the emergency situation they are facing.

Borodzicz and van Haperen (2002) propose that a crisis is an ill-structured and complex event that requires the respondents to interactively create a solution. This supports the view that in a crisis the respondents do not have the luxury of following an already laid down course of action in order to recover the situation.

Barnett et al (2002) summarise the difference between a crisis and an emergency as:

> *"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."*

What are therefore the skills required to handle a crisis? They must be skills that enable the crisis manager to think through the situation they find themselves in, to move towards an understanding of the situation, and to respond in creative and flexible ways as the situation develops, in order to facilitate a successful outcome. If the crisis managers have these skills, how are we then to assess their competence in managing a crisis?

The assessment of competence in crisis management could be undertaken in a number of different ways. It has been argued that the way in which any phenomenon is assessed is dependent upon the way that phenomenon is initially conceptualised and defined (Kuhn, 1970). If competence is conceptualised as consisting of knowledge and skills, and the ability to apply these in the context where the competence is required, then its assessment must also relate to these dimensions. Although some form of examination can be used to assess the level of underpinning knowledge and the level of knowledge of skills, and even the level of knowledge about the application of these skills, it cannot assess the ability to practically apply these skills. In order to be able to assess the practical application of the underpinning knowledge and skills of competence, some form of performance-based assessment is required.

In order to be able to undertake a performance-based assessment for competence in crisis management it is first necessary to define the set of knowledge and skills that are required for a candidate to be a competent crisis manager. It is against this "standard" that any assessment of performance will be made. It is then necessary to design some form of assessment regime in which the candidate being assessed can demonstrate their ability to apply their knowledge and skills, in order to show they can perform to a level that is acceptable when measured against the predefined standard of competence. It is the capability of the assessment regime to elicit a performance by the candidate that resembles as closely as possible their performance in the real world that will ultimately determine if the assessment regime is successful or not. However, the assessment regime not only needs to elicit this 'real' performance, but it also has to capture and evaluate the markers against which this performance will be assessed.

The construct validity of any assessment framework is very important. Does the method of assessment actually measure that which it is trying to measure? In the case of competence in crisis management, could a paper and pencil test, a computer based assessment exercise, or viva voce type examination, be a valid method of assessing competence?

In order to answer this question it is first necessary to define what is meant by competence. There are many various definitions of competence, but most relate to the application of knowledge to perform a required skill to a predetermined standard. The UK Health and Safety Executive (2003) states that:

> *"competence is equated to an ability to perform to the expected standard. Hence, competence assessment entails measuring a person's performance against a standard."*

If it is a person's 'performance' that needs to be measured, the assessment environment must allow the person being assessed to 'perform'. This 'performance' will have to include the elements that are set down within the standard against which the competence is being assessed.

Gipps (1994) argues that "*there is an intimate connection between skills and the contexts in which they are used*". If competence in crisis management is a skill, and it is this skill that is to be assessed, the context in which this skill will be measured against a standard would appear to be very important.

Resnick and Resnick (1992) argue that:

> *"We cannot teach a skill component in one setting and expect it to be applied automatically in another. This means, in turn, that we cannot validly assess a competence in a context very different from the context in which it is practiced or used."*

In order to ensure the construct validity of the assessment, it would therefore seem appropriate to assess competence in crisis management in the context of a simulated crisis scenario. This would provide a context that was as close as possible to the context in which the competence of crisis management was used.

Following an extensive study into competence assessment methods used in different hazardous industries the UK Health and Safety Executive (2003) concluded that in relation to the competence assessment of interpersonal skills, team management skills and safety behaviours and attitudes, whilst psychometric personality tests may provide a prediction of these skills, behaviours and attitudes:

> *"Observation of actual behaviour in the work setting using behavioural markers tends to provide a more valid measure."*

It will be seen from the discussion later in this chapter that the management of crises requires the application of interpersonal skills and team management skills, as well as safety behaviours and attitudes. Because of this, the observation of actual behaviour in a setting that is as close to the work setting as possible is thought to be the best way of assessing competence in crisis management.

If we are to assess competence in the management of crises, we need to be able to in some way observe the crisis manager at work. We need to be able to ascertain what actions the crisis manager takes as elements in a process towards the outcome of the crisis scenario. The actions taken can be characterised by the behaviours exhibited during their enactment. It is therefore feasible that by relating these overt behaviours to the outcome of a crisis scenario, some measure of competence in crisis management may be possible. The following sections will discuss the relative merits of the use of behavioural markers, verbal protocols, and cognitive analysis, as metrics for the assessment of competence.

2.2 THE USE OF OVERT BEHAVIOURAL MARKERS FOR ASSESSMENT

Behavioural marker systems were first developed in the aviation industry for the training and assessment of flight crews (Helmreich et al, 1994). Behavioural markers systems have subsequently been developed in other safety critical industries and organisations such as nuclear power generation, the military and medicine. In the report from the Kolleg Group Interaction in High Risk Environments Behavioural Markers Workshop (Klampfer & Jochum, 2001), a series of statements were made proposing what the workshop participants considered to make a good behavioural marker:

- *it describes a specific, observable behaviour, not an attitude or personality trait, with clear definition (enactment of skills or knowledge is shown in behaviour);*
- *it has demonstrated a causal relationship to performance outcome;*
  - *it does not have to be present in all situations;*
  - *its appropriateness depends on context;*
- *it uses domain specific language that reflects the operational environment;*
- *it employs simple phraseology;*
- *it describes a clear concept.*

Although helpful when considering whether a particular behaviour should be used as a marker, these statements are open to some interpretation. Within the context of a crisis scenario, whether a particular behaviour can be demonstrated to have a direct causal relationship to the overall performance outcome of the crisis may be difficult. Whereas, the causal relationship between a particular behaviour and the performance outcome of a discrete task or subgoal within the overall management of the crisis, might be more easily established. One of the statements makes the point that behavioural markers have a contextual specificity, implying that behavioural marker systems are not directly transferable between domains. The workshop report goes on to state, *"behavioural marker systems do not transfer across domains and cultures without adaptation"*. This is because behavioural markers are directly related to the overt behaviours exhibited by the persons undertaking tasks within their working environment, and these tasks are context specific.

Klampfer and Jochum (2001) also discuss the characteristics of good behavioural marker systems, and propose the following:

- *Validity: in relation to performance outcome;*
- *Reliability: interrater reliability, internal consistency;*
- *Sensitivity: in relation to levels of performance;*
- *Transparency: the observed understand the performance criteria against which they are being rated, availability of reliability and validity data;*
- *Usability: easy to train, simple framework, easy to understand, domain appropriate language, sensitive to rater workload, easy to observe;*
- *Can provide a focus for training goals and needs;*
- *Baselines for performance criteria are used appropriately for experience level of rater (i.e., ab initio vs. experienced ratees);*
- *Minimal overlap between components.*

The relationship between the validity of the behavioural marker system and the performance outcome raises the interesting question of whether this performance is overall performance, or performance related to subgoals. If using a behavioural marker system to assess performance in the workplace, then relating the assessment to overall performance outcome may indeed be worthwhile and indeed necessary. However, if using a behavioural marker system to assess competence in a simulated environment, could the validity of the system be related to the successful completion of subgoals, and not to overall performance outcome? This issue will be discussed later in section 2.15.

Any behavioural marker system that is used must be reliable, objective and unambiguous if it is going to provide consistent assessments when used by different assessors. As with all assessment frameworks, the interrater reliability and agreement are powerful metrics for measuring this reliability. The reliability of the assessment framework will stem from its usability. If raters find the system is easy to understand and use, it will be reliable.

The sensitivity of any assessment framework to the level of performance is also very important, especially when assessing competence. On the one hand, the assessment framework must ensure that those who are not competent are not assessed as being competent, and on the other hand, those who are competent should not be assessed as being not competent. Many competency assessment frameworks only have two assessment levels, “competent” and “not yet competent”. Therefore, in order for these systems to give a fair and objective assessment, they have to be very sensitive in relation to the levels of performance of candidates. Also, if candidates of differing levels of experience are to be assessed, the assessment criteria used must be shown to be valid for those differing levels of experience.

The transparency of the performance criteria is also an important issue, and it can be argued that it would be unfair, and unethical, for a candidate to be unaware of the criteria against which they were to be assessed. However, in relation to a behavioural marker based assessment system, it does raise the issue of whether candidates can learn how to exhibit the correct discrete behaviours that are the behavioural markers, rather than learning to be wholly competent in the task being assessed. The argument against this issue being a problem is that, if a candidate exhibits the behaviours that are the behavioural markers of a competent performance, it follows that they are competent. However, this also goes to show how necessary it is that the behavioural markers used are indeed valid, and can be shown to be valid.

The characteristics proposed by Klampfer and Jochum (2001) for the provision of focus for training goals and needs, and the minimal overlap between components of the behavioural marker system, are generic characteristics of any good assessment system.

Klampfer and Jochum (2001) go on to suggest that behavioural marker systems are limited because they *"cannot capture every aspect of performance and behaviour"* due to the:

- *limited occurrence of some behaviours*
  - these are important but infrequent behaviours, such as conflict resolution;

- *limitations of human observers*
  - distraction, overload (e.g., in complex situations, or when observing large teams).

It is true that during any competence assessment based in the workplace it may not be possible to observe all of the behaviours that make up the assessment framework because the opportunities for those behaviours to be exhibited may not arise. However, if the assessment is undertaken in a simulated environment, the assessment exercise can be designed to ensure that opportunities are presented for all of the behaviours within the assessment framework to be exhibited.

The limitations of the assessors to be able to fully capture, and interpret, all of the behaviours being exhibited during the assessment is a serious limitation to the use of behavioural markers systems, in both workplace and simulated environments. However, at least in simulated environments it may be possible to capture most aspects of candidates' performance and behaviour, with the use of video recordings. As long as time and resources then permit, even the behaviours exhibited during complex situations, and amongst large interactive teams, could then be captured and assessed.

The behavioural marker assessment 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. This must include the interaction between the candidate being assessed and the rest of the team members within the assessment environment. It must also capture any aspect of the assessment environment that might impact upon the candidate's performance, such as the cues being generated within the environment, the hardware that makes up the environment, and the complexities of the assessment scenario. Klampfer and Jochum (2001) give an example where the candidate being assessed is part of a team. They make the point that the behaviour of any one of the other team members within the candidate's team, could have either an adverse or positive affect on the assessment of the candidate. They argue that behavioural marker based assessment frameworks should be designed to detect and record such effects.

The argument that if a candidate is being assessed when performing as part of a team, then their assessment may be affected by the actions of the other team members, is based upon the premise that the behavioural markers being used within the assessment framework may be affected by the other team members. Some behaviours of the candidate will surely be affected by the actions of other team members, however, it might be possible to determine some behavioural markers that are not affected by the actions of other team members. It is these markers that would be the most efficacious in forming an assessment framework for assessing a candidate performing as part of a team, as they could lead to an assessment that was independent of the actions of other team members. However, dealing with difficult or disruptive team members is a reality, and so could be treated as just another element of the assessment environment. In fact, the actions of other team members could be viewed as bringing more 'real-world' fidelity to the context in which the assessment is being undertaken.

Oser et al (1989) determined a set of overt behaviours that discriminated between more and less effective teams. This set of indicative behavioural markers comprised:

- *Thanking another team member for pointing out a mistake;*
- *Helping other team members having difficulty;*
- *Making motivating statements;*
- *Praising other team members;*
- *Suggesting ways to find an error.*

This study by Oser et al. is evidence that a relatively small set of behavioural markers can be used to successfully assess the level of effectiveness of a team. If these overt behaviours are metrics that can be used to discriminate between more or less effective teams, they could also be useful in assessing competence in crisis management. A team that is effective is more likely to be good at managing a crisis, and more likely to have a competent team leader.

The use of behavioural markers is generally accepted as a valid approach to the assessment of competence in many safety critical industries and organisations. However, the way in which these various industries and organisations both use and validate their behavioural marker based competence assessment frameworks can be quite different. A number of these assessment frameworks, and their associated problems, are discussed in chapter 3.

Throughout this literature review, any potential behavioural markers that could be used in the assessment of competence in crisis management will be recorded and evaluated in more detail in chapter 5.

The use of overt behavioural markers as assessment criteria is a way of directly observing the outputs of the cognitive processes taking place within the minds of those being assessed. However, there are also assessment methods that try to assess the cognitive processes themselves, by using techniques to infer the cognitive processes that lead to the decisions that produce the observable behaviours. One such technique is the analysis of verbal protocols.

2.3 THE USE OF VERBAL PROTOCOLS FOR ASSESSMENT

Another method of assessing competence in crisis management is to try to assess the cognitive processes used by crisis managers to reach their decisions. Although we do not have direct access to the cognitive processes of crisis managers (Miller, 1962), verbal protocols have been used to infer the cognitive processes that have led to decisions within crisis situations. Competence is then assessed by subject matter experts analysing the inferred cognitive processes and comparing these to the cognitive processes expected to be used by an expert. However, the use of concurrent verbal reporting, or 'thinking aloud', during team based crisis management exercises can be very disruptive to the management task. Team members can become confused between the communication required to undertake the task and the concurrent verbal reporting. During a crisis situation there can often be a greater level of communication between team members, and it can be more critical that this communication is effective. Therefore, the use of concurrent verbal reporting to facilitate assessment can have a very detrimental affect on the performance of the team, by getting in the way of essential communications that are required in order to effectively manage the crisis situation.

Bainbridge (1974) argues that in order to infer cognitive processes from concurrent verbal reports, it is necessary to assume that there is an undistorted mapping between the cognitive processes and the verbal protocols. Two of the cognitive processes that might be inferred from the use of verbal protocols are “feedback control”, a reactive management approach, and “predictive control”, a proactive management approach. Bainbridge proposes that inexperienced process operators use feedback rather than predictive control when undertaking a process task. Experienced operators use a greater degree of predictive control. If it is assumed that predictive control is necessary for the effective management of a crisis, then the ratio of the degree of feedback control exhibited to the degree of predictive control exhibited may be a behavioural marker that can be used to assess competence in crisis management.

Bainbridge also observed that even when some operators have correctly verbalised the next action to take, they have difficulty in committing to taking the action. It is suggested that the depth of consideration given to the decision to take an action is related both to the amount of processing time available and to the level of knowledge of the task. If time is short, such as during a crisis situation, there may be trade-off between the speed of decision-making and the efficacy of the decision. It could therefore be argued that the use of concurrent verbal reporting will inevitably use up some of the time available to the crisis manager for decision-making, and could therefore affect the efficacy of the decisions made, possibly resulting in a poorer outcome and consequently a poorer assessment.

In order to overcome the problems of concurrent verbal reporting mentioned above, some researchers have used a methodology that only analyses the communication between team members, and does not require team members to also try and concurrently verbalise their thought processes. Serfaty et al (1994) proposed a system for analysing and assessing team communication and coordination by mapping message destinations against message type and content. The system utilised a trained observer who recorded the direction, type and content of each utterance onto a matrix. Instances of type referred to whether, for example, the message was a request or a transfer. Instances of content referred to whether, for example, the message was providing information, calling for a specific action or calling for a complete task to be undertaken. These metrics were analysed to see where improvements could be made in the team’s performance through what the researchers called “Team Adaptation and Coordination Training” (TACT).

Communication patterns were measured by the TACT research team in both high and low stress scenarios. The areas where it was found that teams were weak and required training were:

- lack of information flow towards the team leader from other team members;

- low level of anticipation of future required actions and the needs of other team members.

The research team cited these deficiencies as support for their hypothesis that:

> *"Implicit team coordination mechanisms are essential to effective adaptation to stress."*

Communication direction diagrams, such as that shown in Figure 1., were used to map information flow within teams.

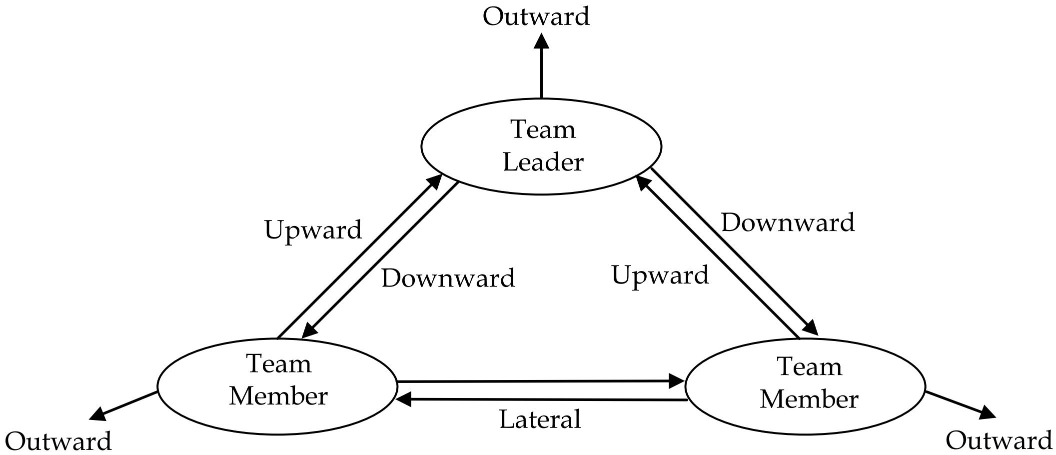


Figure 1: Communication Direction Diagram (Serfaty, Entin and Deckert, 1994)

Serfaty and Entin (1997) proposed that reciprocal monitoring and backup behaviours between team members can lead to each team member learning about the other team members' needs, proclivities, and tasks, and thereby develop highly congruent mutual mental models of the situation they are working in. Once a congruent mental model has been established the team members' communications can shift from being explicit to implicit, and they can start to anticipate the needs of the other team members. By giving periodic situation assessment updates to the other team members, a team leader can improve the congruence and validity of the mental model shared by the team. They go on to argue that such behaviours increase team performance and efficiency, particularly during emergency or crisis situations.

The level of anticipation of a team member, or team leader, to the needs of other members of the team, and of future action and task requirements, can therefore be seen as a metric that could be used for the assessment of competence in crisis management. This metric could consist of the amount of relevant information communicated in anticipation of need, as a ratio of the amount of relevant information available, and could be used as a behavioural marker. However, this type of behavioural marker may be found to be difficult to use during live assessments, where the assessor is making observations solely during the assessment scenario and not from a recording of the scenario, because of the amount of data the assessor would have to keep track of.

Breaking a complex task down into easily managed sections, each with its own goal, has been shown to be a cognitive process employed when a complex task is undertaken, and has been termed 'subgoal formation' (Resnick, 1987). Empirical evidence for this has come from studies of industrial plant operators actually undertaking complex dynamic tasks where records have been made of the verbal interaction between the operators (Pew, Miller, & Feeher, 1981; Rasmussen, 1986; Reinartz & Reinartz, 1989). In these studies the authors propose that the moment-to-moment thinking of the operators can be inferred by their verbal commentary as the operators interact with the industrial process to achieve the goals of the task.

By an analysis of the verbal transcripts of operators responding to a process fault, Pew et al. (1981) inferred that a number of different types of cognitive processes occurred such as interpretation, expectation, intention, and action. A distinction was made between the task goals and the cognitive goals involved in meeting the task goals. For example, the task goal may be to maintain a steam pressure of 20 bar. The cognitive subgoals involved in meeting this may be to find out what the pressure is now, to determine if this is within acceptable limits, and to determine, if required, how the pressure can be restored (Bainbridge, 1992). The overall task goal is therefore broken down into manageable cognitive goal modules (i.e. subgoal formation takes place). It is the identification of these cognitive processing modules that the studies of operator's verbal commentaries have tried to achieve. Although the specific goals associated with these modules may vary from task to task, these studies have proposed that the cognitive processing within the modules can be seen to be applicable to any task.

The analysis of verbal protocols is one way of analysing the cognitive processes of the crisis manager. However, as discussed above, this analysis is complex, and in order to infer the cognitive processes, extremely detailed studies of the verbal protocols need to be undertaken. This means that the assessment of competence in crisis management by the use of verbal protocols could only be achieved through a detailed analysis of a recording of the assessment scenario, and could not be carried out in real time.

A further method of assessment is to infer the cognitive processes by analysing the metacognition of a person. This method is known as cognitive analysis, and uses interviews to codify the knowledge a person has about their own cognitive processes and the factors that may have influenced these processes.

2.4 THE USE OF COGNITIVE ANALYSIS FOR ASSESSMENT

In order to evaluate the possible use of cognitive analysis for the assessment of competence in crisis management, a number of theoretical perspectives related to cognition need to be reviewed. This section will first discuss theories of cognitive processing and will then discuss the cognitive skills in which competence may be required to allow complex dynamic tasks to be undertaken successfully. The discussion will then move on to our understanding of the way we structure and use knowledge to meet our cognitive goals when undertaking a complex dynamic task. The final part of this section will look at the way in which these cognitive skills might be used for the assessment of competence in crisis management.

2.4.1 Cognitive Processes

The analysis of any cognitive processes involved in the management of a crisis is made difficult by virtue of the fact that cognitive processes are covert. There have been a number of theories that place the cognitive processes within an information-processing model of input-processing-output (McCormick, Jeanneret, & Mecham, 1969; Miller, 1974; Cunningham, Boese, Neeb, & Pass, 1983). These theories attempt to distinguish the cognitive processes occurring at different stages of the information processing sequence. Reinartz and Reinartz (1989) observed process operators undertaking complex dynamic tasks during the operation of a nuclear power plant. These authors attempted to relate the overt activities of the operators to the covert cognitive processes behind them. They categorized the cognitive processes into five activity groups:

- *data/information collecting;*
- *information processing;*
- *planning and strategy;*
- *actions;*
- *team specific: information distribution, task allocation, and management.*

If the overt behaviours relating to the exemplary performance of these activities can be identified, they could be used to assess competence in undertaking complex tasks. When different complex tasks, such as the management of crises, are analysed, it has been found that the same types of cognitive processing skills are used (Simon, 1976). This suggests that if it were possible to train crisis managers to improve their performance in using these cognitive processes, operational performance could be improved across a range of different complex tasks. Some of the common cognitive processes employed when a complex dynamic task is undertaken have been identified as:

means-end analysis: comparing the desired outcome of a task with that which may be achieved by current actions (Resnick, 1987);

subgoal formation: breaking the task down into easily managed sections each with its own goal (Resnick, 1987);

generate and test routines: generating actions to test hypothesis (Resnick, 1987);

symptom matching: matching symptoms to a related condition (Rasmussen, 1984);

application of heuristics: applying known rules or procedures (Patrick, 1989).

These information-processing models are highly discriminating, breaking down the cognitive processes into discrete psychological aspects. Because of their discrete nature, these aspects are not dependent upon the context of the task in which they are involved. The failure of these models is that they do not identify any interaction between the discrete psychological aspects of the overall cognitive process. This interaction is thought to be controlled by some higher-order cognitive activity (Sternberg, 1980; Feuerstein, 1980). For example, the psychological aspects associated with responding to an industrial process alarm may be:

- *activation (detection of the need for action);*
- *orienting response;*
- *information seeking;*
- *hypothesis testing* (Reinartz & Reinartz, 1989).

These aspects follow a logical sequence, but to direct this sequence it seems reasonable to assume that some controlling influence will be required. The processes that make up this controlling influence have been termed ‘metacognitive’ processes. When attempting to analyse any complex dynamic task it is therefore necessary not only to identify the discrete psychological aspects, but the controlling metacognitive processes as well (Patrick, 1992). Brown, Bransford, Ferrara, & Campione, (1983) view a person's ability to control the discrete psychological aspects of a task in terms of metacognitive skills. They propose that these skills are used to keep track of one's own understanding. They also relate these skills to the effectiveness of learning, by the organisation of attention. However, some theories have suggested that these metacognitive skills appear to be dependent upon specific knowledge (Voss, Greene, Post, & Penner, 1983; Glaser, 1984). If these theories were correct, then it would be best to develop these skills within the context of their use where the specific knowledge is available. These theories therefore also bring into question the degree to which metacognitive skills can be developed within one context of use and then be transferred for use within a different context.

So what are the cognitive skills required to effectively undertake a complex dynamic task such as the management of a crisis, and how do we go about the analysis of these skills in order to facilitate their assessment? The following section will discuss these issues.

2.4.2 Cognitive Skills

If the performance of a team leader undertaking a complex dynamic task, such as the management of a crisis, is to be accurately assessed, it will first be necessary to analyse the cognitive skills required by the operator. Patrick (1992) proposes three areas of study for this analysis:

- *identifying the specific knowledge associated with the task;*
- *determining alternative representations of the same knowledge;*
- *assembling types of knowledge and their representations into a complete and coherent model of expert performance.*

Knowledge can be categorized into two types, declarative and procedural (Anderson, 1985). Declarative knowledge is factual; it can be stated and made explicit and is not context related. Procedural knowledge relates to how a task is performed and is often implicit and is context specific. Anderson (1985) proposes that cognitive skills are acquired by the transformation of declarative knowledge into procedural knowledge. He suggests that this transformation initially uses general problem solving procedures and analogies to allow the plant operator to develop appropriate procedures while actually doing the task. He, further, proposes that the acquired cognitive skills are improved through the practice of these procedures. During this practice, the procedures are tuned, and each time they are successfully used makes them more likely to be activated again in the future. The implication from this process is that any assessment of a candidate's specific procedural knowledge of the elements within a complex dynamic task, such as crisis management, can only be obtained by an 'in-context' practical demonstration of competence.

The procedural knowledge is a set of rules relating to how a task is performed. The construction of these rules is accomplished by doing the task (Anzai & Simon, 1979). Even if a plant operator is told in advance the rules that relate to the performance of a task, until that task is performed, the operator will not be able to internalise them. Cognitive skills are developed by "doing" (Patrick, 1992).

Depending upon the domain in which a task is undertaken, there will be many different types of declarative knowledge. This knowledge will be internally structured by the plant operator in some way. Certain theories have indicated that this structure takes the form of a mental model (Gentner & Stevens, 1983; Johnson-Laird, 1983). As this declarative knowledge is transformed into procedural knowledge, the mental model becomes dynamic (Bainbridge, 1993). The operator uses knowledge of plant constraints, inferences of process state to response required, process time constants, plant variable associations, and typical event sequences to build this dynamic model. Gentner and Stevens (1983) propose that these mental models will provide predictive and explanatory abilities for the operator to understand their interaction with the plant. These abilities will be especially important if the operator finds himself or herself leading the response to a crisis situation.

Depending upon the context in which the crisis manager structures any knowledge and their own perspective of that knowledge, the same knowledge may be represented in different ways (Ohlsson, 1986). Wittgenstein (1963) proposed that the aspect of an object that we perceive is not a property of the object, but an internal relation between it and other objects. This can be graphically demonstrated by looking at reversible figures such as that shown in Figure 2. How we internally relate particular elements of this drawing affects the way in which we perceive it. The drawing offers two aspects, one of an ugly old woman and one of a beautiful young woman. By internally relating the individual elements of this drawing in two different ways, our perception is radically changed, to the extent that what we perceive to be the eye of the ugly old woman, becomes the ear of the beautiful young woman.

Figure 2: The 'wife/mother-in-law' reversible figure.

The way in which we perceive cues from the environment may, therefore, be said to be influenced by the context in which we see them. The same could be said to be true of the crisis management process. How the crisis manager structures the cues from the environment that he or she retrieves will determine what interpretation he or she may place upon them. The crisis manager must, therefore, be aware of this so that he or she can prevent any misinterpretation by guiding their transformation of knowledge and questioning their interpretation of the knowledge gained.

Observations during machinery space simulator training courses have shown evidence of the benefit of crisis managers questioning their interpretations of knowledge. When working as a team, team members can be taught to review their interpretations of knowledge. Through their individual perceptions, the members of the team will have different interpretations. By discussing these differences of interpretation the trainees often arrive at an improved shared interpretation. Fewer misinterpretations have been observed when this team review is carried out (Gatfield, 1999).

This team review process uses two alternate representations of the same knowledge, the pattern-type representation of the visual stimuli from the plant instrumentation, and the semantic type representation of the discussion between the team members (Bainbridge, 1988). Both representations are used together to arrive at an interpretation of knowledge upon which an action may be based. The idea that the way in which we group stimuli and knowledge, causes us to interpret and structure them in a certain way, has been put forward as part of the Gestalt Theory (Wertheimer, 1959). One of Wertheimer's hypotheses is that in order to be a successful problem solver, a person requires the skill to be able to see the overall structure of a problem. He proposes that the problem solver first characterises a certain region of the problem structure as being crucial. By focusing on the interaction of the different parameters of the problem with the crucial region, the structure of the problem becomes better understood. This understanding enables the problem solver to make reasonable predictions to test his or her hypotheses for a solution. The cognitive skill of being able to structure a complete and consistent picture of the problem situation is seen as essential to the success of problem solving behaviour. Gestalt Theory is, therefore, a way of looking at how cognitive processes are organised and structured overall, and how this overall organisation and structure affects the cognitive skill of a person.

In the process of analysing cognitive skills, reference can be made to a model of best performance. This is the model of how an expert in a particular field performs within that field. The expert will use numerous types of knowledge in many different ways when undertaking the skilled performance of a complex dynamic task (Singleton, 1989). In an example of an operator diagnosing a fault with an industrial process, Singleton observes the following:

- *most of the information used during the diagnosis does not come from stimuli or cues, but from their absence;*
- *the operator relies on a hybrid model of the plant using heuristics, mental pictures, and symbolism;*
- *there is little conscious guidance to the cognitive process, it takes more the form of a chain of events, each cognitive or physical event leading on one to another.*

The overt nature of operators' interaction with the stimuli, or absence of stimuli, from the plant, has been observed during exercises within a machinery space simulator. Gatfield (1999) observed that plant operators will commonly take actions to test hypotheses concerning the plant, based on the implicit information given by the lack of an abnormal stimuli or cue. Singleton's (1989) observation that most information comes to the operator by virtue of an absence of stimuli can, thus, be confirmed. In an attempt to categorise the types and levels of knowledge used when an expert undertakes a complex dynamic task, Rasmussen and Lind (1981) put forward five levels of abstraction to represent the functional properties of a system:

- *system purpose – objectives, constraints, inputs/outputs*
- *abstract function – representation of system function such as 'mass energy flows'*
- *generalised function – process description – feedback loops, heat transfer*
- *physical function – mechanical, electrical or chemical processes*
- *physical form – actual physical appearance of the system*

They imply that these levels of abstraction can be used when considering the plant in different unit sizes, from a whole plant overview, down to individual equipment component parts, to follow the operator's cognitive processes. During observations of the performance of a complex dynamic task, Rasmussen (1986) plots the level of abstraction being applied to the particular plant unit in an attempt to map the cognitive sequence of events. Although this methodology is useful in showing that different types and levels of knowledge are used when undertaking a complex dynamic task, this map fails to incorporate the metacognitive processes that make the map dynamic. Our observations of complex dynamic tasks are limited to the overt behaviour of the operator and the articulation of the cognitive process by the operator.

From the discussion above it can be seen that the way in which knowledge is structured and used is very important in the management of complex dynamic tasks, as are the cognitive goals that make up the 'waypoints' on the task manager's cognitive sequence map. The following section will discuss theories of knowledge management and cognitive goal modules, and relate these to the assessment of competence in crisis management.

2.4.3 Cognitive Goals

One view of the structures of knowledge is given by Bainbridge (1992). She looks at the structures at a higher cognitive level than Rasmussen, relating them to memory function and metacognition. She proposes three ways in which knowledge is used in cognitive processing:

- knowledge of permanent or potential characteristics of some part of the external world (long term memory or knowledge base);

- temporary inferred knowledge about the present or predicted state of the external world (short term memory, operational memory, working memory);

- knowledge of the outcomes and properties of the operators own behaviour (metacognition).

The concepts of a knowledge base, working memory, and metacognition can be useful in trying to understand the cognitive processes that occur when undertaking a complex dynamic task such as the management of a crisis. Bainbridge (1992) suggests that skilled plant operators build up and maintain an overview of the present state of the plant in their working memory. She proposes that they will form an idea about the way they expect the plant to perform dynamically and will formulate plans for any future actions they may need to take to maintain plant safety and performance. This overview will provide the context for the operator's future decision-making and maintains their situational awareness. The same cognitive processes can be expected to be used during a crisis situation to inform any "critical thinking", the formulation of plans for future actions, leading to the generation of multiple options, from which the 'best' option can be chosen.

In order to maintain this overview, Bainbridge proposes that reference is made to the knowledge base held within the plant operator's long-term memory. This includes experiential knowledge, which is important for testing hypotheses relating to future actions. The operator's previous experience is the basis for the explanatory hypotheses he or she suggests (Bainbridge, 1995).

Because of the many different ways in which complex dynamic tasks within a crisis situation can be undertaken, the crisis manager has a degree of flexibility in choosing the way in which he or she undertakes any given task. There may indeed be different appropriate working methods, even within the framework of applicable procedures. If the knowledge of alternate working methods is available to the crisis manager, their behaviour in choosing one of these methods is an aspect of metacognition. The way an operator will behave is a product of their past experience (Bainbridge, 1992). This has implications for training and assessment, in that by giving crisis managers experience with safe and efficient behaviour, by choosing different working methods, they will have a greater knowledge base to assist them in making future choices.

The studies referred to above have put forward many possible cognitive processing modules, each of which has a particular cognitive goal. This goal is related by Bainbridge (1992) as *"to build up a temporary structure of information about the current state of the task."* Some of the proposed modules are:

- *identify;*
- *predict state;*
- *review task goals;*
- *review present state;*
- *review action availability/effect;*

- *evaluate state;*
- *predict task goals;*
- *infer present state.*

These modules make reference to information available from the operator's knowledge bases and the operator's mental models within their working storage memory, as well as directly from the plant itself. Figure 3. shows the proposed interaction of the 'identify' module with the operator's working storage and knowledge bases. We can see from this model that, in order to identify the stimulus from the plant or to obtain information to use in the testing of hypotheses relating to the stimulus, reference is made to the operator's knowledge bases. Required responses to the state of the plant, and set procedures to be followed to make these responses, are also referenced from the knowledge base. Information about the previous and predicted state of the plant is also used by the processing module, drawn from the operator's working storage memory, where a mental model of the plant is maintained.

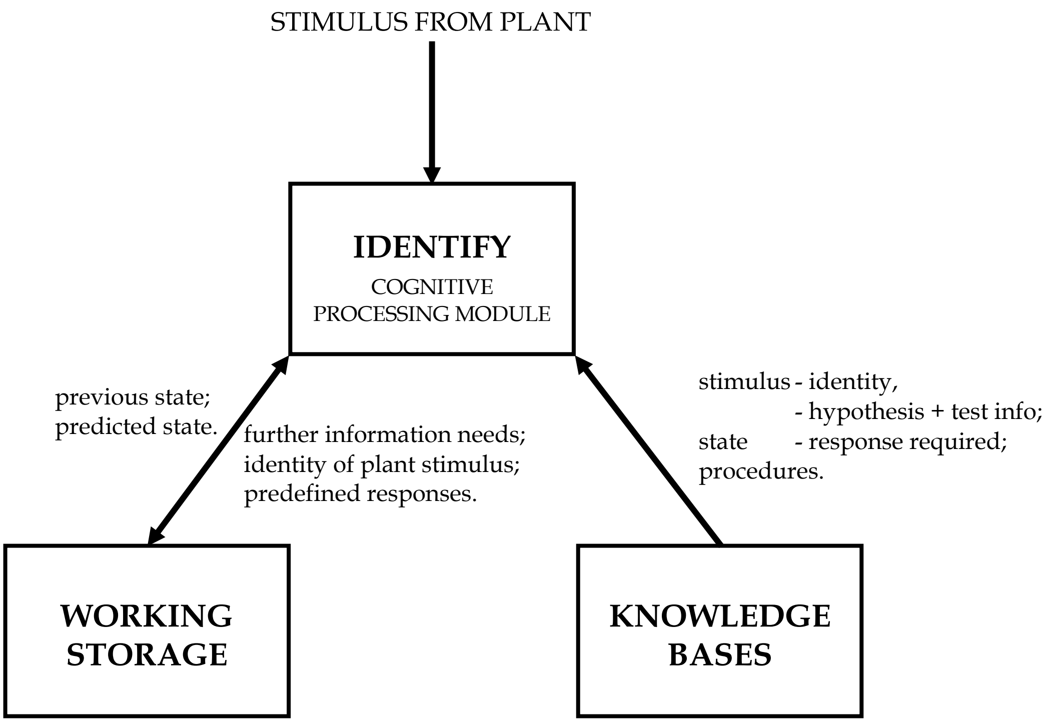


Figure 3: Cognitive processing module relationships with working storage and knowledge bases (after Bainbridge, 1992).

Depending upon the knowledge held concerning the stimulus received, it may be necessary to draw upon further information from the working storage to complete the identification of the stimulus. If the stimulus is identified from the knowledge base, this information can be passed to the working storage to be incorporated into the dynamic mental model of the state of the plant. Also, any predefined responses to the stimulus held within the knowledge bases can be incorporated into the model.

The way in which these cognitive processing modules are structured together into some rational sequence will constitute the working method of the operator. In order that this structure of modules will achieve the necessary task performance the operator must have the associated knowledge bases and be able to maintain an overview of the state of the task (Norman, 1986).

