CAA PAPER 98005

•

BEHAVIOURAL MARKERS FOR CREW RESOURCE MANAGEMENT

CIVIL AVIATION AUTHORITY, LONDON, PRICE £7.00

CAA PAPER 98005

BEHAVIOURAL MARKERS FOR CREW RESOURCE MANAGEMENT

Rhona Flin Lynne Martin

•

•

•

.

REPORT PREPARED BY DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF ABERDEEN AND PUBLISHED BY CIVIL AVIATION AUTHORITY, LONDON, JULY 1998 © Civil Aviation Authority 1998

•

•

•

•

•

•

•

•••••

•

•

•

•

•

ISBN 0 86039 735 1

Printed and distributed by Westward Digital Limited, 37 Windsor Street, Cheltenham, England

Executive Summary

This study was commissioned in January 1997, by the Safety Regulation Group (SRG) of the UK Civil Aviation Authority (CAA), with the aim of examining current practice in the development and use of behavioural markers for training and assessing non-technical or Crew Resource Management (CRM) skills both in the UK and abroad. The project objectives were:

- 1. To establish how CRM behavioural marker systems are used in practice for CRM training and skills assessment by major airlines in the UK, mainland Europe and the USA.
- 2. To compare the principal CRM behavioural marker systems in current use.
- 3. To determine the empirical basis for CRM marker systems.
- 4. To recommend the adoption of an existing CRM behavioural marker system or to produce an outline for a new marker system based on observed good practice.

1 How CRM behavioural marker systems are used in practice

European aviation requirements for flight operations and flight crew licensing (JAR-FCL; JAR-OPS) are introducing certain requirements for the training and assessment of pilots' non-technical skills.

A survey of 11 UK airlines showed that only 5 of them had developed a CRM behavioural markers list and none of these were used for formal CRM assessment. Instead they had been designed for CRM training and to structure LOFT feedback sessions. These airlines were aware that they might be required to evaluate CRM skills more formally at some future date. Respondents' views on the introduction of behavioural markers systems reflected concern regarding training of instructors and examiners, and the need for a reliable and valid marker system before formal CRM assessment could be introduced.

A survey of 14 non-UK airlines using behavioural marker systems indicated that these systems were used for CRM training and assessment in most cases, but only in 6 airlines were they used for a formal examination of CRM skills. Again respondents emphasised the importance of training those evaluating CRM skills and the need to produce a simple and usable marker system.

2 Comparing the principal CRM behavioural marker systems

An analysis of current marker systems showed considerable differences between the categories and elements listed, although these seem to be more apparent than real. That is, although there are considerable differences in terminology, the concepts are essentially the same. Examples of 31 checklists from the two surveys and from research teams working on CRM skills were compared to produce a table showing the most common categories and elements of CRM skills. There is a core set of CRM elements represented under categories such as team work, communication, leadership, workload management, decision making, and situation awareness. These concepts are used fairly consistently but labelling differs across research studies, airlines and fleets. There is no standard method of rating employed, formats ranged from checklists to ten point scales.

3 The empirical basis for CRM marker systems.

0

.

A literature review revealed that only a limited amount of research had been conducted into the development of behavioural marker systems for CRM skills training and assessment. Most of this

focused on the rating of crews rather than individual pilots. There is very little information available on the psychometric properties of rating scales being used for assessment purposes.

The available work is predominantly from the USA, and the Line/LOS Checklist (LLC) developed by Helmreich et al (1990; 1997) (or derivative versions) is the dominant marker system for rating a flight crew's CRM skills. This instrument was developed for research rather than assessment purposes, although it is now widely used for line audits. There is a developing body of research, also from the USA, relating to the Advanced Qualification Programme (AQP) where CRM skills analysis and evaluation is required.

• • • • • • •

•

•

•

•

•

•

•

•

•

•••••

Issues of rater reliability have begun to be researched and results highlight the importance of rater training to achieve calibration against established standards, to ensure fair and accurate evaluation of non-technical skills. Raters conducting licence evaluations will themselves need to be assessed and qualified.

A number of studies have considered the differences between ratings based on entire sessions compared to ratings for given stimulus events or event sets. In the case of the latter, specific behaviours can be specified and research teams have argued for the merits of this approach, although in practical terms, scenario development costs have to be taken into account. Others have suggested that it is advantageous to rate crew behaviours in relation to phase of flight.

For any airline moving to CRM skills assessment, this will involve a number of stages: identification of critical CRM skills; rating scale development and trialling, training of assessors, system evaluation. The research indicates that an organisation moving through all these stages will need to ensure pilots' familiarity with the concepts and allow the time necessary for cultural adjustment. The terminology and design of a rating scale will influence pilots', instructors' and examiners' acceptance of it as a valid and reliable basis for the assessment process.

4 Recommending a CRM marker system

It is recommended that the European NOTECHS (1997) behavioural markers framework (or some comparable version), is tested for possible use as an individual pilot's CRM skills assessment instrument. This framework was developed by a group of psychologists and pilots working together on an EC funded project to design a behavioural marker system for the evaluation of pilots' non-technical skills. It should be emphasised that the NOTECHS system has not been empirically evaluated, although this work began in January 1998 with the new EC JARTEL project.

Glossary

•

•

•

•

•

•

•

•

• • •

•

•

•

•

•

•

•

•

•

•

•

•

•

•

AIC	Aeronautical Information Circular
ALPA	Airline Pilots Association (USA)
AQP	Advanced Qualification Programme
BALPA	British Airline Pilots Association
CAA	Civil Aviation Authority (UK)
CMAQ	Cockpit Management Attitudes Questionnaire
CRM	Crew Resource Management
DLR	Deutsches Zentrum fur Luft- und Raumfarht (D)
FAA	Federal Aviation Administration (USA)
ICAO	International Civil Aviation Organisation
IFALPA	International Federation of Airline Pilots Associations
IMASSA	Institut de Medecine Aerospatiale (France)
JAA	Joint Aviation Authorities (Europe)
JAR	Joint Aviation Requirements (Europe)
JARTEL	JAR Translation and Elaboration (EC project)
JAR-OPS	Joint Aviation Requirements - Flight Operations
JAR-FCL	Joint Aviation Requirements - Flight Crew Licensing
KLM	Royal Dutch Airlines
LLC4	Line/LOS checklist version 4
LOE	Line Oriented Evaluation
LOFT	Line Oriented Flight Training
LOS	Line Oriented Simulation
MCC	Multi-pilot Co-ordination Course
NASA	National Aeronautics and Space Administration (USA)
NLR	National Aerospace Laboratory (NL)
NOTECHS	Non-Technical Skills group (EC project)
SHAPE	(KLM scale)
TADMUS	Tactical Decision Making Under Stress (US project)
UK	United Kingdom
USA	United States of America
UT	University of Texas
WMCL	workload, management, communication & leadership scale (KLM)



Contents

.

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

2

3

4

1 Introduction

1.1	Project Objectives	1
1.2	Crew Resource Management	1
	1.2.1 Development of CRM	2
	1.2.2 Evaluation of CRM	3
	1.2.3 Regulation of CRM	4
1.3	European Joint Aviation Requirements	4
1.4	NOTECHS Project (1997)	5
1.5	JARTEL Project (1998-1999)	5
1.6	Summary	5
Liter	ature Review - Assessment of CRM Skills	
2.1	Introduction	6
	2.1.1 Central Issues	6
2.2	Research into CRM skills	7
	2.2.1 NASA/UT/FAA Line/LOS Checklist (LLC)	8
	2.2.2 CRM Behaviour Marker Systems and Associated Research	9
	2.2.3 Analysing CRM Skills in AQP	11
2.3	Rating CRM Skills	11
2.4	Training Instructors and Examiners	13
2.5	Standards of Competence	13
2.6	Summary	14
Meth	od and Procedure	
3.1	Introduction	16
3.2	Sample	16
3.3	Survey Method	16
Inter	view Survey	
4.1	Introduction	17
4.2	Method	17
	4.2.1 Procedure	17
	4.2.2 Sample	17
4.3	CRM Courses and Development in UK	17
4.4	Development of Behavioural Markers Checklists	18
	4.4.1 Opinions of LLC	18
	4.4.2 Training for Behavioural Markers Checklists	19
4.5	Assessment of CRM Skills	19
	4.5.1 Retraining on Failure of Non-Technical Assessment	20
	4.5.2 Use of Behavioural Markers Data	20
4.6	Support for CRM Assessment	21
	4.6.1 The Role of the Regulator	21
	4.6.2 Management Commitment	22
4.7	Summary	22

5	Que	Questionnaire Survey, non-UK Airlines			
	5.1	Introduction	23		
	5.2	Method	23		
		5.2.1 Procedure	23		
		5.2.2 Sample	23		
	5.3	Responses to the Questionnaire	23		
		5.3.1 Details of the Behavioural Marker Systems	23		
		5.3.2 Design and Content	25		
		5.3.3 Training Instructors/Examiners	25		
		5.3.4 Use of the Behavioural Marker System	25		
		5.3.5 Pilots' Views on the Use of Behavioural Markers	27		
		5.3.6 Future Developments	28		
	5.4	Summary	28		
6	Ana	Analysis of Behavioural Markors Lists			
	6.1	Introduction	20		
	6.2	Method	29		
	0.2	6.2.1 Sample	29		
		6.2.2 Collation of Categories	29		
	63	Analysis of CRM Elements	29		
	6.4	Comparison with Established CDM Skill Lists	30		
	6.5	Format of CDM Skill List	34		
	6.6	Summary	35		
	0.0	Summary	35		
7	Disc	Discussion			
	7.1	Introduction	36		
	1.2	CRM Programmes	36		
		7.2.1 CRM Training	36		
		7.2.2 CRM Assessment	36		
	7.3	Behavioural Markers	37		
		7.3.1 Review of Existing Marker Systems	37		
		7.3.2 Designing a Marker System	38		
	7.4	Identifying a Suitable CRM Marker System	39		
	7.5	Training for CRM Assessment	40		
	7.6	Guidance from Regulator	40		
	7.7	Summary	40		
8	Cone	clusions	41		
9	Reco	ommendations	42		
	Refe	rences	43		
	A also		45		
	ACKI	nowledgements	46		
	App	Appendices:			
	1	(Perminted with normination of NOTECHE	47		
	2	NASA (IT Lim (I OS Charlie I Ct d			
	2	2 NASA/UT Line/ LOS Checklist LLC4.4			
	2	(Reprinted with permission of Professor Helmreich)			
	5	The interview schedule	55		
	4	List of control in the	57		
	5	List of contributors	63		
	0	Example behaviours for elements of cognitive categories	65		

Chapter 1 Introduction

1.1 **PROJECT OBJECTIVES**

This study was commissioned in January 1997, by the Safety Regulation Group of the UK Civil Aviation Authority (CAA). The aim was to examine current practice in the development and use of behavioural markers for training and assessing Crew Resource Management (CRM) skills in both the UK and abroad. The project objectives were:

- To establish how CRM behavioural marker systems are used in practice for CRM training and skills assessment by major airlines in the UK, mainland Europe and the USA.
- To compare the principal CRM behavioural marker systems in current use.
- To determine the empirical basis for CRM marker systems.
- To recommend the adoption of an existing CRM behavioural marker system or to produce an outline for a new marker system based on observed good practice.

The report is set out in nine chapters. In this opening chapter, the background to the project is outlined in terms of the development and current status of CRM training, and the European position relating to the assessment of CRM skills. The relationship of this project to the European Commission (DGVII) projects on the assessment of pilots' non-technical skills (NOTECHS and JARTEL) is also explained. Chapter 2 reviews current research related to CRM behaviour marker systems. Chapter 3 outlines the methods used for the two surveys. Chapter 4 presents the principal themes from the interview survey on the use of behavioural marker systems for CRM in a sample of UK airlines. Chapter 5 describes the findings from a questionnaire survey of the use of behavioural markers in a sample of international airlines based abroad. In Chapter 6, a synthesis of behavioural marker schemes is reported. Chapter 7 discusses these findings, with conclusions presented in Chapter 8, and recommendations in Chapter 9.

1.2 CREW RESOURCE MANAGEMENT

The concept of Crew Resource Management (originally Cockpit Resource Management) originated twenty years ago and was defined by aviation psychologist John Lauber (1984, p. 20) as, 'using all the available resources – information, equipment, and people – to achieve safe and efficient flight operations.' By the late 1970s, many of the international airlines had become concerned with the human factors aspects of their flight operations and had introduced training in flight crew co-operation. The impetus for the increased interest in non-technical skills was a growing awareness that a significant percentage of aircraft accidents were attributable to flight crew failures, rather than technical problems (Boeing, 1994; NTSB, 1991, 1994).¹

¹ It is acknowledged that the pilots act as the last line of defence in the aviation safety management system and that while the immediate causes of accidents may be attributable to the human operator's errors, the root causes or conditions for accidents are also likely to implicate managers and designers located elsewhere in the organisation, see Reason (1997).

Concern on both sides of the Atlantic had been heightened as a result of several specific accidents which had a clear human factors cause, such as the 1977 Tenerife runway collision between two Boeing-747 aircraft. Moreover, psychological research using accident analysis, pilot interviews, flightdeck voice recordings and simulator observations had confirmed the need for non-technical skills (CRM) training which would focus on pilots' leadership, decision making, situation awareness, communication and team work (Green et al, 1996; Hawkins & Orlady, 1993; Orasanu, 1994; Wiener, Kanki, Helmreich, 1993; Wiener & Nagel, 1988).

•

•

.

•

•

•

•

•

•

•

•

•

•

•

• • • •

1.2.1 Development of CRM

Against this background of accident analysis and research, CRM has been crafted. Helmreich (1996) charts its development as a progression through four stages to date and predicts that it will enter a fifth. While his analysis is primarily based on airlines in the USA, it essentially mirrors the history of CRM in Europe and beyond. He describes the first generation of CRM courses in the early 1980s as intensive seminars which had a strong similarity to business management training courses and aimed to enhance interpersonal skills. There was an emphasis on personality and many of the exercises used were not specifically designed for pilots or flightdeck situations. Hence, some pilots rejected these courses as irrelevant 'charm school' training. The use of simulators for training and practising CRM in Line Oriented Flight Training (LOFT) began to be introduced. Second generation CRM in the late 1980s focused more on the 'crew' and the importance of team skills. General acceptance increased although training material not tailored to the pilots' concerns was (probably justifiably) rejected as 'psychobabble'. By the early 1990s, a third generation of CRM began to appear, addressing emerging safety issues such as company culture and flight deck automation. More advanced human factors training was now being provided for instructors and examiners, and the scope of CRM programmes was being widened by many airlines to include other teams such as ground staff and cabin crews.

The current, fourth generation of CRM is behaviour oriented and focuses on integration of CRM skills in technical training. This is principally a result of the introduction of the voluntary Advanced Qualification Programme (AQP) in the USA which allows airlines to develop more innovative and flexible aircrew training tailored to their operational requirements (see Birnbach & Longridge, 1993). Companies who have shifted onto an AQP are required to undertake very detailed task analysis for each aircraft type specifically addressing CRM skills, as well as psychomotor skills, and to establish proficiency objectives which serve as the foundation for training and assessment (Lanzano, Seamster & Edens, 1997). Evaluation of crews' performance is undertaken using Line Oriented Evaluation (LOE) in full mission simulation. Thus airlines with an AQP have developed core lists of CRM knowledge and skills and may include key CRM behaviours on their flightdeck checklists (Helmreich, 1996).