The sequence of cognitive processing modules a crisis manager uses in undertaking a task shows their strategy for achieving the task. If these sequences can be mapped through the observation of crisis managers within a plant simulator, the effectiveness of the chosen strategy should become apparent to an assessor and this could form the basis of an assessment framework. Initially, novice crisis managers may follow set rational processing module sequences to achieve set task goals. More experienced crisis managers may reduce the number of processing modules in their sequences by the use of appropriate preset responses and prelearned cues, allowing certain modules to be circumvented (Rasmussen, 1986). If the processing module sequence can be mapped, the crisis manager’s choice of subgoals could be analysed and compared against models of expert behaviour for assessment. Any significant deviation from expert behaviour could be addressed as part of a training programme. Patterns of errors that become evident through viewing crisis manager’s performance at a macroscopic level are useful when designing these training programmes (Patrick, 1992).

Samurcay & Rogalski (1988) argue that the strategies used by crisis managers to achieve their subgoals could also be analysed by the observation of crisis managers within a plant simulator. By analysing a crisis manager’s strategy when using a cognitive processing module, such as retrieving the current state of the plant, it may be possible to infer the level of the competence of the crisis manager. The next section will explore this potential use of cognitive skills as part of a competence assessment framework.

2.4.4 Assessing Cognitive Skills

There have been a number of training programmes produced that aim to improve the higher order cognitive skills of the trainees within a specific context (Woods, 1983; Wales & Nardi, 1985).

These programmes use a number of techniques to try and improve the trainees higher order cognitive skills such as problem solving. Some of the techniques used are:

- extensive practice of solving problems;
- the use of heuristic strategies;
- use of graphical representations to show the structure of problems;
- having team members justify their solutions to one another;
- having team members evaluate other team members solutions. (Resnick, 1987).

The use of these techniques to effectively manage a crisis situation may be evidence of a crisis manager’s competence in using higher order cognitive skills.

Other studies have been directed at trying to generate training programmes to improve general problem solving skills that would be transferable into different contexts of application (de Bono, 1985; Covington, 1987). These programmes aim to improve the trainees planning of their cognitive and metacognitive strategies. They use a different set of techniques to the contextualised programmes, namely:

- considering multiple sides of an issue (lateral thinking);
- considering consequences;
- selecting goals and planning strategies;
- prioritising factors involved in a situation;
- generating and evaluating evidence;
- using perceptual rather than logical thinking (Resnick, 1987).

The idea behind the use of perceptual thinking is to make problem-solving strategies more practical and less theoretical. However, system faults do not always present adequate stimuli to allow a successful problem solving strategy to be formed by perceptual thinking alone. In these cases, a strategy involving a logical thinking process using theoretical knowledge bases may be required to solve the problem by developing a new working method from first principles. Despite the claims made by these programmes, there has been little research to provide the empirical evidence necessary to prove their efficacy in facilitating the learning of transferable cognitive skills. However, these techniques, as with those used in the contextualised training programmes, could be used as assessment criteria for assessing competence in crisis management; the criteria being evidence of the effectiveness of the candidates’ cognitive and metacognitive strategies.

In order to be able to analyse the cognitive processes of a person by getting that person to report on their metacognition, it must first be assumed that people have an awareness of their own metacognition. However, there have been many hypotheses proposing that humans are unaware of their own cognitive processes. If this is the case, the use of post assessment scenario interviews to try to determine the cognitive processes that led to a person making a particular decision could be misleading and certainly not suitable for assessing competence.

Miller (1962) states that, *"it is the result of thinking, not the process of thinking, that appears spontaneously in consciousness".* If this is the case, any analysis by a person of their own metacognition will only result in them reporting their own pre-conceived ideas about their thought processes and the factors affecting their thought processes, not an account of their actual thought processes and the factors affecting them. This is also a powerful argument against the use of verbal protocols for the assessment of competence in crisis management, as any verbalisation of cognitive processing would probably just reflect the result of thinking, and would not necessarily describe the processes of thinking.

Neisser (1967) agues that the constructive processes of encoding perceptual sensations never appear in consciousness, it is only the products of these processes that appear. Mandler (1975) concurs with this view, stating that the analysis of situations, and the appraisal of the environment in which these situations occur, goes on mainly at the non-conscious level. If this is correct, we could never hope to deconstruct any situational analysis, as we would be unconscious of the processes that constructed it.

Nisbett and DeCamp Wilson (1977) argue that humans apprehend many more stimuli than can be stored in short-term memory or transferred to long-term memory. This infers that our conscious memory of an event is the result of some sort of selective filtering within our subliminal perception. They go on to propose that:

> *"Some stimuli may affect ongoing mental processes, including higher order processes of evaluation, judgement, and the initiation of behaviour, without being registered in short-term memory, or at any rate without being transferred to long term memory."*

If this were the case, then if subjects were asked to report on the reasoning behind their decisions, their report would necessarily contain omissions. Nisbett and DeCamp Wilson make the following proposal relating to this type of report:

> *"When people are asked to report how a particular stimulus influenced a particular response, they do so not by consulting a memory of the mediating process, but by applying or generating causal theories about the effects of that type of stimulus on that type of response."*

They go on to further argue that even if a subject's subjective report about their own cognitive processing is correct, this would only be due to the *"incidentally correct employment of a priori causal theory"*, not due to any *"direct introspective awareness"*.

However, it has been argued that certain metacognitive skills are required for proficient decision-making. Cohen et al (1996) propose that in order to be a good decision maker in a time stressed situation the following metacognitive skill set is required:

- *"going beyond pattern matching in order to create plausible stories for novel situations;*
- *noticing conflicts between observations and a conclusion;*
- *elaborating a story to explain a conflicting cue rather than simply disregarding or discounting the cue;*
- *having sensitivity to problems in explaining away too much conflicting data;*
- *attempting to generate alternative coherent stories to account for data;*
- *having a refined ability to estimate the time available for decision-making."*

Although these metacognitive skills in themselves are unobservable, the application of these skills may lead to some directly related overt behaviours. For example, a crisis manager may advise their team of a number of alternative hypotheses that could account for the available data, or they might advise their team of the time available before a decision has to be made about a particular situation. Once these related overt behaviours have been determined, they could then be used as markers for the assessment of competence in decision-making. It is widely accepted that one of the attributes of a crisis manager should be competence in decision-making.

2.4.5 Cognitive analysis for assessment - Conclusions

It can be concluded from the above discussions that cognitive analysis, either through the use of verbal protocols, or ‘reporting on metacognition’, is neither a very suitable or effective way of undertaking the assessment of competence in crisis management. Both of these techniques for inferring cognitive processes can be seen to interfere with the assessment process. The verbal protocol technique reduces the time available for decision-making, thus affecting any assessment. The ‘reporting on metacognition’ technique raises questions about what is actually being reported, introspective information about the cognitive processes, or a conscious formation of a pre-conceived causal theory about the cognitive processes that might have taken place. However, both of these techniques have been shown to generate overt behavioural markers that may be suitable for use in the assessment of competence in crisis management. An analysis of verbal protocols has been used to highlight overt patterns of communication between team members, and the application of metacognitive skills has been shown to generate certain overt behaviours, such as generating alternative coherent stories to account for data.

In summary, the concept of cognitive goal modules has been shown to be a useful tool in the analysis of the competency requirements of crisis managers who undertake complex dynamic tasks during the management of crises. The cognitive skill of a crisis manager has been shown to be dependent upon the crisis manager’s ability to select goals and to formulate subgoals as part of an overall task strategy. By the observation of crisis managers undertaking these tasks, an analysis of the crisis manager’s sequence of cognitive processing modules and subgoal formation can be made. From this analysis, it has been proposed that an assessment framework could be designed to assess the competence of crisis managers in relation to expert performance.

Theories have been discussed proposing that the cognitive skill of a crisis manager relies upon the crisis manager’s contextualised knowledge base and their ability to maintain a dynamic mental model of the state of the crisis situation within working storage memory. Although the same types of higher order cognitive processing skills have been shown to be used across different complex tasks, because of the reliance upon specific contextualised knowledge bases, transferability to different contexts is not assured unless other relevant knowledge bases are already present. This would tend to suggest that the assessment of competence in crisis management, a competence based upon the application of contextualised knowledge bases, would also have to be context specific.

A number of common cognitive processes used when undertaking complex dynamic tasks have been identified. The nature of these processes, such as the application of heuristics, would appear to make them suitable for use in assessment frameworks. However, caution must be drawn to the fact that these processes are discrete psychological aspects and as such, any assessment dealing with these aspects in isolation would appear too reductionistic to achieve an overall assessment of cognitive skills. For this to happen, the controlling metacognitive processes would have to be considered as well, and it has been shown that these would be difficult to assess. Cognitive skills, such as problem solving, have been shown to be built up from a whole network of interacting cognitive processes from the discrete psychological aspects such as subgoal formation, to the metacognitive processes such as cognitive goal module sequencing, and are thus very difficult to assess directly.

It would appear from the evidence discussed in this section that cognitive analysis could be used to assess the competence of crisis managers, but only if that assessment framework is designed within the correct scope. If too holistic, the relevant cognitive processes may be overlooked. If too reductionistic, the controlling metacognitive processes may be overlooked. However, considerable research is still required before cognitive analysis could produce any certainty about the cognitive processes involved when an operator undertakes a complex dynamic task. Until there is this certainty, any assessment framework would have to rely upon assessing performance against models of expert behaviour to measure competence.

Therefore, in themselves, verbal protocols and cognitive analysis have not proven to be successful techniques for the assessment of competence. However, the related overt behaviours arising from the analysis of these techniques may possibly be useful metrics for the assessment of competence.

In order to be able to determine which behavioural markers might be relevant for assessing competence in crisis management, some discussion should first take place to determine the attributes of a crisis manager that make them competent in managing crises. The following section will discuss these attributes.

2.5 THE ATTRIBUTES OF A CRISIS MANAGER

There are a number of attributes that have been shown by researchers to be exhibited by competent crisis managers. Regester and Larking (1997) propose that although the *"human participative manager"* is generally the most effective leader, this leader's participatory style can sometimes inhibit the rapid decision-making necessary during a crisis. Whereas, the rapid decision-making of an authoritarian leader can lead to the demotivation and stifling of creativity amongst other team members, which may in turn lead to a lack of feedback of information to the leader, thus restricting the leaders overall awareness of the situation. This lack of situational awareness has been shown to lead to poor decision-making (American Bureau of Shipping, 2004; Pekcan et al, 2005). Regester and Larking suggest that a competent crisis manager will recognise their team members' different attributes and values, and will utilise these to maximum advantage during a crisis. This may require a form of adaptive leadership style, the leader recognising the type of follower being interacted with, and adapting their style of leadership in order to maximise each individual team members performance (Blanchard and Hersey, 1969).

Similarly, Flin (1996) argues that to be effective, a leader needs to be able to diagnose not only the tasks or problems associated with the situation they are in, but also the mood, competence and motivation of the team they are leading. She argues that an effective leader should have a range of leadership styles available, such as delegative, consultative, coaching, facilitative or directive, and should be able to match the style they employ to the situation they are in and the team they are leading.

Fredholm (1997) argues that in order to be effective at making decisions a leader needs to have the ability to use both of the main accepted forms of decision-making; the more immediate recognition primed decision-making process, and the consciously analytical processing type of decision-making. Orasanu (1997) provides a comprehensive definition of the recognition primed decision-making 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."*

Fredholm states that a leader also requires the ability to switch between these two types of decision-making processes, as there is a need to function in both these ways depending upon the situation. Fredholm calls this a cognitively parallel approach to decision-making. The time available to make a decision plays an important part in which decision-making process the leader uses, with the analytical processing type of decision-making generally accepted as taking more time to reach a decision than the recognition primed decision–making process.

The manager of a crisis situation within a safety critical industry can be said to be operating as part of a complex system. Rasmussen (1990) argues that individuals who are in control of complex systems need to be flexible in their resolution of problems. This flexibility can be brought about by the individual having "*multiple degrees of freedom for action*", and "*many possible 'right' answers*". Rasmussen goes on to argue that because humans have a tendency to reduce degrees of freedom by reducing the available alternative choices, occasional errors are bound to result. It therefore follows that in order to resolve those conflicts of resource demand that occur during a crisis situation, the crisis manager needs the skill to be able to dynamically shift between alternative strategies. The method of generating multiple explanations for situations, and generating multiple courses of action to manage those situations, is well accepted as a way of ensuring effective decision-making. This method is often given the term 'critical thinking' and is regarded as a skill required for effective decision making in atypical situations (Cohen et al, 1997). It therefore follows that these should be attributes exhibited by a competent crisis manager.

Cohen et al (1996) argue that during a time stressed situation, proficient decision-makers will dedicate more resources to resolving any uncertainty about the situation. They argue that this is done by the decision-maker '*explicitly asking how much time they have before they must commit to a decision.'* They further argue that the proficient decision-maker can estimate this 'time to commit' more precisely. Cohen et al also propose that the more experienced a decision-maker is, the more sophisticated will be the way that they analyse and critique their current mental model of a situation. They suggest that the experienced decision-maker will question what is wrong with their mental model, especially what might be missing from it. He or she will also try to modify their mental model to explain any conflicting evidence, and then evaluate any assumptions required by alternate models. This leads to them constructing a more plausible model of the situation by revising the most unreliable assumptions in their current model.

Research into emergency decision-making in the offshore oil and gas industries noted that much of an offshore installation manager's information gathering activity was confined to monitoring the information exchange between other team members, rather than from direct requests for information (Flin et al, 1996). The degree to which a crisis manager monitors the information exchange between members of their crisis management team could therefore be a possible overt behavioural marker for assessing competence in crisis management.

McIntyre and Salas (1995) showed that effective team leaders exhibit model team-working behaviours of effectively backing-up other team members, both accepting and providing feedback, and a willingness to listen to others.

The following individual attributes could therefore be considered necessary for the competent crisis manager:

- the ability to recognise their team members' different attributes and values;
- the use of an adaptive leadership style to maximise team performance;
- the ability to switch between the consciously analytical and the recognition primed types of decision-making process;
- the ability to maintain multiple degrees of freedom for action;
- the skill to be able to dynamically shift between alternative strategies;
- the ability to improve their mental model of a situation through analysis and critiquing;
- the use of model team-working behaviours.

It can be seen from the above that the interactions between the crisis manager and the crisis management team play an important part in the effective management of a crisis. The following section looks at these interactions and how the behaviours exhibited within the team environment may be used as markers for assessing the team leaders competence in managing crisis situations.

2.6 THE ATTRIBUTES OF A CRISIS MANAGEMENT TEAM

It has been stated that superior crisis management teams have one common quality:

"the ability to adapt to task demands" (Serfaty and Entin, 1997)

Under increasing levels of stress, it has been argued that high performing teams will adapt their decision-making and co-ordination strategies, and even their organisational structure, in order to maintain a high level of performance, and to keep perceived stress levels within tolerable limits. The co-ordination mechanisms that support any adaptation by the team may either be 'explicit', based on specific communications, or 'implicit', based on a shared mental model of the situation (Serfaty and Entin, 1997).

Serfaty and Entin (1997) propose that the following features are characteristic of high performing teams:

- *the team structure is 'adaptive' to changes in the task environment. The same individuals may play completely different roles under different circumstances;*

- *the team maintains 'open and flexible communication' lines;*

- *team members are extremely 'sensitive to other members' workload and performance in high tempo situations.*

Serfaty and Entin further propose that highly effective teams will have developed a high level of congruence between the mental models of the individual team members. The team members are able to make use of these models to anticipate the way the situation will evolve and the needs of the other team members, as well as to generate expectations about how other team members will behave. They argue that both the explicit and implicit co-ordination mechanisms will generate observable communication patterns. Serfaty, Entin and Volpe (1993) hypothesised that highly effective teams would adapt to stressful situations by changing their co-ordination strategy from explicit co-ordination mechanisms to implicit co-ordination mechanisms as the workload increased. Serfaty and Entin (1997) state that:

> *"Communications that provide information to a team member in the absence of requests for that information indicate an implicit co-ordination mechanism at work."*

The relevance and timeliness of this information to the situation would seem to be a good indication of the degree of congruence of the mental models of the individual team members.

Urban et al (1996) showed that teams with a high workload communicated less than did teams with a low workload. They suggested that this reduction in communication was a coping strategy employed to allow teams to cope with high workloads. It could therefore be suggested that some teams could remain highly effective under stressful conditions, despite their level of communication falling, by switching instead to implicit co-ordination mechanisms. However, other teams might become ineffective under stressful conditions because as their level of communication fell, there was no corresponding increase in their implicit co-ordination; the team falls apart.

Urban et al (1996) proposed a method of measuring communication by counting the frequencies of communication behaviours that were relevant to resource allocation in a team task. The definitions of these behaviours are shown in table 1 below:

Behaviour	Definition	Example
Interrogative supply.	Questions regarding the availability of resources.	Alpha, how many resources do you have?
Declarative supply.	Statements regarding the availability of resources.	I have 4Xs and 5Ys
Interrogative demand.	Questions regarding the resources required to complete the task.	How many Xs do we need?
Declarative demand	Statements regarding the resources required to complete the task.	Task needs 2Xs.
Request/interrogative exchange.	Questions or requests relating to the transfer of resources among team members.	I need someone to send me 2Xs.
Respond/declarative exchange.	Statements relating to the transfer of resources among team members.	I am sending you the 2Xs.

Table 1: Communications Behaviours (Urban et al. 1996)

Stout et al (1994) state that a consequence of high time pressure within a situation will lead an ineffective crisis manager to focus on their individual tasks rather than the teams' tasks. Further research in this area suggested that, during demanding high workload situations, team members revert to working as individuals at the expense of the team task requirements (Urban et al, 1995). This may be seen as another coping strategy. It also implies that during a crisis situation there must be some optimum distribution of effort between individual tasks and team tasks in order to best manage the crisis.

In a research programme studying how teams use shared mental models the researchers codified the teams' verbal interaction in order to determine the degree of overlap of the team members' mental model and the organisation of the division of the model between team members (Banks and Millward, 2000). Each verbal interaction between team members was coded using one of the following designators:

- *request*
- *offer*
- *question*
- *propose*
- *hone*
- *support*
- *widen*
- *other*
- *not relevant*

Some research has shown that there appears to be a range for the level of useful communication between team members within which the team is most likely to be effective at managing a crisis situation (Tollcott, 1992; Urban et al, 1996). The team becomes ineffective if the level of useful communication between the team members is either below or above this level. If this is the case, then an analysis of the verbal interaction between team members in order to determine if the level of useful communication lies within the effective range may be an overt behavioural marker that can be used to assess competence in crisis management. It may therefore be useful to codify the verbal interaction of the team, using designators such as those used by Banks and Millward, in order to determine the level of useful verbal interaction between the crisis management team. This level of useful interaction could then be used to determine if there is an optimal range of communication level related to the successful management of crises, which if shown, could be used as a behavioural marker.

Oser et al (1989), in their study of critical team behaviours, showed a significant correlation between their indices of *"team member confidence in other team members"* and *"team members' liking of others"* with the effectiveness of the team. This again shows that it is possible, even with only a small set of behaviour markers, to make a valid assessment of the effectiveness of a team.

McIntyre and Salas (1995) discuss the behaviours exhibited by effective teams. They propose that the development of teamwork is encouraged by interdependence between all of the members of the team, so that the team views their success as being dependent upon their own interaction. They argue that teamwork involves effective communication, and that this often needs to be closed-loop communication, with relevant feedback to the originator of any communication. They also argue that effective teamwork requires team members to have a flexible repertoire of behavioural skills that they can call upon, depending on the circumstances they are presented with.

It may be useful to further deconstruct the crisis management team so that the individual team member attributes can be identified. This should give further insight into the interaction between team leader and team member, so leading to a better understanding of the behaviours exhibited by the team leader during a crisis situation. The following section will discuss the individual attributes of crisis management team members.

2.7 THE ATTRIBUTES OF A CRISIS MANAGEMENT TEAM MEMBER

There are some generally accepted types of team members, whose attributes can be used to advantage by the team leader during a crisis. Regester and Larking (1997) propose that the styles of team members may include

"The Ideas Person:	*a creative member who is constantly injecting new ideas and suggestions. Some of these may be far-fetched but some may have real merit. It is vital for the leader to filter out the viable ideas and discard the remainder without discouraging the flow;*
The Communicator:	*the individual who helps the flow of information both within and outside the team (not necessarily the team leader, although the team leader should possess strong communications skills also);*
The Doom Merchant:	*the devil's advocate who brings out the negative aspects of each proposed idea or solution;*
The Book Keeper:	*the neat and orderly member who wants the records and logs maintained to perfection. This individual is more comfortable in such a role than as a decision maker. Nonetheless, it is a vital role;*
The Humanist:	*the people-oriented member whose solutions always focus on the human aspects of the problem – an important visionary in the heat of the moment."*

During training exercises within the engine room simulator at Warsash Maritime Academy the author has noted that some of these attributes, far from being an advantage to the team leader, have been seen to be very disadvantageous.

If an assertive "*ideas person*" is in the team of an inexperienced and unassertive leader, then 'verbal disruption' can occur. This type of behaviour has been noted where one team member has been so vocal during the simulation exercise that they have disrupted the team task. In an effort to be helpful to the team, these team members will continually verbalise the situational assessment, as they perceive it. They will offer many varied hypotheses for any perceived problem, they will continually offer advice to the other team members and they have also been observed thinking aloud. By continuously talking, they make it very difficult for other members of the team to communicate, and consequently the performance of the team suffers. It also been observed that an inexperienced leader, who is possibly seeking assistance from his team to manage the situation, will sometimes take up the suggestions made by the "*ideas person*" even though the idea they put forward is erroneous, just because of the level of conviction with which the idea is proposed. When the team member who was causing the verbal disruption was given a non-participative role of observer within a simulator exercise, the performance of the remainder of the team was seen to improve dramatically.

In a similar way the "*doom merchant*", although valuable for highlighting the risks associated with any intended course of action, if they are both inexperienced in assessing the risks, and assertive in the way they communicate the possible risks, may adversely influence the decisions of the team leader.

Regester and Larkin (1997) suggest that members of a team subject to a crisis situation can sometimes exhibit a misplaced belief in the competence of their leader or leaders, and this can assist in the development of the crisis. However, it has also been noted by the author, during simulator based crisis management exercises, that if a leader is seen to lack the competence to manage the situation, another team member will often come forward and actively assist the leader, or in some cases, take over the leadership role.

Brannick et al (1993) showed that 'cooperation' and 'giving suggestions' were behaviours that showed a high degree of correlation with successful team performance, and as such are important attributes for a crisis management team member. They also showed that when these behaviours were used as assessment metrics, assessors consistently rated them. They also showed that these behaviours were useful metrics because assessors found that they could assess them relatively independently of one another, promoting a higher degree of interrater reliability. However, 'cooperation' is a very 'high level' behavioural term, consisting of a complex set of behaviours. Although assessors may have been observing overt 'cooperation' within a team, which of the simple 'low level' behaviours that go to make up 'cooperation' they were in fact observing, is not made clear. The consistency of the rating of the behavioural marker termed 'cooperation' could in fact have been made up of many very inconsistent sets of much simpler lower level behavioural markers.

Table 2 below shows how the complex 'high level' behaviour of 'cooperation' might be decomposed into a number of less complex 'lower level' overt behaviours. If assessments are undertaken based on criteria that are comprised of complex 'high level' behaviours, it is likely that there will always be a level of subjectivity, and interrater reliability and agreement is likely to always be adversely affected.

'High Level' Behaviour	Associated 'Mid Level' Behaviours	Associated 'Low Level' Behaviours
Cooperation	Cordial intra-team communication;	Smiling when communicating; Even tone of voice; Normal voice level;
	Consonance with commands;	Pleasant acknowledgement; Prompt response; Nodding of head;
	Attentive listening;	Forward leaning posture; Nodding of head; Quiet mouthing of affirmations;

Table 2: The decomposition of a complex 'High Level' behaviour.

After discussing some of the attributes of the crisis management team, the following sections will discuss some of the behaviours arising from these attributes that can influence the decisions made by a crisis manager within a team based environment.

2.8 BEHAVIOURAL BIAS AND OPTION GENERATION

In analysing the USS Vincennes incident, where a US naval vessel shot down an Iran Air commercial airliner, Collyer and Malecki (1998) argue that given a belief that an attack was underway, the team onboard the USS Vincennes may have misinterpreted the facts and selectively attended to facts that supported this hypothesis. This behaviour they term an 'expectancy bias'; the hypothesis of what is happening becomes a self-fulfilling prophecy.

Tollcott (1992) suggests that in the decision-making process hypotheses are generated to account for the information being received, and then alternate actions are generated and evaluated. The number of alternative hypotheses and actions communicated to team members could therefore be considered as a metric for assessing competence in crisis management. It would seem reasonable to suppose that if a team leader only came up with one possible hypothesis to account for the information being received, this may not be correct, and the actions taken on the basis of this incorrect hypothesis would not lead to a resolution of the problem. Conversely, if a plethora of hypotheses were proposed, it may take too much time to generate and evaluate the possible actions. There is probably a range of numbers of alternative hypotheses and actions that, when evaluated, lead to the most efficient and effective outcome of a crisis situation.

Simon (1957) demonstrated that in many situations people do not attempt to evaluate all available response choices, but rather consider only as many alternatives as needed to discover one that satisfies them. Simon refers to this process as *'satisficing'*. Within a team environment the level of 'satisficing' is an overt behaviour and might therefore be a suitable metric for the assessment of competence in crisis management.

Another form of behaviour that leads to a limitation of the generation of options, or even to the complete lack of options, is termed 'framing bias' or 'paradigm paralysis'. This behaviour is characterised by the decision maker placing an artificial boundary around the problem, and only generating hypotheses and actions relating to what they perceive as being within this artificial boundary, even if the solution to the problem can in reality only be found outside the boundary they have imposed. A clear example of this behaviour was exhibited by the engineer officers onboard the vessel Green Lily during the incident that led to her grounding in November 1997. Given a situation where water was splashing onto the electrical control panel of the main engine control system, the engineer officers wrongly assumed that this was the cause of the main engine tripping, and proceeded to concentrate all of their efforts into this option. By placing an artificial boundary around the problem as they saw it, i.e. water on the electrical control panel, the engineer officers failed to generate any other options to explain the tripping of the main engine that lay outside of their self imposed boundary. In fact, the investigation into the grounding found that the vessel's main engine had tripped due to the operation of the main engine mechanical over-speed trip, which was totally independent from the electrical control panel (Marine Accident Investigation Branch, 1999).

In order to be able to observe the overt behaviours that are associated with the decision-making processes of crisis management, it is important to have the correct research environment. The following section will discuss the requirements for the decision-making research environment.

2.9 THE DECISION-MAKING RESEARCH ENVIRONMENT

Competence in crisis management requires good decisions to be made within apposite time limits. Therefore, factors that either assist or hinder the decision-making process will affect competence in crisis management. The study of these factors is therefore very relevant to the field of crisis management.

Current models of decision-making have moved away from the classic model of the reasoned optimisation of choices to arrive at the 'best' option, to models that better reflect the way decisions are made by decision makers in the real world, i.e. more naturalistic models. The most influential of these models is the naturalistic decision-making model, which has been defined as follows:

> *"The study of naturalistic decision-making 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 organization in which they operate."* (Pruitt et al, 1997)

This definition reveals a number of characteristics of the situations in which naturalistic decision-making takes place:

- 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. (Barnett et al, 2003)

It has been suggested by many naturalistic decision-making theorists that in order to conduct research into how decisions are made, the study participants, their experimental tasks, and the research environment in which these tasks are undertaken, must be selected or constructed so that the performance within the research environment can closely resemble how performance occurs in the real world (Cannon-Bowers et al, 1996; Cohen, 1993; Zsambok and Klein, 1997).

Cannon-Bowers and Bell (1997) assert that naturalistic decision-making relies on the appropriate cue strategy relationships being embedded in the task. Therefore, when a particular cue, or set of cues, is presented within the task environment, the decision-maker adopts a decision strategy from memory that they identify as satisfying the needs of the task. They argue that naturalistic decision-making skills are contingent upon the context in which they are applied, with the decision-making environment required to provide the necessary cues. It could therefore be said that the overt behavioural markers that indicate competence in crisis management are also contingent upon the context in which they are exhibited, as these behaviours are the overt expression of the decision strategy adopted. It therefore follows that when undertaking research that involves participants making decisions, it is necessary that as much of the context of the real world task environment as possible is provided to the participants in order for these cues to be present.

It is generally accepted that performance within a 'full mission simulator' during competence assessments will not predict exactly performance during an actual incident scenario. A full mission simulator being a simulated representation of an actual working environment in which a significant number of the cues that are within the real environment are recreated. However, it is also generally accepted that simulator based competence assessment does enable the identification of those individuals who would not be effective emergency or crisis managers (Flin, 1996).

It can therefore be argued that when undertaking research that relies upon the abilities of participants to make decisions, the research environment should be as realistic as possible, providing as many of the cues that would be in the real world as possible to facilitate the choice of decision strategies. The use of a high fidelity simulator would seem to be an appropriate way to provide such a research environment.

In order to be able to assess "natural" decision-making, it is first necessary to understand the factors that typify this process.

2.10 FACTORS TYPIFYING DECISION-MAKING IN NATURALISTIC ENVIRONMENTS

Orasanu and Connolly (1993) propose that it is the following factors that typify decision-making in naturalistic environments:

- *ill-structured problems;*
- *uncertain dynamic environments;*
- *shifting, ill defined or competing goals;*
- *multiple event-feedback loops;*
- *time constraints;*
- *high stakes;*
- *multiple players;*
- *organisational norms and goals that must be balanced against the decision-makers personal choice.*

In order to undertake the assessment of naturalistic decision-making, the decision-making environment in which the assessment will be undertaken should provide as many of these factors as possible. Simulator exercise scenarios should be developed that can provide these factors.

Pruitt et al (1997) contend that naturalistic decisions can be characterised on the basis of at least three sources of variables:

- *those associated with the decision or task;*
- *those associated with the decision maker;*
- *those associated with the environment* [in which the decisions are made].

From the above it is clear that many domain specific factors and variables affect the decision-making process. Research that seeks to analyse and assess the naturalistic decision-making skills of persons that work in a specific domain should therefore ensure that the variables associated with the task and the decision-making environment are as representative as possible of those in the real domain. Without the domain specific cues being available, a domain expert may not be able to call upon the appropriate decision strategies that their experience within the domain has given them.

Breznitz and Ben-Zur (1997) state that decision-making strategies that utilise all available information within the environment are necessarily more time consuming, and demand greater effort, than those that are based on a few dimensions only. They go on to argue that under time pressure, such as during an emergency or crisis situation, subjects tend to emphasise the importance of negative dimensions, and if the information received is too discouraging, insufficient resources are deployed, the situation can turn into a self-fulfilling prophecy. Their research showed that with a combination of high workload and time pressure, decisions became based almost exclusively on short-term considerations. The behaviours exhibited when handling 'discouraging' information could therefore be important markers when assessing competence in crisis management. The making of decisions with long-term considerations when under conditions of high workload and time pressure could also be seen as a criterion for the assessment of competence in crisis management. However, any assessment of such a criterion would have to be made by an evaluation of the decision itself, as it would probably not be possible to link any specific overt behavioural markers to this criterion.

Cohen et al (1996) argue that proficient decision makers are *'recognitionally skilled'*, in that they have, in long-term memory, a large repertoire of familiar situations, for which they have appropriate responses. This relates closely to other theories of naturalistic decision-making (see Zsambok & Klein, 1997). The response to 'discouraging' information outlined above could possibly be explained by the decision-makers lack of repertoire of familiar situations. If this large repertoire of familiar situations exists, then it is more likely that the decision-maker will be able to come to a good decision, using less information from the environment, and in a shorter time. The behavioural marker of quick decision-making, if it leads to good decisions, could therefore be an indication of the decision-makers large repertoire of familiar situations in memory.

The previous sections have discussed the environment in which naturalistic decisions are made, and the factors that typify these types of decisions, and the consequences for research. However, even if the decision-making environment is conducive for naturalistic decision-making to occur, it is still necessary to understand how the crisis manager can control the crisis situation. The following section will discuss this.

2.11 CONTROL OF COMPLEX DYNAMIC TASKS

Research undertaken by Kerstholt et al (1996) showed that the complexity of a task, as defined by the occurrence of multiple events, as occurring in a crisis situation, degraded the performance of those undertaking the task in a number of ways. They found that successive system disturbances were detected significantly later, and some disturbances were completely ignored. They also found that there was a tendency to concentrate on single disturbances, while ignoring the rest of the system. This is often termed 'tunnel vision' and leads to a lack of overview of the situation by the crisis manager. Kerstholt et al argue that in order to ensure that a subject's mental model of the system state is kept updated, and that any disturbances are detected, those subgoals under their control should be regularly sampled, and they call this action '*sampling behaviour*'. However, by limiting the sampling to only those subgoals that are under the control of the crisis manager, only a limited overview would be achieved. The sampling of any other tasks, systems and goals that are associated with the crisis, but not directly under the control of the crisis manager, should also be considered if a true overall awareness of the situation is to be maintained.

Kerstholt (1997) goes on to suggest that people are not very good at controlling dynamic systems because they:

- *ignore feedback delays;*
- *focus too much on the reduction of uncertainty by requesting information;*
- *are not very accurate in timing the right moment of intervention.*

Rapid and reliable feedback is essential for good decision-making. If feedback delays are ignored, errors may not be picked-up, understood or corrected (Croskerry, 2003). If, in order to reduce the level of uncertainty, there is an undue level of information sought before a decision is made, the situation may have escalated before any action is taken. Similarly, unnecessary delays in decision-making can lead to intervention actions being undertaken at inappropriate times, so rendering them less effective, or in the worst case, totally ineffective. All three of these behaviours have overt aspects, which could be used as markers for the assessment of competence in crisis management.

Kerstholt proposes that there are two reasons for sub-optimal performance of people controlling complex dynamic tasks. The first being '*cognitive lockup*', the tendency to focus on one sub-system at a time, thereby ignoring the dynamics of the complete system. The second being a lack of knowledge concerning the relation between symptoms and causes. Therefore, although the crisis manager may well perceive the majority of cues from the environment that are relevant to the crisis situation, they cannot readily relate those cues that are symptomatic of a problem, to the problem itself.

Endsley (1995a) suggests that:

> *"In complex dynamic environments operators must do more than simply perceive the state of their environment. They must understand the integrated meaning of what they are perceiving in the light of their goals."*

Endsley goes on to suggest that the first step in achieving situational awareness is to perceive the status, attributes and dynamics of the relevant elements within the environment. In a further paper (Endsley 1995b), she goes on to codify three levels of situational awareness:

> *"Level 1: the perception of relevant information;*
>
> *Level 2: the integration of various pieces of data in conjunction with operator goals to provide an understanding of the meaning of the perceived information;*
>
> *Level 3: based on this understanding, future events and system states can be predicted, allowing for timely and effective decision-making."*

It has been suggested that if the realism of a simulation is altered it can fundamentally change the way the operator conceptualises the underlying information (Manktelow and James, 1987). If this is the case it follows that this change in conceptualisation of information could lead to change in the understanding of the information, and hence to a change in the level of situation awareness. The integrity of the realism in any simulation used to study behavioural markers of crisis managers would therefore appear to be of the highest importance.

Endsley (1995b) also notes that an operator's perception of a situation is only overtly observable through their actions and imbedded or elicited verbalisations by the operator. Although the level of situational awareness is a very important factor in the control of complex dynamic tasks, much of this awareness does not necessarily lead to specific overt behaviours being displayed. This makes the use of the 'level of situation awareness' a difficult metric to use for the assessment of competence in crisis management, as it can only be subjectively implied by the observer.

So, if it is accepted that competence in decision-making is one of the important criteria for assessing competence in crisis management, what methodologies can be employed in order to determine how a situation can affect the decision-making skills of a crisis manager?

2.12 EXPERIMENTAL METHODOLOGIES FOR RESEARCH INTO DECISION-MAKING

During a crisis, there will be many stressors placed upon the crisis manager, and the effect of these stressors on the decision-making process is of particular interest if competence in crisis management is to be determined. When researching the impact of complex stressors on decision–making Johnston et al. (1998) used a full-mission simulator in the following way:

Stage 1. The participants were given a ninety-minute simulator familiarisation training session, which included team role familiarisation, practice with control systems, and two by fifteen-minute training scenarios for team practice. The participants were then given a short written test to establish their level of knowledge of the simulator. A list of procedural artificialities of the simulator was also given to each participant.

Stage 2. The participants received a pre briefing on the assessment exercise.

Stage 3. The participants undertook a sixty-minute assessment exercise.

Stage 4. The participants were debriefed on their performance.

Stage 5. The participants were asked to complete a stress perception questionnaire that applied Likert scale ratings to:

- scenario workload
- time pressure
- ambiguity
- overall stress level

Johnston et al. state that simulator scenario development should consist of designing 'events' that elicit the behaviours of interest. They propose that in order to effectively assess the impact of complex stressors on tactical decision-making, scenario 'pairs' should be developed that use the same events, but which are carried out under more stressful conditions. In this way, they propose that comparisons of performance can be made under similar, but more stressful conditions. One way of increasing the level of stress during a simulator scenario is to present the team with some imminent danger. For example, if the simulated environment is a ships engine control room, following a total power loss, a call from the Master of the vessel to the engine control room stating that at the present rate of drift the vessel will be aground in thirty minutes will increase the workload, time pressure and overall stress level of the Engineer Officers.

Salas, Driskell and Hughes (1996) define stress as:

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

If a situation arises where the actual, or perceived, demand for resources exceeds the available resources, the situation could be said to be uncontrollable, and beyond the scope of any laid down emergency procedures. In this case, the situation could be defined as a crisis. From this it could be argued that stress may be a precursor to a crisis, and stressors the building blocks of the crisis situation.