There is some concern that whilst CRM training courses have improved during this progression, with the tailoring of courses to specific needs within airlines, the original safety aims of CRM training may have been lost along the way. Helmreich (1996) advocates that the next (fifth) generation of CRM should return to its 'roots' and concentrate on the management of error within the cockpit. 'Underlying the fifth generation of CRM is the premise that human error is ubiquitous and inevitable – and a valuable source of information. If error is inevitable, CRM can be seen as a set of error countermeasures with three lines of defence.'(p.7). The defences are avoiding errors, trapping errors before they become consequential and mitigating the consequences of errors that do occur. As Reason (1997) points out, it is not possible to remove all error from a system as complex as an aircraft and therefore the focus of study should be on the detection and correction of CRM may have changed over twenty years, the essential issues it addresses have not.

CRM is now used almost universally by the major international carriers, typically taking the form of initial two or three day training courses covering topics such as situation awareness, decision making, stress, leadership, communication (CAA, 1993; ICAO, 1989). Teaching methods include lectures, practical exercises, case studies, films and personal assessment questionnaires. Continuation or refresher training is also advised; usually these are half or full day courses focusing on a particular CRM topic. In addition, ongoing training and monitoring of CRM skills normally takes place during LOS (line-oriented simulation), such as LOE and LOFT. According to Wiener et al (1993, p.xxi), 'LOFT provides the organisation with a means of creating conditions requiring the practice of effective crew co-operation to resolve complex emergency situations. It is also the instrument for reinforcing and evaluating the concepts learned in the CRM classroom.' What is critical for effective LOFT is the debriefing skills of the trainer who can enable the crew to evaluate and improve their CRM and technical skills (see Dismukes, Jobe & McDonnell, 1997; McDonnell, Jobe & Dismukes, 1997; Wilhelm, 1991). For a full account of the development of CRM and associated research see Jensen (1995) and Wiener et al (1993). Details of CRM programmes in companies such as KLM, Lufthansa and Qantas can be found in Johnston et al (1995) and McDonald et al (1995) or for helicopter pilots in David (1996). Good descriptions of research and current practice are given in the collected papers from the International Association of Aviation Psychologists' biennial conferences.

1.2.2 Evaluation of CRM

Early assessment of the value and impact of CRM consisted of standard training evaluation techniques based on measures of pilots' opinions on the quality and relevance of the programmes (Helmreich & Foushee, 1993; Mashcke et al., 1995; Naef, 1995). These have generally been favourable, especially with the increased use of aviation-specific course materials. This is important but it is not a robust test of the effectiveness of CRM which was introduced with the aim of improving flight safety and minimising accident rates. The required proof that aviation accidents and incidents have been reduced as a result of CRM remains elusive, as accident rates are very low, making changes related to CRM difficult to detect. Moreover, many other factors also affect aviation safety (Gregorich & Wilhelm, 1993). While these outcome data may be available in the future, evaluation efforts are now targeted on the practice of CRM skills. As CRM is intended to train skills shown to improve flight safety, then these can be assessed to guage the effectiveness of CRM and LOFT.

Pilots' attitudes to factors affecting safety performance (such as personal limitations, command, and team working) have been measured before and after CRM training using the Cockpit Management Attitudes Questionnaire (CMAQ) (Gregorich, Helmreich & Wilhelm, 1990). Results were generally encouraging with more positive attitudes reported following CRM training (Gregorich & Wilhelm, 1993). Whilst a change in behaviour is likely to be linked to a change in attitude, a change in attitude does not necessarily lead to a change in behaviour (Azjen & Fishbein, 1980), and the latter is of prime importance. There is some evidence that crews who had completed CRM training showed better CRM knowledge, skills and attitudes when observed in a LOFT session than crews who had not received CRM training (Helmreich, Wilhelm, Gregorich & Chidester, 1990). However, these were aggregate data rather than individual pilot or crew scores. In fact what was required was a psychological measure for assessing crew's or pilot's team skills which would permit an assessment of whether CRM training was actually transferring to the flightdeck. The need for such research has been driven by the introduction of the AQP system in the USA and forthcoming European legislation (see below). In response, current CRM research efforts are focused on techniques for identifying and assessing core CRM skills which can be observed during line or simulated flight operations. This literature is reviewed in the next chapter.

1.2.3 Regulation of CRM

The Federal Aviation Administration (FAA) in the USA has promoted the development of CRM training, although this is not mandatory (FAA, 1991). In the Advanced Qualification Programme (AQP) being adopted by most of the major US airlines, carriers have to undertake a full skills analysis, provide CRM and LOFT for all flight crews and undertake evaluation of CRM skills (Birnbach & Longridge, 1993; Lanzano et al, 1997).

•

•••••

•

.

•

•

•

•

•

•

•••••

In the UK, human factors training and examination are required for a Flight Crew Licence, and CRM training has been a mandatory requirement for commercial pilots since January 1995. Guidance from CAA on the subject of CRM training was issued in the form of two Aeronautical Information Circulars (AIC) in 1993 and 1995. As of April 1995, all UK flightdeck crew have been trained in CRM to an initial awareness level. Although the CAA approves all CRM courses, it does not set the syllabus, giving airlines the freedom to develop their own training. This has permitted airlines to develop courses to suit their organisation and within their own means but has also resulted in distinct differences between courses. Individual operators have designed their own CRM programmes either using in-house expertise or by purchasing programmes from larger airlines and consultants. There is no requirement to formally assess CRM skills, and any evaluation and feedback during training and checking is currently done on a voluntary basis by the airline concerned.

Nevertheless the assessment of non-technical skills has been part of an aircraft commander's assessment since the early days of flying. However, only in certain organisations, such as the military, has there been a formal marking system for non-technical skills such as leadership. It is apparent from the disquiet shown by pilots' unions and associations that the current assessment of non-technical skills can be somewhat arbitrary. This indicates a need for a more formal and objective system involving universally agreed markers to ensure an equitable system.

The approach of the mainland European regulatory authorities has been similar to the UK, with CRM training being supported but not, as yet, prescribed. However, forthcoming European Joint Aviation Requirements (JAR) will introduce a requirement for the introduction of CRM training and for more formal evaluation of CRM skills in multi-crew operations.

1.3 EUROPEAN JOINT AVIATION REQUIREMENTS

In Europe, the Joint Aviation Authorities (JAA) based in Hoofddorp, The Netherlands, produce Joint Aviation Requirements (JAR). In relation to CRM training and its evaluation, the proposed rules can be found in two sets of civil aviation requirements – Flight Operations (JAR-OPS) and Flight Crew Licensing (JAR-FCL) which are due to be implemented in 1998 and 1999 respectively.

During 1997 there has been some debate in International Federation of Airline Pilots' Associations (IFALPA) over the interpretation of these JAA requirements and whether they mean that CRM skills must be assessed for licence re-validation (see for instance, *Flight International*, 1997, 152, 4599, p. 30). The CAA understands that the JAR-FCL and JAR-OPS requirements will require the assessment of CRM (non-technical) skills for certain examinations.

1.4 NOTECHS PROJECT (1997)

In January 1997, a European project called 'The Evaluation of Non-Technical Skills of Multi-Pilot Aircrew in Relation to the JAR-FCL Requirements' (NOTECHS) was initiated by the JAA Research Committee Human Factors Project Advisory Group on Human Factors to provide background information for the JAR-FCL in relation to the evaluation of a pilot's non-technical skills. Non-technical skills were defined as pilots' attitudes and behaviours in the cockpit not directly related to aircraft control, system management, technical consequences and standard operating procedures. The goal of the project was to develop a methodology for assessing pilots' non-technical skills during flight and simulator checks. The scope of the project related principally to JAR-FCL, part 1, Subpart F, paragraph 240 (see above) as well as relevant sections of JAR-OPS. The project aim was to provide background material to support development of :

- 1. Interpretative and Explanatory Material for JAR-FCL 1.240, and
- 2. Acceptable Means of Compliance for JAR-FCL 1.240

The NOTECHS project was undertaken by a consortium of psychologists from four partner teams – NLR (Netherlands), DLR (Germany), IMASSA (France) and University of Aberdeen (UK), as well as pilots from KLM, associated with the NLR group. In essence, the project objectives were to review the use of non-technical skills marker systems and to either provide a preliminary endorsement of one particular system or to develop a draft non-technical marker system based on existing systems and previous research. Due to our membership of the NOTECHS project, part of the survey research from this University of Aberdeen CAA project and materials collected by the other members of the NOTECHS group were exchanged on a reciprocal basis. On completion of the review of existing systems and previous research, the NOTECHS group decided to produce a new draft standard for the assessment of individual pilot's non-technical skills (see NOTECHS, 1997 for details). The resulting framework was developed on a theoretical basis against a number of principles, such as 'no evaluation of personality traits'. This framework is shown in Appendix 1 and its constituent elements are compared with other systems in Chapter 6.

1.5 **JARTEL PROJECT (1998-1999)**

The draft non-technical skills standard (behavioural marker system) (NOTECHS, 1997) will be evaluated on a cross-cultural basis in a new phase of research sponsored by EC DGVII. This project which began in 1998, involves the NOTECHS consortium plus additional research organisations and two airlines, British Airways and Alitalia. The project is called JARTEL (Joint Aviation Requirements Translation and Elaboration) and the group will report on an empirical study of the use of the NOTECHS (1997) marker system by 1999.

1.6 SUMMARY

This opening chapter outlined the background to the project, which is designed to examine the use of behavioural marker systems for CRM. The relevant European aviation requirements covering CRM training and assessment (JAR-OPS; JAR-FCL) were presented, as well as details of current European projects (NOTECHS; JARTEL) which are developing methods for the assessment of pilots' non-technical skills.

Chapter 2 Literature Review – Assessment of CRM Skills

2.1 INTRODUCTION

The background to CRM was briefly described in the previous chapter. A general consensus appears to prevail regarding the core categories or elements of CRM skills, which typically include situational awareness, leadership, decision making, team working, communication, error management, and awareness of personal limiting factors, such as stress and fatigue (CAA, 1993, 1995; FAA, 1989, 1991; ICAO, 1989; Wiener et al, 1993). However, labelling and inclusion of even the main CRM categories varies across research teams, regulators and airlines. Brannick and Prince (1997) point out that while there are clear similarities in taxonomies of team process factors, there remains a need to resolve the finer differences which are apparent when the categories are scrutinised more carefully. This becomes even more apparent at the level of specifying critical behaviours for those categories, as demonstrated in Chapter 6 below.

•

•

•

•

•

•

•

•

•

•

•

•

•••••

•

There is a fairly large body of research into the main elements of CRM, such as pilot communication (Connelly, 1997), leadership (Pettitt & Dunlap, 1997), situational awareness (Endsley, 1995), decision making (Orasanu, 1997) and team work (Brannick, Salas & Prince, 1997). This work is not necessarily conducted under the umbrella of CRM but is part of the wider human factors research in aviation underpinning CRM development. To give an example, Jentsch, Sellin-Wolters, Bowers and Salas (1995) used a PC-based simulation with military aircrew and found certain crew co-ordination behaviours distinguished between crews who quickly and correctly identified a typical problem and those teams that required more time. The key behaviours they identified were standard communications, leadership statements and situation awareness observations. They commented on the value of identifying specific behaviours related to desired task performance for training aircrew co-operation, in that instructors will be able to train these specific behaviours rather than just awareness of the relevant CRM skills.

This literature review does not cover the basic aviation human factors and CRM research, but is instead concerned with the behavioural markers for CRM skills and their application for assessment. It concludes with a brief discussion of the use of competence-based systems for assessing job performance.

2.1.1 Central Issues

Houle (1995) of the American Airline Pilots' Association, reviewing current developments in CRM, advocated that CRM evaluation should develop from its objective of a safer flying environment. At a practical level this translates into an assessment of CRM skills that can be defined, observed and measured. Once skills are identified, a training programme can be constructed around these and CRM behaviours can be evaluated as long as some ground rules are observed. For instance, only those behavioural markers that can be translated into a specific skill which can be taught and used on the flightdeck should be evaluated. Houle felt developers had become overloaded trying to attach an evaluation to every manoeuvre in the Line Oriented Simulation (LOS). This method was time consuming for grading and allowed too much subjectivity. He pointed out that check pilots dislike systems like this because they did not quantitatively distinguish between high and low performing crews. Houle favoured using observable behaviours because they quantify the grading system in an objective manner, removing much subjectivity in evaluation.

The term 'behavioural markers' refers to a prescribed set of behaviours which have been identified as indicative of some aspect of skilled human performance. The typical behaviours or 'markers' are listed in relation to the component skills and are then used for selection, training and competence assessment. In the case of CRM, the focus is on the pilot's non-technical skills rather than the technical skills required to operate the aircraft. While there is considerable interest in the UK aviation community in the identification and assessment of critical CRM skills, little British research has been published on this topic. There are reports of CRM behaviour marker systems in European airlines, such as KLM (Antersijn & Verhoef, 1995) and there is European work in progress, for example the NOTECHS and JARTEL projects mentioned above. The bulk of available research appears to have been undertaken in the USA, funded primarily by the FAA Office of the Chief Scientific and Technical Advisor for Human Factors, who produce regular, informative resumes of their sponsored projects (Edens, 1996;1997). The following review draws mainly on reports from investigations conducted under this programme.

In any discussion on the assessment of CRM skills, a number of salient issues arise:

- 1. What is the unit of analysis? Is it the crew or the individual pilot? CRM is about crew cooperation but airline crews are not fixed units and pilots for multi-crew operations are required to possess 'portable' team skills.
- 2. What are these skills? Is it possible to develop a generic set of CRM categories and associated behaviours or should these be type-specific? Moreover, for flightdeck training and assessment purposes, these behaviours must be observable by an instructor or examiner.
- 3. Having defined the unit of analysis and the observable CRM behaviours, how are they to be assessed? On a pass/ fail; competent/ not yet competent; present/absent basis, or on some more finely grained rating scale which measures frequency or levels of performance? If the latter, how many scale points should there be 3,4,5,6 etc? Should there be behaviours stipulated at each point or just for the end points? These design features will influence the quality of measurement obtained. Furthermore, for this type of psychological performance assessment, it is necessary to ensure that the measurement instrument is both valid and reliable. Validity is the extent to which the scale measures what it purports to measure, and reliability is a measure of consistency or stability, within and between raters. (For a more detailed discussion of these psychometric properties, see Aiken, 1996; Dickinson and McIntyre, 1997).
- 4. Who will do the assessment and how will they be trained and assessed?
- 5. How will the assessment be carried out on the line or in a simulator? As part of a technical assessment or as a separate exercise?
- 6. How will the stimulus events be designed to elicit the required behaviours? And, will the rating be made on the basis of the overall session (flight), for phases of flight or for each scenario/event?

2.2 RESEARCH INTO CRM SKILLS

Some of these questions have been addressed in psychological studies conducted in the last ten years. An intensive programme of research on flight crew co-ordination behaviours has been undertaken by Helmreich and his colleagues at the University of Texas, Austin. A summary of their work is presented, followed by a number of other approaches to CRM skill assessment. Then the issues of rater reliability and training are considered.