Cannon-Bowers and Salas (1998) define a number of stressors that can affect decision-making:

- *multiple information sources*
- *incomplete or conflicting information*
- *rapidly changing or evolving scenarios*
- *requirement for team coordination*
- *adverse physical conditions*
- *performance pressure*
- *time pressure*
- *high workload or information overload*
- *auditory overload or interference*
- *threat*

All of these stressors can be presented to teams within a full mission simulator at varying levels. Therefore, by the control of these stressors, it is possible to control the overall level of stress presented to the team leader during a simulator exercise. Cannon-Bowers and Salas propose that certain stressors are most suitable to be used for the manipulation of subjects in experimental studies, namely:

Task-related Stressors:

- *workload / time pressure*
- *uncertainty / ambiguity*
- *auditory overload*

Ambient Stressors:

- *auditory interference*
- *performance pressure*
- *fatigue / sustained operations*

The use of stressors will be discussed in Chapter 4, Research Methodology.

In order to make any valid judgements on the affect of certain stressors on the decision-making capabilities of a crisis manager, assessments of these affects have to be validated by showing consistency when using a number of different observers. Fowlkes et al (1994) argue that team performance measuring techniques that use human observers to make ratings historically had chronically low interrater reliability. They argue that assessment techniques are required that minimise the observers judgement, so that more objective and hence more reliable and useful assessments may result. They go on to propose an experimental methodology that first identifies stimulus events that can be used as cues to initiate task responses from team members that have been identified as being exemplar responses within a certain scenario. By having a set of previously identified exemplar responses to specific scenario events as a checklist for the observer, they argue that more reliable assessments can be carried out. The problem with this methodology is that it limits the assessment to a pre determined task response set, thus precluding any successful novel or atypical responses from the team members' assessment rating.

Instead of assessing exemplar responses to stimulus events, it may be possible to determine behavioural markers that give a measure of competence that is not only applicable to typical, exemplar responses, but also to the atypical responses as well. It is how factually based a behavioural marker is that will determine its objectivity during use in an assessment. The characteristics of behavioural markers could be said to be inter-dependent and on different ends of a scale of interrater agreement, as shown in Figure 4 below.

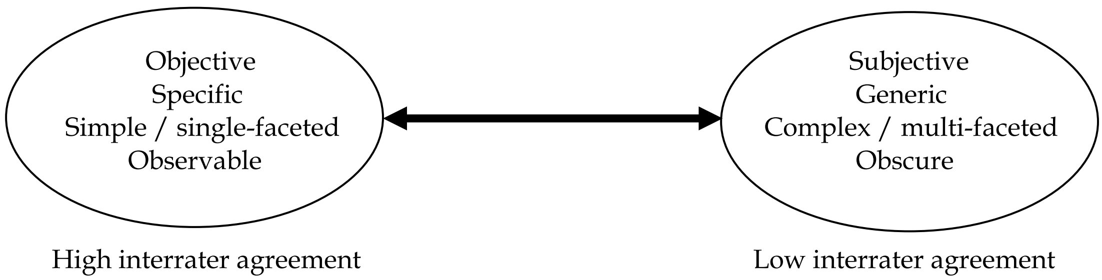


Figure 4: Relationship between behavioural marker characteristics and interrater agreement.

One method of validating any assessment framework is to show that it has a high degree of interrater agreement when used by a number of assessors to assess the same evidence.

2.13 NON-TECHNICAL SKILLS (NOTECHS)

During 1998, the first stage of a European Union funded research project was completed that had been tasked to develop a feasible and efficient method for assessing the non-technical skills of multi-pilot aircrew during flight and simulator checks (Avermaete and Kruijsen, 1998). This research project was given the title 'NOTECHS'. The study proposed that there were four primary categories that could be used for the assessment of competence of pilots' non-technical skills. These categories were co-operation, leadership and managerial skills, situation awareness and decision-making. The study went on to further sub-divide these primary categories as shown in Table 2. below.

Primary Category	Elements
Co-operation	Team Building and Maintaining Considering Others Supporting Others Conflict Solving
Leadership and managerial skills	Use of Authority and Assertiveness Providing and Maintaining Standards Planning and Co-ordination Workload Management
Situation Awareness	Systems Awareness Environmental Awareness Anticipation
Decision-making	Problem Definition / Diagnosis Option Generation Risk assessment / Option choice Outcome Review

Table 2: NOTECHS framework (Avermaete and Kruijsen, 1998)

The NOTECHS project put forward five guiding principles related to the way in which non-technical skills should be assessed. These principles were:

i) Non-technical skill cannot be rated and cannot provoke a FAILED condition out of the context of a related objective technical consequence leading to compromised safety in the short or long term.

ii) The assessment framework for non-technical skills should use a two-point rating scale of either ACCEPTABLE or UNACCEPTABLE.

iii) For each unacceptable rating on a non-technical skill category, the assessor should provide the candidate with an explanation in free text.

iv) A certain leitmotif of similar detrimental behaviours during an assessment exercise would have to be recognisable to conclude that a candidate had a problem in this area.

v) Any assessment should only be based on observable behaviours. The assessment should not make any reference to the candidates' personality or emotional attitudes.

These principles can be seen to be relevant to any assessment framework that uses behavioural markers as assessment criteria. However, the first principle could be seen to pose a difficult philosophical issue, i.e. whether the demonstration of a leitmotif of detrimental behaviours, within a process that leads to a successful outcome, should in any way lead to an unacceptable competency rating. It may be the case that although in a particular scenario the team leader demonstrated such a leitmotif, it was only due to the actions of the other team members that the team leader's behaviours did not lead to a unsuccessful outcome. The same team leader, in the same situation, might well exhibit the same detrimental behaviours, but with a different set of team members, the outcome would be unsuccessful. This issue will be further explored in section 2.15.

By 2001, the NOTECHS research project had defined a range of behavioural markers that it used within an assessment framework to assess the non-technical skills of candidate aircrew pilots (JAR TEL Consortium, 2001). This research highlighted one of the difficulties in using behavioural markers to assess competence in crisis management. The results of the experiments to validate the proposed NOTECHS assessment framework showed very good interrater agreement when assessing competence in straightforward, non-complex scenarios using a 2-point scale (pass/fail). When assessing pilots within these scenarios, the mean within-group interrater agreement coefficient (r_{wg})[1] varied between approximately 0.95 and 1. However, when using the proposed assessment framework to assess pilots within complex scenarios, using the same 2-point scale, there was much poorer interrater agreement. When using these complex types of scenario the within-group interrater agreement coefficient (r_{wg}) varied between zero and 0.2. The NOTECHS

[1] The within-group interrater agreement coefficient (r_{wg}) is defined as the proportional reduction in error variance of a distribution of obtained ratings compared to a distribution representing a random response pattern (James, Dameree and Wolf, 1984 & 1993). An r_{wg} of 1 indicates complete agreement between raters, whereas an r_{wg} of zero indicates that the raters agree no more than they would by chance.

project authors propose the following explanation for the poor interrater agreement measured when assessing the non-technical skills of pilots within complex scenarios:

> *"Although the original experimental plan demanded a mix of simple and more challenging scenarios for the Raters to evaluate, this requirement has created difficulties in interpreting some of the data. It would be expected that marginal performances of either technical or non-technical skills would generate a broad spectrum of responses from raters." (JAR TEL Consortium, 2001)*

The implications from this statement are that if a scenario is challenging, such as a crisis scenario, one can expect marginal performances from some of the candidates being assessed. When presented with such marginal performances, raters require highly objective and discriminatory assessment criteria in order to maintain a high degree of interrater reliability and agreement. The broad spectrum of responses from raters using the NOTECHS assessment framework for the assessment of pilots within complex scenarios would tend to suggest, that for this type of scenario, the assessment criteria are not sufficiently discriminatory.

The following behavioural marker is an example of those used for the assessment of non-technical skills within the NOTECHS assessment framework:

> *"Talks about possible risks for course of action in terms of crew limitations"*

It could be argued that this behavioural marker may be interpreted in many different ways. Different raters may well have different ideas about the possible risks, biased possibly by their own personal experience. The raters may also have differing views about the limitations of the crew. Together, these different interpretations can lead to the "*broad spectrum of responses*" alluded to by the NOTECHS project authors. The use of the NOTECHS assessment framework will be discussed in more detail in Chapter 3 of this thesis.

The NOTECHS project report also states that one of the basic tenets of resource management is that team members should communicate in a manner that reveals their mental models and thinking processes to the other members of the team. Assuming that competence in crisis management requires the effective management of resources, the ability to communicate in a manner that reveals ones mental models could be a possible behavioural marker for the assessment of competence in crisis management.

The NOTECHS project was "*designed to build on existing knowledge in the field of human factors, enhancing flight safety through improved non-technical skills training*" (Avermaete and Kruijsen, 1998). It is generally accepted, in many different domains, that human errors, arising from human factors issues, are involved in 80% of incidents and accidents. Therefore, training that attempts to reduce the frequency of human errors is addressing the biggest threat to safety. Assessment frameworks that provide for the valid assessment of competence in the skills that reduce human errors can be used to ensure that this training is effective. These human errors form part of the causal chain of events leading to an incident or accident. The next section will discuss how these causal chains can be related to an assessment framework for assessing competence in crisis management.

2.14 CAUSAL CHAINS

Research has shown that most incidents or accidents do not occur because of some erroneous high-risk decision, or some catastrophic system failure, but because of a chain of relatively minor errors and faults, that together form the causal chain that can lead to some disastrous incident (Reason, 1990). Crises have often been shown to be propagated by the complexity of a situation characterised by a large number of minor faults occurring simultaneously leading to the team being overwhelmed, rather than by a single serious fault. (Wagenaar & Groeneweg, 1987).

If a crisis can be categorised by its complexity arising out of a large number of simultaneous disturbances that depart from normal operations, the way in which the crisis manager copes with these disturbances could be seen as a way of assessing the crisis manager's competence. Kerstholt et al (1996) proposed a number of behavioural markers that could be used to determine how effectively a control task supervisor coped with such disturbances:

- the disturbance detection time, defined as the time between the onset of a disturbance and the first time a related action was carried out or a related information request was made;
- the time span between the first information request or action related to a disturbance and the correction of that disturbance;
- the number of information requests made about a particular disturbance;

- the number of times a supervisor inspected systems through their graphical control interface, as indicated by the number of mouse clicks on the control screen; a measurement of monitoring behaviour.

Kerstholt et al (1996) went on to discuss how the performance of a control task supervisor degraded when a situation became complex. They defined a complex situation as one in which there is an occurrence of multiple events or multiple simultaneous disturbances. Their experiments showed that within a complex situation, successive disturbances were detected significantly later and that some disturbances were ignored altogether. They argued that these findings supported the cognitive lock-up hypothesis in that there is a tendency for task supervisors to focus on a single system disturbance, to the exclusion of the rest of the system. They further argued that in order to meet the overall goal, a task supervisor requires an efficient strategy that allows their time and effort to be divided between different sub-goals. They suggest that some task supervisors adaptively switch between the processing of information about different sub-goals, whereas others have to completely finish one sub-goal before considering another.

Incident or accident causal chains consist of multiple events in series, or if very complex, multiple events both in series and at times in parallel. How the crisis manager copes with these events to ensure a safe outcome can be seen to be measures of his or her competence. However, what if the crisis manager manages most of the individual events in a causal chain competently, but does not achieve a successful overall outcome? Can a reductionist approach to the assessment of competence, which looks at discrete elements of performance, ever be a valid method of assessment if the overall outcome of the crisis manager's performance is not successful? The next section will discuss this issue further.

2.15 PROCESS VERSUS OUTCOME

When endeavouring to determine competence in the management of a complex environment such as that presented by a crisis within the machinery spaces of a vessel, arguments can be raised as to whether it is competence in the crisis management process that should be assessed, or competence in achieving a successful outcome. If using a behavioural marker system to assess competence in a simulated environment, could the validity of the system be related to the successful completion of subgoals, and not to overall performance outcome?

Smith-Jentsch et al (1998) discuss the differences between process measure and outcome measures in relation to the measurement of team-related expertise in complex environments. They put forward the following definitions for process and outcome measures:

Process Measures:	Measures that describe the strategies, steps, or procedures used to accomplish a task.
Outcome Measures:	Measures that assess the quantity and quality of the end result.

They go on to describe process measures as measures of performance, often used to evaluate the human factor involved in complex systems. They describe outcome measures as measures of effectiveness that are influenced by much more than just human performance. They argue that this means that outcome measures do not specify which elements of human performance are deficient. It could also be argued that outcome measures do not specify which elements of human performance were effective. The outcome of any crisis scenario may be influenced by factors that are beyond the direct control of the crisis manager, such as the actions of outside agencies, or the weather. If, as Smith-Jentsch et al argue, outcome measures are indeed influenced by much more than just human performance, it would seem to follow that if human performance is being measured, it is the process that is more important than the outcome. An assessment framework that uses human behavioural markers to assess process measures should be able to measure competence in crisis management, as this is an aspect of human performance.

Smith-Jentsch et al stated that they were not able to determine any consistent relationships between processes and outcomes, even though they systematically evaluated both at multiple events within a scenario. This may be because, as others have argued (Reason, 1990; Besnard and Baxter, 2003), there need not be consistent relationships between processes and outcomes.

It is of course possible for a crisis management team to be ‘lucky’ and to achieve a successful outcome despite the use of incorrect processes. If an assessment were to be only based upon outcome, these incorrect processes would not be considered. Conversely, although a crisis management team may have used the correct processes, the outcome may be unsuccessful.

Incorrect processes that lead to a successful outcome may be viewed as beneficial violations of correct processes (Reason, 1990; Besnard and Baxter, 2003). The question is should these beneficial violations be assessed positively or negatively. On the one hand, they are a departure from the accepted 'best practice' and therefore could be assessed negatively; on the other hand, they have contributed to a successful outcome and so could be assessed positively.

Besnard and Greathead (2003) argue that it does not matter whether a process is correctly carried out, or violated, but it is the cognitive environment in which the process is undertaken that is important. They propose that this cognitive environment is comprised of a number of dimensions, namely:

- *"the knowledge that operators have about the limitations of the system;*
- *how compatible their mental model is with the functioning of the system;*
- *the extent to which future events can be anticipated;*
- *the understanding of the consequences of the actions performed."*

  (Besnard and Greathead, 2003)

They go on to argue that it is the accuracy of the operators' mental model of a situation that plays a key role, allowing some process violations to improve system safety during a crisis. Therefore, if there are no consistent relationships between processes and outcomes, and some process violations are beneficial, the use of complete models of the expert behaviour exhibited when undertaking a process, as assessment criteria, would not seem to be a viable option. Instead, the behavioural markers that support and facilitate successful processes, beneficial process violations and successful outcomes, would seem to be the most relevant assessment criteria.

Kirijan and Mancuso (1995) stated that situational awareness was a process. They argued that by specifically focusing on the processes within crew resource management, rather than the outcome, two distinct benefits could be gained. Firstly, it is possible to train crews to perform processes that can be applied to different situations. Secondly, it is possible to design assessment instruments that isolate process deficiencies, regardless of the outcome. However, when dealing with a crisis, an atypical situation, it may be the case that atypical solutions are required that are outside the normal process repertoire of the crisis manager. If such an 'atypical' response is used to reach a successful outcome of a crisis scenario, where does that leave the assessment instrument that was designed to isolate 'typical' process deficiencies? The next section discusses the characterisation of behaviours that are related to these 'atypical' responses and how these can be used to develop a more flexible assessment instrument, one that does not just try to isolate 'typical' process deficiencies.

2.16 BEHAVIOURAL MARKERS AND THE ASSESSMENT OF CRISIS MANAGEMENT

Regester and Larkin (1997) propose that during a crisis there are some specific behaviours that are commonly exhibited. Persons within a crisis situation may see the situation around them as some surreal image rather than as the reality they are experiencing. They may have a belief that because this situation has not happened in the past, it cannot be happening to them now. They may have the hope that because emergency procedures have been written, the situation will rapidly be brought under control. They may also find that, due to the stress of the situation, their communication skills fail them, just when they need them the most.

Within the assessment process the assessor has to make a judgement about the competence or otherwise of the candidate being assessed. Although some crisis management skills, such as situational awareness and decision-making, are cognitive, Seamster et al. (1997) argue that cognitive skills can be analysed using non cognitive methods and that these skills can be identified and linked to observable behaviours. Other crisis management skills, such as leadership and co-operation, are social skills, and as such can be characterised by observable behaviours. Tourville (1997) determined key overt behaviours that were exhibited by team members who successfully managed 'crisis' situations. These behaviours have been characterised as follows:

- Delegation of tasks based upon known crewmember strengths.

  In order to ensure the most effective response to a crisis situation the crisis manager must know the strengths and weaknesses of their individual team members in order that this can inform their decision-making when delegating tasks.

- High degree of crewmember integration within the team.

  Crewmembers participate fully and actively within the team's tasks and are not left isolated from other members of the team. Crewmembers are kept fully aware of what is happening, and are regularly updated on the status of the team's task.

- Information backup provided during periods of ambiguity.

  When a team leader tries to assimilate ambiguous information, their level of situational awareness is affected, which in turn can lead to poor decision-making.

Therefore, during periods of ambiguity, the team leader must seek alternate sources of information in order to reduce the degree of ambiguity towards a degree of certainty.

- Regular soliciting of information between crewmembers.

  In order that the level of awareness of the situation is maintained across all members of the team, information should be regularly shared between team members. This should lead to better decision-making and a better acceptance of decisions by the team.

- Integration of real-world operations experience, rather than a reliance on procedures.

  If one of the definitions of a crisis is that there is no emergency procedure for the situation presented, the crisis manager must be able to call upon their experience in order to aid their decision-making and guide their actions. This is the basis of recognition-primed decision-making, where experts use their experience to recognise a situation as a certain type of situation and then, through a process of internal evaluation, retrieve from their memory an adequate stored pattern of response (Orasanu, 1997).

- Extensive "what-if" analyses undertaken.

  The crisis manager should be generating multiple options for explaining the situation they are in, how the situation might develop in the near future, and courses of action to take to address the situation. If this is not done, and the crisis manager follows a course of action that is based upon only one interpretation of the situational cues, extensive time and effort can be wasted, if this sole option is either incorrect, or unsuitable.

- Task prioritisation carried out.

  During the crisis situation the crisis manager should be prioritising the tasks to be undertaken so as to manage the crisis most effectively. Secondary tasks should not hinder primary task completion.

- Adaptation of plan to take into account evolving events.

  As the crisis develops, the crisis manager's awareness of the situation should be sufficiently good so as to allow changes to be made to the planned course of action after taking into account how the situation is likely to develop in the near future.

- Distracting information coded and assigned level of urgency for attention.

  As the crisis manager assimilates information about the situation from different sources, the information is prioritised, so that information of little current significance does not distract attention from information required for more immediate higher priority decision-making.

- Maintain a focused attitude to task objectives.

  During a crisis, the crisis manager's attention is not drawn away from the primary task objectives.

- Contingency plans developed on an ongoing basis.

  Although contingency plans should be in place for many emergency situations, by its very nature a crisis is a situation for which a specific contingency plan will not have been developed. It is therefore necessary for the crisis manager, through a high level of situational awareness, to predict possible future events, and to develop contingency plans to handle these events on an ongoing basis.

- Overt awareness of time as a finite resource by questioning time management status.

  The crisis manager will overtly question and update other team members in relation to the time criticality of tasks in order to maintain a shared temporal overview of the crisis amongst the team.

- Demonstrates awareness of 'big picture' context.

  The communications and decisions made by the crisis manager show an understanding of the whole crisis situation and an ability to maintain and update this overview throughout the crisis.

- Closely co-ordinates actions with others.

  The crisis manager maintains an awareness of the actions of other parties involved in the crisis situation and takes these into account when making decisions about future actions.

- Gives consideration to the needs of others.

  The crisis manager maintains an awareness of the activities and state of other team members in order to be considerate towards them when making further decisions.

- Utilise team briefings for specific threats.

  During the course of the crisis situation the crisis manager will call the team together to discuss specific threats that have arisen.

- Provide a large range of response options to threats.

  When faced with specific threats the crisis manager will use a team review in order to propose a large range of possible response options.

- Crewmembers performance thoroughly questioned and crosschecked.

  The crisis manager will ensure that all of the decisions and actions taken by team members are crosschecked, and if necessary questioned by other team members. This should include the crisis manager's own decisions and actions.

- Prompts for timely response to correct any crewmember-induced errors.

  Through a high level of awareness of the actions being taken by other members of the team, the crisis manager is able to crosscheck their actions and intervene in a timely manner to prevent any errors made from having a detrimental impact on the situation.

- Prompts for timely response to any change in situation.

  Through a high level of situational awareness the crisis manager is able to make decisions to ensure that the required responses to any changes in the crisis situation are effective.

- Keeps crewmembers focused on the accomplishment of the task.

  The crisis manager maintains the team members focus on achieving the required tasks.

- Communication kept succinct.

  The crisis manager maintains the succinctness of all communications between members of the team.

- Effectively increases work rate to cope with escalation of crisis.

  The crisis manager motivates the team to effectively meet the challenges presented by any escalation of the crisis situation.

Although these behaviours were observed during simulated air combat mission training sessions, they would appear to be generic in nature, and may well be found to be relevant to the merchant marine domain. It is assumed that this list is not exhaustive. If, as suggested by Tourville, overt behavioural markers can be used as the basis of an assessment framework for assessing competence in crisis management in the air combat mission domain, how such markers could be used to assess competence in crisis management within other safety critical organisations would need to be established.

One of the factors that may be important in determining if these behaviours can be used to assess competence in crisis management in other domains is 'time'. The time available to the crisis manager during the crisis situation, in which it might be possible to exhibit the behaviours described above, is an important consideration for the validity of any assessment framework. If crisis situations, within the domain for which assessments are to be undertaken, occur over a very short timescale, then it will be difficult to provide an assessment framework based on observable behaviours. Within the short time of the crisis situation, too few overt behaviours may be exhibited to make a valid assessment of competence.

Within the merchant marine domain it can be noted from the industry's casualty report data that crises usually occur over many hours, or even days. It is therefore anticipated that there will be sufficient time within the simulation scenarios within this research for assessors to observe a sufficient number of relevant overt behaviours in order to assess a candidate's competence in crisis management. However, this does raise the question of how many discreet observable behaviours are required to be assessed before a valid overall assessment of competence in crisis management can be made.

It is commonly argued that behavioural markers cannot be used to assess competence because the candidate for assessment will just learn the behaviours that are required to be exhibited in order to achieve a good assessment, and not the skills that should underpin and elicit those behaviours. The generally accepted counter argument to this is that if exemplar behaviours are being exhibited, then the candidate must be using the skills required to elicit those behaviours. If competence in crisis management skills leads to exemplar crisis management behaviours, it could be argued that, exemplar crisis management behaviours arise from competence in crisis management skills.

However, when an assessor undertakes an assessment based on behavioural markers, the assessment is usually not solely based on the exhibition of behaviours, but on how the exhibited behaviours relate to the success of the processes and/or outcomes of the situation. It could be argued that a single behavioural marker cannot be taken in isolation as an indication of good performance, but that an assessment should be based upon the exhibition of a related set of indicative behavioural markers, observed in the context of the assessment scenario. This again raises the question of how many behavioural markers need to be in this assessment set. Theoretically, it might be possible to find a single behavioural marker that was indicative of competence in crisis management.

In research studying the assessment of teamwork dimensions of teams working in complex environments (Smith-Jentsch et al, 1998), it was shown that dimensions requiring raters to infer a state of mind, such as team monitoring and situation awareness, were less reliable than those that were based on the observation of overt behaviours. The research showed that a number of teamwork behaviours were distinct and objectively observable measures of performance. These behaviours were:

"Information exchange	-	*Seeking information from all available sources* *Passing information to the appropriate persons before being asked* *Providing big picture updates*
Communication	-	*Using proper phraseology* *Providing complete internal and external reports* *Avoiding excess chatter* *Ensuring communications are audible and ungarbled*
Supporting Behaviour	-	*Correcting team errors* *Providing and requesting backup or assistance when needed*
Team Initiative and Leadership	-	*Providing guidance or suggestions to team members* *Stating clear team and individual priorities."*

(Smith-Jentsch et al, 1998)

In experiments, Smith-Jentsch et al showed that for some behaviours, there were high correlations between the ratings of a team across a range of scenario events. They argued that this was evidence of the relatively stable nature of these behaviours as an assessment tool. They referred to this as '*convergent validity*' and proposed that this was an indication that these behavioural markers were diagnostic measures.

They argued that correlations between different behaviours within a particular scenario event should be low. If these correlations are high, they concluded that the ratings were biased because the raters were probably limiting the range of their assessment criteria. They proposed that for a set of behavioural markers to be truly diagnostic, multiple ratings of these behaviours should be relatively independent of one another. They referred to this as '*discriminant validity*'.

Smith-Jentsch et al's research showed that it was difficult to rate some behavioural markers consistently at both the individual and team levels. They state that:

> *"Raters are believed to have a limited cognitive ability to discriminate among multiple performance dimensions. Raters are often prone to decision biases, such as halo error, in which judgements of one dimension are coloured by impressions formed on other dimensions."*

They hypothesised that any correlation between different behavioural markers within a scenario exercise may represent true '*exercise effects*' that mirror stable dispositions towards behaviours exhibited in particular situations in the real world. They found that the behaviours of information exchange, communication and team initiative/leadership were directly related to the quality of team decisions across different scenario events.

Following their research, Smith-Jentsch et al recommended that team performance assessment dimensions should be "*defined in terms of specific behaviours that are objectively observable and or audible to raters*". They state that regardless of the number of dimensions assessed, three to five dimensions are generally sufficient to account for subject variability.

They propose that raters should evaluate instances of specific behaviours within an assessment dimension. These ratings, they suggest, can then be combined mathematically to give an overall rating. They recommend this assessment methodology as a way of preventing the bias that they state is associated with trying to give an overall rating directly. Using this methodology, they were able to find a strong relationship between accurate decision-making and effective information exchange, communication and team initiative/leadership. They stated that this relationship remained strong even with contextual differences across the scenario events.

Schwab et al (1975), however, cautioned against the use of any behaviourally anchored rating scale if that scale was not defined by a large number of examples. They argued that if the rating scale is not defined clearly enough by using a large number of examples of behaviour the assessor may have difficulty in assigning observed behaviours to specific assessment dimensions. They also argued that without sufficient examples the assessor would have difficulty in deciding the scale value of the effectiveness of the observed behaviour. Errors by the assessor in either assigning behaviours to dimensions, or rating the effectiveness of the behaviour, could both be significant sources of assessment variance. Williams et al (1997) stated that there is a dilemma associated with any assessment task and that is to distinguish bias or cognitive distortion from truth.

If assessors are to give consistent, reliable and accurate ratings, then they will need to agree on both their behavioural observations and their performance ratings when they assess the same assessment scenario. This is because assessors should be interchangeable; the assessment rating should in no way be dependent upon which particular assessor undertaking the assessment (Brannick & Pearce, 1991; Baker et al, 2001).

Baker et al (2001) define two types of rater accuracy:

Observation Accuracy	-	the extent to which raters can correctly identify and record behavioural information.
Rating Accuracy	-	the extent to which raters assign the correct rating (i.e. on a defined rating scale) to the particular level of performance that was observed.

Baker et al point out that there is a critical distinction between the process of observation, that involves detection, perception and recall, and the process of assessment, that involves categorising, integrating and evaluating. Both of these processes can be a source of assessment error. They go on to advocate a regime of assessor training which they suggest should reduce assessment error and lead to high levels of interrater reliability and agreement. The forms of assessor training they discuss are:

Rater Error Training - used to familiarise raters with common rating errors such as:

Halo	-	where the judgement of one dimension is affected by the impressions formed on other dimensions;
Leniency	-	where the judgement is always overly lenient;
Severity	-	where the judgement is always overly severe;
Central tendency	-	where the judgement is always average.

Performance Dimension Training – used to train raters to recognise and use the appropriate performance dimensions and rating scales, and to rely on these dimensions and scales when making observations.

Behavioural Observation Training – used to provide raters with strategies that focus on observing and recording behaviour. Raters are trained to take detailed notes during observations.

Frame-of-Reference Training – used to provide raters with different standards of performance on dimensions. This training includes rating practice and feedback so that raters can adopt a common frame of reference. Raters are given samples of varying levels of performance on behaviours that represent each dimension. These samples are provided on videotape and the raters can then practice, and get feedback on, the use of the performance standards. Subject matter experts are used to give a 'true' rating of performance against which the rater can gauge their own assessment performance.

Baker et al argue that any videotapes used for rater practice should be of actual performance, as opposed to a scripted performance. They suggest that this is because actual performance typically contains more subtle variations that are harder for raters to observe and distinguish.

In a review of the use of behavioural markers for the assessment of crew resource management skills within the civil aviation industry, Flin and Martin (2001) categorise resource management skills into either cognitive or inter-personal. Within the cognitive category, they place the skills related to situation awareness, workload management, planning and decision-making. Within the inter-personal category, they place the skills related to crew coordination, communication, leadership/followership, and group-climate. They go on to discuss the wording of behavioural markers for use in assessing these categories of resource management skills. They argue that behavioural markers should be concise and use simple wording, and that the verb of any behaviourally anchored assessment criteria should relate to a '*clearly observable*' behaviour such as '*monitor*' or '*ask*'. They give an example of an unobservable behavioural marker - '*made a decision*'; and an example of an observable behavioural marker – '*communicates a decision*'. However, when developing any assessment framework relating to resource management skills, it should be noted that Seamster et al (1994) argued that a significantly greater proportion of the problems associated with the assessment of resource management skills came from the cognitive category rather than the inter-personal category. One reason for this is that inter-personal skills are more easily identified with specific overt behaviours than cognitive skills.

Birnbach and Longridge (1993) suggested some guidelines for enhancing the reliability of any crew resource management (CRM) assessments. Firstly, they suggest that any measurement of CRM should be based on observable behaviours that could be clearly defined and explicitly identified so that no ambiguity can exist in the assessment process. Secondly, they suggest that

CRM assessments should occur within standardised environments to ensure that all assessments are evaluated in the same way, under the same conditions. Lastly, they suggest that assessors should be 'calibrated' so that they are consistent in their assessments. They suggest that assessors should be trained and periodically assessed to ensure that high levels of interrater reliability are achieved.

The degree of awareness that a candidate has of the skills and associated behaviours being evaluated during an assessment exercise has been given the term '*skill transparency*'. It has been argued that during an assessment exercise, the higher the level of skill transparency a candidate has, the more they will introduce characteristics into their performance to meet the demands of the assessment criteria and thereby remove the variability in their performance that is due to their natural disposition (Kleinmann, 1993).

Smith and Smith-Jentsch (2001) suggest that even candidates for assessment who, for example, would not normally use assertiveness to handle a particular type of situation, may do well in an exercise where they knew they were being assessed on their ability to be assertive. They suggest that the resulting measure may be an excellent indicator of the candidate's pure skill at using assertiveness, or what one **can** do, while being a poor indicator of typical real world performance in similar situations, or what one **will** do. They argue, "*the degree of skill transparency associated with a situational exercise should be considered when interpreting behavioural assessments*". They suggest that when developing situational exercises for assessment using behavioural markers, the purpose of the assessment should be carefully considered. If the purpose of the assessment is to determine the candidate's level of skill or ability, then they argue that the behavioural markers can be transparent to the candidate, and that this type of exercise may be useful to assess the effectiveness of a training programme in developing skills. However, if the purpose of the assessment is to determine how a candidate might typically perform on-the-job in a real situation, they argue that the behavioural markers should not be transparent to the candidate. If the result of any assessment were to be only either '*competent*' or '*not yet competent*', it would not be necessary to disclose the behavioural markers used in the assessment process to the candidate.

Seamster et al (1995) proposed that when assessing competence related to crisis management the overall assessment scenario should be broken down into discrete "event sets". They describe event sets as:

> *"a group of complex related events designed into an assessment scenario to stimulate the candidate to perform specific crew resource management skills."*

They argue that assessors should rate candidate's performance in each of the event sets, and not their overall performance. Their research showed that if assessors rate performance for the whole assessment scenario, they give a much higher mean rating, relating to better performance, than if they assess the discrete event sets. They argue that the lower mean ratings, relating to poorer performance, resulting from the "event set" method of assessment may more accurately reflect the actual performance in the real world.

From the literature, it would appear there are certain overt behaviours that can be directly related to the skills required by successful crisis managers. Even though these skills may be both cognitive and non-cognitive, they can be identified and linked to observable behaviours. It would also appear that some of these behaviours could be used to objectively assess competence in crisis management, with good levels of interrater reliability and agreement, as long as raters are suitably trained. For competency assessment that is relevant for the work place, it would appear that behavioural marker based assessment frameworks should not be made transparent to the candidate.

2.17 SUMMARY OF RELEVANCE OF LITERATURE REVIEW TO RESEARCH

The use of behavioural markers is generally accepted as a valid approach to the assessment of competence in many safety critical industries and organisations.

The use of overt behavioural markers as assessment criteria is a way of directly observing the outputs of the cognitive processes taking place within the minds of those being assessed. However, considerable research is still required before cognitive analysis could produce any certainty about the cognitive processes involved when an operator undertakes a complex dynamic task.

The interactions between the crisis manager and the crisis management team play an important part in the effective management of a crisis. Therefore, it is the behaviours that define these interactions within the team environment that may be used as markers for assessing the team leaders competence in managing crises. By deconstructing the crisis management team, so that the individual team member attributes can be identified, further insights into the interaction between team leader and team member can be obtained, leading to a better understanding of the behaviours exhibited by the team leader during a crisis situation.

The literature has shown that when undertaking research that relies upon the abilities of participants to make decisions, the research environment should be as realistic as possible, providing as many of the cues that would be in the real world as possible to facilitate the choice of decision strategies. The integrity of the realism in any simulation used to study behavioural markers of crisis managers would therefore appear to be of the highest importance.

The literature would appear to support the argument that because there are no consistent relationships between processes and outcomes, and some process violations are beneficial, the use of complete models of the expert behaviour exhibited when undertaking a process, as assessment criteria, would not seem to be a viable option. Instead, the behavioural markers that support and facilitate successful processes, beneficial process violations and successful outcomes, would seem to be the most relevant assessment criteria. It would also appear that it is how factually based a behavioural marker is that will determine its objectivity during use in an assessment.

Throughout the literature review, many overt behavioural markers have been proffered that would seem to be potentially useful in the assessment of competence in crisis management. These behaviours are shown in Table 3 below. They will be evaluated in Chapter 5 to determine their suitability for inclusion within a behaviour based assessment framework for assessing competence in crisis management within the domain of a merchant naval vessel's engine room control room environment.

	Behavioural Marker
1	Ratio of degree of feedback control to degree of predictive control
2	Integration of team member attributes and values to maximum advantage
3	The number of alternative hypotheses and actions communicated to team members
4	The level of 'satisficing' exhibited
5	Communicating in a way that reveals ones mental models
6	Relevance and timeliness of unsolicited information passed between team members as measures of the degree of congruence of their mental models
7	The level of anticipation of the needs of other team members and of future action and task requirements
8	The amount of relevant information passed in anticipation of need, as a ratio of the amount of relevant information available
9	The periodicity of situation assessment updates to the other team members
10	Ignoring feedback delays
11	Focus too much on the reduction of uncertainty by requesting information
12	Inaccuracy in timing the right moment of intervention
13	Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system
14	Lack of knowledge concerning the relation between symptoms and causes
15	The number of dimensions upon which a decision is based
16	The length of term of considerations upon which a decision is based
17	Amount of sampling behaviour exhibited
18	Time between the onset of a disturbance and the first time a related action was carried out or an information request was made; defined as the disturbance detection time.
19	The time span between the first information request or action related to a disturbance and the correction of that disturbance
20	The number of information requests made about a disturbance
21	The number of times a supervisor inspected systems through their graphical control interface, as indicated by the number of mouse clicks on the control screen giving a measurement of monitoring behaviour

22	The division of time and effort over different essential sub-goals
23	The ability to switch between the processing of information about different sub-goals, as opposed to the need to completely finish one sub-goal before considering another
24	Explicitly asking their team how much time they have before they must commit to a decision
25	Communication behaviours (see Table 1)
26	The degree to which a crisis manager monitors the information exchange between members of their crisis management team
27	Seeking information from all available sources
28	Passing information to the appropriate persons before being asked
29	Providing big picture updates
30	Using proper phraseology
31	Providing complete internal and external reports
32	Avoiding excess chatter
33	Ensuring communications are audible and ungarbled
34	Correcting team errors
35	Providing and requesting backup or assistance when needed
36	Team leader focuses on individual tasks rather than the teams'
37	The level of useful communication amongst the team
38	Providing guidance or suggestions to team members
39	Stating clear team and individual priorities.
40	Delegation of tasks based upon known crewmember strengths.
41	High degree of crewmember integration within the team.
42	Information backup provided during periods of ambiguity.
43	Regular soliciting of information between crewmembers.
44	Integration of real-world operations experience, rather than a reliance on procedures.
45	Extensive "what-if" analyses undertaken.
46	Task prioritisation carried out.
47	Adaptation of plan to take into account evolving events.