2.2.1 NASA/ UT/ FAA Line/LOS Checklist (LLC)

The seminal research on behavioural markers comes from the group of aviation psychologists led by Professor Helmreich at the University of Texas/ NASA/ FAA Aerospace Crew Research Project in Austin. They have many years experience studying crew behaviour in flight operations and in the late 1980s developed a data collection form called the 'LINE/LOS CHECKLIST' (LLC) to collect information on flightcrews' CRM performance (Helmreich, Wilhelm, Kello, Taggart & Butler, 1990). This checklist is very widely cited in the research literature and it has been used as the basis of many airlines' CRM behaviour marker lists (see Chapter 5). The behaviours included on the LLC have their origin in the analysis of accidents and incidents with identifiable human factors causation (e.g. Connelly, 1997) as well as supportive evidence from psychological research. The LLC has been refined over a number of years into a diagnostic tool which is used during in-flight observations to evaluate non-technical (CRM) skills in human factors line-audits carried out for major airlines (Helmreich, Hines & Wilhelm, 1996; Taggart, 1995). The current (Helmreich et al, 1997) version, LLC4.4 (see Appendix 2), elicits ratings under six categories of behaviours for four phases of flight. The categories are: Team Management and Crew Communications; Situational Awareness and Decision Making; Automation Management; Special Situations; Technical Proficiency; and Overall Observations. This gives a total of 28 behavioural marker elements and two overall evaluation measures, all of which are rated on a four point scale (poor, minimum expectations, standard, outstanding). It should be noted that the LLC is used to evaluate the crew's performance, rather than that of an individual pilot. (There is a separate section of the form where comments on a particular flight crew member can be recorded).

•

•

•

•

.

•

••••

.

Using an earlier version of the LLC, Helmreich, Wilhem, Gregorich and Chidester (1990) found high degrees of variation in CRM performance ratings for crews flying different types of aircraft within the same airline. Across the two airlines observed, different behaviours were linked through their ratings to superior performance. In airline one – inquiry, technical skills, advocacy and decision making were correlated with ratings of above average performance; whilst for airline two – superior performance was associated with briefings and concern for the group. The researchers could not establish from their data whether these differences were due to true organisational differences or to different emphasis on particular aspects of the CRM courses. Butler (1991) used the LLC and compared four US airlines through 108 observations on overall technical efficiency and overall crew effectiveness and found a wide range of performance as well as significant differences between airlines. He also reported that during training to use the LLC, trainees' evaluations of the same crew's performance could vary widely. Such findings led Helmreich et al to conclude that standardisation across raters was vital before the validity of CRM assessment could be properly gauged.

There is also a problem of the standardisation of scenarios or events during which behaviours are observed (Wilhelm, 1991; Law & Wilhelm, 1995), namely, that there are large differences between airlines in the CRM behaviour patterns they advocate. This was attributed to different emphasis being placed on the CRM concepts across airlines. Law and Wilhelm (1995) suggest that crew behaviour changes during a flight and that data should be collected for each phase of flight. They used the LLC4 to collect 1495 instructor observations from two airlines. Some elements were rated across all phases of flight whilst others were rated only during certain phases or were rarely rated. The absolute number of ratings also varied with the phase of flight. They found specific crew behaviours were differentially related to crew effectiveness at varying phases of the flight. Significant differences were found between the ratings in the two airlines on 19 of the elements and also between fleets within the same company. Law and Wilhelm conclude that the LLC4 is sensitive enough to detect reliable teamwork and performance differences between and within different organisational settings.

2.2.2 CRM Behaviour Marker Systems and Associated Research

Besides the LLC, a number of other behavioural marker systems for assessing crew performance have been developed. Fowlkes, Lane, Salas, Franz and Osler (1994) produced a team performance measurement approach called TARGETs (Targeted Acceptable Responses to Generated Events or Tasks) for US military cargo helicopter teams. This was based on a set of critical aircrew co-operation behaviours, grouped into seven basic skill areas: Mission Analysis; Adaptability/ flexibility; Leadership; Decision making; Assertiveness; Situational awareness; and Communication. In this system, for each stimulus event in a scenario, there is a predefined set of acceptable behaviours (task responses, i.e. the TARGETs). Each is rated as present or absent. They tested the TARGETs approach in a training and evaluation study of six military aircrews and found the measure to have sensitivity (discrimination between crews) and an acceptable degree of rater reliability.

To collect material from which to specify behavioural markers, Seamster et al (1995b) recommend collecting instructor remarks collected from proficiency checks; asking instructors where crews have the most problems during training; and fleet specific task analyses conducted under AQP. Kaempf and Klinger (1992) advise caution when using general task analyses because they may not cover all performance elements and should therefore only be used as a starting point. They used a supplemental form of interview known as the Critical Decision Method which can elicit additional behaviours in the context of specified conditions.

With the aim of designing a prototype expert system for CRM assessment, Seamster and Edens (1993) analysed how experienced instructors identify and categorise crew behaviours for CRM assessment. They asked six instructors to sort 60 LOFT concepts (based on work by Helmreich et al, 1990), identifying their reasons for grouping them together. The result was three clusters of assessment concepts: Two related to CRM assessment- cognitive (problem identification, task prioritisation and workload management) and interpersonal (teamwork, communication, group climate and leadership/followership) and one to technical assessment (procedures, technical skills, system knowledge and manoeuvres). Seamster and Edens suggested that this framework has applications not only for a possible expert system but also for the training of CRM assessors. 'One of the most difficult aspects in becoming proficient in CRM assessment is not in learning the individual elements, but in compiling those elements into a meaningful hierarchy so that their relationship is understandable as well as usable.' (p126).

In the second phase of this project, Seamster, Edens, McDougall & Hamman (1994) used observable behaviours associated with crew problems in Proficiency Checks and First Look sessions. When 703 instructor remarks on CRM were categorised using slightly different labels, they showed that the four cognitive categories (situation awareness, workload management, planning and decision making) made up a substantially greater percentage of crew problems (68%) than the four interpersonal categories (crew co-ordination, communications, leadership/followership, group climate). From these results, they argued 'that in both scenario development and scenario evaluation, there should not be an evenly distributed emphasis on CRM categories.' (p3). With the help of eight instructors and captains, they were able to link subsets of observable crew behaviours to scenario event sets for transition or qualification training on the Boeing-737-300. This exercise showed that 'when scenario event sets are specified and listed with likely crew behaviors, experienced pilots with some familiarity with the LOE concept can show substantial agreement on the primary observable behaviors to properly assess the related tasks. Therefore, it is likely that making CRM assessments based on observable behaviors will produce reliable assessments.' (p10).

In a subsequent paper, Seamster, Hamman and Edens (1995) emphasised the importance of the instructor/evaluator's role in identifying the observable behaviours related to CRM and how these behaviours can be represented in a check tool. On this basis, they advocated, (as in the TARGETs system) the use of 'event sets' within LOFT/LOE, for which specific behavioural markers are written. These can then be carried forward to assessments to increase inter-rater reliability. Identification of observable behaviours can be better focused if a clearly definable unit of action or time is specified and used to delimit the observable crew behaviours. They give an example of an LOE worksheet showing different event sets and specific CRM behaviours for each set which are rated on a 4-point scale (unacceptable, minimally acceptable, standard, above standard). The scenario event set allows instructors/evaluators to focus on particular CRM categories at given times in the session depending on the phase of flight and the objectives of the event set. This may reduce instructor workload by allowing attention to be directed on a few key CRM categories rather than having to monitor for all categories continually. Although Seamster et al (1995b) focused on the development of behavioural markers for Line Oriented Evaluation (LOE) and LOFT scenario design, the principles they discussed may also be relevant to the general development of tools to assess non-technical skills. They recommended that the classification used for a marker system should divide CRM behaviours into groupings that facilitate the assessment process. Broadly CRM behaviours divide into two - interpersonal elements and mental activities. The former are directly observable whereas for the latter, the instructor/evaluator will often have to make inferences based on interpersonal or technical behaviours to assess mental activities. Therefore, according to Seamster et al, instructor/evaluators should not be asked to observe where the crew is engaged in mental actions (e.g. making a decision) rather, they should be asked to observe actions (e.g. specific crew communications) which indicate that a decision has been made. Interpersonal factors can be directly observed through crew communication and co-ordination of tasks.

•

•

•

•

•

A computer based representation of how experienced instructors assess CRM was developed by Dutra, Norman, Malone, McDougall and Edens (1995) which they called the Crew Resource Management Assessment Expert System Tool. This was based on instructor remarks about crew problems observed during 'first look' and proficiency check simulator sessions for B-737-300 and B-767 aircraft, for two air carriers. They segmented and coded 1298 remarks into four assessment categories: (i) Cognitive (decision making, situation awareness, workload management, planning); (ii) Interpersonal (communications, group climate, crew co-ordination, leadership/ followership); (iii) Technical; (iv) Other. A total of 579 segments were coded as CRM-related (i.e. Cognitive or Interpersonal) with an associated 931 observable crew behaviours. They then linked these behaviours to ten event sets they had chosen and through this reduced the observable behaviours to 90. After instructors had rated the centrality of the behaviours, this list was reduced to 54 elements. For each event they had chosen, they were able to specify five or six key CRM behaviours which the instructor should focus on during that stage of the LOFT flight. They found when event sets were clearly specified, subject matter experts showed agreement on the CRM behaviours that need to be observed to assess the related tasks. Thus it was concluded that reliable and valid evaluations of CRM can be conducted using clearly defined observable behaviours.

Finally, Seamster et al (1995b) identified a number of problems with existing behaviour lists which make them more difficult to use. When a marker contained more than one behaviour it was difficult to rate, to use their example: 'team concept *and* environment for open communications established *and/or* maintained'. They recommend that only one behaviour should be contained in each behaviour statement and a second should only be added if it is (absolutely) necessary to qualify the first. It is important that the wording of markers is concise and simple and that the verb of the statement refers to a clearly observable behaviour such as 'monitor' or 'ask'. This means that 'made a decision' is not observable whereas 'communicates a decision' is. They emphasised that the designer should always remember that the tool has to

be understood and used by instructor/evaluators who have a high workload during the assessment process and therefore should make each element as simple as possible.

2.2.3 Analysing CRM Skills in AQP

The process of upgrading crew training for US airlines using AQP has raised the profile of CRM skills in curriculum development, crew training and crew performance assessment. Seamster, Prentiss and Edens (1997) reviewed methods for identifying and specifying the primary CRM skills, using standard behavioural skill analysis methods. They suggested that carriers need to improve their methods of analysis in order to produce a complete skill list with appropriate behaviours which can be observed for CRM assessment. In a related study, Lanzano, Seamster and Edens (1997) examined the results of a comprehensive task analysis for one fleet of a carrier involved in an AOP. This revealed 2500 unique knowledge and skills entries in the Program Audit Database. Two main categories of skills are in such databases, psychomotor and cognitive. Their review showed that only 13 unique CRM elements were categorized as 'cognitive skill' (3%) and they concluded that 'carriers are identifying a very limited number of cognitive skills as being related to CRM' (p3). The single largest group of CRM elements (48%) were associated with unique knowledge components. This finding supported their view that the focus on CRM training has been primarily knowledge-based rather than skill-based. From this result, they make a number of recommendations for the AQP, such as 'When working with CRM skills, specify and adhere to a common level of detail' and 'CRM skills restated as performance objectives should be linked to observable behaviors for training and assessment purposes during LOFT and LOE' (p5).

2.3 RATING CRM SKILLS

As Helmreich and his colleagues have demonstrated (see above, p9), one of the fundamental concerns for ensuring the quality of any system for rating pilots' CRM behaviours is the reliability of the raters' judgements. Hamman (1997) identified several reasons for pilots' rating errors: Personal interpretation; memory errors; scale use; biases due to motivation. A degree of bias or systematic error can be expected in any performance rating task and Furnham (1997, p.507) lists some of the most common:

- central tendency providing a rating of average or around the midpoint for all items.
- *contrast effect* the rater assesses performance against that of other individuals recently rated instead of against the established standard.
- *halo effect* giving a high rating on one scale which influences the rater to give that individual similar or higher ratings on other scales.
- *horn effect* giving a low or unsatisfactory rating on one scale which influences the rater to give low ratings on other scales.
- *initial impression* rating based on behaviour early in the session, rather than from the entire appraisal session.
- *latest behaviour* rating influenced by the most recent behaviour, rather than from the entire appraisal session.
- *lenient tendency* rating consistently higher than the expected norm or average.

Different judges may be particularly susceptible to certain biases which raises a second problem, that of inter-rater reliability. Brannick and Prince (1991, p1) explain, '..if judges cannot be trained to be interchangeable, then feedback to air crews will depend more upon the particular instructor than on the team's behavior.' These measurement errors would also apply in the case of an examiner rating the performance of a pilot's technical skills. They do underline the importance of ensuring the validity and reliability of any system for rating non-technical skills, where the markers are likely to be rather less specific than instrument readings or control positions.

Williams, Holt and Boehm-Davis (1997) looked at the inter-rater reliability of instructor/ evaluators (instructors and examiners) rating pilot performance. They found that instructors' and examiners' inter-rater reliability was very low prior to training. However, training during which instructors and examiners discussed consistency and other aspects of reliability led to an improvement of inter-rater reliability, although Williams et al do not quantify this. Training was of three types, familiarisation with the rating scales, frame of reference training, where instructors were informed about organisational standards to provide them with a common baseline from which to work, and awareness training. They report the third method was the least effective. Williams et al had 60 subjects rate videotaped LOEs using three and four point CRM and technical behaviour scales. The researchers looked at agreement, systematic differences, congruency, consistency and sensitivity between raters. Generally they found all measures of agreement pre-training were low but also that the four point rating scale was related to more stable and consistent differences amongst subjects.

•

•

•

•

•

•

•

•

•

Law and Sherman (1995) note the sparseness of reliability and validity data for CRM skill assessments. Because CRM skills can be observed, it is possible to evaluate the effectiveness with which CRM is practised. Trained observer ratings show reduced halo bias, increased rater accuracy and a reduction in errors, which indicates that it is important to train trainers to use the instruments if they are to be employed consistently and yield results of any value. However, these positive effects do fade with time, although Law and Sherman suggest that if the skills are frequently practised this fade may not occur, as refresher training can ameliorate increases in errors over time. They suggest an index of agreement through which the reliability of CRM ratings can be evaluated and compared. In a genuine training situation, 24 evaluator trainees viewed a video of a two person crew in a full-fidelity simulated line flight. Ten other evaluator trainees viewed a different crew in the same simulation. Trainees used Helmreich's LLC4 (see above) to provide judgements of 27 elements for four phases of flight and gave two overall ratings. The agreement between raters was high with the two groups showing overall effectiveness ratings which correlated at 0.82 and 1. Technical rating agreement was lower, with the correlations between ratings being 0.78 and 0.94. Thus raters were in more agreement about crew performance effectiveness than they were in crew technical proficiency. Law and Sherman discuss some limitations of their findings, including the absence of a control group for the experiment. Another concern is that whilst trainees were in high agreement over the overall performance of crews, their ratings of the 27 CRM elements varied. This suggests that raters do not necessarily agree on exactly when a specific CRM behaviour should be rated for a given phase of flight.

Brannick and Prince (1995) wanted to investigate the degree to which raters' scores of aircrews correlate across scenarios. With 51 military aircrews (102 pilots), they had 17 raters use the US Navy's co-ordination dimensions to rate crews on two scenarios. Raters were carefully allocated to remove a possibility of a halo effect They found convergent validity. In a second study, they had six instructor pilots rate the performance of 50 pilots on a further two scenarios using four dimensions of assertiveness, situation awareness, communication and decision making. In this second study, inter-rater reliability was high (usually greater than 0.8). Brannick and Prince found that aircrew behaviour was not very consistent across events, nor in their reactions to specific events. Whilst it is easy to focus on poor inter-rater reliability as a problem for using scales, it must not be forgotten that the participants being observed are not 'fixed', and respond variously according to the situation.