48	Distracting information coded and assigned level of urgency for attention.
49	Maintain a focused attitude to task objectives.
50	Contingency plans developed on an ongoing basis.
51	Overt awareness of time as a finite resource by questioning time management status.
52	Demonstrates awareness of 'big picture' context.
53	Closely co-ordinates actions with others.
54	Gives consideration to the needs of others.
55	Utilise team briefings for specific threats.
56	Provide a large range of response options to threats.
57	Crewmembers performance thoroughly questioned and crosschecked.
58	Prompts for timely response to correct any crewmember-induced errors.
59	Prompts for timely response to any change in situation.
60	Keeps crewmembers focused on the accomplishment of the task.
61	Communication kept succinct.
62	Effectively increases work rate to cope with escalation of crisis.
63	Thanking another team member for pointing out a mistake.
64	Helping other team members having difficulty.
65	Making motivating statements.
66	Praising other team members.
67	Suggesting ways to find an error.

Table 3: Behavioural markers potentially related to competence in crisis management.

Chapter 3. REVIEW OF EXISTING ASSESSMENT PRACTICE

This chapter will review some of the existing practice for the assessment of competence in crisis management from a range of safety critical industries and organisations. The review will detail some of the assessment criteria used and will discuss the behavioural markers related to these criteria, and how objectively these markers can be assessed. A discussion on the current state of assessment of competence in crisis management within the commercial shipping industry will also be presented. The chapter will conclude with some direct comparisons between the different assessment frameworks reviewed, and a discussion of their relevance to any assessment framework developed for assessing competence in crisis management of merchant marine engineering officers.

3.1 MILITARY

In the literature review it was argued that the assessment of competence of crisis management should be undertaken in an environment that resembles the real world situation as closely as possible, in order to ensure that, as far as possible, all of the required cues for decision-making are available. Through their use of war games, utilising real hardware, the military attempt to provide this environment.

The observation of behavioural markers in such large and complex environments is a great challenge. To meet this challenge the military 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. This gives the advantage of a moderation process within the assessment.

The assessment is for the command team as a whole, and not for individuals within the team. Within military navies, the practice is for the command and control team to be kept together for a tour of duty onboard a vessel, which may be as long as two years. This therefore allows the team to be assessed together, rather than there being any necessity for individuals to be assessed within the team.

The following observations were made by the author during the assessment of a warship's command and control team taking part in a major war game exercise:

- There were twenty assessors used throughout the vessel.

- The assessment was divided into specific areas; initial actions, firefighting, containment, re-entry, mechanical aspects, electrical aspects, command and control.

- Assessors attempted to get candidates to verbalise their thought processes and to state priorities during the assessment exercise. This seemed at times to hinder the candidates' performance.

- Assessors would give 'hints' to candidates at times during the assessment exercise in order to prevent one mistake affecting too many other aspects of the exercise. However, there were also indications that some of these 'hints' were an extension of the training, and that training and assessment were therefore at times being undertaken simultaneously.

- There was at times some dissatisfaction exhibited by the assessors towards the set assessment criteria. This dissatisfaction was borne out of the openness to interpretation of some of the criteria used. An example of this was if a failure was critical or not critical to the task.

- The assessment had to be made under strict time limitations. This led to single point interpretations by the senior assessor if the moderation discussion between assessors was taking too long.

- On some occasions there was no positive outcome from the moderating process and this led to an 'average' assessment being given.

Some examples of the behavioural markers used as assessment criteria by a military navy are shown in Table 4 below. 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 complex nature of the assessment criteria, subjective assessments are inevitable. However, because of the large number of assessors used, fair and effective assessments can be achieved through moderation. Support was given to the overall assessment process by a number of candidates who said that in general they thought that the assessment of their performance was fair.

Command and Control
Did the Command team quickly close up at HQ1.
Was a comprehensive set of check cards / aide memoirs provided and used.
Were smoke boundaries established and effectively maintained.
Was there a good flow of information into the control position at all stages.
Was the incident picture well kept.
Was an alternative control position considered.
Were hands piped to emergency stations in good time.
Was there an organisation for accounting for personnel and did it work effectively.
Did the Damage Control Officer have firm control of the situation.
Was there a good flow of information between the Damage Control Officer / Electrical / Propulsion Managers regarding the state of essential services.
Between them, did they provide sound support to the Marine Engineering Officer.
Was there a good flow of information to the Command.

Table 4: Examples of Military Command and Control assessment criteria (Royal Navy, 2002)

It can be seen that some of these assessment criteria are descriptions of very complex sets of overt behaviours and others infer some form of covert cognitive processing. Each of these criteria could be broken down into more simple and discrete overt behaviours. For example the criteria *"Was the incident picture well kept"* could be broken down into the overt behavioural markers of:

- Verbal feedback about incident situation given by team members to team leader.
- Team leader gives clear and concise command orders after receiving feedback.
- Team leader updates incident status board.
- Reference is made to incident status board by team members.
- 'Command Huddle' technique used to elicit information from team members.

By observing these more simple and discrete overt behavioural markers, the complex cognitive processing assessment criteria of *"Was the incident picture well kept"* could be implied.

The criteria within the military Command and Control assessment framework are used to give an overall assessment of the team's performance based on a seven-point rating scale. This rating scale is shown in Table 5. below. This assessment is used within the individual officer's appraisal reports as part of the officer's overall assessment of their potential, so cannot be considered as a non-jeopardy assessment.

Assessment	**Definition**
VERY GOOD	Outstanding results achieved. Execution of the highest order. High standard throughout. Innovative solutions to problems. Equipment, fittings and systems well maintained and free from defects.
GOOD	Good results achieved in almost all respects. Execution requires only a few minor improvements. Few material / organisational /administrative deficiencies. Prepared effectively.
VERY SATISFACTORY	Execution, equipment, performance and organisation / administration at a strong standard, but a number of shortcomings, although minor, does not merit a Good assessment.
SATISFACTORY	Adequate results achieved with some shortcomings (one or two of which could be major (significant)) with scope for improvement. Execution safe and steady but could be improved. Some organisational / administrative deficiencies. Equipment generally providing effective performance.
JUST SATISFACTORY	Although not poor enough to merit a Below Standard assessment, execution, equipment, performance or organisational / administrative need further improvement before a Satisfactory standard is achieved. Several significant weaknesses.
BELOW STANDARD	One or two critical weaknesses. A number of other significant weaknesses. Considerable wide improvement necessary. Execution inadequate.
UNSATISFACTORY	Several critical weaknesses. Unacceptable execution / material / organisation / administrative deficiencies.

Table 5. Military Command and Control assessment rating scale. (Royal Navy, 2002)

It can be seen that the definitions within the overall assessment rating scale above are very broad. With the rating of the individual behavioural marker criteria having been assessed as either 'met' or 'not met', the process of then obtaining an overall assessment based on the definitions above was observed to be very difficult and very subjective. However, it appeared to achieve a valid assessment due to the process of moderation.

The use of complex assessment criteria is bound to lead to more subjective assessment outcomes, as evidenced during the observation of the military navy assessment exercise. However, the assessment of competence in crisis management by the military, even with the use of high-level complex behavioural markers, has been found to be fair and effective because of their use of large numbers of assessors and a process of moderation.

3.2 Civil Aviation

Within the civil aviation industry, the training of crew resource management skills has been introduced as a way of improving safety performance. The civil aviation industry has recently been undertaking research into the best way to assess the non-technical skills of aircrew (JAR TEL Consortium, 2001). Non-technical skills can be defined as those skills, in addition to technical skills, required for competence in crisis management. There are four main categories of resource management skills, or non-technical skills, being used within behavioural marker systems within the civil aviation industry:

- Co-operation
- Leadership and Management
- Situation Awareness
- Decision Making (Flin & Martin, 1998)

The European Union research project 'Joint Aviation Requirements Translation and Elaboration of Legislation' (JAR TEL Consortium, 2001) has evaluated the use of such a behavioural marker system for the assessment of resource management skills of commercial flight crews. Through the JAR TEL research project, a methodology for assessing the non-technical skills of aircrew, by observing individual overt behaviours, has been proposed. The JAR TEL project concluded that such an assessment framework: "*is capable of proving itself a valid and reliable method for assessing non-technical skills*."

However, the reservations of civil aviation pilots to the concept of the assessment of non-technical skills are also discussed in the JAR TEL report. One of the pilots main concerns was that they felt that the criteria on which any assessment is based are largely subjective, and thus cannot easily be monitored for fairness and accuracy (JAR TEL Consortium, 2001). Some examples of the behavioural markers used as criteria within the JAR TEL assessment framework are shown in Table 6 below.

Behavioural Markers indicating good practice:	*Behavioural Markers indicating poor practice:*
Reflects on the suggestions of others.	Ignores suggestions of others.
Demonstrates will to achieve top performance.	Does not care for performance effectiveness.
Takes condition of other crewmembers into account.	Does not take account of other crewmembers.
Secondary operational tasks are prioritised to retain sufficient resources for primary flight duties.	Secondary operational tasks interfere with primary flight duties.
Allocates enough time to complete tasks.	Ignores signs of stress and fatigue.

Table 6. Examples of Civil Aviation Non-Technical Skills assessment criteria (JAR TEL Consortium, 2001)

It can be seen from these assessment criteria that the reservations held by some pilots may be justifiable, especially when viewed in the light of the interrater agreement data discussed previously in section 2.13. Behaviour such as "reflection" would be difficult to observe, unless this reflection resulted in some further related behaviour. In addition, assessments based upon metrics such as "enough" and "sufficient" could be seen as possibly leading to subjective outcomes.

The cockpit environment is very different to that of the military's 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. As with the military assessment framework discussed in section 3.1, the JAR TEL criteria are assessed as having either been passed or failed. The assessments are carried out within high fidelity full mission aircraft simulators, using a single assessor. However, this time the assessment environment is a non-jeopardy environment. A number of scenarios can be used within the simulator in order to assess the performance of the candidates. These may relate to either normal or abnormal flight conditions.

Pilots are also subject to assessment by the use of Line Operations Safety Audits (LOSA), which are non-jeopardy observations of crews conducting normal line flights. The basis for many of the behavioural markers rating scales in use for this type of assessment is the 'University of Texas Behavioural markers Rating Scale', which is detailed within the UK Civil Aviation Authority's guidance document CAP 737 (CAA, 2003) and within the International Civil Aviation Organization's guidance document 9803 AN/761 (ICAO, 2002).

The focus of the LOSA assessment criteria is on the management of threats and errors. Examples of the behavioural markers used as criteria within the LOSA assessment framework are shown in Table 7. below.

	Behavioural Markers	
PLANS STATED	Operational plans and decisions were communicated and acknowledged.	Shared understanding about plans; Bottom lines were established.
WORKLOAD MANAGEMENT	Roles and responsibilities were defined for normal and non-normal situations.	Workload assignments were communicated and acknowledged.
VIGILANCE	Crewmembers remained alert of the environment and position of the aircraft.	Crewmembers maintained situational awareness.
AUTOMATION MANAGEMENT	Automation was properly managed to balance situational and/or workload requirements.	Automation set-up was briefed to other members; Effective recovery techniques from automation anomalies.
INQUIRY	Crewmembers asked questions to investigate and/or clarify current plans of action.	Crewmembers not afraid to express a lack of knowledge; "Nothing taken for granted" attitude.

Table 7. Examples of Line Operations Safety Audits (LOSA) assessment criteria (CAA, 2003; ICAO, 2002)

Unlike the JAR TEL assessment framework that uses a two-point assessment scale, pass or fail, the LOSA criteria are assessed against a four-point rating scale.

1 = poor	**2 = marginal**	**3 = good**	**4 = outstanding**
Observed performance had safety implications	Observed performance was barely adequate	Observed performance was effective	Observed performance was truly noteworthy

Table 8. Line Operations Safety Audits (LOSA) assessment rating scale (ICAO, 2002; CAA, 2003)

Non-technical skills assessments within civil aviation are currently not used to obtain an assessment of competence that can affect the issuance of a pilot's flying licence. They are only used to obtain data about the issues related to the use of non-technical skills during flight operations, and to inform pilot training programmes, to try to alleviate any problems found. The assessment of non-technical skills within civil aviation is therefore considered to be a non-jeopardy assessment.

A major difference between the assessment of competence in crisis management within the military context, and within 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.

3.3 OFFSHORE INDUSTRY

The following observations were made by the author during the assessment of an Offshore Installation Manager.

Within the offshore industry, the crisis manager of an offshore installation would be the offshore installation manager (OIM). The assessment of OIM's is undertaken in a high fidelity full mission offshore installation control centre simulator, using two assessors. The OIM is assessed as the team leader. The team working for the OIM during the assessment exercise is supplied by the assessment centre. This team is essentially "dumb", competently undertaking what they are asked to do and providing appropriate feedback to the team leader, but not using their initiative.

The OIM candidate is assessed over the course of three simulator exercises. After each exercise, the candidate receives feedback from the assessors on his or her performance. The candidate also has the opportunity to clarify their actions, decisions and intentions to the assessors. This interview therefore forms a sort of debrief of each assessment exercise.

Table 9. below gives some examples of the performance criteria used to assess OIM's.

OIM Performance Criteria
Appropriate resources are utilised throughout emergency.
Valid interpretations are made and valid decisions are taken throughout.
Appropriate actions are taken as quickly as possible.
Emergency response teams are coordinated and directed effectively.
Valid decisions are taken on which activities should be delegated in the light of the circumstances of the moment.
Actions and behaviours contribute to the confidence and effectiveness of the team at all times.
The appropriate degree of detachment is maintained throughout.

Table 9. Examples of Offshore Industry assessment criteria (RGIT Montrose, 2002)

It can be seen that these criteria relate to complex sets of behaviours and as such will be open to subjective interpretation. The use of adjectives such as *appropriate*, *valid* and *effective* within the criteria also means that there may be differences of opinion between assessors as to how appropriate, or how effective, a particular decision or course of action has been. In order to ensure that these differences do not adversely affect the assessment, two assessors are used to moderate the outcome.

Within the offshore industry, it is accepted practice that OIM candidate performances under simulated conditions are viewed as opportunities to identify areas for development and improvement (RGIT Montrose, 2002). Unlike the military and some civil aviation assessment frameworks, where the individual assessment criteria can have only one of two ratings, competent or not yet competent, the offshore industry assessors use the following rating scale for assessing the candidate:

1. *Fully meets the required standards in every respect.*
2. *Sufficiently meets the required standards – but with areas identified for improvement (requiring some advisory comment by the assessors).*
3. *Not yet demonstrating the required capabilities. (Within reason, the assessors will comment on those areas requiring substantial improvement).*

(RGIT Montrose, 2002)

Although individual elements of the assessment framework are assessed against this rating scale, the final overall assessment may not just be the arithmetic average of all of the individual assessment criteria ratings. If the assessors consider that the candidate has a low rating in a "critical area", the candidates overall rating may be given as lower than the arithmetic average.

The assessment criteria are derived from the Offshore Petroleum Industry Training Organisation (OPITO) approved standard of performance for an OIM controlling emergencies (OPITO, 2005). A candidate has to pass this assessment before they can be employed as an offshore installation manager. Therefore, as with the military assessment system, this is not a non-jeopardy assessment environment. Some of the companies that operate offshore installations require existing OIM's to undertake this assessment process at least once every three years, in order to prove their continuing competence.

3.4 FIRE BRIGADE

The following observations were made by the author during the assessment of a Fire Brigade Incident Commander.

Incident Commanders within the Fire Brigade have to manage crisis situations on an almost daily basis. The Fire Brigade therefore has a rigorous assessment procedure in place to ensure that all of their Incident Commanders are competent in the management of crises. The Fire Brigade has comprehensive simulation facilities in which these assessments are undertaken. These consist of a high fidelity full mission incident command vehicle simulator and incident scene simulations, presenting visual scenes of the incident. Interaction with third parties who may be involved in the incident, such as the police, ambulance service and members of the public, is also provided. Through this range of simulations, the majority of the cues that would be in the real world incident environment can be presented to the candidate.

The Incident Command Assessment process is part of the overall assessment of the Fire Brigade Officers used to assess their potential for promotion, and as such is not a non-jeopardy assessment. The candidates are fully briefed about the assessment process before the simulation exercise is started. The information presented to the candidate during the simulation exercise follows a standardised script to ensure that all candidates experience the scenario in the same way. In order to ensure this standardised experience, the team working with the Incident Commander during the assessment exercise is supplied by the assessment centre. As in the Offshore Installation Manager assessment described in the previous section, the team supplied by the assessment centre is once again essentially “dumb”. The assessment centre ensures that this team competently undertakes what they are asked to do by the candidate, and that they provide appropriate feedback to the candidate, but that they do not use their initiative.

Two assessors are used to assess the performance of each candidate. The assessors each have an assessment sheet, which has a separate page for each area of competence to be assessed. The assessors are trained to note clear examples of the candidate’s performance that relate to the demonstration of each particular competence. The assessors then also have to rate the candidate’s performance using a five-point rating scale that provides behavioural markers as indicators of both effective performance and less effective performance. Some examples of these behavioural markers are shown in Table 10. below.

Incident Commander Performance Indicators	
More Effective Performance Indicators	*Less Effective Performance Indicators*
Recognises relationships between different pieces of information.	Considers issues in isolation from each other and fails to recognise relationships between information.
Develops a structured process by which decisions are made.	Decision-making is haphazard and inconsistent.
Is proactive in order to deal with dynamic situation.	Is purely reactive to the dynamic situation.
Seeks to develop working relationships and rapport with others.	Is uninterested in building relationships with others.
Prioritises objectives based on identified risks and resources required.	Fails to prioritise objectives or priorities unclear or arbitrary; priorities do not reflect identified risks and resources required.

Table 10. Examples of Fire Brigade assessment criteria (London Fire Brigade, 2002)

The five-point Likert type rating scale used by the assessors to rate the individual performance of the Incident Commander candidate is reproduced in Table 11. below.

5	4	3	2	1
Strong Tendency To **More Effective**	Tendency To **More Effective**	No Strong Tendency Either Way	Tendency To **Less Effective**	Strong Tendency To **Less Effective**

Table 11. Fire Brigade Incident Command assessment rating scale (London Fire Brigade, 2002)

At the end of the assessment exercise, the two assessors provide the candidate with an overall assessment summary, highlighting the key elements of the candidate's performance. This feedback is used for the candidate's continuing professional development.

Some of the behavioural markers used in this assessment framework, such as '*develops a structured process by which decisions are made*', relate to cognitive processes. The presence of these types of marker can only be inferred by any related overt behaviours, such as the sequence of command orders, the updating of a status board, or the verbalisation of intended actions as feedback to other team members. However, other behavioural markers, such as *'seeks to develop working relationships and rapport with others'*, do relate to more simple overt behaviours.

3.5 ANAESTHESIOLOGY

There have been a number of behavioural marker based assessment frameworks developed to assess the competence of anaesthetists (Gaba et al, 1998; Fletcher et al, 2004). These frameworks have been used to assess the competence of anaesthetists working within a surgical team. The framework developed by Gaba et al was used to assess clinical performance during simulated crisis situations, including cardiac arrest and malignant hyperthermia.

These assessments frameworks were developed from the crew resource management assessment frameworks used in aviation and therefore have some similarities to these. They are currently only used in a non-jeopardy environment to gather data on performance issues related to the competence of anaesthetists in different clinical situations.

The assessment framework developed by Gaba et al utilised a patient simulator as the assessment environment. An anaesthesia team of two were assessed as part of a surgical team of four. The other members of the team were provided by the assessment centre, and acted as the surgeon and operating room nurse. Some examples of the behavioural markers used for the assessment of the anaesthesia team are shown in Table 12. below.

Crisis Management Behavioural Performance Markers	
Crisis Management Behaviour	Behavioural Markers
Inquiry/Advocacy/Assertion	Inquiry is encouraged, and questions answered openly; Crewmembers seek information from others and speak up with appropriate persistence; Someone other than the main anaesthetist assumes command when necessary; Anaesthetist(s) insists on suspension of surgery with appropriate assertiveness.
Communications	Crewmembers notify each other of necessary information; Messages stated precisely and to specific individuals; Crewmembers acknowledge communications and verify ambiguous communications; Efforts are made to establish and maintain an open atmosphere; Tone of voice is appropriate to the situation.
Feedback	Errors in management are identified and corrected; Feedback addresses positive as well as negative performance; Is given and accepted objectively and non-defensively; Is given in appropriate manner and times.
Group Climate	Crew and team members remain calm; Inappropriate behaviours are ignored or countered properly; Atmosphere in the Operating Room is relaxed but escalates as appropriate; Group attention is focused on patient care; Distractions are actively modulated (e.g. music turned down or off when workload increases).
Vigilance	Demonstrates awareness of special characteristics of patient or situation; Monitors and cross-checks all sources of information; Considers abnormalities to be real until proven to be false.

Table 12. Examples of Anaesthesiology assessment criteria (Gaba et al, 1998)

These assessment criteria are rated using a five-point rating scale. There is also a rating to indicate that no behavioural markers relating to a certain crisis management behaviour category were observed. However, if none of the behavioural markers for a particular crisis management behaviour category were observed, the assessment framework requires both the non observance rating and a rating from the five-point scale to be recorded. The assessment rating scale is shown in Table 13. below.

0	1	2	3	4	5
Not Observed (ALSO select a rating from 1-5)	Poor Performance	Minimally Acceptable Performance	Standard Performance	Good Performance	Excellent Performance

Table 13. Anaesthesiology assessment rating scale (Gaba et al, 1998)

Many of the criteria used in this assessment framework could be open to subjective interpretation. Adjectives are used within some of the behavioural markers, such as *appropriate* and *necessary*, that could lead to disagreement between assessors. In addition, the use of adverbs such as *objectively, properly* and *actively* can also lead to differences of opinion amongst assessors.

However, other behavioural markers are quite simple and uncomplicated, which would make them easy to observe. These are markers such as *'inquiry is encouraged'* and *'messages stated precisely'*. The omission of unobservable cognitive processes as markers also helps to make this assessment framework easier to use. This view is supported by the interrater agreement values obtained by Gaba et al when using this assessment framework. The within group interrater agreement coefficients (r_{wg}) for each of the crisis management behaviour categories assessed using the cardiac arrest scenario ranged between $r_{wg} = 0.65$ to $r_{wg} = 0.8$

3.6 COMMERCIAL SHIPPING

Unlike all of the other safety critical industries previously discussed in this chapter, the commercial shipping industry is unique in having a competence standard for ships officers that is applied throughout the world. All of the Flag State administrations of the world, through which vessels are registered and regulated, have agreed to assess the competence of their seafarers against the standards laid down in the Seafarer's Training, Certification and Watchkeeping (STCW) Code, (International Maritime Organization, 1995). Within this Code, a mandatory standard has been included that specifies the minimum standard of competence required in "*crisis* *management and human behaviour*" for those officers who have responsibilities for passengers. These competencies and their related assessment criteria are set out in table A-V/2 of the STCW Code. Examples of these assessment criteria are shown in Table 14. below.

Competence	Criteria for Evaluating Competence
Organize shipboard emergency procedures.	The shipboard emergency procedures ensure a state of readiness to respond to emergency situations.
Optimize the use of resources.	Contingency plans optimise the use of available resources; Allocation of tasks and responsibilities reflects the known competence of individuals; Roles and responsibilities of teams and individuals are clearly defined.
Control response to emergencies.	Procedures and actions are in accordance with established principles and plans for crisis management onboard; Objectives and strategy are appropriate to the nature of the emergency, take account of contingencies and make optimum use of available resources; Actions of crewmembers contribute to maintaining order and control.
Establish and maintain effective communications.	Information from all available sources is obtained, evaluated and confirmed as quickly as possible and reviewed throughout the emergency; Information given to individuals, emergency response teams and passengers is accurate, relevant and timely; Information keeps passengers informed as to the nature of the emergency and the actions required of them.
Control passengers and other personnel during emergency situations.	Actions of crewmembers contribute to maintaining order and control.

Table 14. Merchant Navy assessment criteria from the *Specification of minimum standard of competence in crisis management and human behaviour*. (International Maritime Organization, 1995)

Although entitled *'competence in crisis management'*, the competencies within table A-V/2 of the STCW Code relate primarily to the management of emergencies. From the discussion within section 2.1 of this thesis it was proposed that one of the ways a crisis differs from an emergency is due to a lack of pre-defined responses available for the situation being experienced. The competencies within table A-V/2 cite the use of procedures and actions in accordance with established principles and plans as a criterion for evaluating competence, and therefore are clearly related to the management of an emergency, not a crisis. In addition, the criterion for assessing competence in organising shipboard emergency procedures purely relates to the ability to prepare emergency procedures, not the ability to enact those procedures.

The assessment criteria given in Table A-V/2 of the STCW Code are also highly subjective and open to the interpretation of the assessor. The use of adjectives such as "*accurate, relevant and timely*" may lead to poor interrater reliability and agreement when assessing complex dynamic crisis scenarios. Similarly, the use of verbs such as "*contribute*" in assessment criteria invariably leads to differences in assessor interpretation. To 'contribute' to something is by definition only to play a part in the overall process that achieves an outcome. The significance of the part played by the candidate to the overall outcome of the competence being assessed, is left to the assessor to interpret.

The competencies within Table A-V/2 of the STCW Code are used to assess individual performance within a team-working environment. Assessment is usually carried out by an individual assessor observing the candidate perform during a simulated emergency, either at a shore based assessment centre, or onboard a vessel. The simulations are usually tabletop type simulations related to the command and control activities onboard the vessel. The assessment process must:

> *"provide evidence that the required standard of competence has been achieved in accordance with the methods and the criteria for evaluating competence" (International Maritime Organization, 1995)*

The methods for demonstrating competence are also set out within table A-V/2 as:

> *"Assessment of evidence obtained from approved training, exercises and practical demonstration"*

The approval of the training, exercises or demonstrations comes from the maritime administration of the country in which the assessment centre is based, or in which the vessel is registered.

3.7 SUMMARY OF RELEVANCE OF ASSESSMENT PRACTICE IN OTHER DOMAINS

So, what can be learnt from an evaluation of the assessment practice in other safety critical industries or organisations, and how could this inform the development of a behavioural marker based assessment framework for merchant navy engineering officers?

From the discussion in the previous sections of this chapter, it can be seen that safety critical organisations undertake the assessment of competence in crisis management in very different ways. Based upon observations within various safety critical organisations, Table 15 below provides a summary of their use of crisis management assessment frameworks.

Although using complex, high-level, assessment criteria, the assessment framework used by the military appeared to produce valid assessments. However, this is achieved using a very large number of assessors and some very expensive resources. The assessment environment used was virtually the real operational environment, to the extent that it was not entirely devoid of jeopardy. This type of assessment system would not be possible in the commercial shipping industry where the use of multiple assessors in order to moderate any assessment would not be viewed as economically viable. However, the concept that assessment criteria are either 'met' or 'not met' is an effective way of ensuring that assessors make a decision about the candidates' competence.

It may be feasible to apply the civil aviation assessment framework for non-technical skills within a merchant marine context, as the competences would appear to be transferable to other domains. However, due to the complex and in some cases cognitive nature of the assessment criteria, the assessment framework has not yet shown itself to be reliable when assessing competence in managing crisis scenarios. When using this assessment framework to assess pilots undertaking complicated 'crisis' scenarios, the within-group interrater agreement coefficient (r_{wg}) varied between zero and 0.2, or in other words, the agreement between assessors was not much greater than there would have been by chance. For this reason, the civil aviation non-technical skills assessment framework does not appear to be useful for the assessment of competence in crisis management within the commercial shipping domain. However, once again the concept that a candidate can only either pass or fail an assessment criteria, may be useful for the commercial shipping domain to ensure the clarity of assessments.

Context	Assessment Environment	Assessor(s)	Assessed	Assessment Criteria	Remarks
Military	High fidelity War Game within real environment.	Multiple Assessors distributed throughout assessment environment.	Team	Specific task orientated completion criteria.	Complexity of assessment environment leads to subjective interpretation. Fairness achieved through moderation.
Civil Aviation	High fidelity simulator.	Single	Individual working within a 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.
Offshore Oil/Gas	High fidelity simulator and simulations onboard.	Two	Individual working within a team.	Specific task orientated completion criteria and some overt behavioural markers.	Two assessors used to moderate subjectivity of assessment.
Fire Service	High fidelity simulator.	Two	Individual working within a team.	Specific task orientated completion criteria.	Two assessors used to moderate subjectivity of assessment.
Anaesthes-iology	High fidelity patient simulator	Single	Anaesthesia team of two working in a surgical team of four.	Overt behavioural markers with examples given of good practice.	Achieved acceptable interrater reliability.
Merchant Marine	Simulations onboard and tabletop.	Single	Individual working within a team.	Prescriptive, but very open to subjective interpretation.	Assessment framework too open to interpretation by assessing authority.

Table 15. Comparison of the Assessment Frameworks for Assessing Competence in Crisis Management within different Safety Critical Organisations

The offshore industry uses some complex high-level assessment criteria for the assessment of their Offshore Installation Managers. By the use of two assessors to moderate assessment outcomes, the use of debriefing interviews and the use of multiple assessment scenarios for each candidate, a fair and valid assessment appears to be achieved. However, this could not be corroborated, as no evidence was available to evaluate the level of interrater reliability and agreement achieved by the offshore industries OIM assessment framework. The offshore industries assessment process is quite resource intensive, not only because it involves two assessors, multiple assessments and candidate debrief interviews, but because the other members of the OIM's team are also supplied by the assessment centre to ensure the consistency and repeatability of OIM's supporting team's input into the assessment exercises. Once again, the large resource requirement for this type of assessment framework would make it unfeasible for use within the commercial shipping industry.

Although some of the behavioural markers used by the Fire Brigade in their assessment framework relate to cognitive processes, other behavioural markers used relate to more simple overt behaviours. It is these simple, low-level, overt behavioural markers that may be relevant to the commercial shipping domain. This is because by using this type of behavioural marker as assessment criteria, it may be possible to undertake an assessment by using only one assessor, and this assessor might not need to be a domain expert. This is because low-level overt behavioural markers are more easily observable, and distinguishable, than complex high-level behaviours. Using this type of behavioural marker might also mean that a high level of interrater reliability and agreement was achievable. The development of any assessment framework that could be reliably used with a minimum of resources would be attractive to commercial shipping companies, as it would mean that the cost of any assessment could be kept relatively low, when compared to how assessments in crisis management are currently undertaken in other safety critical domains.

The assessment framework used in anaesthesiology does have some criteria that could be open to subjective interpretation, but it also has some behavioural markers that are quite simple and uncomplicated, and therefore easy to observe. This assessment framework also omits any unobservable cognitive processes as behavioural markers, which also helps to make it easier to use. In use, this framework showed good interrater agreement coefficients (r_{wg}) ranging from 0.65 to 0.8. This is therefore good evidence to support the use of low-level, uncomplicated and easily observable behavioural markers as criteria for the assessment of competence in crisis management.

In summary, it can therefore be said that the assessment criteria within most assessment frameworks currently in use by safety critical industries contain a high proportion of behavioural markers that consist of complex sets of behaviours. The evidence from the interrater agreement coefficients would also suggest that these criteria are difficult to assess objectively, especially in real time during a crisis scenario, when there are a great many coincidental behaviours occurring in a short space of time.

If a new assessment framework is to be developed for the assessment of competence in crisis management for merchant marine engineering officers, the evidence above would tend to suggest that the following attributes are necessary:

- the assessment criteria should consist of low-level, uncomplicated behaviours;
- the behavioural markers should be both overt and easily observable;
- the criteria used need to be validated to show they are true indicators of competence in crisis management;
- the criteria should be assessed against a two-point scale of either ‘passed’ or ‘failed’;

The following chapter will detail the methodology used to develop a new assessment framework for assessing competence in crisis management for merchant marine engineering officers.

Chapter 4. RESEARCH METHODOLOGY (PHASE ONE)

4.1 BACKGROUND AND HYPOTHESIS

On starting work within the simulation section at Warsash Maritime Academy, the author of this thesis soon became aware that the simulator instructors would make their own personal assessments of the officers who were being trained within the simulator. Some instructors would make an assessment after only observing the officer in question for a relatively short period, sometimes as short as only thirty minutes. It was also observed that these initial assessments were quite often not modified, even after much longer periods of observation of up to four and a half days, because they were shown to be correct.

A degree of validity was afforded to these assessments when the officers were presented with a crisis situation to manage. Invariably, the officers who had been originally assessed as being 'good', were the most competent at managing the crisis situation.

The question was therefore raised as to the criteria that the simulator instructors were using in order to carryout their personal assessments of the officers being trained. When asked, the simulator instructors could not codify the criteria they were using, they would just say that it was "obvious" who was competent and who was not. It was this lack of ability to codify the criteria used to make these informal assessments that was one of the seeds for this research.

One thing that was 'obvious' was that these formal assessments were being made solely based on observable and overt behavioural markers exhibited by the officers during the exercises within the simulator. There were no other interactions between the officers and the simulator instructors that influenced the assessments. The instructors must have been observing very simple and uncomplicated behaviours, as assessments were made very quickly, without any apparent analysis or difficulty. In order to be able to assess competence in crisis management both simply and objectively, the assessment criteria used by these instructors needed to be codified. A reliable and valid assessment framework, based on these simple behavioural markers, would be a valuable tool for enhancing maritime safety.

From the observations detailed above, it is possible to propose the following hypothesis:

> It is possible to use simple overt behavioural markers to objectively assess competence in crisis management within the context of a simulated merchant vessel engine room control room.

In order to prove, or disprove, this hypothesis, an experiment was designed that recreated the conditions under which the simulator instructors had observed the performance of the marine engineering officers, before making their informal assessments.

During the literature review, evidence of the use of behavioural markers for the assessment of competence in managing crises within various safety critical organisations was recorded. Empirical evidence for the relevance, or irrelevance, of these markers to the commercial shipping domain was also required to be determined by experimentation. The crisis scenario experiment was designed to allow this empirical evidence to be collected and to provide evidence of any other behavioural markers that were relevant to the assessment of competence in crisis management.

The experiment is detailed in the following sections.

4.2 ETHNOGRAPHIC ANALYSIS

In order to capture empirical evidence of the behaviours of the research experiment participants the following methodological attributes were required:

Unobtrusiveness – so as not to adversely affect the natural behaviour of the participants;
Repeatability – so that data from a number of experiments could be collected in the same way;
Holistic – so that all relevant behaviours were recorded;
Permanence – so that the data could be analysed as often as required;
Contextual – so that the data reflect the setting in which the behaviours take place, as well as the behaviours themselves;
Accuracy – so that the data are a precise record of both the audio and visual elements of the observed behaviour;
Speed – so that none of the observable behaviour is missed;
Objectivity – so that the data captured is an objective record of the behaviours observed.

There are a number of problems associated with the assessment of competence through observation. If these observations are undertaken in real time, with no permanent record being taken, there may always be contradictions between what the observer believed happened during the assessment period, and what actually happened. In some cases, this happens because of defective or selective memory functions of the observer. In other cases, the observers' recollections and perceptions may be biased by their own experiences and beliefs. However, it may also happen because the activity being assessed may be very complex, with many behaviours being exhibited simultaneously, so making it impossible to adequately observe and interpret all of these behaviours in real time. There may also be many tacit behaviours exhibited, which although overt, are sometimes abstruse and difficult to observe, and need to be analysed in depth in order to interpret their significance.

When observing tasks being performed in a complex technological environment, such as the engine room control room of a vessel, the social interaction of those being assessed is often affected by the individuals' interactions with the technology that surrounds them. These interactions may cause subtle influences on the behaviours of those being observed, which are difficult to observe and interpret in real time (Luff et al., 1993).

The use of questionnaires and interviews to analyse and assess performance has also been shown to be fraught with problems, not least because the questions used in these methodologies, and their scope, are often based on the preconceptions of the assessor, and as such cannot produce an objective assessment. In addition, those being assessed will display many tacit behaviours of which they are consciously unaware, and therefore unable to describe in response to a questionnaire or during an interview. Due to the complexity of tasks, especially the management of crises, many of the behaviours exhibited are performed automatically, without the need for conscious thought. The close coordination of the activities of the team members in the engine room control room relies on mutual monitoring, as much by the tacit behaviours of overseeing and overhearing, as by any more direct forms of communication. As these behaviours are not normally described, they are also often difficult to elicit by just using questionnaires or interviews (Jirotka, 1998).

Ethnography is a qualitative research methodology that seeks to provide thorough descriptions of social interactions. Katz (1997) describes ethnography as

> *"telling the story of how people, through collaborative and indirectly interdependent behaviour, create the ongoing character of particular social places and practices"*

Ethnographic research methods focus on the behaviours of the members of a community or team, by studying them in natural, ongoing settings (DuFon, 2002). An ethnographic approach allows for behaviours to be investigated in the context in which they are produced, and to be interpreted and explained in relation to the environment of which they are a part (Watson-Gegeo, 1988). Ethnographic methods using video recordings try to alleviate the problems discussed above by capturing all of the behaviours exhibited in the naturalistic setting, from which a highly detailed description of how tasks are undertaken can be produced. In this way, not only can all of the explicit behaviours be analysed, but the implicit, tacit behaviours can be analysed as well (Jirotka, 1998). A video recording of the whole of the engine control room environment in which the crisis scenario takes place can provide all of the methodological attributes described above.

There are many different forms of ethnographic study, some of which involve the long term integration of the researcher into the community or team being observed, so as to try and limit the influence of the researcher on the natural behaviours of those being observed. Due to the limited time that the research participants were available to participate in the research experiments, and the fact that within the commercial shipping domain, teams are often only together for relatively short periods, it was decided that the ethnographic research methodologies where the researcher becomes embedded within the team, would be both impracticable and unnecessary. Discrete video recording of the teams undertaking the research experiments was seen to be a way of capturing the teams behaviours, without unduly influencing those behaviours.