Seamster, Edens and Holt (1995) evaluated CRM ratings at a session level compared to event level, to assess how experienced instructors and check pilots assess crew performance of videoed LOE sessions. A total of 32 aircrew instructors used four instruments containing CRM and technical elements to rate five scenario event sets. The highest correlations were obtained when instructors rated the entire scenario using a five point scale (0.47), as opposed to the lowest correlation from rating of event sets (0.19). They suggest that the higher

correlation may indicate a halo effect, making the event set rating more discriminatory. They conclude that the effectiveness of a particular method of assessment may depend on the exact rating scales being used, making it critical to specify both the assessment method and the rating being used when conducting CRM research.

Hamman (1997) describes inter-rater reliability training in a major carrier which involves comparison with established benchmarks and individual and group feedback. His training programme had the following objectives: (i) Congruency between each rater's distribution of judgements and the group's distribution of judgements; (ii) No systematic rater differences; (iii) Consistent shifts of evaluations upward for better performance and down for poorer performance; (iv) Sensitivity to discriminating the performance differences of better and worse crews; (v) High rater agreement on each item.

There are other psychometric issues which should be considered in the design of the rating scales: for example, the number of scale points, whether these have behavioural anchors, how the scale should be scored and standardised (see Aiken, 1996; McIntyre & Dickson, 1997). These issues have not been tested systematically at this stage for CRM behaviour scales, although four and five point scales tend to be favoured by both the research teams and the airlines (see Chapter 5) and Aiken (1996, p.55) advises that '4-5 rating categories are best for unipolar scales.'

2.4 TRAINING INSTRUCTORS AND EXAMINERS

Whatever type of behaviour rating system is used, the resulting effectiveness will depend on how it is employed by instructors and examiners. This means that an important factor in the implementation of any rating device is the training of those tasked to use the system (Taggart, 1991). Antersijn and Verhoef (1995) discuss the process of gaining acceptance for a new non-technical skills rating instrument within the pilot community. The non-technical skills were subdivided into five main categories with the acronym WILSC- work attitude; information management; leadership; stress management; co-operation. They stress that such tools should be practical and visible, and the users should be part of the development of the system. When KLM introduced their Feedback and Appraisal system they prepared their instructors through an advanced instruction course where the system was presented in practical sessions. During assessments, the definitions and descriptions of the behavioural markers are available to all those taking part, from instructor to flight engineer. Antersijn and Verhoef concluded from a survey of 118 instructors and 194 pilots/ flight engineers that the Feedback and Appraisal system KLM has developed is a success. They also found that some of the elements in the system were used more often in different environments, that is some elements are more useful during normal flights and others more useful during simulator training.

In the UK, a sub-committee of the Royal Aeronautical Society's Human Factors Group and the Air Training Association have received funding from the Department of Education and Employment to develop competence and performance standards for CRM instructors as the basis for an accreditation scheme. This study is due to be completed in early 1998.

2.5 STANDARDS OF COMPETENCE

To conclude this chapter, a brief look is taken at the use of performance standards. In the UK competence or proficiency standards have been introduced for a wide range of occupations, including civil aviation. Competence assurance systems have been developed for most British occupations in the last ten years. In the late 1980s, there was a government initiative to

develop an improved system of national vocational qualifications, based on universal standards of competence: competence being defined as the ability to perform consistently within an occupation to the standards expected in employment. The national standards were produced by industry-led committees of subject-matter experts using task analysis. Where there is an associated vocational qualification, the competence standards are used to assess performance in the workplace (Fletcher, 1991). For many occupations these competence standards deal with technical skills, however for some managerial and command positions. the emphasis is on cognitive and interpersonal skills. The standards developed for emergency commanders in the offshore oil industry or in the fire service (see Flin, 1996 for examples) bear some similarity to core CRM behaviours as they are also concerned with situation assessment, decision making, resource allocation and communication. Commanders in organizations who have adopted a competence assurance system will have their competence formally assessed on the basis of their performance when in command of a simulated emergency. In neither occupation is this mandated by the regulator, nevertheless a 'not yet competent' rating will require further training before the individual returns to the command role. CRM training has been advocated for command teams in both the oil industry and the fire service, and there may be useful cross-fertilization of research initiatives on identification and assessment of command skills between these domains and aviation in the future.

•

•

•

•

•

•

•

•

•

.

•••••

•

•

•

As part of this national exercise in the UK, vocational standards of competence for air transport pilots were also developed; Level 4 Key Role D deals with flight management (Aviation Industry Vocational Standards Council, 1995). These do not appear to be being used to any significant extent by either the regulator or the airlines, perhaps because competence assurance systems for pilots were already well established in civil aviation due to licensing requirements.

2.6 SUMMARY

CRM has developed through a number of generations, and the industry is now at the point of ensuring that CRM training is transferring from the classroom back to the flightdeck to improve safety performance, with a particular focus on error management. The CRM literature suggests that there is a general acceptance of CRM training and a growing acknowledgement of the need to assess these non-technical skills, (although this appears to be mainly in the larger carriers, especially those in the USA, operating an AQP). Using a rating system based on behavioural markers to assess CRM skills is a constructive way forward as it will allow a greater degree of objectivity in the assessment. However, the research to date has only begun to provide answers to the questions raised earlier in this chapter.

Unit of assessment

The unit of assessment in most research studies, particularly the work of Helmreich et al (1996, 1997), is the flightdeck crew, rather than an individual pilot. It is obviously of interest to airlines how their crews are performing but for individual licensing, individual assessments will be required, as they are currently for technical skills.

Identification of CRM skills and associated behavioural markers

The assessment of CRM has to be founded on an established set of CRM skills. Core CRM concepts are often subdivided into two categories: cognitive skills (decision making, situation awareness, workload management) and social skills (e.g. leadership, team work). These concepts appear to be used fairly consistently but labelling differs

across research studies, airlines and fleets. Devising a behavioural markers system for those skills presents a number of challenges. The sheer number of non-technical behaviours that could potentially be assessed has to be reduced to match a set of critical CRM skills. These can be condensed by being more specific about the event or the scenario, as each airline and fleet will choose different scenarios. Thus the behaviours being recorded in the industry could cover a very wide range, even allowing for synonyms and equivalent terms. This is examined in Chapter 6.

Assessment method

While there is an extensive psychological literature on the measurement of individual and team performance (Brannick et al, 1997), there have been very few studies which actually systematically compare rating scale designs for CRM performance, particularly for the rating of individual pilots. Issues of rater reliability have begun to be examined and results highlight the importance of rater training to achieve calibration against established standards and minimise bias effects.

A number of studies have considered the differences between ratings based on entire sessions compared to ratings for given stimulus events or event sets. In the case of the latter, specific behaviours can be specified and research teams have argued for the merits of this approach (Fowlkes et al, 1994), although in practical terms scenario development costs would have to be taken into account. Helmreich et al (1996) suggest that it is advantageous to rate crew behaviours in relation to phase of flight.

There are other questions relating to assessment methods which have not been examined in the research studies, for example comparison between line and simulator based evaluations and the extent to which technical and non-technical skills assessment can be integrated, although work is beginning in this area (Lanzano et al 1997).

• Rater training

As mentioned above, training in facilitation, debriefing and the use of CRM rating scales is critical in order to ensure fair and accurate assessments of non-technical skills. A number of research projects have been undertaken in this area and advisory publications released, for instance the use of the LLC crew rating system (Helmreich et al 1990b) or the conduct of LOFT and LOE sessions (Dismukes et al, 1997; McDonnell et al, 1997). Raters conducting licence evaluations will themselves need to be assessed and qualified.

Chapter 3 Method and Procedure

3.1 INTRODUCTION

The aim of this phase of the study was to gather, by means of two surveys, the opinions of those involved with CRM and to establish an overview of the use of behavioural marker systems.

• • • • • • •

•••••

•••••

•

The methods used were mainly qualitative, namely, a literature survey (reported in Chapter 2), an interview survey with UK participants (Chapter 4) and a questionnaire survey sent to airlines abroad (Chapter 5). Other valuable sources of information were the CRM developers group on the internet, the ninth International Aviation Psychology Symposium (April 1997 in Columbus, Ohio) and the NOTECHS project group, described in Chapter 1.

3.2 SAMPLE

About 40 organisations were identified as potential users of CRM marker systems, in the main by the CAA project steering group or through the NOTECHS project, with some participants volunteering information. A total of 33 organisations took part in the study at a variety of levels (airlines, plus representatives from FAA, ICAO, ALPA, BALPA and a CRM training company were also interviewed). Whilst not comprehensive, it is considered that the sample did give a representative view of the development of behavioural markers worldwide. All participants were helpful, and willing to give their opinions on markers and CRM.

3.3 SURVEY METHOD

The questionnaire and interview both asked open-ended questions about CRM, and in particular behavioural markers (see Appendices 3 & 4).

UK Survey. Participants in the UK were initially contacted by letter and then by telephone. Interviews (14) were conducted by telephone or face to face, the former taking about 45 minutes and the latter just over an hour.

Non-UK Survey. Foreign participants were initially contacted by telephone, letter or at the 1997 Aviation Psychology Symposium. Following this, interested parties were sent the questionnaire which was designed for completion in about 30-45 minutes. A total of 9 airlines completed the questionnaire. Additional interviews were carried out by NOTECHS group members (3 airlines) and information was obtained from a further 2 airlines, giving a total sample of 14 airlines represented.

A more detailed outline of the methods and results for the interviews and questionnaires are in Chapters 4 and 5 respectively. Frequencies of responses were taken from the questionnaire and a content analysis was performed on the interviews. Chapter 6 presents a more detailed analysis of the behavioural marker systems collected from the survey.

Chapter 4 Interview Survey

4.1 **INTRODUCTION**

To gain a representative view of the use of CRM behaviour markers in the UK, interviews were conducted with a cross section of those operating aircraft in the UK, including large multi-fleet operations and smaller operators, both fixed wing and rotor wing.

4.2 METHOD

4.2.1 Procedure

A semi-structured interview schedule was developed, consisting of 15 questions (Appendix 3). Notes were taken during the interviews and these were transcribed at a later date. Analysis of the interviews is qualitative – a broad content analysis. Four major themes were drawn from the transcripts; each of these is dealt with below.

4.2.2 Sample

The sample consisted of eleven commercial operators (Air Atlantique, Bond Helicopters, Bristow Helicopters, Britannia, British Airways, British International Helicopters, British World Airways, Magec, Monarch, Shell Aircraft, Virgin) and the Royal Air Force (RAF). Six interviews were conducted face to face and six were by telephone. Participants were involved in the development and delivery of CRM in their company. An additional meeting was held with representatives from British Airline Pilots' Association (BALPA) and information was provided by a CRM training company (LMQ).

4.3 CRM COURSES AND DEVELOPMENT IN THE UK

Most participants' CRM courses were two or three days long and involved a combination of presentations and exercises. The concepts presented included teamwork, communication, decision making, situation awareness, leadership, as well as topics specific to the organisation such as customer services.

Where organisations differed, was in the degree to which they had advanced their course into an ongoing training programme. Some airlines had stopped at the point of having an initial CRM course, that is they were planning refresher courses but had not yet introduced these. Other organisations had completed two rounds of refresher training to date and were embarking on their third. For these organisations their concerns were to develop the programme onto its next stage. For most, the third stage of the CRM programme is to develop the practice of CRM – usually through LOFT.

Putting together a scenario training package for LOFT involves a series of developments, including training facilitators to run exercises and debriefings, and constructing a system to guide the facilitators' discussion. This last guideline is usually in the form of a behavioural markers checklist. Airlines who were at this stage of CRM programme development anticipated that the next logical step would be to begin to assess the CRM skills of their pilots. This anticipation formed a second reason for these participants beginning to develop behavioural markers checklists. They wanted to be prepared when formal assessment is introduced.

There were two reasons given by participants whose organisations were not developing their CRM programme, for not doing so. The smaller companies may not have the resources or the personnel to devote to developing CRM training that is not mandated. The second reason was if CRM had not been readily accepted by the pilots in the company, developers were reluctant to push the concept further than the initial course. It was beyond the remit of this study to probe for the reasons underlying the non-acceptance of CRM. For this, a further specific study would be necessary. However, some participants did comment that the concept of CRM has matured over its 20 year history. When it was initially introduced, CRM was developed from management training courses and therefore was presented in a way which was relevant for managers but not pilots. Early style approaches may have contained psychological or managerial jargon, and presented concepts in a way that pilots could not identify with, which has led to adverse opinions of CRM skills training.

•

.

•

•

•

•

•

•

•

The focus of the present study is on the development of behavioural markers and the associated training and procedures. Discussion of these topics is presented below.

4.4 DEVELOPMENT OF BEHAVIOURAL MARKERS CHECKLISTS

Of the 12 organisations surveyed, only 50% (5 airlines, plus the RAF) had any kind of behavioural marker scale. (One of which has just abandoned its use). None of these were used for formal CRM assessment. Instead they had been designed for CRM training and to structure LOFT feedback sessions. In one system they were described as 'an aide memoire of a range of behaviours'. The five airlines were aware that they might be required to evaluate CRM skills more formally at some future date. Those organisations who had CRM skills checklists had consulted a number of examples before beginning to develop their own. For some, this development was minimal, adopting an existing list and changing the odd word or phrase. Others had carried out in-house research to produce their own behavioural markers framework. All participants developing checklists had reviewed Helmreich et al's LLC (see Appendix 2) and their comments on this system are presented below.

Copies of five of the six marker frameworks were obtained. Figures in brackets represent (categories/ elements or behaviours). They were called 'LOFT – Human Factors' (8/37); 'Team Skill/ Pointer'(8/58); 'LOFT – HF/CRM Performance Markers' (6/36) 'Performance Markers' (8/32) and 'CRM Behavioural Markers' (3/16). The CRM skill categories typically included situational awareness, decision making, communication, leadership/followership, crew relations, workload management or some variation of these. But the five systems were structured differently and contained different categories, elements and behaviours. Only one had a rating scale which listed 3 behaviours for each element, one unsatisfactory, one standard and one above standard. In two systems the markers were for an individual pilot's behaviour, in the other three they were worded in terms of crew skills, but could also be used to give feedback to a particular pilot.

Thus it appeared that behavioural markers in UK airlines are not in widespread use and those that are available are at a fairly early stage of development and used only for debriefing and feedback purposes.

4.4.1 Opinions of Line/LOS Checklist (LLC)

The general opinion of the LLC (Helmreich et al 1990;1997) was that it was an excellent research tool but it did not entirely meet the needs of practitioners. Specifically, participants thought the LLC was too complex to be used in LOFT or during other forms of training – facilitators need a simple checklist that will provide an 'aide-memoire' rather than a comprehensive scheme which demands full-time attention. Facilitators, especially in simulator sessions, have many things to do, i.e. run the simulator, assess the pilots' technical

skills, run the scenario, and could not devote the time to look for specific behaviours to complete the LLC usefully. The LLC would be suitable if facilitators had the luxury of a co-facilitator to look specifically at CRM.