DuFon (2002) argues that if a researcher is to determine the structure or organisation of an event it is necessary to capture the whole event, or at least complete sequences of activity within an event. She argues that recording only parts of an event makes it difficult to interpret the exhibited behaviours, as their interpretation is influenced by what has come before. Recognising the boundaries of what constitutes a 'complete event' can be difficult. However, in the experimental protocol described below it can be seen that the research participants are taken from a period of normal and benign operation, into a period of crisis. The recording of the crisis continues until a natural concluding outcome is reached, whether good or bad. In this way, a complete event was recorded for each team of research participants.

There is a general acceptance that the behaviours exhibited by groups under observation may be affected by their knowledge of being observed. It is also generally accepted that in order to counter this affect, ethnographic researchers need to become embedded within the community they are observing. In this way, the researcher can build trust with the participants and an understanding of the behaviours that are exhibited. It can also be hoped that by becoming embedded within the community, the researcher's presence will exert less influence on the natural behaviours of the participants. Within the case of the crisis scenario experiment the observer is, by the nature of their previous experience within the industry, already embedded into the community of the participants. The researcher has twenty years experience of the commercial vessel engine room environment as a marine engineer officer and was therefore able to readily become trusted and embedded within the engine room team and to understand the behaviours exhibited by the participants. Anecdotal evidence from the participants suggested that the presence of video cameras within the simulated engine room control room environment did not affect their natural behaviours.

4.3 ETHICAL CONSIDERATIONS

There are ethical concerns when undertaking research involving human participants. The United Kingdom's Economic and Social Research Council (ESRC) has a research ethics framework (ESRC, 2006) that provides a number of principles of ethical research that are relevant to the research undertaken for this thesis. They state that:

- *"Research staff and subjects must be informed fully about the purpose, methods and intended possible uses of the research, what their participation in the research entails and what risks, if any, are involved;*
- *The confidentiality of information supplied by research subjects and the anonymity of respondents must be respected;*
- *Research participants must participate in a voluntary way, free from any coercion;*
- *Harm to research participants must be avoided." (ESRC, 2006)*

These principles are reflected in Southampton Solent University's own Ethics Policy (Section 2S of the Academic Handbook) and have been complied with throughout the experimental research undertaken for this thesis. The following sections detail how these ethical principles have been complied with.

4.3.1 Informed Consent

All of the volunteers that took part in this research were asked to read and sign a consent form, a copy of which is presented in Appendix 1. This form detailed the purpose of the research study, why they, in particular, had been asked to volunteer to take part in the research study, and the procedures that they would be asked to participate in as part of the research. It was judged by the researcher that this was sufficient information for the volunteers to make an informed decision as to whether or not they wished to participate in the research. The opportunity was given to all volunteers to ask questions of the researcher if they required further clarification of the information provided on the consent form.

4.3.2 Confidentiality

The consent form (Appendix 1) stated that volunteers would not be personally identified in any reports or publications that may result from the research study. It also stated that the use of the video recordings taken during the study would be strictly limited to the researcher named on the form.

4.3.3 Right to Refuse to Participate or Withdraw

It was clearly stated on the consent form (see Appendix 1) that the volunteer could refuse to participate or withdraw from the study at anytime.

4.3.4 Risk Assessment

The experiments undertaken to obtain data for this thesis can be defined as social research, and as such have the potential to involve psychological risk. It is therefore necessary for the level of risk to any participant to be determined, and if appropriate, for risk mitigation measures to be put in place.

The level of risk to any participant within a simulated work environment that models closely the work environment onboard a commercial vessel is considered to be considerably less than the actual working environment onboard a commercial vessel. This is because, unlike the actual work environment, if errors in operation are made within the simulated work environment, the participant cannot be physically harmed. Because of this, it is likely that the level of psychological risk to participants within the simulated work environment is also less than it could be within the real work environment.

As all of the research study participants were experienced serving merchant navy engineering officers, undertaking continuing professional development training, they were obviously able to cope with the level of psychological risk in their real working environment. It was therefore concluded that they would also be able to cope with the reduced level of psychological risk within the simulated work environment, and that no additional risk mitigation measures were required.

4.3.5 Complaints Procedure

The consent form clearly stated the volunteer could report, anonymously if they chose, any complaints or comments regarding the manner in which the study was being conducted to the Quality Manager of Warsash Maritime Academy. The address of the Quality Manager was also given on the consent form.

4.4 EXPERIMENT PARTICIPANTS

In order to be able to assess competence in crisis management it would be necessary for any research participant to at least be competent in normal operations of engine room watchkeeping onboard a merchant vessel motorship. It was also thought to be beneficial that, if possible, within each group of research participants, there should be a senior officer who would take the role of team leader during the research exercise scenarios. This was to ensure that there was a level of management experience within each of the participating teams.

The volunteers for the research study were drawn from the oil, chemical and gas tanker safety programmes run as professional development short courses at Warsash Maritime Academy. These courses run for five days and are attended by serving merchant navy officers of all ranks and experience. About half of the delegates for each course were serving marine engineering officers, all of whom had current valid certificates of competency issued by their national maritime administration or by the flag administration of the vessels they were serving on. At the start of each course, the researcher was permitted to talk to the course delegates about the research and to ask for volunteers to participate in the research experiment exercise.

Three volunteers were requested from each course and informed that they would have to undertake two sessions within the engine room simulator. The first session of one and a half hours duration being a simulator familiarisation session, and the second session of one and a half hours duration, in which they would take part in the research exercise scenario. The first session was carried out on the Tuesday evening of their course, from 1730 hrs until 1900 hrs. The second session was carried out on the Wednesday evening of their course, also from 1730 hrs until 1900 hrs. The days and times of the sessions were kept the same for all teams of volunteers.

Familiarisation with the simulator is necessary before the research exercise is undertaken in order that aspects of the 'simulation' do not hinder the participants' performance when the research exercise is undertaken. This is generally accepted as usual practice within the simulator community. It was important that the research participants understood how to operate the simulated vessel and understood the layout of the engine room plant. The familiarisation session followed the same schedule for each group of research participants so that no team had any advantage or disadvantage compared to any other. The familiarisation schedule for every team of volunteers was as detailed in Table 16 below.

Period	Activity
1730 hrs – 1750 hrs	Simulator and engine room plant orientation led by researcher with simulator in halted condition.
1750 hrs – 1810 hrs	Simulator and engine room plant orientation led by researcher with simulator in running condition. Demonstration of paralleling of alternators in manual, semi-automatic and automatic modes of operation. Demonstration of engine room control and bridge control functions for main engine. Demonstration of the use of the distributed control system.
1810 hrs – 1900 hrs	Self directed familiarisation of simulated engine room plant, with assistance from researcher if required.

Table 16. Research Participant Simulated Engine Room Environment Familiarisation Schedule.

It was made clear to the research participants that during the research exercise scenario, the researcher would take on the role of Extra Second Engineer, another member of the team who was available only via the use of the walkie-talkie radios supplied to each team member. There was also a mechanic available who would work as directed by the Extra Second Engineer. These two extra team members are necessary within the simulated engine room environment to undertake tasks that are not available to the research participants through the simulation, but that they may require to be undertaken. These tasks are such things as opening up a piece of machinery for inspection or visually checking the position of a valve. The research participants were told that the Extra Second and the Mechanic would always undertake any task given to them both efficiently and effectively and that they would clearly feedback the outcome of any task they undertook to the team leader. However, the research participants were also advised that the Extra Second Engineer and the Mechanic would not use their own initiative to undertake tasks, they would only work as directed.

In order to ensure that a range of behaviours and performance was observed during the research exercise scenarios it was decided to run twelve identical research exercises, with a different team of three serving marine engineering officers for each exercise. Thirty-six research participants were therefore required for the research study. The next section will provide details of the research exercise scenario.

4.5 EXPERIMENTAL PROTOCOL (PHASE ONE)

In order to allow for the observation of overt behavioural markers exhibited by research volunteers during a simulated crisis scenario, a repeatable crisis scenario had to be designed that was outside the normal sphere of experience of most marine engineering officers. Repeatability was important so that each team of volunteers could face the same crisis scenario. By ensuring the scenario was outside of the normal sphere of experience of the research experiment participants, the scenario would not be dealt with as an emergency situation, with action being taken to regain control in accordance with a set documented emergency procedure. The scenario was designed to be atypical, and therefore required those responsible for handling the crisis to think through the situation, and respond in creative and flexible ways.

A crisis scenario was designed based upon the incident involving the motor tanker ‘Braer’ on 5 January 1993 (MAIB, 1993). This vessel sailed from Mongstad, Norway on 3 January 1993, bound for Quebec. As soon as the vessel cleared Mongstad, she experienced severe southerly gales. On the following morning, a number of steel pipes that had been secured on the after deck of the vessel, broke free from of their lashings and started to roll about on the deck, between the port engine casing and the port ship’s side railings. The pipes eventually caused damage to the air vent pipes of the diesel oil bunker storage tanks. This in turn allowed seawater that was washing over the deck to enter the air vent pipes and run down into the diesel oil storage tanks. This in turn led to severe contamination of the diesel oil fuel supply systems to the vessel's boiler, main engine and diesel alternators. At 0440 hrs on 5 January 1993, due to the severe contamination of the diesel oil with seawater, the Braer's main engine stopped, shortly followed by the failure of the vessel's diesel alternator engines.

At this point the vessel was ten miles off the southern tip of Sheltland, and was drifting north towards Sumburgh Head. Three and a half hours after stopping the vessel had drifted within half a mile of land and was thought to be in imminent danger of grounding. The decision was made at this point to abandon ship, and the complete crew was safely evacuated. Despite attempts by volunteers to set up a tow with an attending tug, the vessel eventually grounded on Garths Ness at 1119 hrs on 5 January 1993. The vessel was a total constructive loss, spilling 84,700 tonnes of crude oil into the sea.

The aspect of this incident that was modelled was the seawater contamination of the diesel oil storage tanks. If introduced as a fault on the engine room simulator whilst a diesel alternator is running and supplying power to the vessel, the electrical power network voltage and frequency values start to fluctuate. This fluctuation increases over a period of fifteen minutes until a diesel alternator engine exhaust gas temperature deviation alarm sounds. If no action is taken to correct the situation, after a further five minutes the diesel alternator engine will stop, causing a loss of electrical power capacity on the electrical distribution network. Depending upon the decisions made by the Chief Engineer, and the actions taken by the engine room team, this could lead to the main engine stopping and the vessel being left drifting without power.

An incident involving serious contamination of the fuel supply system with seawater is not a common occurrence, and therefore is not the type of incident that would be within the sphere of experience of most seafarers. It is also not the type of incident for which there is likely to be a documented emergency response procedure. In this respect, this type of incident can be categorised as a crisis situation.

The crisis scenario was developed to allow two recovery actions to be available that would ensure that the crisis could be effectively managed. These actions were:

- The use of the residual steam pressure to immediately reinstate the electrical power supply to the vessel by use of the steam driven turbo alternator, and to in turn fire the boiler to maintain the necessary steam pressure. This would then allow the main engine to be restarted;

- To use the emergency alternator to run all of the required main engine auxiliary machinery supplied from the emergency switchboard, and to close the switchboard interconnecting tiebreaker in order to run the main engine auxiliary scavenge air blower from the main switchboard. This would then in turn allow the main engine to be restarted.

4.5.1 Initial Plant Condition

The plant condition at the start of the simulated crisis scenario was as follows:

- Main Engine running at 110 rpm on heavy fuel oil and in Bridge Control, Full Away on Passage;

- All main engine No.2 pumps running with all main engine No.1 pumps set to automatic standby;

- Boiler online firing on heavy fuel oil. Combustion control in automatic. Steam pressure 16 bar;

- No.1 Diesel Alternator running in parallel with the Turbo Alternator. Total electrical load 650kW;

- Emergency Alternator and emergency lighting set to automatic standby condition;

- Interconnecting Tiebreaker between Main and Emergency Switchboards closed;

- Interconnecting Tiebreaker Main Consumer Overload Trip set to 100kW.

The research participants were made aware of these conditions in a briefing before the start of the exercise.

4.5.2 Initial Situation

The vessel situation at the start of the simulated crisis scenario was as follows:

- The vessel is a single hulled oil tanker of 90,000 tonnes deadweight built in Korea in 1986;
- The vessel is carrying a cargo of 84,700 tonnes of light crude oil;
- The vessel has been in heavy weather for the past 24 hours;
- The Master has advised the Chief Engineer that they will soon be entering a traffic separation system and going to Standby condition.

The research participants were made aware of the situation outlined above in a briefing before the start of the exercise.

4.5.3 Personnel Roles

The roles of the personnel involved in the simulated crisis scenario were as follows:

- One research volunteer was designated as Chief Engineer;
- One research volunteer was designated as Second Engineer;
- One research volunteer was designated as Third Engineer;
- Simulator Instructor was designated as Extra Second Engineer (only available to the engine room team via radio communication);
- A mechanic was available, but working through the Extra Second Engineer.

The most senior of the volunteers taking part in each research exercise was chosen to be the Chief Engineer. All of the research participants were asked if they were happy to take on their designated role before the start of the exercise.

4.5.4 Crisis Scenario Interventions

As discussed previously in section 2.12 of this thesis it can be argued that stress is a precursor to a crisis, and that stressors are the building blocks of a crisis situation. In order to ensure that a crisis situation was presented to the research participants a number of interventions were made throughout the exercise scenario that constituted stressors to the team leader. These included:

- Multiple information sources – Master, Extra Second Engineer and other team members;
- Rapidly changing or evolving scenario – problems with electrical power generation systems and vessel in close navigable situation;
- Requirement for team coordination – problems with electrical power generation systems;
- Performance pressure – need to restore electrical power as soon as possible;
- Time pressure - vessel drifting into towards shoals;
- High workload or information overload – diagnosis of problems experienced with electrical power generation systems;
- Threat – vessel without power in a close navigable situation and with the possibility of running aground.

By controlling the timing of the application of these stressors, it was possible to control the cumulative level of stressors presented to the team leader during a simulator exercise. It was therefore possible to ensure that each team leader was subjected to the same stressors at the same times during each experiment.

The interventions made during the simulated crisis scenario were as follows:

Start +5 minutes

Master informs Chief Engineer that vessel is now entering the traffic separation system, but traffic density is expected to be light, so he does not anticipate the need to reduce the main engine revolutions to manoeuvring speed. Master rings Standby condition.

Start +10 minutes

Master informs Chief Engineer that traffic density is increasing so will go into Engine Room Control and Chief Engineer can start to reduce main engine revs to maximum manoeuvring speed, 85 rpm.

Start +15 minutes

Master informs Chief Engineer that due to the close approach of a cross channel ferry he will have to slow the vessel down to half ahead. Master informs Chief Engineer that the vessel is now in very close navigable waters. The Master asks the Chief Engineer if all is okay down below.

Start +20 minutes

The fault of 'water contamination of the diesel oil service and settling tanks' is introduced.

Start +40 minutes

If vessel is without power, Master informs Chief Engineer that vessel is drifting towards shoals and may run aground.

4.5.5 Recording of New Observed Behavioural Markers

The twelve research exercises were run and a video recording of each exercise was taken using a fixed camera position and fixed microphone positions. The camera field of view and sound recording levels of each exercise were maintained the same. Each recording was analysed to determine if any behavioural markers relevant to the management of crises were observable that had not been found during the literature review.

4.5.6 Rating of Observed Behavioural Markers

For a behavioural marker to be an effective assessment metric it needs to be relevant to the competence being assessed, easily evaluated as a demonstration of good or poor behaviour, easily observed, and should occur reasonably frequently. The behavioural markers observed during the twelve experiments were therefore rated using metrics of ease of observation, ease of evaluation, frequency of occurrence and relevance to competence in crisis management, to determine the most efficacious sub-set of markers to be used within any experimental assessment framework. A five point Likert scale was used to rate each behavioural marker as follows:

A) Ease of Observation	Easy	1	2	3	4	5	Difficult
B) Ease of Evaluation	Easy	1	2	3	4	5	Difficult
C) Frequency of Occurrence	Often	1	2	3	4	5	Rarely
D) Relevance to Competence in Crisis Management	Relevant	1	2	3	4	5	Irrelevant

The relevance of a particular behavioural marker to competence in crisis management was evaluated by the observed affect that behaviour had on the management of the crisis.

An overall index value was assigned to each behavioural marker. This was determined by multiplying together the four Likert scale ratings given to each behavioural marker. The rating for frequency of occurrence was weighted at two times the other ratings. This is because if a behavioural marker does not occur very often, it will not be a useful metric. This is because however relevant it is, and however easy to evaluate and observe, if it rarely occurs, it will rarely contribute to the assessment of competence.

$$\text{Behavioural Marker Index} = A \times B \times 2C \times D$$

This behavioural marker index was therefore an indication of the suitability of a particular behavioural marker for inclusion within an assessment framework for assessing competence in crisis management. The lower the index value the more suitable a behavioural marker was considered for inclusion in the assessment framework.

All of the behavioural markers observed within the twelve experiments run were rated using the Behavioural Marker Index formula.

4.5.7 Determination of Experimental Assessment Framework

In order to be usable by an assessor for an assessment of competence in crisis management in real time, any assessment framework has to be made up of a limited number of behavioural markers. If there are too many behavioural markers within the assessment framework, it will become difficult for the assessor to use. For this reason it was decided to limit the experimental assessment framework to include only those behavioural markers that had an index rating of eight or less.

Once these behavioural markers had been determined, examples of good and poor behaviour associated with these markers were codified by observation of the research exercise recordings. These example behaviours would be included within the guidance for assessors using the experimental assessment framework.

4.5.8 Expert Evaluation of Experimental Assessment Framework

An experiment was undertaken where two expert crisis management assessors were asked to rate the behavioural markers within the experimental assessment framework for ease of observation, ease of evaluation, frequency of occurrence and relevance to competence in crisis management. In order to undertake this experiment, the expert assessors were given a copy of one of the recorded crisis scenario experiments run. The results of this experiment were compared with the ratings given by the researcher for all of the recorded experiments. This was done to show the degree of agreement between the expert crisis management assessors and the researcher on the effectiveness of the behavioural markers included within the previously determined experimental assessment framework.

If there was agreement between the expert assessors and the researcher on the most effective behavioural markers to use for the assessment of crisis management, the experimental assessment framework could be taken through to phase two of the research methodology unaltered. However, if there was any major disagreement, the experimental assessment framework may have to be amended to reflect the differences.

The expert crisis management assessors who took part in this experiment were two of the lead assessors from the London Fire Brigade Incident Commander Assessment Centre.

Chapter 5. RESULTS AND ANALYSIS (PHASE ONE)

5.1 BEHAVIOURAL MARKERS ARISING OUT OF EXPERIMENTS

Following the running of the twelve crisis scenario experiments; all of the video recordings were studied to determine if any overt behavioural markers, relevant to competence in crisis management, could be observed that had not been highlighted during the literature review. Ten new behavioural markers were found that had the potential to be useful in the assessment of competence in crisis management. These are detailed in Table 17 below.

Overt Behavioural Marker	**Comments**
Sharing of workload	How well the crisis manager shares the tasks to be undertaken between the team members. Number of tasks directly undertaken by the crisis manager, possibly leading to a loss of situational overview.
Unfinished sentences	Crisis manager not completing sentences when communicating to other team members.
Patterns of movement	Uses measured movements within a defined area where crisis manager is able to maintain overview of the situation.
Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	
Team leader using different types of decision making strategies	➢ Analytical ➢ Intuitive (Recognition Primed Decision-making) ➢ Critical Thinking
Thinking aloud in order to gain reassurance from other members of the team	

Not returning to task in hand after a distraction	
Team Leader 'grasping' at suggestions made by other team members	Crisis manager does not reflect on the suggestions of others before acting on those suggestions.
Measured movements in response to stimuli	Crisis manager makes measured, reasonable, considered and controlled movements in response to stimuli.
Making a series of truncated movements	Crisis manager moves between task locations in a truncated or hesitant manner. Moves between two task locations without, in the short term, reaching either. Presents an on the spot rocking motion.

Table 17. Behavioural markers arising out of experiments.

5.2 EVALUATION OF BEHAVIOURAL MARKERS

With the behavioural markers identified from the literature review, and the additional behavioural markers identified during the crisis scenario experiments, a total of seventy-seven behavioural markers were evaluated for ease of observation, ease of evaluation, frequency of occurrence and relevance to competence in crisis management. A behavioural marker index was also calculated for each behaviour observed using the method detailed in section 4.5.6.

The evaluation was carried out through a detailed study of the video data collected during the experiments. The results of this evaluation are presented in Table 18 below.

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Ratio of degree of feedback control to degree of predictive control.	Situational Awareness	5	2	3	1	60
Integration of team member attributes and values to maximum advantage.	Leadership Workload Management	5	5	2	1	100
The number of alternative hypotheses and actions communicated to team members.	Sharing Mental Models Option Generation Team Building	2	1	1	1	4
The level of 'satisficing' exhibited	Option Generation Decision–making	3	3	1	1	18
Communicating in a way that reveals ones mental models	Sharing Mental Models Team Building	2	2	1	1	8
Relevance and timeliness of unsolicited information passed between team members as measures of the degree of congruence of their mental models	Sharing Mental Models Teamwork Team Building Situational Awareness	3	2	1	1	12
The level of anticipation of the needs of other team members and of future action and task requirements	Situational Awareness Teamwork	4	4	2	1	64
The amount of relevant information passed in anticipation of need, as a ratio of the amount of relevant information available	Situational Awareness Teamwork	5	5	3	1	150
The periodicity of situation assessment updates to the other team members	Sharing mental Models Situational Awareness Management	2	2	2	1	16

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Ignoring feedback delays	Situational Awareness	3	4	3	2	144
Focus too much on the reduction of uncertainty by requesting information	Decision-making	2	5	3	3	180
Inaccuracy in timing the right moment of intervention	Decision-making Situational Awareness	1	2	2	2	16
Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	Situational Awareness	1	2	2	1	8
Lack of knowledge concerning the relation between symptoms and causes	Situational Awareness	4	5	2	1	80
The number of dimensions upon which a decision is based	Decision-making	4	4	1	2	64
The length of term of considerations upon which a decision is based	Decision-making	5	5	1	4	200
Amount of sampling behaviour exhibited	Situational Awareness	2	2	1	1	8
Time between the onset of a disturbance and the first time a related action was carried out or an information request was made; defined as the disturbance detection time	Situational Awareness	3	3	1	1	18
The time span between the first information request or action related to a disturbance and the correction of that disturbance	Decision-making	3	3	1	1	18

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x2 C x D
The number of information requests made about a disturbance	Situational Awareness Decision-making	2	3	2	2	48
The number of times a supervisor inspected systems through their graphical control interface, as indicated by the number of mouse clicks on the control screen giving a measurement of monitoring behaviour	Situational Awareness	2	2	1	4	32
The division of time and effort over different essential sub-goals	Workload Management Co-ordination	5	4	1	1	40
The ability to switch between the processing of information about different sub-goals, as opposed to the need to completely finish one sub-goal before considering another	Workload Management	5	5	1	1	50
Explicitly asking their team how much time they have before they must commit to a decision	Co-operation Situational Awareness Decision-making	1	1	5	1	10
Team leader focuses on individual tasks rather than the teams'	Leadership Workload Management	1	2	2	1	8
The level of useful communication amongst the team	Co-operation Teamwork Sharing Mental Models	3	4	1	1	24
The degree to which a crisis manager monitors the information exchange between members of their crisis management team	Situational Awareness	5	5	1	5	250

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Seeking information from all available sources	Situational Awareness. Decision-making	2	3	2	1	24
Passing information to the appropriate persons before being asked	Situational Awareness Co-operation	3	3	3	1	54
Providing big picture updates	Sharing mental models Situational Awareness	1	1	3	1	6
Using proper phraseology	Communication	1	3	1	4	24
Providing complete internal and external reports	Sharing mental model	2	3	4	1	48
Avoiding excess chatter	Communication	1	2	5	3	60
Ensuring communications are audible and ungarbled	Communication	1	2	2	1	8
Correcting team errors	Leadership Co-operation	1	3	5	3	90
Providing and requesting backup or assistance when needed	Co-operation Situational Awareness	1	2	5	1	20
Providing guidance or suggestions to team members	Co-operation Teamwork	1	2	2	2	16
Stating clear team and individual priorities	Leadership Communication	1	1	3	1	6
Delegation of tasks based upon known crewmember strengths	Workload Management	2	4	4	2	128
High degree of crewmember integration within the team	Teamwork Co-operation	1	1	1	2	4

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Information backup provided during periods of ambiguity	Communication Situational Awareness	2	3	3	2	72
Regular soliciting of information between crewmembers	Communication Situational Awareness	1	1	2	1	4
Integration of real-world operations experience, rather than a reliance on procedures	Decision-making Situational Awareness	3	3	2	2	72
Extensive "what-if" analyses undertaken	Risk Assessment Decision-making	1	1	3	2	12
Task prioritisation carried out	Decision-making	3	1	3	1	18
Adaptation of plan to take into account evolving events	Situational Awareness	3	3	2	1	36
Distracting information coded and assigned level of urgency for attention	Situational Awareness Decision-making	1	1	2	1	4
Maintain a focused attitude to task objectives	Leadership	2	2	3	2	48
Contingency plans developed on an ongoing basis	Situational Awareness	3	3	3	1	54
Overt awareness of time as a finite resource by questioning time management status	Situational Awareness	2	2	4	1	32
Demonstrates awareness of 'big picture' context	Situational Awareness	1	1	2	1	4
Closely co-ordinates actions with others	Co-operation	2	2	2	2	32
Gives consideration to the needs of others	Co-operation Teamwork	2	3	2	3	72

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Utilise team briefings for specific threats	Sharing mental model	1	2	2	1	8
Provide a large range of response options to threats	Decision -making	2	3	4	1	48
Crewmembers performance thoroughly questioned and crosschecked	Teamwork	3	3	3	2	108
Prompts for timely response to correct any crewmember-induced errors	Teamwork Co-operation	3	3	3	2	108
Prompts for timely response to any change in situation	Teamwork Co-operation	3	3	3	2	108
Keeps crewmembers focused on the accomplishment of the task	Leadership	1	2	2	1	8
Communication kept succinct	Communication	1	1	2	2	8
Effectively increases work rate to cope with escalation of crisis	Workload Management	1	2	4	1	16
Sharing of workload	Workload Management	1	2	1	1	4
Unfinished sentences	Communication	1	1	1	3	6
Patterns of movement	Leadership	1	1	1	3	6
Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Situational Awareness	1	1	3	1	6

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Team leader using different types of decision-making strategies	Decision-making	3	4	2	3	144
Thinking aloud in order to gain reassurance from other members of the team	Co-operation Teamwork	1	2	5	4	80
Not returning to task in hand after a distraction	Situational Awareness	1	2	3	3	36
Team leader 'grasping' at suggestions made by other team members	Situational Awareness Decision-making	1	1	2	2	8
Measured movements in response to stimuli	Leadership	1	1	2	2	8
Making a series of truncated movements	Leadership	1	1	2	2	8
Thanking another team member for pointing out a mistake	Co-operation Teamwork	1	1	5	4	40
Helping other team members having difficulty	Co-operation Teamwork	1	1	3	4	24
Making motivating statements	Leadership	1	1	4	3	24
Praising other team members	Leadership Teamwork	1	1	4	3	24
Suggesting ways to find an error	Co-operation Teamwork	2	3	3	3	108

Table 18. Evaluation of Behavioural Markers.

By using the Behavioural Marker Index, it is possible to start to differentiate between those markers that might be effective criteria assessing competence in crisis management, and those that would not. By applying the cut-off value for the index described in section 4.5.7, a sub-set of markers was produced. This sub-set is made up of those markers that have the greatest potential to be effective assessment criteria. The number of markers within this sub-set is twenty-two and this is a more manageable number of criteria for an assessment framework that is to be used for a real time assessment. This assertion is supported by the fact that the number of assessment criteria used within the assessment frameworks of other safety critical domains reviewed in Chapter 3 were as shown in Table 19 below.

Safety Critical Domain	Number of Assessment Criteria
Military	28
Civil Aviation	43
Offshore Industry	24
Fire Brigade	19
Anaesthesiology	21
Commercial Shipping	27
Average Number of Assessment Criteria	27

Table 19. Number of assessment criteria in different assessment frameworks.

Those behavioural markers with an index of eight or lower that were to be used in the experimental assessment framework are shown below in Table 20.

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
High degree of crewmember integration within the team	Teamwork Co-operation	1	1	1	2	4
Regular soliciting of information between crewmembers	Communication Situational Awareness	1	1	2	1	4
Distracting information coded and assigned level of urgency for attention	Situational Awareness Decision-making	1	1	2	1	4
Demonstrates awareness of 'big picture' context	Situational Awareness	1	1	2	1	4
Sharing of workload.	Workload Management	1	2	1	1	4
The number of alternative hypotheses and actions communicated to team members.	Sharing Mental Models Option Generation Team Building	2	1	1	1	4
Unfinished sentences	Communication	1	1	1	3	6
Patterns of movement	Leadership	1	1	1	3	6
Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Situational Awareness	1	1	3	1	6
Providing big picture updates	Sharing mental models Situational Awareness	1	1	3	1	6
Stating clear team and individual priorities	Leadership Communication	1	1	3	1	6

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
Utilise team briefings for specific threats	Sharing mental model	1	2	2	1	8
Keeps crewmembers focused on the accomplishment of the task	Leadership	1	2	2	1	8
Ensuring communications are audible and ungarbled	Communication	1	2	2	1	8
Communication kept succinct	Communication	1	1	2	2	8
Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	Situational Awareness	1	2	2	1	8
Communicating in a way that reveals ones mental models	Sharing Mental Models Team Building	2	2	1	1	8
Team leader 'grasping' at suggestions made by other team members	Situational Awareness Decision-making	1	1	2	2	8
Measured movements in response to stimuli	Leadership	1	1	2	2	8
Making a series of truncated movements	Leadership	1	1	2	2	8
Amount of sampling behaviour exhibited	Situational Awareness	2	2	1	1	8
Team leader focuses on individual tasks rather than the teams'	Leadership Workload Management	1	2	2	1	8

Table 20. Behavioural markers with the lowest indices.

5.3 CHARACTERISATION OF BEHAVIOURAL MARKERS

Before sending this experimental assessment framework to the expert crisis management assessors for a further evaluation, it was necessary to characterise these behavioural markers by giving examples of good and poor behaviours. This was done in order to clarify the assessment rating scale for the "ease of evaluation" criteria. The characterisations for the behavioural markers in the experimental assessment framework are detailed in Table 21 below.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
1.	High degree of crewmember integration within the team	Involves other crewmembers in planning, diagnostic and decision making processes. Accepts feedback and advice from crewmembers.	Isolates a crewmember or crewmembers. Only utilises crewmember or crewmembers by giving them a command order. Ignores feedback or advice from crewmembers. Specific case may be cultural isolation.
2.	Regular soliciting of information between crewmembers	Regularly asks other crewmembers for information.	Does not ask other crewmembers for information.
3.	Distracting information coded and assigned level of urgency for attention	Gives a low priority to information that is of a secondary nature with respect to the primary task.	Gives a high priority to information that is of a secondary nature with respect to the primary task.
4.	Demonstrates awareness of 'big picture' context	Verbalises task priorities in anticipation of future events. i.e. "We need to do *xxxxx* now or in *yy* minutes *zzzzz* will happen."	No anticipation of future events evident. Acts surprised as events happen.
5.	Sharing of workload	Delegates individual tasks in order to maintain situation overview.	Gets too involved in individual tasks to the extent that situation overview is lost.
6.	The number of alternative hypotheses and actions communicated to team members.	Proposes alternative hypotheses and actions to team members.	Does not propose any alternative hypotheses or actions to team members.

Ref:	**Overt Behavioural Marker**	**Good Behaviours**	**Poor Behaviours**
7.	Unfinished sentences	Communicates to team with complete and coherent sentences.	Uses unfinished sentences in communications with team.
8.	Patterns of movement	Uses measured movements within a defined area where team leader is able to maintain overview of situation.	Uses very fast movements. No defined area of movement, so overview of instrumentation is not possible. Focuses too much on one item of instrumentation. Leaves control room for extended periods.
9.	Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Asks team members questions about the situation in order to improve situational awareness.	Does not ask team members questions about the situation.
10.	Providing big picture updates	Provides team members with updates on the overview of the situation.	Does not provide team members with any updates on the overview of the situation.
11.	Stating clear team and individual priorities	States clear team and individual team member task priorities.	Does not state team or individual team member task priorities.
12.	Utilise team briefings for specific threats	Briefs team about specific threats.	Does not brief team about specific threats.
13.	Keeps crewmembers focused on the accomplishment of the task	Provides encouragement and motivates team to keep them focussed on the accomplishment of the task. Tasks completed in good time.	Fails to monitor task progression. Team does not complete tasks in good time.

Ref:	Overt Behavioural Marker	Good Behaviour	Poor Behaviour
14.	Ensuring communications are audible and not garbled	Communications clear and easy to understand.	Communications inaudible and garbled. Communications not understood by team members.
15.	Communication kept succinct	Communications kept succinct.	Communications too long and over complicated.
16.	Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	Maintains overview of all systems, thereby maintaining awareness of system interactions.	Focuses on one system to the exclusion of all others. Is surprised by system interactions.
17.	Communicating in a way that reveals ones mental models	Communicates thoughts on the situation and how it is developing to team members.	Does not communicate thoughts to team members.
18.	Team leader 'grasping' at suggestions made by other team members	Reflects on suggestions from team members before accepting or rejecting them.	Acts immediately on suggestions from team members without any prior reflection.
19.	Measured movements in response to stimuli	Makes measured and controlled movements in response to stimuli.	Makes rapid and uncontrolled movements in response to stimuli.
20.	Making a series of truncated movements	Moves without hesitation. Moves smoothly between tasks.	Movement truncated. Movements hesitant. Moves between two task locations without, in the short term, reaching either. Presents an on the spot rocking motion.
21.	Amount of sampling behaviour exhibited	Undertakes a lot of sampling of information in order to maintain a high level of situational awareness.	Does not undertake much sampling of information.

Ref:	Overt Behavioural Marker	Good Behaviour	Poor Behaviour
22.	Team leader focuses on own individual tasks rather than the teams'	Maintains focus on overall team task objective.	Focuses too much on own individual tasks that are not directly related to the overall team task objective. i.e. team leader spends a lot of time trying to put another alternator onto the main switchboard.

Table 21. Characterisation of Behavioural Markers

5.4 EVALUATION BY EXPERT CRISIS MANAGEMENT ASSESSORS

The experimental behavioural marker assessment framework, including the characterisation of these behavioural markers, detailed in Table 21, was given to two expert crisis management assessors for evaluation. These assessors evaluated the behavioural markers in the assessment framework using one of the crisis scenario experiment recordings. It was appreciated that by only giving the expert assessors one of the twelve recording with which to evaluate the behavioural markers, their assessment would be limited in relation to the frequency of occurrence of any particular marker. However, it was also appreciated that those behavioural markers that were frequently observed when viewing just one crisis scenario recording, would probably be frequently observable across all of the recordings. As each crisis scenario recording was over one hour in length, it was deemed too great an imposition on the expert assessors to ask them to evaluate more than one complete recording for this part of the research study. The ratings they gave for ease of observation, ease of evaluation, frequency of occurrence and relevance to competence in crisis management are shown in Table 22 below.

Overt Behavioural Marker	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant
1. High degree of crewmember integration within the team	1	1	5	1
2. Regular soliciting of information between crewmembers	1	1	4	1
3.Distracting information coded and assigned level of urgency for attention	3	3	5	3
4. Demonstrates awareness of 'big picture' context	1	1	5	1
5. Sharing of workload.	1	1	4	2
6. The number of alternative hypotheses and actions communicated to team members.	1	1	4	1
7. Unfinished sentences	1	3	5	3
8. Patterns of movement	1	1	1	2
9. Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	1	1	2	1
10. Providing big picture updates	1	1	3	1
11. Stating clear team and individual priorities	1	1	5	1

Overt Behavioural Marker	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant
12. Utilise team briefings for specific threats	1	1	4	1
13. Keeps crewmembers focused on the accomplishment of the task	1	1	5	3
14. Ensuring communications are audible and not garbled	1	1	3	2
15. Communication kept succinct	3	3	1	3
16. Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	1	3	2	2
17. Communicating in a way that reveals ones mental models	1	1	3	2
18. Team leader 'grasping' at suggestions made by other team members	1	1	5	1
19. Measured movements in response to stimuli	1	3	2	1
20. Making a series of truncated movements	1	2	1	2
21. Amount of sampling behaviour exhibited	4	4	3	2
22. Team leader focuses on own individual tasks rather than the teams'	1	1	1	1

Table 22. Evaluation by expert crisis management assessors.

The ratings given by the expert assessors were then used to calculate the behavioural marker indices for each behavioural marker. These indices are shown, in ranked order, with the lowest first, in Table 23 below.