Some CRM trainers/facilitators said that the LLC contained too much psychological jargon, too many words and too much detail for it to be accepted or used in practice. The facilitator needs a simple tool that can be used to note down the key points needed to be raised in the debrief, coupled with (possibly) a more detailed sheet that could be completed subsequently. As a result many participants claimed that the LLC whilst being a useful place to start their search for behavioural markers, was not used as the basis for the system they were developing.

Most participants seem to be aiming for a behavioural markers checklist which combines the positive properties of a research tool like the LLC and a note-card reminder tool like a checklist. They emphasised the need for a system which was simple, easy to understand and to use. Many said that behavioural markers tools should look simple but that this does not mean they are simple. These participants felt that successful use of the behavioural markers would rest on the users understanding the concepts underlying behavioural markers and therefore training sessions should accompany the introduction of markers.

4.4.2 Training for Behavioural Markers Checklists

Some participants felt that all pilots should receive training explaining the underlying CRM skills concepts of the behavioural markers checklists under development. Others felt that instructors and evaluators needed to receive this training and that pilots only required a comprehensive manual and an overview to be presented in CRM initial and refresher courses. They felt that as marker systems were developed, these would be absorbed into courses developed earlier.

Other components of facilitator training would include how to generate and conduct debriefs – getting pilots to talk and to share their experience. This training requires a change in approach, from one of giving students direction, to one of encouraging them to reflect on their own performance. The introduction of behavioural markers may therefore be daunting for instructors and examiners as it demands a change of style. This may make the development of facilitator training key in the acceptance of behavioural markers. Some airlines are developing facilitation training for their instructors and examiners whilst others are relying on consultants to provide this training for them. Participants commented that much of the resistance they have encountered from their trainers with respect to using behavioural markers is based on their lack of confidence at being able to use the system. They believe once instructors and examiners have been given background information to the behavioural markers system, become familiar with it and understand its principles and concepts, they will be comfortable using it.

Airlines who are developing behavioural markers have all planned a phase of user verification, where instructors and examiners will be asked to test-run and critique the behavioural markers checklist by using it during training such as LOFT.

4.5 ASSESSMENT OF CRM SKILLS

Currently in the UK, CRM skills are not formally assessed, however, a central issue in the CRM behavioural markers debate is whether the UK is moving towards a point where they will be. The general concerns of participants are that this system should be fair, if a pilot's licence is 'on the line'. This raises a raft of issues about how assessment should take place

and opportunities for pilots to retrain, where necessary. However, most participants felt that assessment was a useful progression of the CRM training. Many participants felt that CRM skills had been assessed for years under the guise of 'airmanship'. If a pilot lacked 'airmanship' skills then he or she could fail his or her licence revalidation test or their operator proficiency check under the present assessment system. •

•

•

• • • •

•

•

•

•

•

•

•

•

•

• • • • • •

The prospective introduction of a jeopardy assessment increases the importance of developing the right behavioural markers system. The need for observations to be concrete and repeatable is emphasised. Representatives from pilots' associations (BALPA, ALPA, IFALPA) were interviewed during this project and it is clear, as discussed in Chapter 1, that these bodies are particularly concerned about unresolved issues relating to jeopardy assessments of non-technical skills.

Some participants felt that outcomes of assessments should not have 'fail' as an option but should instead recommend further training. However, this would only delay the discussion of assessment to how many times does a pilot have to be recommended for further training before the organisation concludes that the pilot does not have the required skills to hold a licence? Participants felt that if CRM becomes a jeopardy assessment then providing retraining and opportunities to resolve bad performance will become central. These issues also pertain to technical skills assessment for licensing.

A further debate is whether a pilot 'passes' a CRM assessment or earns the label 'safe'. Similarly should a pilot be judged 'not yet safe' or 'not yet competent' rather than 'fail'? Whilst this is a 'war of words', the use of a distinction such as safe/not safe explains more and may promote the development of further training for pilots who have not passed the assessment. Some participants also felt that non technical behaviour on a flight deck cannot be prescribed, there are so many individual differences that safe behaviour can be achieved in hundreds of different ways. A behavioural markers list is therefore an indication of the direction in which behaviour should be going, a guide rather than a prescription. The NOTECHS (1997) group has also emphasised that the criteria, against which behaviours should be observed, are to determine acceptable from unacceptable behaviour.

Objections to making CRM a jeopardy assessment were, in the main, those faced by any national assessment programme. Apart from the reliability/validity issues discussed above, another concern raised was whether one checklist could be used to assess pilots who had been trained on different CRM courses in their airlines. This argument underlines the need to develop and test a standard marker system with an accepted set of core CRM behaviours.

4.5.1 Retraining on Failure of Non-Technical Skill Assessment

Some participants were thinking beyond behavioural markers and their use in assessment to how they would manage pilots who failed the check. Should pilots take the training course again, or would they be better served by individual study which would present the materials in a different way? One participant argued that a pilot who failed a non-technical skills check simply reflects that the training course was at fault rather than the individual. Again, as with all national courses, this may be the case if the majority of candidates fail but if they are isolated examples, this implies it is the individual who needs more training.

4.5.2 Use of Behavioural Markers Data

Most respondents believed that the only way one can judge whether pilots have understood enough of what has been taught on CRM courses to put it into practice is to observe their performance, either on the line or during a realistic simulation. This obviously requires an established method of assessment, hence the development of behavioural markers systems. It was also mentioned that while non-technical skills have always been an important element of pilots' performance, only the technical skills have traditionally been formally assessed.

Some participants questioned the necessity to assess CRM skills. One argument against assessment was that if CRM is about error management and safety, then a 'social skills' examination may not be appropriate. (However error management is comprised of core skills, including communication, leadership and teamwork – all of which are encompassed by behavioural markers). Another issue raised was whether it is appropriate to assess an individual pilot's non-technical skills rather than the skills of the flightdeck crew who are working together. It is likely that this issue will need to be addressed by a systematic comparison of individual versus crew CRM ratings at some future date.

Some airlines have plans for using the data they collect from assessments to improve their CRM programmes. By reviewing the performance of pilots during assessment, they hope to identify generally weak areas and be able to tailor their refresher programmes to concentrate on these skills. Plans such as this will 'close the loop' on CRM training and tie the programmes in with the areas of greatest need. This should lead to a continual improvement of CRM skills.

4.6 SUPPORT FOR CRM ASSESSMENT

A factor that participants felt is important in the development and acceptance of behavioural markers is having the support of higher bodies, both within the airline (management commitment) and externally e.g. the regulator (CAA) and the unions. A number of interviewees noted that there seems to be a reluctance in these bodies to give the necessary support.

Some participants emphasised the importance of keeping the instructors and examiners informed and involved with the development of behavioural markers. These pilots hold senior positions in their companies and their views are influential in the pilot community. As the people who will be asked to use the systems both in practice and potentially during assessment, it is vital that they not only know how to use them but that they support their introduction. Therefore, instructors' and examiners' acceptance of behavioural markers is crucial to pilots' acceptance of the system.

4.6.1 The Role of the Regulator

Many participants mentioned the role of the CAA. They felt that the CAA will have to mandate or regulate every stage of CRM development e.g. the use of behavioural markers, before some airlines will use them. Part of the reason for this is the earlier argument that some small aviation companies cannot afford to develop CRM if it is an 'optional extra'.

Participants commented that they were looking to the CAA for advice and guidance when they were considering using behavioural markers. They felt that the CAA should provide recommendations on the background checklists to review, the way to structure a behavioural markers checklist and how it should be used. Participants were concerned that because some airlines have independently decided to develop their behavioural markers systems in the absence of advice from the CAA, that the systems developed will not be comparable. This could lead to confusion over the level of training that some pilots have received. In addition, they are concerned that if they fully develop a behavioural markers programme, and the CAA regulates its use, the airline may have to revise their behavioural markers system, taking up further time and resources. Those involved with CRM training were looking to the CAA to give them the confidence to pursue development of behavioural markers. Since this has not seen to be forthcoming it has made putting a system together difficult. This arises for two reasons: firstly they have to justify the value of CRM and behavioural markers to their management; secondly, they also have to demonstrate to their fellow pilots that such a system will be a necessary component of CRM in the future. •

•

•

•

•••••

•

.

•

•

•

•

4.6.2 Management Commitment

According to the respondents, airline management varies in its support of CRM development. For most, CRM entails a substantial investment, which does not yet have a proven link to crew effectiveness. This is partly because CRM skills are not assessed and therefore an improvement over time cannot be measured. Participants also felt that, without management commitment, staff in that airline would not fully accept introduction of behavioural markers.

4.7 SUMMARY

Participants raised a number of issues concerning the introduction of behavioural markers. The central point is that they should not be introduced in isolation. For markers to be understood and useful they have to be part of an education programme which explains their role in CRM training and assessment. Whilst CRM awareness training is mandated in the UK, which provides a baseline training for all UK pilots and ensures a basic level of awareness, this does not mean that all airlines have chosen to develop the programme any further than this. A number of airlines with more advanced CRM programmes, have already introduced marker systems and are using these in LOFT feedback sessions.

Moving to CRM skills assessment involves a number of stages: checklist development, checklist trialling, training of assessors, system evaluation. An organisation moving through all these stages will need to ensure pilots' familiarity with the concepts and allow the time necessary for cultural adjustment. The terminology of a checklist and the way it is set out will influence pilots', instructors' and examiners' acceptance of it as a valid component of the assessment process. This was reflected in the comments of the questionnaire survey participants who reported adjusting their behavioural markers systems to simplify the language (see next chapter). The needs of the users must be taken into account in the design of any checklist which should be viewed as a tool and come complete with a 'user manual'.

Some CRM instructors/ examiners who will be using behavioural markers systems appear to lack confidence in their ability to use or develop these tools. In many cases, this stems from an unfamiliarity with both the style of behavioural markers presentation and their practical application. Another concern expressed was the lack of guidance provided by official bodies and, in some cases, a lack of support from management.

The survey results showed that in the UK the development of behavioural markers for CRM skills is not as advanced as it is in some European or USA airlines, especially those in an Advanced Qualification Programme (AQP). Some international airlines have fully developed behavioural markers systems for assessment and the following chapter examines the use of these systems in airlines outside the UK.

Chapter 5 Questionnaire Survey, non-UK Airlines

5.1 INTRODUCTION

A second survey was conducted to assess the use of behavioural markers for CRM in airlines outside the UK.

5.2 METHOD

5.2.1 Procedure

A questionnaire was constructed on the basis of the literature review and the UK interviews (Appendix 4). It had seven sections with a total of 32 questions which focused on the development and use of behavioural markers within an airline and how these were related to its CRM programme. This was approved by CAA and sent to international airlines who were known to have a behavioural marker system as part of their CRM programme. The questionnaire was designed to gather qualitative information in place of an interview. Therefore the responses are reported below as a content analysis rather than a quantitative summary.

In addition, several interviews (based on the questionnaire) were conducted with representatives from a number of major airlines and from aviation organisations at the Aviation Psychology Symposium (1997) or represented on the NOTECHS group.

5.2.2 Sample

The questionnaire was sent to thirteen airlines, nine of whom returned it (69% response rate). These were Aer Lingus, Air Canada, Ansett, Atlantic Coast, Braathens SAFE, Cathay Pacific, Delta, Northwest, and SAS. In response to an earlier letter, information on behavioural markers had been received from American Airlines. Some information from Continental was obtained from a workshop at the Ohio meeting. The NOTECHS group members provided information for European airlines (KLM, Lufthansa, Air France; see the NOTECHS (1997) report for details of these systems). This gives a final sample of 14 airlines, although some of these are incomplete responses (hence the changing sample numbers in the discussion to follow).

At the Ohio meeting interviews were also conducted with representatives from ICAO, FAA, and ALPA. Their views have been taken into account in the following discussion.

5.3 **RESPONSES TO THE QUESTIONNAIRE**

5.3.1 Details of the Behavioural Marker Systems

All 14 airlines had some type of CRM behavioural marker system used in training and in some cases for assessment. Copies were provided by 13 of them. Some systems were labelled as mnemonics which represented the first letters of the behavioural markers categories in the system, presumably because this made them easy to remember and from this the categories can be recalled. Sample names of other systems were 'Crew Effectiveness Pointers'; 'Crew Effectiveness Markers'; 'Observable Crew Behaviors'; 'Essential Skill List'; 'Check Assessment System'; 'Crew Performance Indicators'.

These marker systems had been developed from 1979 onwards, the range of dates implying that airlines have had different approaches to the use of behavioural markers within their CRM systems; with some focusing on them from their first use of CRM concepts, to other airlines who were just introducing them now. Most systems had been developed and used from 1990 onwards.

•

.

•

•

•

•

•

•

•

•

•

.

•

•

Almost all systems had been developed using in-house expertise from their pilots, trainers and psychologists. When asked for 'the research basis of their marker system', only eleven airlines provided this information, nine of whom (82%) had used or referred to the NASA/UT Line/LOS Checklist (LLC) (Helmreich et al, 1997, see Appendix 2 for the latest version). Some companies had adopted the LLC with minimal alteration. Other sources were also mentioned – several airlines had collected in-house data from pilot performance, pilot opinions, and incident data. Some airlines had studied marker systems from other airlines or military air forces before designing their own version. This suggests that the majority of marker systems had been based on the LLC, although airlines tend to adapt this for their own purposes. About 25% of airlines surveyed had developed their own systems from in-house research, although they also considered other systems during the development phase.

In their general comments, two airlines mentioned the importance of developing a generic system. Reasons for this were twofold, firstly where there were international regulations, (e.g. in Europe), then the behavioural markers systems in use should assess pilots on an equivalent basis. Secondly, confusion can arise when different behavioural markers systems are in circulation – raising questions in the pilot community as to the respective validity of these systems.

Behavioural markers systems had been introduced within airlines in the main through courses, either CRM training, LOFT, other courses or a mixture of methods (see Table 5.1 for data on 13 airlines). However, written reference manuals were also a frequent method. The courses ensured pilot understanding of the concepts whilst the manuals allowed the pilots to refresh their memory of the concepts at any time.

Table 5.1 Media for introduction of Behavioural Markers

Medium of introduction of behavioural markers	Total (n=13)
CRM courses	9
LOFT	8
other courses (e.g. instructors)	7
flight manuals	1
notes/ handouts/ posters	3
examination sessions	1

Having introduced behavioural markers systems, nine airlines had modified the original system to reach the one in use at present. There was no consistent way in which these systems had been modified. Airlines reported a range of modifications from minor adjustments, like making certain terms in the system more comprehensible, to complete overhaul and reworking of all the markers in line with company-wide opinions. One airline had to improve its training for instructors and evaluators and they incorporated the markers onto the instructor's gradesheet. Only two airlines reported having to make changes because the system was rejected by the pilots. Six airlines reported that they still had problems with their current system, of these, four have modified the system from the original, and two are planning further modifications. A problem which had been addressed in some airlines was the use of too much psychological terminology in the behavioural markers system. Other

problems mentioned, included the system being too complex and in another airline they felt it had been introduced too quickly.

One company provided very detailed information on the development of their event sets used in LOFT and LOE for CRM training and assessment. This is clearly an important element of the CRM skills assessment process, but was not included as a topic in this survey.

5.3.2 Design and Content

The behavioural markers lists provided varied widely in design, and included from 7 to 30 behaviours. Nine of the systems considered technical skills with non-technical skills. The main CRM categories were broadly similar, and were sometimes subdivided into social (interpersonal) and cognitive (mental) skills. The principal categories typically included leadership; communication; team co-ordination; decision making; situation awareness; workload management, although the labels for these categories often differed. Some lists included additional categories such as stress, automation management or crew self evaluation. The actual item content (listed behavioural markers) is analysed in Chapter 6 and so will not be covered in this section. No two systems had exactly the same content or format.