Overt Behavioural Marker	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
22. Team leader focuses on own individual tasks rather than the teams'	1	1	1	1	2
8. Patterns of movement	1	1	1	2	4
9. Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	1	1	2	1	4
10. Providing big picture updates	1	1	3	1	6
2. Regular soliciting of information between crewmembers	1	1	4	1	8
6. The number of alternative hypotheses and actions communicated to team members.	1	1	4	1	8
12. Utilise team briefings for specific threats	1	1	4	1	8
20. Making a series of truncated movements	1	2	1	2	8
1. High degree of crewmember integration within the team	1	1	5	1	10
4. Demonstrates awareness of 'big picture' context	1	1	5	1	10
11. Stating clear team and individual priorities	1	1	5	1	10

Overt Behavioural Marker	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
18. Team leader 'grasping' at suggestions made by other team members	1	1	5	1	10
14. Ensuring communications are audible and not garbled	1	1	3	2	12
17. Communicating in a way that reveals ones mental models	1	1	3	2	12
19. Measured movements in response to stimuli	1	3	2	1	12
5. Sharing of workload	1	1	4	2	16
16. Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	1	3	2	2	24
13. Keeps crewmembers focused on the accomplishment of the task	1	1	5	3	30
15. Communication kept succinct	3	3	1	3	54
7. Unfinished sentences	1	3	5	3	90
21. Amount of sampling behaviour exhibited	4	4	3	2	192
3. Distracting information coded and assigned level of urgency for attention	3	3	5	3	270

Table 23. Behavioural marker indices from expert assessor evaluation.

In order to make a comparison between the behavioural marker indices as rated by the researcher, and the indices as rated by the expert crisis management assessors, all indices are shown in Table 24 below. This table also shows the difference between the researcher's indices and the expert assessors' indices, as well as the average index for each behavioural marker.

Overt Behavioural Marker	Behavioural Marker Index Rated by Researcher	Behavioural Marker Index Rated by Expert Assessors	Difference Between Indices	Average Index
1. High degree of crewmember integration within the team	4	10	6	7
2. Regular soliciting of information between crewmembers	4	8	4	6
3. Distracting information coded and assigned level of urgency for attention	4	270	266	137
4. Demonstrates awareness of 'big picture' context	4	10	6	7
5. Sharing of workload	4	16	12	10
6. The number of alternative hypotheses and actions communicated to team members.	4	8	4	6
7. Unfinished sentences	6	90	84	48
8. Patterns of movement	6	4	2	5
9. Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	6	4	2	5
10. Providing big picture updates	6	6	0	6
11. Stating clear team and individual priorities	6	10	4	8

Overt Behavioural Marker	Behavioural Marker Index Rated by Researcher	Behavioural Marker Index Rated by Expert Assessors	Difference Between Indices	Average Index
12. Utilise team briefings for specific threats	8	8	0	8
13. Keeps crewmembers focused on the accomplishment of the task	8	30	22	19
14. Ensuring communications are audible and ungarbled	8	12	4	10
15. Communication kept succinct	8	54	46	31
16. Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	8	24	16	16
17. Communicating in a way that reveals ones mental models	8	12	4	10
18. Team leader 'grasping' at suggestions made by other team members	8	10	2	9
19. Measured movements in response to stimuli	8	12	4	10
20. Making a series of truncated movements	8	8	0	8
21. Amount of sampling behaviour exhibited	8	192	184	100
22. Team leader focuses on individual tasks rather than the teams'	8	2	6	5

Table 24. Comparison of behavioural marker indices.

It can be seen from the table above that the researcher's indices and the expert assessors' indices of fifteen of the behavioural markers are in very close agreement, being within six index points of each other. This indicates that these behavioural markers are likely to be effective assessment criteria for assessing competence in crisis management. This agreement shows the reliability of these behavioural markers as assessment criteria, but not their validity. For this reason, these fifteen behavioural markers (1, 2, 4, 6, 8, 9, 10, 11, 12, 14, 17, 18, 19, 20 and 22) were passed through to the validation phase of the research.

Due to the larger differences in the indices of the remaining seven behavioural markers (3, 5, 7, 13, 15, 16 and 21), further analysis was required before a decision could be made as to whether to pass these markers through to the validation phase of the research, or not. In order to facilitate this further analysis, Table 25 below sets out all of the evaluation ratings and indices as assessed by the researcher, with, in brackets, all of the evaluation ratings and indices as assessed by the expert assessors.

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
1. High degree of crewmember integration within the team	Teamwork Co-operation	1 (1)	1 (1)	1 (5)	2 (1)	4 (10)
2. Regular soliciting of information between crewmembers	Communication Situational Awareness	1 (1)	1 (1)	2 (4)	1 (1)	4 (8)
3. Distracting information coded and assigned level of urgency for attention	Situational Awareness Decision-making	1 (3)	1 (3)	2 (5)	1 (3)	4 (270)
4. Demonstrates awareness of 'big picture' context	Situational Awareness	1 (1)	1 (1)	2 (5)	1 (1)	4 (10)
5. Sharing of workload.	Workload Management	1 (1)	2 (1)	1 (4)	1 (2)	4 (16)
6. The number of alternative hypotheses and actions communicated to team members.	Sharing Mental Models Option Generation Team Building	2 (1)	1 (1)	1 (4)	1 (1)	4 (8)
7. Unfinished sentences	Communication	1 (1)	1 (3)	1 (5)	3 (3)	6 (90)
8. Patterns of movement	Leadership	1 (1)	1 (1)	1 (1)	3 (2)	6 (4)
9. Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Situational Awareness	1 (1)	1 (1)	3 (2)	1 (1)	6 (4)
10. Providing big picture updates	Sharing mental models Situational Awareness	1 (1)	1 (1)	3 (3)	1 (1)	6 (6)
11. Stating clear team and individual priorities	Leadership Communication	1 (1)	1 (1)	3 (5)	1 (1)	6 (10)

Overt Behavioural Marker	Associated Crisis Management Skill	Ease of Observation A 1 = Easy 5 = Difficult	Ease of Evaluation B 1 = Easy 5 = Difficult	Frequency of Occurrence C 1 = often 5 = rarely	Relevance to competence in crisis management D 1 = relevant 5 = irrelevant	Behavioural Marker Index A x B x 2C x D
12. Utilise team briefings for specific threats	Sharing mental model	1 (1)	2 (1)	2 (4)	1 (1)	8 (8)
13. Keeps crewmembers focused on the accomplishment of the task	Leadership	1 (1)	2 (1)	2 (5)	1 (3)	8 (30)
14. Ensuring communications are audible and ungarbled	Communication	1 (1)	2 (1)	2 (3)	1 (2)	8 (12)
15. Communication kept succinct	Communication	1 (3)	1 (3)	2 (1)	2 (3)	8 (54)
16. Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	Situational Awareness	1 (1)	2 (3)	2 (2)	1 (2)	8 (24)
17. Communicating in a way that reveals ones mental models	Sharing Mental Models Team Building	2 (1)	2 (1)	1 (3)	1 (2)	8 (12)
18. Team leader 'grasping' at suggestions made by other team members	Situational Awareness Decision-making	1 (1)	1 (1)	2 (5)	2 (1)	8 (10)
19. Measured movements in response to stimuli	Leadership	1 (1)	1 (3)	2 (2)	2 (1)	8 (12)
20. Making a series of truncated movements	Leadership	1 (1)	1 (2)	2 (1)	2 (2)	8 (8)
21. Amount of sampling behaviour exhibited	Situational Awareness	2 (4)	2 (4)	1 (3)	1 (2)	8 (192)
22. Team leader focuses on individual tasks rather than the teams'	Leadership Workload Management	1 (1)	2 (1)	2 (1)	1 (1)	8 (2)

Table 25. Comparison of all evaluation ratings and indices.

The following section will further evaluate those behavioural markers where the behavioural marker index rated by the researcher differed from that rated by the expert assessors by more than 6 index points.

Behavioural Marker No. 3 - Distracting information coded and assigned level of urgency for attention

The ratings for this behavioural marker show significant differences, of more than one index point, for all four assessment criteria. This therefore shows that this marker is unreliable, and for this reason it was rejected from the experimental assessment framework.

Behavioural Marker No.5 - Sharing of Workload

For this behavioural marker, the index for ease of observation was rated as 1 by both the researcher and the expert assessors. The indices for ease of evaluation and relevance to competence in crisis management were only rated differently by one index point. It is the rating given by the expert assessors for 'frequency of occurrence' that introduces the main difference between their index and the researcher's index. As the expert assessors only received one crisis scenario recording to evaluate, for the reasons discussed earlier in this section, their low frequency of occurrence rating for this behavioural marker may not be giving a true reflection of the markers overall frequency of occurrence across all of the experimental recordings. If the term for 'frequency of occurrence' were omitted from the behavioural marker index formula, the revised index for this marker, as evaluated by the expert assessors, would be 2. It was therefore decided that this behavioural marker should be included within the experimental assessment framework.

Behavioural Marker No.7 - Unfinished Sentences

As with behavioural marker No.5, the rating for which there was the greatest disagreement between the researcher and the expert assessors was for the 'frequency of occurrence'. Again, if this term were omitted from the behavioural marker index formula, the revised index for this marker, as evaluated by the expert assessors, would be 9. This marker was therefore also included within the experimental assessment framework.

Behavioural Marker No.13 - Keeps crew members focused on accomplishment of task

As with behavioural markers No.5 and 7, it is the rating for 'frequency of occurrence' that differentiates between the researcher's index and the expert assessors index for this marker. If this term were omitted from the behavioural marker index formula, the revised index for this marker, as evaluated by the expert assessors, would be 3. This marker was therefore also included within the experimental assessment framework.

Behavioural Marker No.15 - Communications kept succinct

For this behavioural marker the researcher's ratings and the expert assessors' ratings significantly disagree, by more than one index point, for both the ease of observation and the ease of evaluation criteria. Even if the term for the 'frequency of occurrence' were omitted, the revised index for this marker, as evaluated by the expert assessors, would still be 27. Therefore, this behavioural marker was rejected from the experimental assessment framework.

Behavioural Marker No.16 - Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system

For this behavioural marker all of the ratings given by the researcher and the expert assessors were within one index point. By omitting the term for 'frequency of occurrence', the revised index for this marker, as evaluated by the expert assessors, would be 6. Therefore, this behavioural marker was retained within the experimental assessment framework.

Behavioural Marker No.21 - Amount of sampling behaviour exhibited

For this behavioural marker, for all but the 'relevance to competence in crisis management' criteria, the researcher's ratings and the expert assessors ratings differ by more than one index point. Therefore, this behavioural marker was rejected from the experimental assessment framework.

Following the above evaluation it was decided that behavioural markers 3, 15 and 21 would be left out of the experimental assessment framework to be validated in phase two of the research. The final set of nineteen behavioural markers that make up the experimental assessment framework to be tested for validity and interrater agreement are shown below in Table 26.

Ref:	**Overt Behavioural Marker**	**Good Behaviours**	**Poor Behaviours**
1.	High degree of crewmember integration within the team	Involves other crewmembers in planning, diagnostic and decision making processes. Accepts feedback and advice from crewmembers.	Isolates a crewmember or crewmembers. Only utilises crewmember or crewmembers by giving them a command order. Ignores feedback or advice from crewmembers. Specific case may be cultural isolation.
2.	Regular soliciting of information between crewmembers	Regularly asks other crewmembers for information.	Does not ask other crewmembers for information.
3.	Demonstrates awareness of 'big picture' context	Verbalises task priorities in anticipation of future events. i.e. "We need to do *xxxxx* now or in *yy* minutes *zzzzz* will happen."	No anticipation of future events evident. Acts surprised as events happen.
4.	Sharing of workload	Delegates individual tasks in order to maintain situation overview.	Gets too involved in individual tasks to the extent that situation overview is lost.
5.	The number of alternative hypotheses and actions communicated to team members.	Proposes alternative hypotheses and actions to team members.	Does not propose any alternative hypotheses or actions to team members.
6.	Unfinished sentences	Communicates to team with complete and coherent sentences.	Uses unfinished sentences in communications with team.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
7.	Patterns of movement	Uses measured movements within a defined area where team leader is able to maintain overview of situation.	Uses very fast movements. No defined area of movement, so overview of instrumentation is not possible. Focuses too much on one item of instrumentation. Leaves control room for extended periods.
8.	Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Asks team members questions about the situation in order to improve situational awareness.	Does not ask team members questions about the situation.
9.	Providing big picture updates	Provides team members with updates on the overview of the situation.	Does not provide team members with any updates on the overview of the situation.
10.	Stating clear team and individual priorities	States clear team and individual team member task priorities.	Does not state team or individual team member task priorities.
11.	Utilise team briefings for specific threats	Briefs team about specific threats.	Does not brief team about specific threats.
12.	Keeps crewmembers focused on the accomplishment of the task	Provides encouragement and motivates team to keep them focussed on the accomplishment of the task. Tasks completed in good time.	Fails to monitor task progression. Team does not complete tasks in good time.
13.	Ensuring communications are audible and not garbled	Communications clear and easy to understand.	Communications inaudible and garbled. Communications not understood by team members.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
14.	Tendency to focus on one system at a time, thereby ignoring the dynamics of the complete system	Maintains overview of all systems, thereby maintaining awareness of system interactions.	Focuses on one system to the exclusion of all others. Is surprised by system interactions.
15.	Communicating in a way that reveals ones mental models	Communicates thoughts on the situation and how it is developing to team members.	Does not communicate thoughts to team members.
16.	Team leader 'grasping' at suggestions made by other team members	Reflects on suggestions from team members before accepting or rejecting them.	Acts immediately on suggestions from team members without any prior reflection.
17.	Measured movements in response to stimuli	Makes measured and controlled movements in response to stimuli.	Makes rapid and uncontrolled movements in response to stimuli.
18.	Making a series of truncated movements	Moves without hesitation. Moves smoothly between tasks.	Movement truncated. Movements hesitant. Moves between two task locations without, in the short term, reaching either. Presents an on the spot rocking motion.
19.	Team leader focuses on own individual tasks rather than the teams'	Maintains focus on overall team task objective.	Focuses too much on own individual tasks that are not directly related to the overall team task objective. i.e. team leader spends a lot of time trying to put another alternator onto the main switchboard.

Table 26. Final Experimental Assessment Framework

Each of the behavioural markers within this experimental assessment framework has been shown, through the analysis of twelve crisis scenarios, to have the best balance of the characteristics for ease of observation and evaluation, frequency of occurrence, and relevance to the assessment of competence in crisis management, of all of the behavioural markers evaluated. Of these nineteen behavioural markers, seven are original markers, solely determined during the ethnographic analysis of the crisis scenario experiment video recordings undertaken within this research. The remainder of the markers were determined from the literature review, and then evaluated for their effectiveness in assessing competence in crisis management, within the context of a merchant vessel engine room control room environment, through the ethnographic analysis. The following chapter will describe the next phase of the research methodology in which the experimental assessment framework, determined by phase one of the research, will be tested for validity and interrater agreement.

Chapter 6. RESEARCH METHODOLOGY (PHASE TWO)

6.1 INTRODUCTION

Phase one of this research resulted in the production of an experimental assessment framework for assessing competence in crisis management for merchant navy engineer officers. Face validity of this assessment framework was sought during phase one of this research by having two expert crisis management assessors use and review the experimental assessment framework in a limited trial. During this trial, the experts were asked to rate the ease of observation, ease of evaluation, frequency of occurrence and relevance to competence in crisis management, of the behavioural markers in the assessment framework. Once the trial had been completed, a comparison was then made between the observations of the expert crisis management assessors and the researcher, from which the experimental assessment framework was derived. The results of the research carried out in phase one gave face validity to the assessment framework. However, in order to prove the validity and reliability of the assessment framework, further research was required.

6.2 RESEARCH HYPOTHESES

The objective of the experimental work and analysis undertaken during phase two of the research was to evaluate the following hypotheses:

1. That the sub-set of behavioural markers that formed the experimental assessment framework determined in phase one of the research can be used as valid markers for the objective assessment of competence in crisis management within the context of a ships engine room control room.

2. That the experimental assessment framework that has been developed can be used with a high degree of interrater agreement.

3. That if a behavioural marker based assessment framework has sufficiently simple and objective assessment criteria, it can be used by non-domain experts to assess competence in crisis management.

4. That if a behavioural marker based assessment framework has sufficiently simple and objective assessment criteria, it can be used successfully, without the need to give specific training to assessors in its use.

The next section will detail the experimental protocol devised in order to undertake the evaluation of these hypotheses.

6.3 EXPERIMENTAL PROTOCOL (PHASE TWO)

Using the video data collected during the experiments carried out in phase one of this research, a video was produced that presented the full range of behaviours used in the assessment framework. This video was comprised of twelve five-minute event sets, one from each of the original experiments. Each event set was temporally anchored, starting from the moment in the crisis scenario when the vessel lost power. This temporal anchoring was important to ensure that, when viewed, each assessor was assessing the behaviours exhibited by the candidates during the same period of the scenario event set. In this way, the number of variables affecting the response of the candidates would be similar in each case. By trying to ensure that the number of variables affecting the performance of the candidate was similar in each case, the assessments should better reflect the individual competence in crisis management of the candidates. The same video was shown to three groups of assessors, with six assessors in two of the groups and seven assessors in the other group.

Currently, the assessment of competence in crisis management is not required for all officers in the merchant navy. Therefore, there is not a pool of maritime crisis management assessors that could be drawn upon to assist within the validation process of the experimental assessment framework. It was therefore deemed important to have two groups of domain expert assessors involved in the validation process, one group of marine engineer officer operations domain experts and one group of crisis management assessment experts.

Group 1

Group 1 consisted of six qualified engineer officers, who had sailed as Chief Engineer Officer, but who were not skilled assessors. The members of this group were therefore marine engineer officer operations domain experts, but not crisis management domain experts. The members of this group were asked, by using only their 'gut' feelings, to rank the Chief Engineers of each of the twelve teams shown in the video event sets from best to worst crisis manager. The mean ranking of this group was determined.

Group 2

Group 2 consisted of six non-domain expert assessors. By this, it is meant that these assessors were skilled at assessing, but that they were not experts in assessing competence in marine engineer officer operations, or crisis management. These assessors were asked to use the assessment framework developed in phase one of this research to assess the overt behavioural markers exhibited by the Chief Engineers of each of the twelve teams shown in the video event sets using a four point rating scale. The rating scale was comprised of the following categories:

- Good
- Towards Good
- Towards Poor
- Poor

The rating scale also included a "not observed" category, to record if a particular behavioural marker from the previously determined assessment framework could not be observed during the event set. The within group interrater agreement of this group of assessors was determined statistically. The mean of the absolute ratings given for each Chief Engineer for each behavioural marker by the assessors in this group was also calculated in order to determine the rank order of the Chief Engineers as measured using the experimental assessment framework.

This group of assessors received no training in how to use the assessment framework. The only documentation they were given, apart from the behavioural tendency response sheets, was a set of examples of good and poor behaviours relating to each of the behavioural markers they were to assess. These behaviours specifically related to competence in crisis management within a merchant navy vessel engine control room environment. This set of example behaviours is shown in Appendix 2.

Group 3

Group 3 consisted of seven experts in assessing competence in crisis management. These assessors were asked, by using only their 'gut' feelings, to rank the Chief Engineers of each team from best to worst crisis manager. The mean ranking of this group was determined.

If the experimental assessment framework is valid, there should be a high degree of correlation between the assessments of groups 1, 2 and 3. Also, the within group interrater agreement of group 2 should be high.

Before sending the example behaviour marker sheet of the final assessment framework that is shown in Table 26, to the group 2 assessors, a review of how the behavioural markers were presented was carried out. As written in Table 26, some of the example behavioural markers were written as being positive behaviours, some were written as being negative behaviours and some were written as neutral behaviours. It was decided that this might be confusing when asking assessors to rate these behaviours using a rating scale. In order to make the assessment framework as clear as possible to the assessors, all of the behavioural markers were re-written so that all gave an indication of a positive behaviour. For example, the neutrally worded behavioural marker "Patterns of movement" became "Uses measured patterns of movement", and the negatively worded behavioural marker "Team leader 'grasping' at suggestions made by other team members" became "Team leader reflects on the suggestions made by other team members". The example behavioural marker sheet that was used during the phase 2 experiments is shown in Appendix 2.

6.4 METRICS

6.4.1 Interrater Reliability and Interrater Agreement

Following the completion of two multitrait/multirater assessment studies, Lawler (1967) concluded that different raters often fail to converge in their assessments because they are observing different things. This implies that the criteria that were being used for these assessments were too open to subjective interpretation, and did not focus the attention of the assessors on the overt behaviours required for successful completion of the task. This may be true of many assessment frameworks where behavioural markers are left too broadly categorised to produce truly objective assessments. Even where the behaviour is well defined, different raters may rely on different perceptions of that behaviour, from different times during the observation period, when they make their final assessment.

LeBreton et al (2003) propose another hypothesis to explain the apparent inconsistency between raters, which they term the Restriction of Variance Hypothesis. This hypothesis posits that the apparent uniqueness of ratings is largely artificial, brought about by the restriction of variance in the observed performance of the candidates. This restricted variance may arise from interventions such as recruitment and selection practices, training and organisational culture, as well as rating biases, such as leniency or central tendency. This restriction of variance in the candidates can lead to a distribution of performance ratings that are negatively skewed, with assessments being mostly good, mostly average, or mostly poor. LeBreton et al (2003) argue that assessing rating similarity using correlation-based analyses, such as intraclass correlation coefficients (ICC's) or Pearson correlations, requires that sufficient between-target variance exists to produce meaningful inferential statistics. They argue that:

> *"without substantial between-target variance, any correlation based index of rating similarity will likely be low." (LeBreton et al, 2003)*

Therefore, it would appear to be unwise to try to determine the level of interrater reliability and interrater agreement solely based on a statistical analysis using a correlation-based index. This is because low correlations could occur either due to substantial rating discrepancies, or due to a lack of between-target variance. In order to ensure that any restriction of variance in the candidates does not result in a misleading statistical analysis, it would be better to undertake such an analysis using a statistical index that is not affected by a lack of between-target variance. LeBreton et al (2003) propose that the r_{wg} index (James, Demaree & Wolf, 1984) is such a statistic.

In order to be clear about what is being measured, a distinction should be made between interrater reliability and interrater agreement. LeBreton and Senter (2007) define interrater reliability (IRR) as:

> *"the relative **CONSISTENCY** of ratings provided by multiple judges of multiple targets."*

The IRR index is used to show if, when rank-ordering targets, judges order with a relative consistency to each other. IRR is not concerned with the absolute equivalence of rating scores given by judges, but with the equivalence of relative rankings.

The interrater agreement (IRA) index is defined by LeBreton and Senter (2007) as:

> *"the **ABSOLUTE CONCENSUS** in scores furnished by multiple judges for one or more targets."*

This means that the interrater agreement index is useful to determine if an assessment framework can produce scores that are interchangeable or equivalent in terms of their absolute values. The higher the level of interrater agreement between multiple judges, the higher can be said to be the discriminatory attributes of the behavioural markers that compose the assessment framework used by the judges.

The IRA index can be used when multiple judges rate a single target on a single variable using an interval scale of measurement, such as a Likert scale. The r_{wg} index defines agreement between the scoring of judges in terms of the proportional reduction in error variance:

$$r_{wg} = 1 - \frac{S_x^2}{\sigma_E^2}$$

where:
S_x^2 is the observed variance on the variable x taken over k different judges or raters;
and
σ_E^2 is the variance expected when there is a complete lack of agreement amongst the judges.
(LeBreton and Senter, 2007)

Brown and Hauenstein (2005) argued that interrater agreement indices are dependent upon both the number of points on the rating scale being used and the number of raters. However, in their study of interrater disagreement and dispersion, Roberson et al (2007) found that there was a high degree of convergence amongst interrater agreement indices. They showed that the average of correlations between the most widely used interrater agreement indices, including r_{wg}, was 0.95, with a range of 0.93 to 0.98. With this high degree of interrater agreement indices convergence, it is considered acceptable to use the r_{wg} index within this research study.

When using the r_{wg} index, if all raters perfectly agree on the rating given to a particular assessment criterion, the observed variance between the raters would be equal to zero, i.e. $S_x^2=0$

Therefore as: $r_{wg} = 1 - S_x^2/ \sigma_E^2$

If $S_x^2 = 0$ $r_{wg} = 1 - 0/ \sigma_E^2$

$r_{wg} = 1$

Conversely, if all raters lack any level of agreement on the rating given to a particular assessment criterion, with their responses being effectively random in nature, the observed variance would equate to that based on the theoretical null error distribution with S_x^2 being equal to σ_E^2.

In this case:

$$\text{If } S_x^2 = \sigma_E^2 \qquad r_{wg} = 1 - \sigma_E^2 / \sigma_E^2$$

$$r_{wg} = 0$$

However, it may be the case that any lack of agreement by the raters may not be random in nature, but may be affected by some sort of response bias, such as leniency, central tendency, or severity. When estimating r_{wg} indices, James et al (1984) recommended that multiple null error distributions should be used to account for any possible affects of rater bias.

If the raters were to exhibit no systematic bias, then the null distribution, or response equating to no agreement, would be uniform or rectangular. However, if a central tendency bias was suspected, the null distribution would be expected to be triangular in form, with a greater tendency for raters to use central rating values, than those at the extremes of the rating scale. Central tendency might be expected from those raters who are cautious, or who are not comfortable or accomplished with the use of assessment criteria. Although the raters forming Group 2 of this study had not been trained in the use of the study's specific crisis management competence assessment framework, they were already expert competence assessors in other domains, and as such would not be expected to exhibit any central tendency bias.

The raters in Group 2 of this study knew that they were participating in a research programme and that the outcome of their assessments was completely confidential, and could not in any way affect the research participants being assessed. Because of this there was no reason to suspect that the raters would exhibit any leniency bias. Also, as none of the raters in Group 2 either personally knew, or had ever met, the research participants being assessed, there was no reason to suspect that the raters would exhibit any severity bias either.

In the light of the discussion in the preceding paragraphs, it is considered acceptable within this study to use a uniform null distribution when determining the within-group interrater agreement for assessor Group 2.

Differential response bias may be another influence on rater response. This bias relates to the situation where raters apply different subjective interpretations to identical points on a rating scale. In order to reduce the risk of this type of bias occurring, a simple four point rating scale was chosen, with clear alternative responses available to the raters, see Figure 5 below:

Good	Towards Good	Towards Poor	Poor
4	3	2	1

Figure 5. Assessment Rating Scale

The option of giving a rating of zero was also available to the raters if they could not observe the behavioural marker being assessed.

Mood, Graybill and Boes (1974) give an equation for the theoretical variance of a uniform null error distribution as:

$$\sigma_{EU}^2 = \frac{A^2 - 1}{12}$$

where A is the number of response options.

Therefore, for a four point rater scale:

$$\sigma_{EU}^2 = \frac{4^2 - 1}{12}$$

$$\sigma_{EU}^2 = 1.25$$

Therefore, in order to calculate the within group interrater agreement r_{wg} index for assessor Group 2, the following equation was used:

$$r_{wg} = 1 - \frac{S_x^2}{1.25}$$

6.4.2 Rank Order

The outcome of the assessments carried out by the different groups of assessors were, in the case of assessor groups 1 and 3, a rank order of the Chief Engineers assessed from best crisis manager to worst crisis manager. The outcome of the assessments carried out by assessor group 2 were also provided in terms of a rank order, by calculating the mean of all of the rating scores given by each assessor for each individual behavioural marker for each Chief Engineer. With a set of three rank order scores, it is possible to use a rank correlation coefficient to calculate the size of the correlation between pairs of score sets. This is a way of determining how well the different groups of assessors agree with each other over the assessment of the Chief Engineers competence in crisis management.

When determining the level of concordance in a series of paired variates there are a number of statistical methods that can be used, such as the Pearson product-moment correlation coefficient or the Spearman rank correlation coefficient. Common practice has been to use the Pearson correlation coefficient, but there are arguments in favour of using the Spearman rank correlation coefficient to determine the level of concordance between the three groups of assessors when assessing the Chief Engineers competence in crisis management.

The first argument is that the Spearman rank correlation coefficient is a distribution free, or non-parametric statistic, which is more suited to correlating non-normally distributed data (McDonald & Green, 1960). The data obtained from the assessment of the Chief Engineers is not expected to have a normal distribution if the level of objectivity and discrimination of the assessment criteria is sufficiently high.

The second argument is that in a number of studies it has been shown that when comparing an analysis of the same set of data there is a high level of agreement between the results obtained using the Pearson correlation coefficient and those obtained using the Spearman rank correlation coefficient (McDonald & Green, 1960; O'Brien & Griffiths, 1965).

For the reasons discussed above it was decided to use the Spearman rank correlation coefficient in order to determine the level of assessment concordance between the different groups of assessors in this experiment.

As each group of assessors is made up of either six or seven assessors, there is always a chance of some level of variability in their assessments, whatever assessment framework they are using. By using a statistical method such as a rank correlation coefficient, a test to determine the statistical significance of the results can be carried out to show whether or not the correlation has a recognised level of significance.

As the rank order is based upon a ranking of 12 Chief Engineers, the generally accepted critical values of the Spearman Rank Order Correlation Coefficient (r_s) are shown below:

Number of Subjects (N)	Level of Significance $p < 0.05$	Level of Significance $p < 0.01$
12	0.503	0.678

Lee (2008)

If a Spearman rank order correlation coefficient is determined that is statistically significant between assessor groups 2 and 3, this will show that the experimental assessment framework can be used by non-domain experts, without the need for specific training in its use, to assess competence in crisis management of marine engineering officers in an engine room control room environment.

If a Spearman rank order correlation coefficient is determined that is statistically significant between assessor groups 1 and 3, this will validate the use of the expert crisis management assessors as a means of providing a benchmark for the level of competence in crisis management for each of the Chief Engineers assessed.

6.4.3 Observability

The effectiveness of any assessment framework based on behavioural markers is in part dependent upon the observability of the behavioural markers used. It does not matter how valid a particular behavioural marker is at measuring a level of competence in crisis management, if that behavioural marker cannot be observed when an assessment is being undertaken.

The design of any assessment scenario is very important when using a behavioural marker based assessment framework. The scenario must be designed to allow the greatest opportunity for all of the behavioural markers within the assessment framework to be observed at some level on the rating scale. If a behavioural marker is consistently not observed, then its worth within the assessment framework would have to be questioned. A behavioural marker may not be observed for a number of reasons:-

i) The behavioural marker is physically difficult to observe.

In order to address this issue the behavioural markers within the experimental assessment framework were all determined and validated by direct observation of the experiment scenario used in phase one of this research.

ii) The assessment scenario does not lend itself to eliciting that behaviour.

Once again this issue was addressed by ensuring that during the development of the experimental assessment framework, all of the proposed behavioural markers were observed when using the experiment scenario.

iii) The candidate being assessed does not exhibit that behaviour.

If more assessors do not observe a particular behavioural marker than is expected by chance, then it is probable that the candidate has not exhibited that behaviour. This is a perfectly valid result and can be interpreted as part of the candidates overall assessment.

iv) The assessor does not recognise the behavioural marker even though it is being exhibited by the candidate.

All of the behavioural markers used within the experimental assessment framework were carefully chosen for their ease of observability. This should make it easy for any assessor to recognise each behaviour if it is being exhibited by the candidate.

v) The behavioural marker has been poorly worded within the assessment framework, making it difficult to interpret whether it has been observed or not.

The assessment framework was evaluated by expert crisis management assessors during its development. Part of this evaluation was to ensure that the behavioural markers were described clearly and concisely in order to make them easy to interpret objectively and without ambiguity.

As discussed in section 6.4.1 the assessment rating scale gave the assessors an option for recording the fact they were unable to observe any particular behavioural marker. It was therefore possible to record the percentage of behavioural markers observed by the Group 2 assessors for each experiment scenario.

The results and analysis of the within-group interrater agreement, the Spearman rank order correlation coefficient and the assessment framework behavioural marker observability, for assessor Group 2, will now be discussed in Chapter 7.

Chapter 7. RESULTS AND ANALYSIS (PHASE TWO)

Three sets of results will be discussed within this chapter:

1) The within-group interrater agreement for the assessments made by the assessors in assessor Group 2 when using the experimental behavioural marker based assessment framework.
2) The rank order of competence in crisis management for the twelve Chief Engineers, as assessed by the three different assessor groups.
3) The observability of the behavioural markers used within the experimental assessment framework.

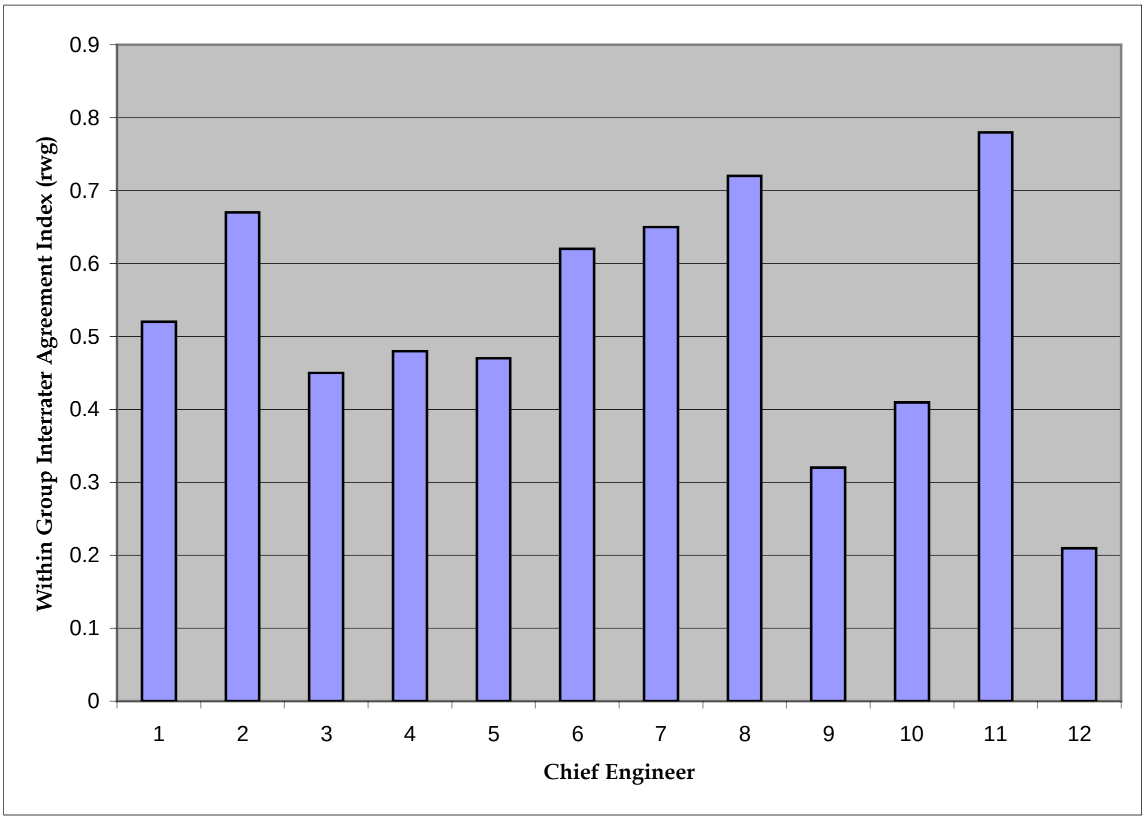
The data recorded by the three assessor groups when viewing the twelve experiment scenarios are shown in Appendix 3.

7.1 INTERRATER AGREEMENT

The results of the within-group interrater agreement for uniform null distribution r_{wg_un} obtained using the assessment scores from the Group 2 assessors when assessing the twelve Chief Engineers are shown in Table 27 below. Mean within-group interrater agreement indices are shown for each Chief Engineer and for each Behavioural Marker.

BM	Within-Group Interrater Agreement for uniform null distribution r_{wg_un}												Mean r_{wg_un}
	C/E 1	C/E 2	C/E 3	C/E 4	C/E 5	C/E 6	C/E 7	C/E 8	C/E 9	C/E 10	C/E 11	C/E 12	
1	0.12	0.47	0.68	0.23	0.55	0.68	0.55	0.76	0.55	0.12	0.79	0.47	**0.50**
2	0.55	0.47	0.68	0.55	0.47	0.87	0.68	0.79	0.76	-0.17	0.87	0.47	**0.58**
3	0.23	0.87	0.44	0.12	0.12	0.44	0.44	0.76	0.04	0.44	0.87	-0.60	**0.35**
4	0.36	0.47	0.12	0.44	0.47	0.47	0.55	0.76	0.87	0.44	0.76	0.44	**0.51**
5	0.84	0.87	0.04	0.23	0.15	0.76	0.80	0.84	0.15	0.04	0.79	0.04	**0.46**
6	0.79	0.47	0.87	0.87	0.47	0.87	0.79	0.79	0.12	0.44	0.87	0.68	**0.67**
7	0.47	0.47	0.47	0.15	0.76	0.79	0.23	0.47	0.12	0.55	0.79	0.15	**0.45**
8	0.12	0.79	0.55	0.60	0.47	0.44	0.84	0.76	0.84	0.44	0.76	0.20	**0.57**
9	0.47	1.00	0.23	-0.28	-0.17	0.68	0.80	0.36	-0.04	0.60	0.79	0.20	**0.39**
10	0.44	0.76	0.44	0.79	0.76	0.55	0.84	0.60	0.36	0.44	0.76	0.23	**0.58**
11	0.44	0.73	0.20	0.44	0.27	0.27	0.80	0.60	0.36	1.00	0.84	0.44	**0.53**
12	0.47	0.60	0.47	0.87	0.79	0.23	0.55	0.79	0.15	0.23	0.76	0.04	**0.49**
13	0.79	0.79	0.79	0.79	0.76	0.68	0.76	0.76	0.47	0.79	0.79	0.68	**0.73**
14	0.44	0.79	0.36	0.44	0.55	0.47	0.55	0.79	0.36	0.79	0.79	0.15	**0.54**
15	0.23	0.79	-0.17	0.55	0.47	0.87	0.47	0.79	0.15	-0.09	0.87	-0.52	**0.36**
16	0.87	0.20	0.47	0.87	0.44	0.84	0.76	0.84	0.47	0.36	0.79	0.47	**0.61**
17	0.87	0.79	0.79	0.23	0.79	0.55	0.68	0.79	0.12	0.47	0.76	0.12	**0.58**
18	0.55	0.79	0.44	0.68	0.23	0.47	0.55	0.79	0.47	0.55	0.76	0.23	**0.54**
19	0.76	0.55	0.68	0.47	0.55	0.79	0.76	0.76	-0.17	0.36	0.44	0.23	**0.51**
Mean r_{wg_un}	**0.52**	**0.67**	**0.45**	**0.48**	**0.47**	**0.62**	**0.65**	**0.73**	**0.32**	**0.41**	**0.78**	**0.21**	

Table 27. Within-Group Interrater Agreement for uniform null distribution r_{wg_un}



Graph 1. Mean Within-Group Interrater Agreement for Chief Engineers

Graph 1 shows the degree of interrater agreement between the assessors for each of the twelve Chief Engineers who were assessed during the crisis management scenario.