Of the 13 behavioural markers systems provided, all included a rating scale, again these showed wide variation. They ranged from two-point to six-point scales, with a five-point scale being the most common, followed by a four-point scale. Typical labels for a five point scale would be: 'unsatisfactory', 'satisfactory', 'average/standard', 'above average', 'excellent' or equivalent terms. In two companies a five-point numerical scale was used. Some airlines used other labels such as +/- (which indicate presence/ absence of a behaviour or pass/fail). The reasons given for choosing the scale in use included that it was a proven design (version of LLC) or that the format provided the information they required.

5.3.3 Training Instructors/ Examiners

Ten out of the thirteen airlines answering this question, provided a course to train their instructors and examiners to use the behavioural markers system. For 50% of the sample this was a dedicated course, for the others it was part of an instructors and examiners training course covering a range of skills and techniques, such as facilitation and feedback. Many courses taught instructors and examiners together. Courses varied from one to three days and tended to be workshop based with many practical exercises and video segments which would be used for practice rating. When using the behavioural markers system, seven airlines reported that their instructors and examiners follow a protocol. No airlines reported that instructors and examiners did not use a protocol. Several respondents were aware of potential difficulties in achieving standardised (calibrated) assessment and emphasised the need for instructors and examiners to be properly trained in the use of behavioural markers to assess CRM skills, particularly where jeopardy assessments were involved.

5.3.4 Use of the Behavioural Marker System

Of twelve airlines, nine report introducing their pilots to the behavioural markers system through training courses, whilst three use written information to do so. This corresponds to a degree with the responses to how the behavioural markers system was introduced into the airline generally, although the use of written material was reported as more prevalent in answer to this question than in the general information question. Markers seem to be introduced in a number of phases, in the same way that CRM generally had been introduced, starting with sessions to raise awareness of the issues, moving to practice and understanding of the concepts and their use in the cockpit.

Respondents were asked to indicate from a list of options how they used their behavioural marker system. From the responses of 11 airlines who completed this part of the form, Table

5.2 was compiled. The respondents reported that their behavioural markers systems were used for training (from ab-initio training to recurrent courses) as well as assessment. Several USA airlines had fleets in the process of transitioning to the Advanced Qualification Programme (AQP) which requires assessment of CRM skills (Birnbach & Longridge, 1993).

•

•

.

•

.

•

•

• • • • •

Use	frequency
ab-initio training	4
AQP (assessment)	2
recurrent training	7
LOFT sessions	7
LOS	1
LOE (assessment)	3
License/line checks	6
instructor training	1

Table 5.2 Uses of behavioural ma	arkers systems in airlines
----------------------------------	----------------------------

Of the total sample of 14 airlines, 12 said that they used behavioural markers for some kind of assessment, 50% of whom (6) reported that the pilots could fail a check based on their CRM skills.

Of the six airlines who said they would not fail on the basis of CRM skills, two reported that they would be moving to this position. Two other airlines stated that they would not fail a check on CRM skills because the airline retrains the pilot in question until they do pass the assessment. Of those six airlines who may fail a pilot on the basis of their assessment, all reassess the pilot at a later date. In the interim, all offered further training although this ranges from repeating the original course, to going through a different course or having training specifically designed in conjunction with the fleet training captain. One airline includes an interview with the fleet manager as part of the review of an unsuccessful assessment.

Some airlines have a concern that understanding of the behavioural markers system and its use is not great enough for it to be used in assessment, that pilots do not yet fully understand what the aim of the system is and have not yet received sufficient training to be assessed using the markers. A more negative outlook on this problem is that assessment can be introduced without sufficient CRM training for pilots and therefore does not follow an increased awareness in or use of CRM skills.

A number of airlines were concerned about retraining, for instance, how to design further training which is different from the original training, as this is, by implication, not presented in a useful format for the pilot who fails the check. This is the issue of how to develop non-technical skills training which will be more suited to the learning style of the individual who has to retake the course. Other airlines were concerned that although behavioural markers are indicators that the individual is having problems, they are not diagnostic and that further work has to be undertaken to specify exactly where the pilot is having difficulties. Only one airline reported that they also use the behavioural markers system to assess the CRM courses they run, with the aim of improving the course through this application of the markers.

Following the use of the behavioural markers system, the pilots were debriefed about the CRM ratings that they had been given and had a chance to discuss the points made. They may be asked to read and sign the assessment sheet and two airlines said that they gave pilots

copies of the assessment sheet. In one airline the pilots were asked to complete a crew self evaluation using the same CRM rating scale. Three airlines did not retain the rating data. The other airlines filed the behavioural markers report, either in a general database or on the pilot's personal file. Generally ratings were de-identified before entry onto a fleet or company database. Following the feedback session, three airlines offered their pilots the opportunity to take further training or practice.

5.3.5 Pilots' Views on the Use of Behavioural Markers

When asked what feedback had been received from pilots who had taken part in courses or assessments where behavioural markers were used, the responses were generally favourable. Eight out of nine airlines reported that their pilots received the behavioural markers system positively; three reported 'high acceptance', five reported 'acceptance' and one said 'rejection'. The following comments are indicative:

'Only positive feedback – I think a lot of our pilots up till now have felt that CRM is difficult to get a grip on, a 'buzz word'. The behavioural markers system shows them that CRM is about concrete, practical skills, something they can do and something that makes sense.'

'CRM LOFTS were non jeopardy so pilots appreciated the opportunity to practise CRM skills. The future assessments will be just the same as other skills assessments.'

'Pilots enjoy and express that they learn more in our LOE/LOFT training. In general they have a positive approach to evaluating crew effectiveness.'

One respondent reported considerable resistance to the introduction of behavioural markers, particularly among the instructors.

'This is mainly because of their lack of confidence in the [name of marker system]. Their feelings seem to have some merit as considerable variation in assessments was found. We also have evidence that building an evaluation consensus amongst instructors after a day of watching a few video tapes for training purposes is by no means the same as achieving an enduring and objective set of performance criteria.'

When asked of any problems perceived by pilots with regard to the behavioural markers system, one respondent mentioned that any problems they did encounter were due to the system being used in a way which emphasised identification and correction of mistakes rather than an acceptance that errors are inevitable and that the critical CRM behaviours are to do with error trapping and mitigation (cf. Helmreich, 1996; Reason, 1997). A different airline commented that they had difficulties balancing emphasis on assessment against flight safety and error management. It seems that using behavioural markers in both jeopardy assessment and training in this airline caused a degree of conflict. Other remarks were:

'Psychobabble words - words used must be clearly understood by all.'

'Complicated markers lead to wide variation between assessors.'

'Had to convince instructors they could grade soft scales.'

'Past training has been insufficient to ensure that instructors and evaluators are proficient in the use of [marker system] to assess crew performance.'

5.3.6 Future Developments
The response to this question was almost uniform, with airlines reporting that they intend to use their behavioural markers systems more, and more widely, in the future. Included in this broadening of perspective is to integrate technical with non technical skills, to make the system more simple and to improve instructors' and examiners' skills. Also some airlines envisaged the need to extend the use of behavioural markers systems to other aviation domains such as maintenance. The following comments were offered: •

•••••

•

•

'I hope, in the future, that knowledge about behavioural markers will be a natural part of every instructor's competence, and I hope they will help them to have better use of video debrief as a tool to teach pilots CRM.'

'We would like to expand it to more specific assessment e.g. to be able to identify more specific areas the candidate would be able to improve on.'

'Measurement of behaviours are critical to addressing human factors issues which feature in all accidents.'

'Continued use as more aircraft fleets train using AQP.'

A final section was included for respondents to add comments and the following were raised:

'I think it is important to develop behavioural markers that different airlines can agree upon, and feel that NASA/UT has done a very good job with their behavioural markers.'

'We have extended the behavioural markers system to other company domains- flight attendants, meterologists and mechanics receive human factors training.'

'We are still developing CRM-wise but we feel that our crews have sufficient understanding and practice that CRM skills can be assessed like all others.'

'There are great differences between behavioural markers systems in use. There seems to be confusion between the use of criteria to help instructors structure their briefings and the notion that cockpit behaviour can be reduced to a checklist of observable behaviours. Behavioural markers may create a new assessment reality without positively affecting pilot performance or impacting upon cockpit management.'

'You must be able to use the markers, they must be observable.'

5.4 SUMMARY

The responses to the questionnaire indicate a range of levels of development and use of behavioural markers systems across the sample studied. Whilst the majority of airlines (13 out of 14) report having a behavioural markers system of some sort which is used in training; fewer (12) have a rating scale as part of this system for assessment of pilots' non-technical skills; still fewer still (6) use this system as part of a formal check. Thus only a third of the responding airlines utilise a behavioural markers system through all stages of training and assessment. This suggests that these systems are still in a development phase. Finally, it should be emphasised that these responses were from a sample of larger airlines and smaller operators were underrepresented in this sample.

Chapter 6 Analysis of Behavioural Markers Lists

6.1 **INTRODUCTION**

With the aim of reviewing CRM behavioural marker systems which could potentially be used for assessment, a sample of CRM skill rating systems was obtained from airlines or research teams. These were analysed to identify the most frequently used CRM skill categories and associated behaviours.

6.2 METHOD

6.2.1 Sample

In total, 23 airlines and research teams from the two surveys (Chapters 4 & 5) provided copies of behavioural markers lists (Appendix 5) which they were either using or putting out to trial. Some airlines provided two lists, using slightly different markers for LOFT and LOE for example, or lists developed for particular fleets, or a current and a protoype list. When available, both lists were included in the survey, giving a total of 31 behavioural markers lists which were sampled. No distinction was made between lists principally designed for crew or for individual pilot assessment as this information was not provided for all lists and some were dual purpose.

To compare these lists for common markers and hence to establish what the aviation community consider to be important CRM skills, the total set of marker terms was examined. Most systems have a small set of categories, subdivided into elements and then may also have specific examples of behaviours set out for each element. The terms, 'categories', 'elements' and 'behaviours' are used in this chapter to describe these three levels of detail. The sample markers lists were scanned to identify the most common terms for the category headings and elements within these categories. For a cultural comparison, this set was then divided into UK, USA and European/world behavioural markers.

Some of the behavioural markers lists did not contain purely CRM behaviours but included technical skills as well. These elements reflect those airlines who are integrating CRM skills with technical skills during assessment.

6.2.2 Collation of Categories

There were 101 separate categories of CRM skills in the 31 checklists (see Table 6.1). Usually the lists had between five and eight categories which indicated there was sometimes little overlap between categories. The UK lists had more in common with the USA lists than the European ones, sharing over 40% of the categories. This finding may have implications for the development and acceptance of a pan-European system, such as NOTECHS (1997).

Table 6.1	Comparison of Categorie	s of Behavioura	I Markers Lists
		UCA	Europal

	UK	USA	Europe/ World
Number of lists	8	8	15
Number of diff.categories	36	26	69
Common with UK	-	11 (42%)	19 (28%)

Whilst most of the categories had general labels, such as 'communication' or 'decision making', others were more specific, e.g. 'fulfil First Officer/Flight Engineer responsibilities' or 'crew self-feedback'. The categories were ranked in order of frequency, and the ten most frequently used for all three data sets (UK, USA, Europe /World) were compared (Table 6.2). The nine categories which were the most frequently used across all the lists were noted in column four of Table 6.2, headed 'Overall'. Following Seamster et al (1995b), these categories can be further classified into social/ interpersonal skills and cognitive skills, the latter being 'decision making', 'situation awareness' and 'workload management'.

•

•

•

•

• • • •

•

•

•

Whilst Table 6.1 emphasised the variation between behavioural markers lists in use because of the sheer number of categories found, Table 6.2 shows a less depressing picture; that the most frequently used categories were very similar. Comparing across the 30 most frequently used, ten each from the UK, USA and Europe, there were only 18 different category labels.

	UK	USA	Europe/world	Overall
1	communication	communication	communication	communication
2	decision making	decision making	leadership	decision making
3	leadership	situation awareness	decision making	situation awareness
4	behaviour	technical proficiency	briefing	workload management
5	workload manag.	team management	situation awareness	teamwork
6	assertion	teamwork	team management	team management
7	relations	automation	workload manag.	leadership
8	situation awarenes	CRM	assertion	assertion
9	teamwork	responsibility	relationships	relationships
10	vigilance	workload manag.	conflict resolution	

Table 6.2 Comparison of most frequent categories (ranked 1-10)

There are certain themes that the UK behavioural markers lists seem to pursue which other nations do not. Fewer of the UK top ten categories are devoted to teamwork and situation awareness themes, whilst the USA behavioural markers lists have fewer categories focusing on leadership and workload. European/world behavioural markers lists focus on team management rather than teamwork. These differences may reflect the emphasis on CRM which differs across nations with UK airlines focusing on leadership as a concept distinct from teamwork.

6.3 ANALYSIS OF CRM ELEMENTS

The same procedure was followed for the elements of the behavioural markers checklists, to assess which were the most common. Within categories there were usually four to six elements, although the range was from one to eleven. The elements from the checklists were compiled into three data sets – one of UK markers, one of USA markers and one of European/world-wide markers. These data were then compared as shown in Table 6.3 below. Comparison of the number of elements in the eight UK lists with the numbers of the USA and Europe/ world lists (rows 2 & 3) indicated that UK lists tend to be longer, with an average of 33 elements per list compared to an average of 22 elements in lists from countries outside the UK.

The third row of Table 6.3 gives the total number of different elements in each dataset whilst the fourth row combines elements where their meanings are the same but the wording is slightly different, for example, 'friendly tone' would be combined with 'tone is friendly'. The distinction serves to illustrate that there is no standard format for wording used by those who compile behavioural markers lists.

In total, there were 498 elements in the 30 behavioural markers checklists. Again this emphasised the lack of agreement between airlines with respect to the crucial behaviours for crew effectiveness. In the UK, 51% of the elements were classed as 'frequent', meaning they were used by more than one airline (row 5, Table 6.3). This was slightly lower than in other countries where the number of elements which were 'frequent' form 72% and 66% of the total. This indicated that UK checklists, when viewed as a whole, covered a broader range of elements, but with less consistency among lists.

		UK	USA	Europe /world
1	Number of lists	8	8	15
2	Average number of elements per list	33	22	22
3	Total number of different elements	296	154	303
4	Number with different meaning	231	118	228
5	Number on more than one checklist	47	34	59
6	Percentage of elements in 'frequent' list	51%	72%	66%
7	Correspondence with UK categories	-	27	49
	(as a percentage)	-	23%	22%
8	Correspondence with 'frequent' UK categorie	-	12	26
	(as a percentage)		35%	44%

Table 6.3 Comparison Statistics for Elements in Lists

From the 498 elements, 68 themes were distilled by eliminating minor differences due to phraseology. This reduction suggested that CRM developers have found many phrases to label identical behaviours. There may also be some confounding because in different countries, and across airlines, interpretation of the phrases used to describe behaviours may differ. This underlines that behavioural markers in a checklist should be accompanied by a full explanation of exactly how they should be interpreted.

To narrow the set further, elements were ranked by their frequency and the 20 most common elements were identified in each of the three data sets. Table 6.4 lists these elements. Whilst 'communication' and 'decision making' remain high as elements on the checklists as well as categories, other elements come to the fore, such as 'conflict resolution'.