The standard for interpreting Interrater Agreement index estimates used in this study is that proposed by LeBreton and Senter (2007):

Level of Interrater Agreement	Substantive Interpretation
0.00 to 0.30	Lack of agreement
0.31 to 0.50	Weak agreement
0.51 to 0.70	Moderate agreement
0.71 to 0.90	Strong agreement
0.91 to 1.00	Very strong agreement

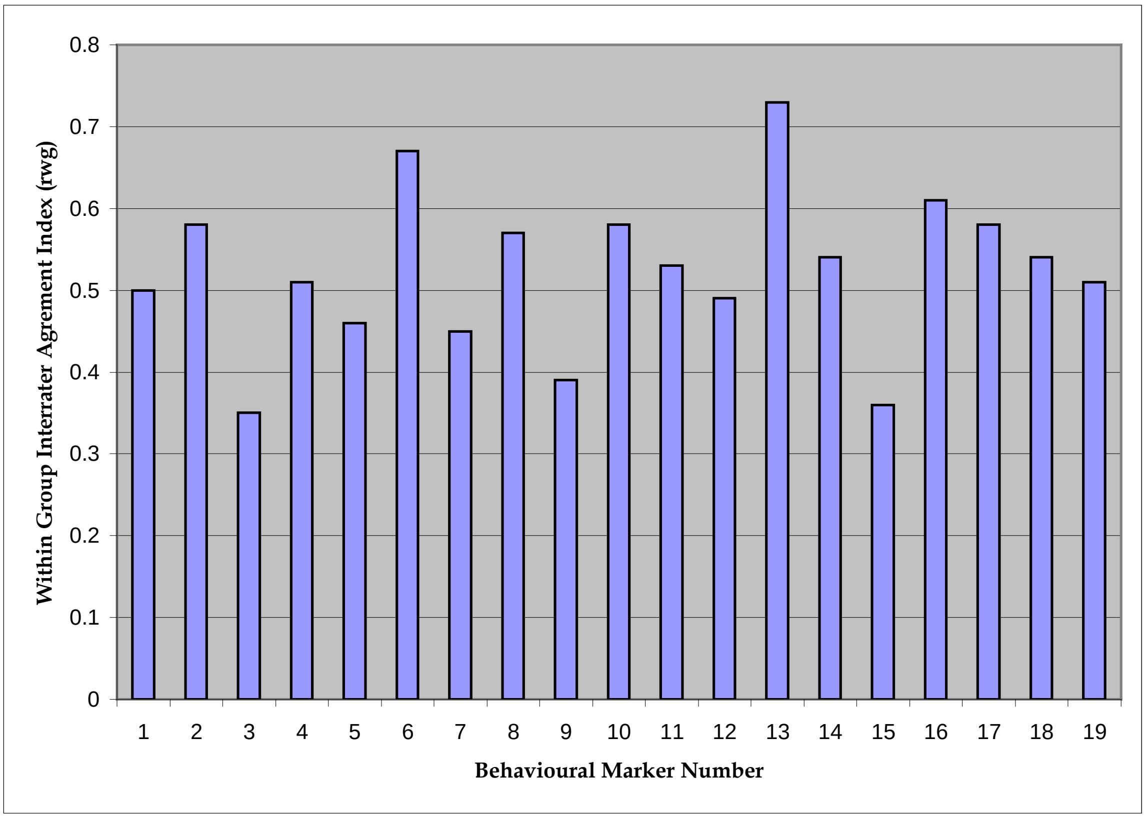
For assessor Group 2 the strongest level of within-group interrater agreement was for Chief Engineers 11, 8, 2 and 7. There was strong agreement between the assessors in their assessment of Chief Engineers 11 and 8. There was moderate agreement between the assessors in their assessment of Chief Engineers 1, 2, 6 and 7. There was only weak agreement between the assessors in their assessment of Chief Engineers 3, 4, 5, 9 and 10. There was a lack of agreement between the assessors in their assessment of Chief Engineer 12.

Williams et al (1997) argue that an acceptable level of interrater agreement when using a behavioural marker based assessment system would be between 0.7 to 0.8. However, this level of agreement was only expected after extensive rater training that would lead to raters being able to calibrate their assessments through comparisons with other assessors. It was one of the objectives of the research undertaken for this thesis to discover whether it was possible to develop a behavioural marker based assessment framework that could provide an acceptable level of interrater agreement without the need for extensive rater training and calibration. The data in Table 27 indicates that, at least in part, this objective has been achieved. The strong interrater agreement in the assessment of Chief Engineers 11 and 8 shows that for certain candidates the experimental assessment framework can achieve an acceptable level of interrater agreement. However, there are also other candidates for which the assessment framework delivers only weak interrater agreement. In the next section on Rank Order, reasons for these differences in interrater agreement will be explained.

These results suggest that this experimental behavioural marker based assessment framework is good at discriminating between the competent and the not yet competent crisis managers.

Graph 2 below shows the Within Group Interrater Agreement Index r_{wg_un} for each of the behavioural markers used within the assessment framework. These results show that there was moderate to strong within group interrater agreement for 12 out of the 19 behavioural markers used within the assessment framework. For those behavioural markers that indicated only weak agreement further research would need to be undertaken to determine if the assessment framework would be improved by their exclusion, or if some refinement of these behavioural markers might be made to improve these levels of interrater agreement. One way of refining these behavioural markers would be to debrief the assessors to determine whether changes to the good and poor behaviour criteria would have made the markers easier to assess.

It should be noted that the interrater agreement results from these experiments have been achieved with no rater training or rater calibration, and yet still compare very favourably with results from research into the use of other behavioural marker based assessment frameworks, for which extensive rater training and calibration was undertaken, such as JAR TEL Consortium, 2001 and Fletcher et al, 2003.



Graph 2. Mean Within-Group Interrater Agreement for Behavioural Markers

7.2 Rank Order

Assessor groups 1 and 3 were tasked to produce a rank order from their assessments. The rank order for assessor group 2 can be determined by calculating the mean of all of the rating scores given by each assessor for each behavioural marker for each Chief Engineer. These mean total ratings are shown in Table 28 below.

Chief Engineer	Mean Total Rating	Rank Order
1	2.59	7
2	1.79	12
3	2.44	8
4	2.60	6
5	2.87	3
6	2.69	4
7	1.97	11
8	3.37	2
9	2.63	5
10	2.08	10
11	3.62	1
12	2.26	9

Table 28. Rank Order of Chief Engineers as assessed by Group 2.

The mean rank order for assessor groups 1 and 3 can be determined by calculating the mean of each of the rank order scores given to each Chief Engineer by each of the assessors in the group. These rank orders are shown in Tables 29 and 30 below.

Chief Engineer	Mean Ranking	Rank Order
1	5.50	4
2	7.50	8
3	8.00	9
4	4.50	3
5	6.33	6
6	5.67	5
7	9.00	10
8	3.33	2
9	7.17	7
10	9.83	12
11	2.00	1
12	9.17	11

Table 29. Rank Order of Chief Engineers as assessed by Group 1.

Chief Engineer	Mean Ranking	Rank Order
1	4.86	4
2	9.14	10
3	5.43	5
4	4.57	3
5	8.71	9
6	7.86	8
7	9.71	11
8	2.86	2
9	6.86	6
10	10.00	12
11	1.14	1
12	7.71	7

Table 30. Rank Order of Chief Engineers as assessed by Group 3.

It is now possible to compare the rank order of the Chief Engineers as determined by each of the three groups of assessors. The three rank orders are shown in Table 31 below.

Rank Order	Group 1 (Chief Engineer)	Group 2 (Chief Engineer)	Group 3 (Chief Engineer)
1	11	11	11
2	8	8	8
3	4	5	4
4	1	6	1
5	6	9	3
6	5	4	9
7	9	1	12
8	2	3	6
9	3	12	5
10	7	10	2
11	12	7	7
12	10	2	10

Table 31. Comparison of the Mean Rank Order of Chief Engineers by each group of assessors.

It is possible to use the Spearman Rank Correlation Coefficient to determine the agreement between two rankings. The rank correlation coefficient is within the interval of –1 to 1 where:

- If there is perfect agreement between the two rankings such that the rank orders are the same, the coefficient has the value of 1;

- If there is perfect disagreement between the rankings, such that one rank order is the exact reverse of the other, the coefficient has a value of –1;

- For all other levels of agreement, the closer the coefficient is to 1 the greater the level of agreement;

- There is one special case where the value of the coefficient is zero. In this case, the rank orders can be said to be completely independent of one another.

Spearman's Rank Correlation Coefficient is given by the formula:

$$\rho = 1 - \frac{6 \Sigma d_i^2}{n(n^2-1)}$$

where d_i = the difference between each rank of corresponding values

and n = the number of pair values

Taking the mean rank order values for non-domain expert assessor, group 2, and the mean rank order value the expert crisis management assessors, group 3, the Spearman rank correlation coefficient between these sets of data can be determined.

Chief Engineer	Rank Order as assessed by Group 2	Rank Order As assessed by Group 3	Difference in rank order between groups d_i	Difference squared d_i^2
1	7	4	3	9
2	12	10	2	4
3	8	5	3	9
4	6	3	3	9
5	3	9	6	36
6	4	8	4	16
7	11	11	0	0
8	2	2	0	0
9	5	6	1	1
10	10	12	2	4
11	1	1	0	0
12	9	7	2	4
			Σd_i^2	92

$$\therefore \rho = 1 - \frac{6 \times 92}{12(12^2 - 1)}$$

$$= 0.6783$$

Null Hypothesis 1:

There is no relationship between the ranking of the candidates by assessor groups 2 and 3.

$\rho = 0.6783 >$ critical value 0.503 ($\alpha = 0.05$, n=12)

So Null Hypothesis 1 is rejected and it is concluded that the relationship between the ranking of the candidates by assessor groups 2 and 3 is statistically significant ($p \leq 0.01$)

This figure of 0.6783 ($p \leq 0.01$) for the Spearman rank correlation coefficient shows that there is a good level of agreement between the mean rank order of the group of expert crisis management assessors and the mean rank order of the group of non-domain expert assessors who used the experimental assessment framework, and that the level of this agreement was significant ($p \leq 0.01$).

As the Spearman rank correlation coefficient is statistically significant between assessor groups 2 and 3, this shows that the experimental assessment framework can be used by non-domain experts, without the need for specific training in its use, to assess competence in crisis management of marine engineering officers in an engine room control room environment.

Now, taking the mean rank order values of the marine engineer officer operations domain experts, group 1, and the mean rank order value of the expert crisis management assessors, group 3, the Spearman rank correlation coefficient between these sets of data can be determined.

Chief Engineer	Rank Order as assessed by Group 1	Rank Order As assessed by Group 3	Difference in rank order between groups d_i	Difference squared d_i^2
1	4	4	0	0
2	8	10	2	4
3	9	5	4	16
4	3	3	0	0
5	6	9	3	9
6	5	8	3	9
7	10	11	1	1
8	2	2	0	0
9	7	6	1	1
10	12	12	0	0
11	1	1	0	0
12	11	7	4	16
			Σd_i^2	56

$$\therefore \ \rho = 1 - \frac{6 \times 56}{12(12^2 - 1)}$$

$$= 0.8042$$

Null Hypothesis 2:

There is no relationship between the ranking of the candidates by assessor groups 1 and 3.

$\rho = 0.8042 >$ critical value 0.503 ($\alpha = 0.05$, $n=12$)

So Null Hypothesis 2 is rejected and it is concluded that the relationship between the ranking of the candidates by assessor groups 1 and 3 is statistically significant ($p < 0.005$)

This figure of 0.8042 for the Spearman rank correlation coefficient shows that there is a very good level of agreement between the mean rank order of the group of expert crisis management assessors and the mean rank order of the group of marine engineer officer operations domain experts, and that the level of this agreement was significant ($p < 0.005$).

As the Spearman rank correlation coefficient is statistically significant between assessor groups 1 and 3, this validates the use of the expert crisis management assessors as a means of providing a benchmark for the level of competence in crisis management for each of the Chief Engineers assessed.

From the comparison of the mean rank order of Chief Engineers by each group of assessors (Table 31), it can be seen that all three groups of assessors agree that the Chief Engineers with the highest level of competence in crisis management were Chief Engineers 11 and 8. It can also be noted that the highest indices for the mean within-group interrater agreement amongst the assessors of assessor Group 2 were also for Chief Engineers 11 and 8, being 0.78 and 0.73 respectively.

There were also high indices for the mean within-group interrater agreement amongst the assessors of assessor Group 2 for Chief Engineers 2 and 7, being 0.67 and 0.65 respectively. There is also close rank order agreement between the expert crisis management assessor Group 1 and the non-domain expert assessor Group 2 for Chief Engineers 2 and 7.

For those Chief Engineers for whom the non-domain expert assessors recorded lower levels of mean within-group interrater agreement there was also some slight discrepancy in the rank ordering of these Chief Engineers between the Group 1 and Group 2 assessors. It is reasonable to suggest that these middle rank order Chief Engineers were neither fully competent crisis managers, nor totally incompetent crisis managers. For this reason, different elements of their performance as crisis managers were assessed slightly differently by the different assessors.

However, these results do clearly show that by using the experimental behavioural marker based assessment framework, even without specific training in its use, or specific calibration of their assessments, the non-domain expert assessors were able to determine, with a good level of agreement, the most competent crisis managers and the least competent crisis managers.

7.3 OBSERVABILITY

The rating scale used by the Group 2 assessors included an option to record that any particular behavioural marker had not been observed during the experiment scenario. It was therefore possible to calculate the percentage of observed responses for each behavioural marker in each of the twelve experiment scenarios. The mean percentage of the total observed responses for each scenario was calculated, as was the mean percentage of total observed responses for each behavioural marker. The results are shown in Table 32 below.

BM	Percentage of Observed Responses for each Scenario (Sc)												Mean
	Sc 1	Sc 2	Sc 3	Sc 4	Sc 5	Sc 6	Sc 7	Sc 8	Sc 9	Sc 10	Sc 11	Sc 12	
1	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
2	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
3	100%	100%	83%	100%	100%	83%	83%	100%	83%	100%	100%	83%	**93%**
4	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
5	83%	100%	83%	100%	83%	83%	67%	83%	100%	83%	100%	83%	**87%**
6	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
7	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
8	100%	100%	100%	83%	100%	100%	83%	83%	83%	83%	83%	83%	**90%**
9	100%	100%	100%	100%	100%	100%	67%	100%	83%	83%	100%	83%	**93%**
10	83%	83%	83%	100%	83%	100%	83%	83%	83%	83%	100%	100%	**89%**
11	83%	67%	83%	83%	67%	67%	67%	83%	83%	83%	83%	83%	**77%**
12	100%	83%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**98%**
13	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
14	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
15	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
16	100%	83%	100%	100%	83%	83%	83%	83%	100%	83%	100%	100%	**91%**
17	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
18	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
19	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	**100%**
Mean	**97%**	**96%**	**96%**	**98%**	**96%**	**96%**	**91%**	**96%**	**96%**	**95%**	**98%**	**96%**	**96%**

Table 32. Percentages of Behavioural Markers being Observed.

The overall mean for the percentage of the total observed responses for the 12 scenarios run was 95.92% with a standard deviation of only 1.78. This shows that there was a very high degree of observability of the behavioural markers used in the assessment framework, across all scenarios. It also shows that the design of the experimental scenario was very good at eliciting the behavioural markers used in the assessment framework, from each candidate being assessed, in a way that was easily observable by the assessors. These results therefore validate the crisis management scenario design in terms of the scenario's ability to elicit the behavioural markers being assessed.

Of the individual behavioural markers being assessed, only three, behavioural marker numbers 5, 10 and 11, were observed in less than 90% of assessments. Of these three behavioural markers, only number 11 was observed in less than 87% of assessments. The overall mean percentage of total observed responses for all 19 behavioural markers was 95.68% with a standard deviation of 6.45.

Eleven out of the nineteen behavioural markers were observed during every one of the 72 assessments undertaken.

These data indicate that there was a very high level of observability of the behavioural markers used within the crisis management assessment framework. These results therefore validate the behavioural markers used within the crisis management assessment framework in terms of the exhibition of these behaviours during the management of crises.

7.4 SUMMARY

It is recognised that there is a problem that is common to all behavioural marker based assessments of competence and that is the problem of aggregation over time. Gaba et al (1998) discuss this at length, pointing out that even over short periods of time, levels of performance may fluctuate between good and poor, and that these fluctuations can occur quite rapidly. Where a group of raters are going to award aggregate scores on a rating scale for a particular behavioural marker that relates to the observed overt behaviour exhibited during a period of performance, there is always the possibility that each rater may aggregate differently because they have each focused on different moments of performance. Gaba et al (1998) suggest strategies for addressing the problem of aggregation over time, such as "*developing specific conventions for aggregating scores over time*". One problem with such strategies is that they require the assessors to receive specific training in order for them to be effective. It was one of the objectives of this research to determine if a behavioural marker based assessment framework could be developed with behavioural markers that were so objective and simple to observe, that specific assessor training would not be required. The results on observability discussed in this chapter would tend to suggest that this objective has been achieved.

The objective of this experimental behavioural marker based assessment framework was to determine competence in crisis management. It is clear from the results that there is good discrimination between those candidates who were competent, the top two, 11 and 8, and those who were not yet competent, the bottom three, 10, 7 and 2. Therefore it can be said that the sensitivity of the measurement system is high for these candidates. If the experimental framework were to be used to assess competence in crisis management, a mean rating value could be set as the cut off point between competent and not yet competent. From the data collected in these experiments a mean total rating cut off point could be set at 3.2, measured against the four point rating scale of the experimental assessment framework, with those scoring above this being deemed to be competent in crisis management.

Chapter 8. CONCLUSIONS

It was the intention of phase one of this research to prove or disprove the hypothesis that simple overt behavioural markers could be used to objectively assess competence in crisis management within the context of a simulated merchant vessel engine room control room. In order to do this, an experiment was designed, using a high fidelity engine control room simulator, which recreated the conditions of a crisis situation at sea, during which the performance of a team of marine engineering officers could be observed.

During the literature review, evidence of the use of behavioural markers for the assessment of competence in managing crises within various safety critical organisations was recorded. Empirical evidence for the relevance, or irrelevance, of these markers to the commercial shipping domain was determined by experimentation. The crisis scenario experiments were designed to allow this empirical evidence to be collected, and to provide evidence of any other overt behavioural markers that were relevant to the assessment of competence in crisis management.

From the empirical evidence collected, a sub-set of simple overt behavioural markers was determined to form an experimental assessment framework. Each of the behavioural markers within this experimental assessment framework were shown, through the analysis of twelve crisis scenarios, to have the best balance of the characteristics for ease of observation and evaluation, frequency of occurrence, and relevance to the assessment of competence in crisis management, of all of the behavioural markers evaluated. Of the nineteen behavioural markers that made up the experimental assessment framework, seven were original markers, solely determined during the ethnographic analysis of the crisis scenario experiment video recordings undertaken within this research. The remainder of the markers were determined from the literature review, and then evaluated for their effectiveness in assessing competence in crisis management through the ethnographic analysis.

The analysis of the results of this research, detailed in Chapter 5, have shown that it is possible to use simple overt behavioural markers to objectively assess competence in crisis management within the context of a simulated merchant vessel engine room control room. By showing, through experimentation, a close agreement between the evaluation of the behavioural markers within the experimental assessment framework given by the researcher, and the evaluation given by the expert crisis management assessors, phase one of this research showed that these behavioural markers were likely to be effective assessment criteria for assessing competence in crisis management.

Although this close agreement between researcher and expert evaluators showed that these behavioural markers were possibly reliable assessment criteria, it did not prove their validity in assessing competence in crisis management. In addition, because the expert assessors' evaluation was only compared to the researcher's evaluation, the interrater agreement of these behavioural markers was not proven. It was in phase two of this research that the experimental assessment framework was tested for validity and interrater agreement.

During a research project into the use of simulators to train and assess merchant navy officers in the handling of escalating emergencies (Habberley et al, 2001), the statement was made that:

> *"Crisis management standards of competence are ill defined and consequently so are their 'behavioural markers' by which the standard may be assessed"*

This research has addressed this gap in our knowledge by proposing, testing and validating a set of behavioural markers that can be used to assess competence in crisis management of merchant navy engineer officers.

A crisis situation is a highly demanding decision-making environment. It is therefore important, within any safety critical organisation, to be able to assess the competence of those tasked with managing a crisis should it arise. This research has demonstrated that, in the domain of merchant navy engineer officers, this type of assessment is possible using a simulated working environment and a behavioural marker based assessment framework.

The original element of this research is that it has shown that it is possible to develop and validate a behavioural marker based assessment framework, for use in assessing competence in crisis management of merchant navy engineer officers, which uses simple, objective behavioural markers. It has also shown that this type of assessment framework can be used successfully by non-domain expert assessors without the need for specific assessor training and calibration.

Other behavioural marker based assessment frameworks reviewed within this thesis, such as those from the military and civil aviation domains, have been shown to use assessment criteria that can be viewed as overly complex and therefore subjective in nature. These assessment frameworks were shown to require the training and calibration of assessors in order for them to achieve valid assessments. Even after this assessor training and calibration, although providing valid assessments when assessing non crisis scenarios, some of these assessment frameworks were found to give ambiguous assessments when used to assess competence in managing crisis scenarios.

This research has provided an understanding of how behavioural markers can be used to objectively assess competence in crisis management of merchant navy engineer officers. The research has also provided an understanding of the methods by which these behavioural markers can be elicited and assessed. The research has shown that through the use of a realistic crisis scenario within a high fidelity simulated ship's engine control room environment it was possible to elicit the objective behavioural markers of a crisis management assessment framework. The observability results show that the crisis scenario was designed to ensure that there were opportunities for all of the behaviours within the assessment framework to be exhibited. This research has defined a set of behavioural markers that constitute a standard of performance for use in the assessment of competence in crisis management.

Within the research it was argued that a competence can only be validly assessed within the context in which it is practiced or used. This argument led to the development of a simulated crisis scenario with a high level of contextual fidelity in which the behavioural markers of the standard performance could be elicited.

The construct validity of the crisis management assessment framework was proven by showing a correlation in the assessment of competence undertaken by the group of non-domain expert assessors and that undertaken by the group of expert crisis management assessors.

The most difficult aspect of developing the assessment framework was determining the behavioural markers to use. These markers not only had to be highly relevant to the competence of managing crises, but also had to very clearly define the enactment of an overt behaviour that could be easily observed. The causal relationship and validity of each behavioural marker with respect to the performance outcome of the crisis manager was determined by the use of the literature review and the validation experiments undertaken during phase one of the research. A number of the markers were synthesised from elements of previous research into the management of crises. Both these synthesised markers, and the remaining markers that were determined through the ethnographic analysis of the video data, were validated by experiment through their evaluation by expert crisis management assessors.

Through the review of the literature it was shown that verbal protocols and cognitive analysis are difficult techniques to use directly in the assessment of competence in crisis management. In fact it was shown that these techniques could, if used in a too holistic or reductionist manner, adversely affect the outcome of an assessment. However, some overt behaviours related to these techniques, such as proposing alternative hypotheses and actions, and communicating in a way that reveals ones mental model of a situation, were shown to be useful metrics for the assessment of competence in crisis management. The review of the literature also showed that the interactions between the crisis manager and the crisis management team play an important part in the management of a crisis.

The differences between 'high level' behavioural markers, consisting of a complex set of behaviours, and 'low level' behavioural markers, consisting of very basic simple behaviours, were discussed. This research provides evidence that the use of 'high level' behavioural markers, as assessment criteria for assessing competence in crisis management, can lead to poor interrater agreement, and therefore inaccurate assessments. Conversely, the research has shown that by using 'low level' behavioural markers for the same purpose, interrater agreement of an acceptable level can be achieved, and therefore accurate assessments.

This research has shown that in order to be able to assess competence in crisis management, an assessment environment needs to be provided that can ensure the presence of those domain specific cues the crisis manager requires to make good decisions. If the assessment environment lacks realism, it can affect the way the candidate conceptualises information from the environment, leading to a change in their level of situational awareness, thereby affecting their ability to manage the crisis. It can be concluded from this that the integrity of the realism in any simulation used to assess competence in crisis management is of the greatest importance.

This research has shown support for the argument that because there need not be consistent relationships between processes and outcomes, and because some process violations are beneficial, the use as assessment criteria comprised of complete models of expert behaviour exhibited when undertaking a process would not seem to be justified. Instead, the behavioural markers that support and facilitate successful processes, beneficial process violations and successful outcomes would seem to be the most relevant assessment criteria.

The evaluation of the assessment frameworks currently in use by a range of safety critical industries and organisations showed that they include a high proportion of behavioural markers that consist of complex sets of behaviours. The evaluation showed that this type of behavioural marker was difficult to assess objectively, especially in real time during a crisis scenario, when there are a great many coincidental behaviours occurring in a short space of time. This evidence therefore suggests that, in order to be effective, any assessment framework developed to assess competence in crisis management should use assessment criteria that consist of simple, uncomplicated, behavioural markers that are overt and easily observable.

In relation to the hypotheses proposed throughout this research, the following can now be concluded:

<u>*Hypothesis 1*</u>

> *It is possible to use simple overt behavioural markers to objectively assess competence in crisis management within the context of a simulated merchant vessel engine control room.*

and

<u>*Hypothesis 2*</u>

> *That the sub-set of behavioural markers that formed the experimental assessment framework determined in phase one of the research can be used as valid markers for the objective assessment of competence in crisis management within the context of a ships engine room control room.*

This research showed that there was a high level of non domain expert assessor within group interrater agreement when assessing the Chief Engineers who are most competent in crisis management. This level of competence in crisis management was validated by both domain expert and expert crisis management assessor groups. It can therefore be concluded from these results that the simple overt behavioural markers that formed the experimental assessment framework can be used as valid markers for the objective assessment of competence in crisis management within the context of a ships engine room control room.

<u>*Hypothesis 3*</u>

> *That the experimental assessment framework that has been developed can be used with a high degree of interrater agreement.*

The results of this research show that, when using the experimental assessment framework, the two Chief Engineers for whom the non-domain expert assessors had strong within group interrater agreement were the two who were the most competent at crisis management. Both the expert crisis management assessors and the domain expert assessors agreed with this assessment. It can therefore be concluded that for those Chief Engineers who are competent in crisis management the experimental assessment framework can be used with a high degree of interrater agreement.

Hypothesis 4

That if a behavioural marker based assessment framework has sufficiently simple and objective assessment criteria, it can be used by non-domain experts to assess competence in crisis management.

The within-group interrater agreement for the non-domain expert assessors was strongest for the most competent Chief Engineers. This result shows that the simple and objective behavioural markers used within the experimental assessment framework have allowed the assessors to clearly discriminate between those Chief Engineers who are competent in crisis management and those Chief Engineers who are not yet competent in crisis management.

Hypothesis 5

That if a behavioural marker based assessment framework has sufficiently simple and objective assessment criteria, it can be used successfully, without the need to give specific training to assessors in its use.

This research has shown that there is a statistically significant Spearman rank correlation coefficient of 0.6783 ($p \leq 0.01$) for the level of agreement between the mean rank order of competence in crisis management as assessed by the group of expert crisis management assessors, and the mean rank order of the group of non-domain expert assessors who used the experimental assessment framework. This correlation shows that the experimental assessment framework can be used by non-domain experts, without the need for specific training in its use, to assess competence in crisis management of marine engineering officers in an engine room control room environment.

This research has also shown that although the non-domain expert assessors received no rater training or rater calibration in the use of the experimental assessment framework, their interrater agreement results compare very favourably with results from research into the use of other behavioural marker based assessment frameworks, for which extensive rater training and calibration was undertaken.

This research was undertaken 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 engineer officers.

The aims of this research were:

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

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

The results of this research have shown that by using uncomplicated, simple, overt behavioural markers as assessment criteria it is possible to objectively assess competence in crisis management of merchant marine engineer officers.

There was a very high degree of observability of the behavioural markers used within the experimental assessment framework when the assessors observed the candidates' performance within the simulated crisis scenario. This high degree of observability has shown that the use of a crisis scenario enacted within a full mission engine room simulator is a valid method for eliciting these behavioural markers.

It was noted in the introduction to this research that, because of the very competitive commercial environment in which the merchant navy operates, any framework for the assessment of competence in crisis management of merchant navy engineer officers would have limited resources available for its implementation. This research has shown that it is possible to have a crisis management competence assessment framework that can be used by a single non domain expert assessor, without any specific rater training or calibration, to produce a valid assessment of a candidate's competence in crisis management.

In the domain of the merchant navy this research has a number of original aspects:

1. The determination of uncomplicated, simple, overt behavioural markers for an experimental crisis management assessment framework, through an ethnographic analysis of crisis manager performance within a crisis scenario enacted within a full mission merchant vessel engine room simulator.

2. The validation of a crisis management assessment framework that can be used by non domain expert assessors, without any prior training or calibration, to make valid assessments of competence in crisis management of merchant navy engineer officers.

3. The development of an assessment tool that can be used within the international maritime industry for the assessment of competence in crisis management of marine engineer officers.

4. This research has moved forward the knowledge of the use of behavioural markers for the assessment of competence.

Chapter 9. REFLECTIONS AND RECOMMENDATIONS

There are elements of this research that could be expanded in order to try to further improve and develop the crisis management competence assessment framework. The phase 2 experiments could be performed again with a new set of assessors for Group 2. These assessors would still be non-domain expert assessors, the same as those originally used in assessor Group 2. The assessment framework for this new set of experiments should only use those behavioural markers that had an original within group interrater agreement index of 0.5 or greater. These new experiments could be used to show if this sub-set of behavioural markers returned a higher within group interrater agreement for Chief Engineer assessments, but also retained a high Rank Spearman Correlation Coefficient between the new assessor Group 2 rank order and the original expert crisis management assessor Group 3 rank order. If this were the case, it would show that this sub-set of thirteen behavioural markers was able to produce a valid assessment of competence in crisis management with a higher overall level of interrater agreement, but with an assessment framework that was easier to use, by having six less behavioural markers to assess.

It might also be of interest to undertake further experiments to determine the minimum number of behavioural markers that could be used within the assessment framework that would still give a valid assessment of competence in crisis management.

The phase 2 experiments could also be performed again, but instead of using the four point rating scale, a two point rating scale of 'competent' and 'not yet competent', could be used instead. If by using a two point rating scale the within group interrater agreement and the Spearman rank correlation coefficient could be improved, this would further validate the experimental assessment framework.

In this research only one crisis scenario was used. It is possible that some candidates responded well, or poorly, to this particular scenario because of their prior experiences. In order to further validate the experimental assessment framework it would be beneficial to assess candidates performance in a number of different crisis scenarios, so as to remove the possibility that a level of prior experience of a particular crisis scenario did not unduly influence the assessment results.

The use of video recordings of candidates' performance, rather than real time live assessment, allowed the assessors time to observe and assess the behavioural markers, even during complex situations, when many behaviours were being exhibited in quick succession.

In order to improve confidence in the validity of the behavioural marker based experimental assessment framework, the crisis scenario experiments could be run again to increase the sample size of Chief Engineers being assessed. The results of the research have shown that the research methodology is robust, but a larger sample size would help to improve confidence in the validity of the experimental assessment framework.

There is always the question of whether an assessment of a candidate's performance carried out within a full mission engine room simulator will truly reflect that candidate's performance onboard a real vessel at sea. Many safety critical industries and organisations, such as civil aviation, medicine, offshore and the fire brigade, use high fidelity full mission simulators to assess the competence of their personnel. It is generally accepted within these industries and organisations that a high fidelity full mission simulator can provide all of the cues from the environment that are necessary for candidates to exhibit behaviours in the same way as they would in their real working environment. It would therefore be expected that a high fidelity full mission merchant vessel engine room simulator could also provide the cues necessary for candidates to perform in a way that reflected their normal performance onboard a real vessel at sea. Anecdotal evidence received by the researcher from hundreds of seafarers who have used the full mission engine room simulator supports this expectation.

This research showed how in many other safety critical industries and organisations training and assessment is undertaken to try to ensure that those personnel who may be called upon to manage crisis situations are competent to do so. It is hoped that this research will be used by others to try to ensure that the international maritime industry also has personnel in place who, if the need arises, are competent to manage crises. In order to try to achieve this aim it is intended that the International Maritime Organization be advised of this research.

REFERENCES

American Bureau of Shipping. (2004). *ABS Review and Analysis of Accident Databases: 1991 – 2002 Data.* Technical Report SAHF 2003-5.1. Houston, TX: American Bureau of Shipping.

Anderson, J.R. (1985). *Cognitive psychology and its implications.* 4th ed. New York: W.H. Freeman and Company.

Anzai, Y., & Simon, H.A. (1979). The theory of learning by doing. *Psychological Review, 86*, 124–140.

Avermaete, J.A.G. van & Kruijsen E.A.C. (1998). (Eds.) *NOTECHS: The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements*. Final Report. NLR Contract Report CR 98443. National Aerospace Laboratory NLR, Amsterdam.

Baker, D.P., Mulqueen, C. & Dismukes, R.K. (2001). Training Raters to Assess Resource Management Skills. In: E. Salas, C.A. Bowers & E. Edens (Eds.) *Improving Teamwork in Organisations: Applications of Resource Management Training.* Mahwah, NJ: Lawrence Erlbaum Associates, 131-145.

Bainbridge, L. (1974). Analysis of Verbal Protocols from a Process Control Task. In: Edwards & Lees (Eds.) *The Human Operator in Process Control.* Taylor & Francis Ltd.

Bainbridge, L. (1988). Types of representation. In L.P. Goodstein, H.B. Anderson, & S.E. Olsen (Eds.), *Tasks, errors and mental models.* London: Taylor and Francis.

Bainbridge, L. (1992). Mental models in cognitive skill: The example of industrial process operation. In: Y. Rogers, A. Rutherford & P. Bibby (Eds.), *Models in the mind.* London: Academic Press. 119–143.

Bainbridge, L. (1993). Planning the training of a complex skill. *Le Travail Humain: Cognitive analysis of work.* Paris: Presses Universitaires de France. 56, 2–3, 211–232

Bainbridge, L. (1995). *Processes underlying human performance.* London: Department of Psychology, University College London.

Banks, A.P. & Millward L.J. (2000). Running shared mental models as a distributed cognitive process. *British Journal of Psychology,* 91, 513-531.

Barnett, M., Gatfield, D. & Habberley, J. (2002). Shipboard Crisis Management: A Case Study. In: *Proceedings of Human Factors in Ship Design and Operation Conference.* London: The Royal Institution of Naval Architects.

Barnett, M., Gatfield, D. and C.Pekcan. (2003). A Research Agenda in Maritime Crew Resource Management. In: *Proceedings of International Conference on Team Resource Management in the 21st Century.* Embry-Riddle Aeronautical University, Daytona Beach, Florida.

Besnard, D. & Baxter G. (2003). *Human compensations for undependable systems.* Technical Report Series, CS-TR-819, School of Computing Science, University of Newcastle upon Tyne.

Besnard, D. & Greathead, D. (2003). A Cognitive Approach to Safe Violations. *Cognition, Technology & Work.* London: Springer-Verlag. 5, 272-282

Birnbach, R.A. & Longridge, T.M. (1993). The Regulatory Perspective. In: E.L. Wiener, B.G. Kanki & R.L. Helmreich (Eds.) *Cockpit Resource Management.* NewYork: Academic Press.

Blanchard, K., & Hersey, P. (1969). Life cycle theory of leadership. *Training and Development.* May, 26–34.

Borodzicz, E., & van Haperen, K. (2002). *Individual and Group Learning in Crisis Simulations.* Discussion Papers in Management. Number M02-5. University of Southampton. ISSN 1356-3548.

Brannick, M.T. & Pearce, C. (1991). Assessment of an air crew rating from within and between scenarios. *Technical Report # DAAL0-3-86-D-001*, Orlando, Florida: Naval Training Systems Center.

Brannick, M.T., Roach, R.M. & Salas, E. (1993). Understanding Team performance: A Multimethod Study. *Human Performance,* Mahwah, NJ: Lawrence Erlbaum Associates, 6(4), 287-308

Breznitz, S. & Ben-Zur, H. (1997). Enhancing effective decision-making by information management techniques. In: R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot: Ashgate.

Brown, A.L., Bransford, J.D., Ferrara, R.A., & Campione, J.C. (1983). Learning, remembering, and understanding. In: J.H. Flavell, & E.M. Markman (Eds.), Cognitive development. (Vol. III), *Handbook of child psychology.* New York: Wiley. 77–166.