	UK	USA	Europe/World	overall
1	feedback	prioritisation	briefing	prioritisation
2	conflict resolution	workload distribution	planning	conflict resolution
3	communication	technical proficiency	assertiveness	communication
4	crew involved	conflict resolution	decisiveness	planning
5	information seeking	using checklists	self critique	briefing
6	situation awareness	communication	system knowledge	inquiry
7	acting decisively	decision making	team concept	checklists used
8	advocacy	preparation	workload manag.	self critique
9	decision making	vigilance	time management	preparation
10	decision communication	briefings	listening	using procedures
11	listening	distraction avoidance	anticipation	situation awareness
12	planning	use of automation	checklists used	time management
13	utilising resources	using procedures	information analysis	decisiveness
14	speaking up	cabin crew in the team	cabin crew in the team	decision making
15	team focus	planning	flight comms.	workload distribution
16	time management	communicate intentions	conflict resolution	decision communication
17	using procedures	cope with stress	acknowledge decision making	listening
18	briefings	self critique	inquiry	cabin crew in the team
19	inquiry	inquiry	prioritisation	procedure orientation
20	prioritisation	situation awareness	procedure orientation	

Table 6.4 Comparison of Most Common Elements in Lists (Ranked 1 – 20)

The most common element in UK checklists was 'feedback', whereas this element did not make the top 20 in other countries. This may reflect different emphasis on particular CRM skills in the UK as opposed to other countries. The USA data set included more elements concerning briefing than the others, whilst the European/world data set emphasised leadership and command. These distinctions may reflect either cultural differences in CRM or the current concerns of the airlines involved in the survey or simply semantic differences in labelling behaviours.

•

•

•

•

•

A complication emerged when the elements were recombined with their categories. In many cases, checklists which had the same elements, categorised them differently. For example, of the nine times that 'communication' was a checklist element, in eight of these it was assigned to a different category. For 'communication' this was not a problem because most of the categories were concerned with communication in some form but for the less frequent elements, the way they were categorised could change the interpretation of the element. For example 'involving crew members' was classified under decision taking in one list and under leadership and teamwork in another. Thus, in one case it was defined as a cognitive process, whilst in the other it was a social element (see Seamster et al, 1995b). It is possible that a given behaviour may be judged differently depending on the category it is listed under.

The most common elements were recombined with the most common categories to show the typical elements and categories of CRM markers. This is presented in Table 6.5 and it

reflects the key themes across the 31 behavioural markers lists surveyed. The aim was to produce a framework into which most of the sampled behavioural markers checklists could be fitted. Thus it represents the most typically used categories and elements from the lists compared in this chapter; for this reason the category labels also appear as elements.

	Social Behaviours		Co Beh	gnitive naviours
TEAMWORK	COMMUNI- CATION	LEADERSHIP	DECISION MAKING	WORKLOAD MANAGEMENT
management	communication	conflict resolution	attention	time management
task distribution	briefing	communicates decisions	information seeking	risk assessment
teamwork	listening	feedback	monitoring	workload management
co-operation	asking questions	authority	decision making	preparation
cabin crew involved		leadership	inquiry	prioritisation
constructive criticism		delegation	situation awareness	planning
		assertiveness	vigilance	

Table 6.5 Typical Categories and	Elements in	CRM Checklists
----------------------------------	--------------------	----------------

The resulting framework covers the two principal components of CRM skills: social and cognitive, in five categories, each with between three and seven elements. This is not intended to be a prototype marker sytem but is a representation of the most common categories (from Table 6.2) and elements (from Table 6.4) found in the lists reviewed. The categories are not distinct but overlap to a degree – 'information seeking' is very similar to 'asking questions' but is used specifically as 'asking questions to find information for decision making'. Other debates about the categorisation of elements could be raised. However as the categories are artificial distinctions, it is suggested these would not be fruitful to pursue at this stage. What is of greater concern is that whilst these elements represent those which the CRM instructors and examiners (CRM researchers) deem to be the most important non-technical skills (indicated by the frequency with which they are used in behavioural markers lists), not all can be observed directly (e.g. risk assessment) and some would have to be inferred from particular behaviours.

As Seamster et al (1995b) discussed, interpersonal skills are directly observable. For these elements, it is a question of trainers selecting the particular behavioural examples that reflect that skill. Cognitive (i.e. thinking) skills are not directly observable. In this case, the assessor must observe a behaviour which indicates that the cognitive skill has been used. For example communicating a decision reveals decision making. In most marker systems, examples of observable behaviours are provided for each of the elements. As a 'worked example', the 13 cognitive elements from Table 6.5 are listed in Appendix 6 with suggested observable behaviours that could indicate the mental activity has taken place. The examples have been drawn from the CRM checklists provided by the airlines and therefore are behaviours that CRM trainers feel are assessable.

•

.

•

•

•

•

•

•

•

•

Two additional components, personal and non-handling technical skills which appeared in some lists are not included in Table 6.5. The personal category included the elements: fatigue, self management, self-critique, which are difficult to assess reliably on the basis of observation. It may be that this category of elements should be incorporated into CRM classroom training and LOFT sessions but not formally assessed. The non-handling skills elements were: using procedures, guidelines and paperwork. It was felt that these would be encompassed within other behaviours, such as workload management, planning or would be assessed with the technical skills.

6.4 COMPARISON WITH ESTABLISHED CRM SKILL LISTS

The framework shown in Table 6.5 is not a suggested list of behavioural markers but a distilled account of the categories and elements which are in use for CRM training and assessment around the world. As discussed in Chapter 2, few of these systems have been fully tested and verified as either reliable or valid, these just represent common opinion. As is illustrated in Appendix 6, there are many possible behavioural indicators of the same activities. It is possible that, when tested, some behavioural markers will be found to be more reliable or valid than others. A programme of empirical testing is recommended before any markers are compiled into a checklist which is used in assessment.

One method of performing an initial check on the typical categories and elements is to compare them with those behavioural markers lists or CRM skill lists which were cited as the basis for current systems. From the airline survey (Chapter 5), the checklist used most often as a basis for airlines' marker systems was the NASA/ UT LLC4 for crew assessment (Helmreich et al, 1997, see Appendix 2). There are other marker systems which have been subject to testing over a period of time, most notably the SHAPE and WMCL/WILSC systems developed by KLM for individual pilot assessment (see Antersijn & Verhoef, 1995; NOTECHS, 1997 for details). In addition, there is standard guidance on 'core' CRM skills such as ICAO (1989) or the European guidelines for a basic multi-crew co-operation course (JAA, 1997, JAR-FCL 1 Subpart F, section 2); there are also CRM behaviour lists from aircrew research projects, such as the TARGETs crew co-ordination checklist (Fowlkes et al, 1994), the list proposed by Seamster et al (1995b) and the new non-technical skills list developed by the European NOTECHS group (Appendix 1). To assess how consistent are the most common behavioural markers, the list in Table 6.5 (plus the six elements of personal skills and non-handling skills, mentioned above) was compared with these eight key lists of CRM behaviours. Of interest, was not only the degree to which these key lists were represented in Table 6.5, but also whether any behavioural markers had been overlooked.

Table 6.6 illustrates that a high percentage of the behavioural markers in each key checklist are among the most common elements generally. The NASA/UT checklist (LLC4) has the highest percentage of its elements represented (93%) which is a reflection of the number of airlines who quoted Helmreich's work as the basis for their checklist development. The NOTECHS (1997) checklist has 81% of the identified common elements.

Key Checklist	Number of behavioural markers in the typical list (Table 6.5)	Percentage of elements covered when key and typical lists are combined
LLC4	26/28 (93%)	96
ICAO	23/32 (72%)	75
SHAPE	16/22 (73%)	77
WMCL	9/11 (82%)	100
TARGET	37/42 (88%)	100
MCC	14/20 (70%)	75
NOTECHS	13/16 (81%)	100
Seamster	20/23 (87%)	96

Table 6.6 Number of common elements in key checklists

From comparison of elements in the typical set (Table 6.5), with elements in key checklists, four were found to be infrequently included in key checklists: task distribution, delegation, briefing and attention. It is possible that these four elements were subsumed under other headings - the former two may have been part of elements such as workload management, briefing being included with communication, and attention being accounted for in monitoring. In total there were 18 elements in the key checklists which were not among the typical elements. The majority of these concern social climate, such as image, culture, company knowledge, organizational issues which were not of concern for the present review. However, four of them are relevant and are part of a number of checklists, namely: making suggestions, task definition, anticipation and team responsibility. If these four elements replace the four elements in the typical list which were not included often in the key checklists, the typical framework now encompasses the views of the key players almost entirely (column 3 of Table 6.6). Of note is that all the elements in the NOTECHS system (Leadership, Cooperation, Decision Making, Situation Awareness : see Appendix 1) are now typical elements which shows by implication that many of the key checklists will also contain these skills. Therefore the NOTECHS marker system, seems to be representative of marker systems in use world wide, which is to be expected as the NOTECHS framework was developed on the basis of a literature review and analysis of established marker systems.

6.5 FORMAT OF CRM SKILL LISTS

This exercise of examining existing lists (see also section 5.3.2) suggests that for practical purposes, an ideal number of elements on a list is around 15- 20, subdivided into 3-6 categories. This number of elements will fit onto one side of A4 size paper, in a clear font, which would make them easy to read. If a rating scale was used for assessment, this was typically a five point scale, although quite a number of systems are simply checklists or two-point pass/fail ratings.

6.6 SUMMARY

In summary, this analysis and comparison of current marker systems indicates that while there are certainly differences between the categories and elements listed, these may be more apparent than real. There seems to be a core set of CRM elements represented under categories such as team work, communication, leadership, decision making and workload management. The NOTECHS prototype system appears to encompass the core categories and elements common to existing airline and research systems.

Chapter 7 Discussion

7.1 INTRODUCTION

The aim of this project was to review the development and use of behavioural markers for the assessment of CRM skills. Data were gathered through literature search, an interview survey of UK airlines and a questionnaire survey of non-UK airlines. This yielded four broad sets of data which were reviewed – the behavioural markers research to date (Chapter 2); the development of behavioural markers in the UK (Chapter 4); the use of behavioural markers world-wide (Chapter 5); and the contents of behavioural markers lists in use at present (Chapter 6). A number of themes have been drawn out of each set of data and these are discussed below in relation to CRM training and assessment, behavioural marker systems and instructor/ examiner training.

• • • •

•

•

•

•

•

•••••

7.2 CRM PROGRAMMES

The sample of 33 airlines and organisations who contributed to the present study demonstrated the range of development of CRM courses and skills evaluation world-wide. Almost all had well established CRM training courses, with LOFT being used for reinforcement of training. The majority had developed or acquired some kind of behavioural marker system but only a few airlines were running formal assessment of CRM skills using these marker systems. The Advanced Qualification Programme (AQP) in the USA is clearly encouraging the larger airlines to undertake detailed CRM skills analysis for assessment purposes.

7.2.1 CRM Training

There appeared to be some differences in the level of CRM programme development in the UK: acceptance and development may not be as advanced generally as is suggested by looking at those airlines with a high profile. Enthusiasm of CRM trainers and developers is clearly important but is not the overriding determinant of the status of CRM in a company. For the smaller operators, levels of resources may influence management commitment to CRM programmes. If the possibility of a national or European CRM behavioural markers system is considered, then the needs and resources of smaller companies to put the necessary training measures in place should be taken into account.

Several respondents from outside the UK mentioned that they were impressed by the mandatory requirement for CRM courses, believing that this provided a common conceptual core and therefore a coherence to the UK CRM programmes. This view contrasted with that of UK participants who felt that there are few commonalties between courses in UK because they had to develop them on their own, with the CAA only recommending a framework rather than specifying a syllabus. On balance, the CAA approach of giving a framework and guidelines for development but not specifying course content in detail, seems to address the problem of providing a system which is generic, but also flexible enough to allow airlines to design courses to meet their own operational requirements.

7.2.2 CRM Assessment

While there have been studies which have charted changes in pilots' attitudes following CRM, there have been few, if any, rigorous studies assessing whether there is an accompanying change in individual pilot behaviour. Nor are there any quantitative results of

CRM effectiveness in terms of accident or incident reduction or operational safety currently available in the wider research literature. Most participants felt that assessment was a useful progression for CRM training in order to gain some measure of its effectiveness. However, our UK respondents had no experience of formal assessment of CRM skills and only six non-UK airlines assessed CRM skills in a licence check. Of the airlines who do not yet assess non-technical skills, some respondents were concerned that behavioural markers are not yet well enough understood to be used for assessment.

7.3 BEHAVIOURAL MARKERS

The theoretical basis for the identification of behavioural markers for CRM skills is limited and there is a need for more empirical research on pilot and crew behaviours in relation to flight safety (Brannick et al 1997). There are research teams in the USA which are examining flight crew behaviours, such as the crew factors group at NASA Ames (e.g. Orasanu, 1997) and other USA research centres (see Edens, 1996,1997). Results from relevant projects were reviewed in Chapter 2. Underpinning research for constructing and testing behavioural markers frameworks for CRM skills assessment is scarce, with the notable exception of the investigations of Helmreich et al (1996; 1997) and Seamster et al (1995, 1997).

7.3.1 Review of Existing Marker Systems

From the airlines sampled, there appeared to be major differences in the development process for identifying markers and the construction of a rating scale and evaluation procedure. Some airlines designed these from in-house research but the majority developed their behavioural markers system with reference to an existing system, usually the LLC.

When comparing behavioural markers lists world-wide, the emphasis on particular CRM skills is seen to vary across nations. The UK tends to emphasise leadership and teamwork as two concepts, whilst European and American checklists have slightly different foci. The elements in UK lists have more in common with USA lists than with those of Europe. These differences could reflect language or cultural differences or the current concern of the airline. Helmreich et al (1990) found differences in CRM behaviours across airlines which they could not attribute definitively to either organisational differences or course emphasis. The UK lists tend to be longer, this may reflect a down side to the benefit of having behavioural markers lists from other airlines as examples. There may be a tendency to try to incorporate the 'good' bits from all the lists – a pitfall reflected in the typical framework generated in the present study.

In reviewing the entire list of behaviours that have been suggested as markers in 31 checklists, the sheer number of elements is daunting. In the UK sample alone (8 lists) there were 36 different categories containing 231 elements. This suggested that there is little correspondence between behavioural markers lists under development in the UK. However, when the entire sample lists are distilled down to the common items, there are only 68 common element themes in 18 categories. This set was further reduced to a framework of 30 markers in five categories (Table 6.5).

In addition, where there were common elements they could be categorised differently. This suggested that there was not an agreed way of labelling different behaviours or a generally accepted 'language' for behavioural markers. Therefore, it is important to clearly define and explain terms in any marker system that is developed. This point is more significant than it at first seems. The terminology may influence pilot acceptance of any scheme that is used. Some respondents to the questionnaire reported that they had reworded their behavioural markers lists to make them more comprehensible by eliminating psychological jargon.

With the number of behavioural markers elements and lists currently available, debates arise over how comparable the systems are – and whether some lists are 'better' than others from a candidate's point of view. One source of confusion over comparability may be the differences in terminology used in checklists, which underlines the importance of behavioural markers being accompanied by full explanations of what they are and how they should be used.

•

•

•

.