Brown, R.D., & Hauenstein, N.M.A. (2005). Interrater agreement reconsidered: An alternative to the r_{WG} indices. *Organizational Research Methods*. 8(2), 165-184.

Cannon-Bowers, J.A. & Bell, H. (1997). Training decision-makers for complex environments: Implications of the naturalistic decision-making perspective. In: C. Zsambok & G. Klein (Eds.) *Naturalistic Decision Making*. Mahwah, NJ: Erlbaum, 99-110.

Cannon-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.

Cannon-Bowers, J.A., Salas, E. & Pruitt, J.S. (1993). Establishing the boundaries for decision-making research. *Human Factors*, 38, 193-205.

Civil Aviation Authority. (2003). *Crew Resource Management (CRM) Training*. CAP 737. London: CAA.

Cohen, M. (1993). The bottom line: Naturalistic decision-making. In: G. Klein, J. Orasanu, R. Calderwood & C. Zsambok (Eds.) *Decision making in action: Models and methods*. Norwood, NJ: Ablex, 265-269.

Cohen, M., Freeman, J. & Wolf, S. (1996). Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting. *Human Factors*, 38(2), 206-219.

Cohen, M., Freeman, J., & Thompson, B. (1997). Training the naturalistic decision maker. In C. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making*, Mahwah, NJ: Lawrence Erlbaum Associates, 257-268.

Collyer, S.C. & Malecki, G.G. (1998). Tactical Decision Making Under Stress: History and Overview. In: J.A. Cannon-Bowers & E. Salas (Eds.) *Making Decisions Under Stress: Implications for individual and team training.* Washington D.C.: American Psychological Society.

Covington, M.V. (1987). Instruction for problem-solving planning. In: S.L. Friedman, E.K. Scholnick & R.R. Cocking (Eds.), *Blueprints for thinking: The role of planning in cognitive development.* Cambridge: Cambridge University Press.

Croskerry, P. (2003). The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them. *Academic Medicine,* 78(8), 775-780.

Cunningham, J.W., Boese, R.R., Neeb, R.W., & Pass, J.J. (1983). Systematically derived work dimensions. Factor analyses of the Occupational Analysis Inventory. *Journal of Applied Psychology,* 68 , 232–252.

de Bono, E. (1985). The CoRT thinking program. In: J.W. Segal, S.F. Chipman & R. Glaser, (Eds.)., *Thinking and learning skills: Vol 1. Relating instruction to research*. Hillsdale, NJ: Lawrence Erlbaum Associates, 363–388.

DuFon, M. (2002). Video Recording in Ethnographic SLA Research: some issues of validity in data collection. *Language and Learning Technology,* 6(1), 40-59.

Economic & Social Research Council. (2006). *Research Ethics Framework.* Swindon: ESRC.

Endsley, M.R. (1995a). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors,* 37(1), 32-64.

Endsley, M.R. (1995b). Measurement of Situation Awareness in Dynamic Systems. *Human Factors,* 37(1), 65-84.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R, Maran, N., Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): evaluation of a behavioural marker system. *British Journal of Anaesthesia.* 90, 580–588

Flin, R. (1996). *Sitting in the Hot Seat: Leaders and Teams for Critical Incident Management.* Chichester: John Wiley and Sons Ltd.

Flin, R. & Martin, L. (1998). Behavioural Markers for Crew Resource Management. *CAA Paper 98005.* London: Civil Aviation Authority.

Flin, R. & Martin, L. (2001). Behavioral Markers for Crew Resource Management: A review of current practice. *The International Journal of Aviation Psychology.* 11(1), 95-118.

Flin, R., Slaven, G. & Stewart, K. (1996). Emergency Decision Making in the Offshore Oil and Gas Industry. *Human Factors*, 38(2), 262-277.

Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T. & Oser, R. (1994). Improving the Measurement of Team Performance: The TARGETs Methodology. *Military Psychology.* 6(1), 47-61.

Fredholm, L. (1997). Decision making patterns in major fire-fighting and rescue operations. In: R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot: Ashgate.

Feuerstein, R. (1980). *Instrumental enrichments: An intervention program for cognitive modifiability.* Baltimore: University Park Press.

Gaba, D., Howard, S., Flanagan, B., Smith, B., Fish, K., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology*. 89, 123–147

Gatfield, D. (1999). Can Cognitive Science Improve the Training of Industrial Process Operators? *Journal of Safety Research*. National Safety Council and Elsevier Science Ltd. 30, 2, 133–142

Gentner, D., & Stevens, A.L. (Eds.). (1983). *Mental Models.* Hillsdale NJ: Lawrence Erlbaum Associates.

Gipps, C.V. (1994). *Beyond Testing: Towards a Theory of Educational Assessment.* London: RoutledgeFalmer.

Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist. 39*, 93–104.

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.

Health and Safety Executive. (2003). *Competence assessment for the hazardous industries.* Research Report 086. Sudbury: HSE Books.

Helmreich, R. L., Butler, R. E., Taggart, W. R. & Wilhelm, J. A. (1994). The NASA/University of Texas/FAA Line/LOS Checklist: A behavioral marker-based checklist for CRM skills assessment. *NASA/UT/FAA Technical Report 94-02.* Revised 12/8/95. Austin, TX: The University of Texas.

International Civil Aviation Organization. (2002). *Line Operations Safety Audit (LOSA).* Doc9803. AN/761. Montreal: ICAO.

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

James, L.R., Dameree, R.G. & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology,* 69(1), 85-98.

James, L.R., Dameree, R.G. & Wolf, G. (1993). r_{wg}: An assessment of within-group interrater agreement. *Journal of Applied Psychology,* 78(2), 306-309.

Janis, L. L. (1982). *Group Think.* Boston: Houghton Miffin.

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.

Jentsch, F. & Smith-Jentsch, K.A. (2001). Assertiveness and Team performance: More Than "Just Say No". In: E. Salas, C.A. Bowers & E. Edens (Eds.) *Improving Teamwork in Organisations: Applications of Resource Management Training.* Mahwah, NJ: Lawrence Erlbaum Associates.

Jirotka, M. (1998). *Video-based requirements elicitation: A Manual.* Oxford: Oxford University Computing Laboratory.

Johnson-Laird, P. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Johnston. J., Poirer, J., & Smith-Jentsch, K. (1998). Decision Making Under Stress: Creating a Research Methodology. In: J.A. Cannon-Bowers & E. Salas (Eds.) *Making Decisions Under Stress: Implications for individual and team training.* Washington D.C.: American Psychological Society.

Katz, J. (1997). Ethnography's warrants. *Sociological Methods & Research.* 25(4), 391-423.

Kerstholt, J.H. (1997). Dynamic decision-making in non-routine situations. In: R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot: Ashgate.

Kerstholt, J.H., Passenier, P.O., Houttuin, K., & Schuffel, H. (1996). The Effect of a Priori Probability and Complexity on Decision Making in a Supervisory Control Task. *Human Factors*, 38(1), 65-78.

Kirijan, F. & Mancuso, V. (1995). AQP Evaluation Strategies and Metrics: The Delta Air Lines Solution. In: *Proceedings of 8th Symposium on Aviation Psychology.* Columbus, Ohio: Ohio State University. 1, 546-550.

Klampfer, B. & Jochum, K. (Eds.). (2001). Group Interaction in High Risk Environments (GIHRE) Project. *Behavioural Markers Workshop*. Berlin: Humboldt Universität

Kleinmann, M. (1993). Are rating dimensions in assessment centres transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*. 78, 482-486.

Kuhn, T.S. (1970). *Structure of Scientific Revolutions* (2d ed.). Chicago: University of Chicago Press.

Lawler, E.E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology.* 51, 369-381.

LeBreton, J.M., Burgess, J.R.D., Kaiser, R.B., Atchley, E.K., & James, L.R. (2003). The Restriction of Variance Hypothesis and Interrater Reliability and Agreement. Are Ratings from Multiple Sources Really Dissimilar? *Organizational Research Methods.* 6(1), 80-128.

LeBreton, J.M. & Senter, J.L. (2007). Answers to 20 Questions about Interrater Reliability and Interrater Agreement. *Organizational Research Methods.* 0: 1094428106296642v1.

Lee, P.M. (2008). *Upper Critical Values for Spearman's Rank Correlation Coefficient* R_s
York: University of York.

London Fire Brigade. (2002). *Incident Command Assessment.* London: London Fire Brigade.

Luff, P., Jirotka, M., Heath, C. & Greatbatch, D. (1993). Tasks and social interaction: the relevance of naturalistic analyses of conduct for requirements engineering. In: *Proceedings of IEEE International Symposium on Requirements Engineering.* San Diego, CA: IEEE. 187-190.

Marine Accident Investigation Branch (1993). *Report of the Chief Inspector of Marine Accidents into the engine failure and subsequent grounding of the Motor Tanker Braer at Garths Ness, Shetland on 5 January 1993.* Southampton: MAIB.

Marine Accident Investigation Branch (1999). *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.* Marine Accident Report 5/99. Southampton: MAIB.

Makins, M. (1994). (Ed.) *Collins English Dictionary* (3rd Ed.) Glasgow: HarperCollins.

Mandler, G. (1975). Consciousness: Respectable, useful and probably necessary. In: R. Solso (Ed.) *Information processing and cognition: The Loyola Symposium.* Hillsdale N.J.: Earlbaum.

Manktelow, K. & James, J., (1987). Principles from the psychology of thinking and mental models. In: M.M. Gardiner & B. Christie (Eds.) *Applying cognitive psychology to user interface design.* Chichester: John Wiley and Sons Ltd., 83-117.

McCormick, E.J., Jeanneret, P.R., & Mecham, R.C. (1969). *A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire.* Report No. 6. Lafayette, Indiana: Purdue University.

McDonald, J.E., & Green, C.R., (1960). A Comparison of Rank Difference and Product-Moment Correlation of Precipitation Data. *Journal of Geophysical Research.* 65, 1, 333-336.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Emerging principles from complex environments. In R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations.* San Francisco: Jossey-Bass., 9-45.

Miller, G.A. (1962). *Psychology: The Science of Mental Life*. New York: Harper & Row.

Miller, R.B. (1974). *A method for determining task strategies.* Technical Report AFHRL-TR-74–26. Alexandria, VA: American Institute for Research.

Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics (3rd ed.)*. New York: McGraw-Hill.

Neisser, G. (1967). *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Nisbett, R.E. & DeCamp Wilson, T. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review,* 84, 3.

Norman, D.A. (1986). Cognitive engineering. In. D.A Norman & S.W. Draper (Eds.), *User centred system design.* Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Brien, J.J. & Griffiths, J.F. (1965). The Rank Correlation Coefficient as an Indicator of the Product-Moment Correlation Coefficient for Small Samples (10 – 100). *Journal of Geophysical Research.* 70, 8, 1995-1998.

Ohlsson, S. (1986). Some principles of intelligent tutoring. *Instructional Science.* 14, 293–326.

Offshore Petroleum Industry Training Organisation. (2005). *OIM Controlling Emergencies.* OPITO Approved Standard. Revision 3. Aberdeen: OPITO.

Orasanu, J.M. & Connolly, T. (1993). The reinvention of decision-making. In: G. Klein, J. Orasanu, R. Calderwood & C. Zsambok (Eds.) *Decision making in action: Models and methods*. Norwood, NJ: Ablex, 3-20.

Orasanu, J.M. (1997). Stress and naturalistic decision-making: Strengthening the weak links. In: R Flin, E Salas, M Strub and L Martin (Eds.) *Decision making under stress: emerging themes and applications*. Aldershot: Ashgate.

Oser, R., McCallum, G.A., Salas, E. & Morgan, B.B. (1989). Toward a definition of teamwork: An analysis of critical team behaviours. *Technical Report No. 89-004.* Orlando, Florida: Naval Training Systems Center.

Patrick, J. (1989). Representation of fault-finding in complex industrial contexts. *In: Proceedings of the Second European Meeting on Cognitive Science Approaches in Process Control.* Siena, Italy: University of Siena. 269–281.

Patrick, J. (1992). *Training: Research and Practice.* London: Academic Press Limited.

Pekcan, C., Gatfield, D. & Barnett, M. (2005). Content and Context: Understanding the Complexities of Human Behaviour in Ship Operation. In: *Proceedings of Human Factors in Ship Operations Conference.* London: Royal Institution of Naval Architects.

Pew, R.W., Miller, D.C., & Feeher, C.E. (1981). *Evaluation of Proposed Control Room Improvements through analysis of Critical Operator Decisions.* Research Project 891. Palo Alto, CA: Electric Power Research Institute.

Pruitt, J.S., Cannon-Bowers, J.A., & Salas, E. (1997). In search of naturalistic decisions. In: R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot: Ashgate.

Rasmussen, J., & Lind, M. (1981). *Coping with complexity.* Riso-M-2293. Roskilde, Denmark: Riso National Laboratory.

Rasmussen, J. (1984). Strategies for state identification and diagnosis in supervisory control tasks and design of computer-based support systems. In W.B. Rouse (Ed.), *Advances in man-machine systems research.* Greenwich, CT: JAI Press. 1, 139–193.

Rasmussen, J., (1986). *Information processing and human machine interaction: An approach to cognitive engineering.* Amsterdam: North Holland.

Rasmussen, J., (1990). Human Error and the problem of causality in analysis of accidents. In: D.E. Broadbent, J. Reason & A. Baddely (Eds.) *Human Factors in Hazardous Situations.* Oxford: Clarendon Press.

Reason, J., (1990). *Human Error.* Cambridge, England: Cambridge University Press.

Regester, M. & Larkin, J. (1997). *Risk Issues and Crisis Management.* Institute of Public Relations. London: Kogan Page.

Reinartz, S.J., & Reinartz, G. (1989). Verbal communication in collective control of simulated nuclear power plant incidents. In *Proceedings of the Second European Meeting on cognitive science approaches to process control.* Siena, Italy: University of Siena. 195–203.

Resnick, L.B. (1987). *Education and learning to think.* Washington, DC: National Academy Press.

Resnick, L.B. & Resnick, D.P. (1992). Assessing the thinking curriculum: New tools for educational reform. In: B. Gifford & M. O'Connor (Eds.) *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction.* London: Kluwer Academic Publishers.

RGIT Montrose. (2002). *Record of OIM Assessment.* Montrose: RGIT Montrose Ltd.

Roberson, Q.M., Sturman, M.C., & Simons, T.L. (2007). Does the Measure of Dispersion Matter in Multilevel Research? A Comparison of the Relative Performance of Dispersion Indexes. *Organizational Research Methods.* 10(4), 564-588.

Royal Navy (2002). *Report of Major Machinery Space Fire at Sea Exercise (F4).* Flag Officer Sea Training Form ME 4B/3. Plymouth: HMS Drake.

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.

Samurcay, R., & Rogalski, L. (1988). Analyses of operator's cognitive activities in learning and using a method for decision making in public safety. In: J. Patrick, & K.D. Duncan (Eds.), *Training, human decision making and control.* Amsterdam: Elsevier.

Schwab, D.P., Heneman, H.G. & DeCotis, T.A. (1975). Behaviorally Anchored Rating Scales: A review of the literature. *Personnel Psychology.* 28, 549-562.

Seamster, T.L., Edens, E.S., & Holt, R.W. (1995). Scenario Event Sets and the reliability of CRM Assessment. In: *Proceedings of the 8th International Symposium on Aviation Psychology.* Columbus, Ohio: Ohio State University. 613-618.

Seamster, T.L., Edens, E.S., McDougall, W. & Hamman, W. (1994). *Observable crew behaviors in the development and assessment of line operational evaluations.* Report No. DTFA01-93-C-00055. Washington DC: Federal Aviation Administration.

Seamster, T.L., Prentiss, F.A., & Edens, E.S. (1997). Methods for the Analysis of CRM Skills. In: *Proceedings of the 9th International Symposium on Aviation Psychology.* Columbus, Ohio: Ohio State University. 1, 500-504.

Serfaty, D., Entin, E.E. & Volpe, C. (1993). Adaptation to Stress in Team Decision-Making and Coordination. In: *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting.* Santa Monica CA: Human Factors and Ergonomics Society.

Serfaty, D., Entin, E.E. & Deckert, J.C. (1994). *Team Adaptation and Coordination Training (TACT).* TR-648-1, Burlington MA: ALPHATEC, Inc.

Serfaty, D., & Entin, E.E. (1997). Team adaption and co-ordination training. In: R. Flin, E. Salas, M. Strub & L. Martin (Eds.) *Decision Making Under Stress: Emerging Themes and Applications.* Aldershot: Ashgate.

Simon, H.A. (1957). *Models of Man: Social and Rational. Mathematical essays on rational human behavior in a social setting.* New York: Wiley

Simon, H.A. (1976). Identifying basic abilities underlying intelligent performance of complex tasks. In L.B. Resnick (Ed.), *The nature of intelligence.* Hillsdale, NJ: Lawrence Erlbaum Associates. 65–98.

Singleton, W.T. (1989). The mind at work. *Psychological ergonomics.* Cambridge: Cambridge University Press.

Smith-Jentsch, K.A., Johnston, J.H., & Payne, S.C. (1998). Measuring Team-Related Expertise in Complex Environments. In: J.A. Cannon-Bowers & E. Salas (Eds.) *Making decisions under stress: implications for individuals and teams.* Washington D.C.: American Psychological Association, 61-87.

Sternberg, R.J. (1980). Sketch of componential subtheory of human intelligence. *Behavioural and Brain Sciences, 3*, 573–614.

Stout, R.J., Salas, E. &Carson, R. (1994). Individual task proficiency and team process behavior: What's important for team functioning? *Military Psychology*, 6, 177-192.

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.

Tourville, S.J. (1997). Identification of key behaviors for effective team coordination in special operations forces combat mission training. In: *Proceedings of the 9th International Symposium on Aviation Psychology.* Columbus, Ohio: Ohio State University. 1, 505-510.

UK P&I Club (1997). *Analysis of major claims – ten-year trends in maritime risk.* London: Thomas Miller P&I Ltd.

United States Coastguard (1995). *Prevention Through People,* Quality Action Team Report. Washington D.C.: USCG.

Urban, J.A., Bowers, C.A., Monday, S.D. & Morgan, B.B., Jr. (1995). Workload, team structure, and communication in team performance. *Military Psychology*. 7, 123-139.

Urban, J.E., Weaver, J.L., Bowers, C.A. & Rhodenizer, L. 1996. Effects of workload and structure on team processes and performance: implications for complex team decision making, *Human Factors*. 38(2), 300-310.

Voss, J.F., Greene, T.R., Post, T.A., & Penner, B.C. (1983). Problem solving in the social sciences. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research theory.* New York: Academic Press. 17, 165–213.

Wagenaar, W. A., & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies.* 27, 587-598.

Wales, C.E., & Nardi, A.H. (1985). Teaching decision-making: What to teach and how to teach it. In: A.L. Costa, (Ed.), *Developing Minds: A resource book for teaching thinking.* Alexandria, VA: Association for Supervision and Curriculum Development.

Watson-Gegeo, K.A. (1988). Ethnography in ESL: Defining the essentials. *TESOL Quarterly,* 22(4), 575-592.

Wertheimer, M. (1959). *Productive thinking.* New York: Harper and Row.

Williams, D.M., Holt, R.W., & Boehm-Davis, D.A. (1997). Training for Interrater Reliability: Baselines and Benchmarks. In: *Proceedings of the 9th Symposium on Aviation Psychology.* Columbus, Ohio: Ohio State University. 1, 514-520.

Wittgenstein, L. (1963). Philosophical Investigations. G.E.M. Anscombe and R. Rhees (Eds), G.E.M. Anscombe, tr. Oxford: Blackwell.

Woods, D.R. (1983). Introducing explicit training in problem solving into our courses. *Higher Education Research and Development, 2* , 79–102.

Zsambok, C., & Klein, G. (1997). *Naturalistic decision-making – Where are we now?* Mahwah, NJ: Erlbaum.

APPENDICES

APPENDIX 1

Research Participant Consent Form

WARSASH MARITIME ACADEMY, SOUTHAMPTON SOLENT UNIVERSITY

CONSENT TO PARTICIPATE IN A RESEARCH STUDY

TITLE OF STUDY: ***Behavioural Markers for the Assessment of Competence in Crisis Management***

RESEARCHER: D.I.Gatfield Tel: 01489 556282 Email: david.gatfield@solent.ac.uk

PURPOSE

You are being asked to participate in a research study. The purpose of this study is to determine how behavioural markers can be used to objectively assess competence in crisis management.

PARTICIPANTS

You are being asked to participate because you are an experienced merchant navy engineering officer who could be called upon to manage a crisis onboard a vessel.

PROCEDURES

If you choose to participate, you will undertake two exercises within the Engine Room Simulator here at Warsash Maritime Academy. Each exercise will last approximately one hour. The first exercise will be used to familiarise you with the simulated engine room environment. During the second exercise, you will be placed in a simulated crisis situation. A video recording will be made of this exercise.

RISKS

There are no known risks associated with the participation in this research study.

BENEFITS

The benefit to you from participating in this study is that you will be able to test your ability to manage a crisis situation onboard a vessel. The benefits to the wider maritime community will be an understanding of how a behavioural marker system can be applied for the assessment of competence in crisis management.

CONFIDENTIALITY

You will not be personally identified in any reports or publications that may result from this study. Use of the video recordings taken during this study will be strictly limited to the investigator named above.

COSTS/COMPENSATION

There will be no cost to you, nor will you be compensated for participating in this study.

Participant's Initials________ Page 1 of 2

RIGHT TO REFUSE OR WITHDRAW

You may refuse to participate or withdraw from the study at anytime.

QUESTIONS

If you have any questions, please ask. If you have additional questions later, contact:
David Gatfield at Warsash Maritime Academy Tel: 01489 556282
Email: david.gatfield@solent.ac.uk

You may report (anonymously, if you so choose) any complaints or comments regarding the manner in which this study is being conducted to: The Quality Manager, Warsash Maritime Academy, Southampton Solent University, Newtown Road, Warsash, Southampton, SO31 9ZL.

CLOSING STATEMENT

MY SIGNATURE BELOW INDICATES THAT I HAVE DECIDED TO VOLUNTEER AS A RESEARCH PARTICIPANT AND THAT I HAVE READ, I UNDERSTAND, AND I HAVE RECEIVED A COPY OF THIS CONSENT FORM.

SIGNATURE OF PARTICIPANT DATE

SIGNATURE OF RESEARCHER DATE

Page 2 of 2

APPENDIX 2

Final Assessment Framework used by Group 2 Assessors

Examples of Good and Poor Behaviours related to Competence in Crisis Management within a Merchant Navy Vessel Engine Room Control Room Environment.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
1.	High degree of crewmember integration within the team	Involves other crewmembers in planning, diagnostic and decision making processes. Accepts feedback and advice from crewmembers.	Isolates a crewmember or crewmembers. Only utilises crewmember or crewmembers by giving them a command order. Ignores feedback or advice from crewmembers. Specific case may be cultural isolation.
2.	Regular soliciting of information between crewmembers	Regularly asks other crewmembers for information.	Does not ask other crewmembers for information.
3.	Demonstrates awareness of 'big picture' context	Verbalises task priorities in anticipation of future events. i.e. "We need to do *xxxxx* now or in *yy* minutes *zzzzz* will happen."	No anticipation of future events evident. Acts surprised as events happen.
4.	Sharing of workload	Delegates individual tasks in order to maintain situation overview.	Gets too involved in individual tasks to the extent that situation overview is lost.
5.	The number of alternative hypotheses and actions communicated to team members	Proposes alternative hypotheses and actions to team members.	Does not propose any alternative hypotheses or actions to team members.
6.	Uses complete and coherent sentences	Communicates to team with complete and coherent sentences.	Uses unfinished sentences in communications with team.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
7.	Uses measured patterns of movement	Uses measured movements within a defined area where team leader is able to maintain overview of situation.	Uses very fast movements. No defined area of movement, so overview of instrumentation is not possible. Focuses too much on one item of instrumentation. Leaves control room for extended periods.
8.	Crisis Manager asking their team questions in order to elicit information so that they can improve their situational awareness	Asks team members questions about the situation in order to improve situational awareness.	Does not ask team members questions about the situation.
9.	Providing big picture updates	Provides team members with updates on the overview of the situation.	Does not provide team members with any updates on the overview of the situation.
10.	Stating clear team and individual priorities	States clear team and individual team member task priorities.	Does not state team or individual team member task priorities.
11.	Utilises team briefings for specific threats	Briefs team about specific threats.	Does not brief team about specific threats.
12.	Keeps crewmembers focused on the accomplishment of the task	Provides encouragement and motivates team to keep them focussed on the accomplishment of the task. Tasks completed in good time.	Fails to monitor task progression. Team does not complete tasks in good time.
13.	Ensuring communications are audible and not garbled	Communications clear and easy to understand.	Communications inaudible and garbled. Communications not understood by team members.

Ref:	Overt Behavioural Marker	Good Behaviours	Poor Behaviours
14.	Focuses on the dynamics of the complete system	Maintains overview of all systems, thereby maintaining awareness of system interactions.	Focuses on one system to the exclusion of all others. Is surprised by system interactions.
15.	Communicating in a way that reveals ones mental models	Communicates thoughts on the situation and how it is developing to team members.	Does not communicate thoughts to team members.
16.	Team leader reflects on the suggestions made by other team members	Reflects on suggestions from team members before accepting or rejecting them.	Acts immediately on suggestions from team members without any prior reflection. 'Grasps' at suggestions of others.
17.	Measured movements in response to stimuli	Makes measured and controlled movements in response to stimuli.	Makes rapid and uncontrolled movements in response to stimuli.
18.	Moves smoothly and without hesitation	Moves without hesitation. Moves smoothly between tasks.	Movement truncated. Movements hesitant. Moves between two task locations without, in the short term, reaching either. Presents an on the spot rocking motion.
19.	Team leader focuses on teams' tasks rather than on own individual tasks	Maintains focus on overall team task objective.	Focuses too much on own individual tasks that are not directly related to the overall team task objective. i.e. team leader spends a lot of time trying to put another alternator onto the main switchboard.

APPENDIX 3

Phase 2 Assessment Data

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Group 1 – Assessor Data

Rank Order of Competence in Crisis Management	Chief Engineer in Video Scenario Number					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
BEST 1	11	11	8	11	11	8
2	8	4	11	1	8	2
3	9	8	6	4	4	4
4	4	1	5	7	6	1
5	1	5	10	5	3	6
6	5	9	12	2	7	11
7	6	6	2	3	9	12
8	3	3	1	6	5	9
9	2	10	9	12	2	3
10	10	12	7	9	1	5
11	12	7	3	8	10	7
WORST 12	7	2	4	10	12	10

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.1

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.1					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	2	1	3	2	3
2	3	2	2	4	3	3
3	3	2	1	3	1	3
4	2	3	1	3	2	1
5	2	0	2	3	2	2
6	4	3	3	4	3	3
7	2	3	3	4	2	2
8	3	1	3	4	2	2
9	3	1	1	2	1	2
10	3	1	2	3	0	2
11	2	1	1	3	0	2
12	3	3	1	3	3	3
13	4	4	3	4	3	4
14	2	2	3	4	2	2
15	4	2	3	4	2	2
16	3	3	2	3	3	3
17	3	3	3	3	2	3
18	4	3	3	4	2	3
19	3	2	2	3	3	2

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.2

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.2					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	1	1	1	2	3	2
2	2	2	1	1	3	1
3	1	2	1	1	1	1
4	2	3	1	1	2	1
5	1	1	1	1	2	1
6	3	1	3	2	3	2
7	2	3	2	1	1	1
8	2	1	2	1	2	2
9	1	1	1	1	1	1
10	1	1	2	2	0	2
11	1	0	1	2	0	2
12	2	3	2	1	0	2
13	3	2	2	3	3	3
14	2	2	1	2	1	2
15	1	2	2	2	2	1
16	2	1	1	3	0	3
17	2	3	2	3	2	2
18	3	3	2	3	2	3
19	1	2	1	2	3	2

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.3

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.3					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	2	3	3	4	3	3
2	2	3	3	4	3	3
3	1	2	1	3	0	2
4	1	3	2	4	3	2
5	1	3	1	3	0	3
6	3	3	3	4	3	3
7	2	2	3	4	3	2
8	2	2	2	3	1	3
9	1	3	1	3	1	2
10	1	2	1	3	0	2
11	2	1	1	3	0	3
12	1	3	2	3	2	3
13	3	3	3	4	3	4
14	2	2	1	3	1	3
15	1	3	2	4	1	3
16	2	3	2	4	2	3
17	2	2	3	3	3	3
18	2	2	3	4	2	2
19	1	2	2	3	2	2

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.4

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.4					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	3	3	3	3	1
2	4	3	3	3	4	2
3	4	3	2	1	3	2
4	3	3	3	2	3	1
5	3	3	1	1	3	2
6	4	3	3	3	3	3
7	4	2	3	1	2	2
8	3	4	2	3	0	3
9	4	3	1	2	1	1
10	3	3	2	2	2	2
11	3	2	3	2	0	1
12	3	3	3	3	3	2
13	4	4	3	3	3	3
14	4	3	2	2	2	2
15	4	3	3	3	2	2
16	3	3	3	2	3	3
17	4	2	2	1	2	2
18	3	2	2	1	2	2
19	4	3	3	2	2	2

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.5

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.5					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	4	3	2	3	3
2	4	4	3	2	4	3
3	3	3	2	2	1	4
4	3	3	3	1	3	3
5	3	3	1	1	3	3
6	4	4	3	2	3	4
7	4	4	3	3	3	4
8	3	4	2	2	2	3
9	2	3	1	1	3	4
10	2	3	2	2	0	3
11	0	3	1	2	0	3
12	2	3	3	2	3	3
13	4	4	3	3	3	4
14	3	3	2	2	3	4
15	3	4	2	2	2	3
16	3	4	3	2	0	4
17	4	3	3	3	3	4
18	4	3	3	3	1	3
19	3	3	2	2	4	3

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.6

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.6					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	3	3	4	3	2	3
2	3	3	4	3	3	3
3	2	3	4	2	0	3
4	3	2	4	3	2	2
5	2	3	3	3	0	2
6	4	3	3	3	3	3
7	2	2	3	2	3	2
8	3	3	3	2	1	3
9	2	2	3	2	1	2
10	2	2	3	2	1	3
11	1	0	3	2	0	3
12	3	3	4	3	1	3
13	4	3	3	3	2	3
14	3	2	3	2	1	3
15	3	3	3	3	2	3
16	3	3	3	3	0	4
17	4	3	3	3	2	2
18	3	3	3	3	1	3
19	3	2	3	3	2	3

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.7

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.7					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	1	2	1	3	2	2
2	2	2	2	3	1	2
3	1	3	3	2	0	2
4	1	3	3	2	2	2
5	1	2	0	2	0	2
6	2	3	3	3	3	2
7	1	2	3	3	1	1
8	2	1	2	2	0	2
9	1	2	0	2	0	2
10	1	2	2	2	0	2
11	0	1	2	2	0	2
12	1	2	3	2	1	2
13	2	3	3	3	2	2
14	1	2	3	2	1	2
15	2	3	3	3	1	2
16	1	1	2	2	0	1
17	2	1	3	2	2	2
18	2	2	3	3	1	2
19	1	1	1	2	2	2

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.8

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.8					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	4	3	4	3	3
2	4	4	3	4	3	4
3	4	3	3	4	3	4
4	4	3	3	4	3	4
5	4	3	3	3	0	3
6	4	4	3	4	3	4
7	4	3	3	4	2	4
8	4	4	3	4	0	3
9	4	3	2	3	2	4
10	3	3	2	4	0	3
11	3	3	2	3	0	4
12	3	4	3	4	3	3
13	4	3	3	4	3	4
14	4	3	3	3	3	4
15	4	4	3	4	3	4
16	3	3	2	3	0	3
17	4	3	3	4	3	3
18	4	3	3	4	3	3
19	4	3	3	4	3	4

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.9

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.9					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	3	3	4	2	3
2	4	3	3	4	3	4
3	3	1	1	3	0	3
4	3	3	3	3	2	3
5	4	3	1	2	3	3
6	4	2	2	3	1	3
7	3	2	2	3	1	4
8	3	3	2	3	0	3
9	4	2	1	3	0	3
10	3	2	1	3	0	3
11	3	3	1	3	0	3
12	3	3	2	3	1	4
13	4	2	2	3	2	3
14	3	2	1	2	1	3
15	4	3	2	3	1	3
16	3	3	3	3	1	3
17	3	2	2	3	1	4
18	3	2	2	3	1	3
19	4	2	2	3	1	4

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.10

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.10					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	1	3	3	2	4	2
2	1	4	3	2	4	2
3	2	3	1	1	1	1
4	1	3	3	2	3	3
5	1	4	0	2	2	2
6	2	3	3	1	3	3
7	1	3	2	1	2	2
8	1	3	2	2	0	1
9	2	3	1	2	0	2
10	1	3	2	1	0	2
11	2	2	2	2	0	2
12	1	3	3	2	1	3
13	2	3	3	2	3	3
14	1	2	2	2	1	2
15	2	4	2	1	1	1
16	1	3	3	2	0	3
17	1	3	2	1	1	2
18	1	3	2	2	1	2
19	1	3	2	2	1	3

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.11

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.11					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	4	3	4	4	4	3
2	4	4	4	4	3	4
3	4	3	4	4	4	4
4	4	3	4	4	3	3
5	4	3	4	4	3	4
6	4	4	4	4	4	3
7	4	3	3	4	3	3
8	4	3	3	4	0	4
9	4	3	4	4	4	3
10	4	3	4	4	3	3
11	3	4	4	4	0	4
12	4	3	4	4	3	3
13	4	3	4	4	4	3
14	4	3	4	4	3	4
15	4	4	4	4	3	4
16	3	3	4	4	3	3
17	4	3	4	4	3	3
18	4	4	3	4	3	3
19	4	3	4	4	2	4

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Assessment Scores from Group 2 Assessors for Chief Engineer in Video Scenario No.12

Behavioural Marker Reference	Behavioural Tendency Score for Chief Engineer in Video Scenario No.12					
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6
1	2	4	3	2	2	3
2	1	3	3	2	2	3
3	1	4	1	1	0	3
4	1	3	3	2	3	3
5	1	3	1	1	0	3
6	2	4	3	3	3	3
7	1	3	3	1	3	3
8	2	3	1	1	0	3
9	2	3	1	1	0	3
10	1	3	1	2	1	3
11	1	3	2	1	0	2
12	1	3	3	1	1	3
13	3	4	3	3	2	3
14	1	3	1	1	1	3
15	2	4	3	1	1	4
16	2	3	3	2	1	3
17	2	3	3	1	2	4
18	2	3	3	1	1	3
19	1	3	3	1	2	3

BEHAVIOURAL MARKERS FOR THE ASSESSMENT OF COMPETENCE IN CRISIS MANAGEMENT OF MERCHANT NAVY ENGINEERING OFFICERS.

Group 3 – Assessor Data

Rank Order of Competence in Crisis Management	Chief Engineer in Video Scenario Number						
	Assessor 1	Assessor 2	Assessor 3	Assessor 4	Assessor 5	Assessor 6	Assessor 7
BEST 1	11	11	11	11	11	8	11
2	9	8	8	8	1	11	4
3	4	6	4	3	8	6	3
4	8	9	1	6	4	4	1
5	1	4	3	1	2	1	5
6	3	12	9	7	3	2	8
7	12	3	12	12	10	5	7
8	6	10	2	9	9	3	12
9	2	1	5	5	12	9	6
10	5	5	6	10	7	12	10
11	10	7	7	4	5	7	9
12 WORST	7	2	10	2	6	10	2

APPENDIX 4

Publications

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

Barnett, M., Gatfield, D. and J. Habberley. 2002. Shipboard Crisis Management: A Case Study. In: *Proceedings of Human Factors in Ship Design and Operation Conference.* London: Royal Institution of Naval Architects.

Barnett, M., Gatfield, D. and C.Pekcan. 2003. A Research Agenda in Maritime Crew Resource Management. In: *Proceedings of International Conference on Team Resource Management in the 21st Century.* Daytona Beach, Florida: Embry-Riddle Aeronautical University.

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

Barnett, M., Pekcan, C. and D. Gatfield. 2004. Recent Developments in Crew Resource Management (CRM) and Crisis Management Training. In: *Proceedings of 2nd LSM Manning and Training in China Conference.* Shanghai: Lloyd's List Events.

Pekcan, C., Gatfield, D. and M. Barnett. 2005. Content and Context: Understanding the Complexities of Human Behaviour in Ship Operation. In: *Proceedings of Human Factors in Ship Operations Conference.* London: Royal Institution of Naval Architects.

Gatfield, D. 2005. Using simulation to determine a framework for the objective assessment of competence in maritime crisis management. In: *Proceedings of the Society for the Advancement of Games and Simulations in Education and Training 2005 Conference.* Portsmouth: University of Portsmouth.

Gatfield, D, Pekcan, C. and Barnett, M. 2006. The Janus Principle in Maritime Safety: Looking Backwards to Look Forward. *In: Proceedings of Learning from Marine Incidents 3 Conference.* London: Royal Institution of Naval Architects.

Barnett, M.L., Gatfield, D.I., & Pekcan, C.H. 2006. Non-technical skills: the vital ingredient in world maritime technology? *In: Proceedings of the International Conference on World Maritime Technology*, London: Institute of Marine Engineering, Science and Technology.