•

•

•

•

•

•

•

•

7.3.2 Designing a Marker System

There are few published guidelines setting out what behavioural markers should be. In some ways the proliferation of different behavioural markers (Chapter 6) shows a lack of agreement over the essential behaviours for crew effectiveness. But as commented above, there is a common core to these markers. The work of Helmreich et al (1990) and Seamster et al (1995) was used as a benchmark for the critical parameters: that elements should be labelled in straightforward language, cover one topic and be based on observable behaviours. CRM trainers reported a 'design by feel approach', using their own experience as aviators, instructors and examiners as well as that of their colleagues to develop suitable marker elements. There are three stages to developing a marker system, firstly deciding exactly what the key CRM skills are, secondly identifying the behaviours which demonstrate those skills and thirdly ensuring that the rating system is psychometrically sound, that is reliable and valid. The importance of consulting instructors and examiners to identify what are practical, observable behaviours was emphasised. The realities of the training/ assessment situations in which a behavioural markers system would be used underlined the need for a workable tool. That is, not just a list, but a marker system which has a clear procedure accompanied by a programme which provides the necessary training.

Both Seamster et al (1995) and Houle (1995) recommend that all markers on the list should be actions. This is because actions are observable and mean that the grading system can be quantified more objectively. Beyond this advice, there is little to guide someone developing a non-technical skills system in the choice of markers. The review above indicated that some elements may have a specific utility but not be appropriate for use under all circumstances. That is, certain markers may be covered in CRM training but are less useful for examination purposes e.g. fatigue. The implication is that a CRM programme may require more than one behaviour list for different purposes – with a more extensive list for training, which can include personal elements such as stress and fatigue, and a more specific list for assessment which comprises observable key skills. It was suggested in the NOTECHS (1997) report that some behaviours may be displayed differently in the simulator compared to during normal line operations, and that this may also have to be taken into account.

Marker checklists, on average, have between 14 and 40 elements. The number of elements that the facilitator has to be aware of is a concern – too many elements will result in a heavy cognitive load and consequent reduction of attention (Seamster et al, 1994). This number can be reduced by taking flight phases and dividing the behavioural markers list to select a few key elements for each phase (as in the LLC4, Helmreich et al 1997). Studies have shown that certain markers were only used by raters at certain times during the flight and therefore the list in use at any one time could be substantially reduced. However, to do this requires definition of when certain markers are useful in a flight and currently not all CRM instructors believe the phase of flight distinction is important. An alternative advocated by Seamster et al, (1994, 1995) and Fowlkes et al (1994) is to determine the key CRM behaviours associated with particular episodes of the scenario and to rate them during these specific event sets. But, there is some debate in the literature as to whether a single flight segment should be the basis for evaluation or whether particular event sets should be designed with behaviours specified for given stimulus (triggering) events. There is a limited and inconclusive research

literature available, and the best advice may be available from airlines such as Atlantic Coast (USA) who have prepared detailed in-house guidance on scenario development for CRM skills evaluation.

7.4 IDENTIFYING A SUITABLE CRM MARKER SYSTEM

One of the primary objectives of this project was to recommend a suitable marker system from those surveyed or to produce an outline based on observed good practice. The survey results demonstrated that the dominant template for airlines' marker systems was the Line/LOS checklist (LLC) designed by Helmreich et al (1997, see Appendix 2). This system has been carefully developed over a number of years, primarily for research purposes, although it is now widely used in airline's line audits of crew performance, particularly in the USA.

Our interviews with UK airlines indicated that those trainers and examiners who had studied the LLC felt that while they recognised that this was an established research tool, it would not fully meet their needs for CRM assessment. They commented, from their experience of observing CRM in LOFT sessions, on the complexity of the rating form and the psychological language used. What they wanted was a rating scale which could be easily read at a glance, in the form of an aide-memoire or checklist. It should be acknowledged that a number of the international airlines had used the LLC as a framework from which to develop a marker system to suit their own requirements and some of these are one page checklist formats. The NOTECHS group also studied the LLC and decided that it did not meet their requirements for a European assessment rating scale for pilots' nontechnical skills (NOTECHS, 1997).

Moreover, the LLC was designed to evaluate crew performance, although it does allow space for comments on individual crew members. The objective of this project was to recommend a CRM marker system which could be used to rate individual pilot CRM performance. In this case there was not a system available in the public domain which could be recommended. Several airlines have fairly well advanced assessment systems but in many cases they have not been specifically designed to assess individual pilot's CRM skills.

For these reasons it is suggested that the prototype marker system which has recently been compiled by the European NOTECHS (1997) group is considered as a possible CRM assessment method. This was designed on a theoretical basis and from a review of current marker systems. It contains four categories: Leadership and Management; Cooperation; Decision Making; Situation Awareness – with 15 component elements (see Appendix 1). A two point pass/fail rating scale is used, with 'acceptable' and 'unacceptable' as the ratings. This is because the evaluation of non-technical skills in NOTECHS has been designed to match the evaluation of technical skills where the judgement is acceptable (pass or unacceptable (fail). See NOTECHS (1997 section 1) for further discussion of this issue and for details of framework development and guidance notes. The NOTECHS system will be tested by a new European project group JARTEL, during 1998 and 1999.

The advantages of the NOTECHS framework is that it was deliberately designed for ease of use and without any requirement for psychological qualifications. In essence it is intended as a practical tool for aircrew rather than for research, which was the original objective of the LLC. Although it is not dissimilar from other systems reviewed, it has been generated by a European aviation research team and is will be tested within European airlines.

One amendment to the design of NOTECHS is advised. Given the predominance of 4 and 5 point rating scales for CRM evaluation (see above at 2.3 and 5.3.2), then it is suggested that

the CAA might wish to consider a more finely grained rating scale (such as 5 points) to be used with the NOTECHS categories and elements. This would allow a more sensitive discrimination of standards of performance and would generate more useful data during the trial phase of instrument development. The scale points could range from unsatisfactory, satisfactory, average/standard, above average, excellent.

7.5 TRAINING FOR CRM ASSESSMENT

As stated above, training for CRM instructors and examiners is a critical component of a CRM evaluation programme. This training needs to encompass scenario design, assessment procedures and facilitation skills. Most of the participants who were developing behavioural markers checklists were introducing new courses in facilitation training for LOFT and LOE (Dismukes et al, 1997; McDonnell et al, 1997). One important research focus for the near future will be to establish how inter-rater reliability and standardisation can be maximised, i.e. what combination of training and behavioural markers tools produces the greatest accuracy and reliability. Where jeopardy assessment is introduced, then the question of training and certification for examiners who will assess CRM skills needs to be addressed.

•

•

•

•

•

•

•••••

7.6 GUIDANCE FROM REGULATOR

To develop a standard framework requires the involvement of a regulatory body i.e. the CAA. Participants commented that they were looking to the UK CAA for advice and guidance on the development of behavioural markers and facilitator training but that it was not always forthcoming. Lack of guidance was cited as a reason for some airlines postponing their development of their CRM programme, whilst for those forging ahead, it is a concern that they will develop a system and then have to revise it if the CAA takes a definitive stance.

7.7 SUMMARY

The discussion of the survey findings highlights the current differences in airlines' development and use of marker systems. The need to establish a common template or framework of core skills and the importance of testing marker systems to ensure that they are psychometrically sound were also emphasised. A central theme to emerge was the need to train instructors and examiners who will be using these marker systems for training and assessment purposes.

Chapter 8 Conclusions

Despite considerable differences in the detail of current marker systems for nontechnical skills, a set of core CRM skills can be identified. These form the basis of almost all the main marker systems and behaviour lists and can be subdivided into cognitive and social skills, with typical elements being leadership, team working, decision making and situation awareness.

Whatever system is adopted, behavioural markers should be targeted at observable actions and should be presented in a direct and simple manner. Consideration needs to be given to how terms are defined and explained, the format of the record sheet and the type of rating scale.

The principal objective of this project was to recommend the adoption of an existing CRM behavioural marker system or to produce an outline for a new marker system based on observed good practice. As discussed in the earlier chapters, there is a wide range of systems being used, but most of them have been based on the crew behaviour observation system Line /LOS Checklist (LLC). For assessing individual pilot CRM performance, there is not a comparable system in the public domain, which has been developed to the same extent as the LLC; although some airlines, such as KLM (Antersijn & Verhoef, 1995, see NOTECHS, 1997), have new marker systems which have been carefully researched and are being trialled. An analysis of existing systems conducted in Chapter 6, produced a similar set of typical CRM elements to the NOTECHS CRM skills framework (see Appendix 1). This will form the prototype for testing as a possible European system for assessing pilots' non-technical skills.

From the interviews and questionnaire responses, the main conclusions are that participants would like to see more commonality between the systems that are developing. In the UK, additional guidance on the design and use of marker systems would be welcomed. It was felt that behavioural markers are not a stand-alone assessment device but should form part of ongoing pilot training, as well as CRM research and development. Moreover, the need to train instructors and examiners in the use of any CRM behaviour markers system was strongly emphasised.

A number of concerns regarding the methods and use of CRM assessment were expressed by pilots, particularly in relation to licence failures based on an evaluation of non-technical skills. This issue is specifically addressed in the NOTECHS (1997) report.

Chapter 9 Recommendations

• The draft NOTECHS framework (Appendix 1) would appear to be the most promising system for the assessment of an individual pilot's CRM skills. Accordingly, it is recommended that the NOTECHS (1997) behavioural markers framework (or some comparable version) is tested for possible use as an individual non-technical skills assessment instrument. It cannot be emphasised too strongly that the NOTECHS system has not been subject to any empirical examination. The evaluation study began in 1998 with the new EC DGVII JARTEL project (see Chapter 1) and the results will be reported in late 1999. •

•

•

•

- Issues relating to the validity and reliability of marker systems should be carefully considered, particularly with regard to inter-rater reliability and into the viability of different elements and checklist formats.
- Behavioural markers should be treated as part of the whole CRM programme and should be introduced at the earliest stages of training.
- Training and accreditation should be provided for instructors and examiners who will use CRM behavioural marker systems for training and assessment.

References

Aiken, L. (1996) Rating Scales and Checklists. New York: Wiley.

- Antersijn, P. & Verhoef, M. (1995) Assessment of non-technical skills: Is it possible? In N. McDonald, N. Johnston & R. Fuller (Eds.), Applications of Psychology to the Aviation System, Aldershot: Averbury.
- Azjen, I. & Fishbein, M. (1980) Understanding Attitudes and Predicting Social Behaviour. New Jersey: Prentice Hall.
- Birnbach, R. & Longridge, T. (1993) The regulatory perspective. In E. Wiener, B. Kanki & R. Helmreich (Eds.) Cockpit Resource Management. San Diego: Academic Press.
- Boeing (1994) Statistical Summary of Commercial Jet Aircraft Accidents. World-wide Operations 1959-1993. Seattle: Boeing.
- Brannick, M. & Prince, C. (1997) An overview of team performance measurement. In M. Brannick, E. Salas & C. Prince (Eds.) *Team Performance, Assessment and Measurement*. New Jersey: LEA.
- Brannick, M. & Prince, C. (1995) Reliability of measures of aircrew skills across events and scenarios. In *Proceedings of the 8th International Symposium of Aviation Psychology*. Ohio: OSU.
- Brannick, M. & Prince, C. (1991) Assessment of aircrew rating from within and between scenarios. Technical Report DAAL0-3-86-D-001. Orlando: Naval Training Systems Center.
- Brannick, M., Salas, E. & Prince, C. (1997) (Eds.) Team Performance, Assessment and Measurement. New Jersey: LEA.
- Butler, R. (1991) Lessons from cross-fleet/cross-airline observations: evaluating the impact of CRM/LOFT training. In *Proceedings of the* 6th International Symposium of Aviation Psychology. Ohio: OSU.
- CAA (1993) Crew Resource Management (AIC 143/1993, Pink 90, September) London: UK Civil Aviation Authority.
- CAA (1995) Crew Resource Management Training (AIC 37/1995, Pink 110, May). London: UK Civil Aviation Authority.
- Connelly, P. (1997) A resource package for CRM developers: Behavioural markers of CRM skill from real world case studies and accidents. Technical Report 97-3. Austin: University of Texas, Aerospace Research Project.
- David, G. (1996) Lessons from offshore aviation: Toward an integrated human performance system. In R. Flin & G. Slaven (Eds.) *Managing the Offshore Workforce*. Tulsa: Pennwell.
- Dickinson, T. & McIntyre, R. (1997) A conceptual framework for teamwork measurement. In M. Brannick, E. Salas & C. Prince (Eds.) Team Performance, Assessment and Measurement. Mahwah, NJ: LEA.
- Dismukes, K., Jobe, K. & McDonnell, L. (1997) LOFT Debriefings: An Analysis of Instructor Techniques and Crew Participation. NASA Technical Memorandum 110442. NASA Ames.
- Dutra, L., Norman, D., Malone, T., McDougall, W. & Edens, E. (1995) Crew Resource Management/assessment: Identification of key observable behaviours. In Proceedings of the 8th International Symposium of Aviation Psychology. Ohio: OSU.
- Edens, E. (1996,1997) Air carrier training research review. Washington: Office of the Chief Scientific and Technical Advisor for Human Factors, AAR-100.
- Endsley, M. (1995) Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32-64.
- FAA (1989) Advisory Circular 120-51: Cockpit Resource Management Training. Washington DC: Federal Aviation Administration.
- FAA (1991) Advisory Circular 120-54: Advanced Qualification Program. Washington, DC: Federal Aviation Administration.
- Flin, R. (1996) Sitting in the Hot Seat: Leaders and Teams for Critical Incident Management. Chichester: Wiley.

Fowlkes, J., Lane, N., Salas, E., Franz, T. & Osler, R. (1994) Improving the measurement of team performance: The TARGETS methodology. *Military Psychology*, 6 (1), 47-61.

Furnham, A. (1997) The Psychology of Behaviour at Work. Sussex: Psychology Press.

- Green, R.G., Muir, H., James, M., Gradwell, D. & Green, R. L. (1996) Human Factors for Pilots. (2 ed). Aldershot: Avebury.
- Gregorich, S., Helmreich, R. & Wilhelm, J. (1990) The structure of cockpit management attitudes. Journal of Applied Psychology, 75, 682-690.

Gregorich, S. & Wilhelm, J. (1993) Crew resource management training assessment. In E. Wiener, B. Kanki & R. Helmreich (Eds.) Cockpit Resource Management. San Diego: Academic Press.

Hamman, W. (1997) Quick Reference Checklist (QRC) United Airlines Validation study. Study report. Washington: FAA.

Hawkins, F. & Orlady, H. (1993) Human Factors in Flight. (2 ed). Aldershot: Avebury.

- Helmreich, R. (1996) The evolution of crew resource management. Paper presented at the IATA Human Factors Seminar, Warsaw, Poland. (Available on Aerospace Project Website http://www.psy.utexas.edu/psy/helmreich/nasaut.htm).
- Helmreich, R., Butler, R., Taggart, W. & Wilhelm, J. (1997) The NASA/ University of Texas/ FAA Line/LOS Checklist: A behavioral-based checklist for CRM skills assessment. (Version 4.4 dated 1.2.97). Austin: NASA/ UT/ FAA Aerospace Group.
- Helmreich, R., Hines, W. & Wilhelm, J. (1996) Issues in Crew Resource Management and automation use: Data from line audits. Paper presented at the 6th CRM Industry Workshop, Charlotte, NC, 18 September.
- Helmreich, R., Wilhelm, J., Gregorich, S. & Chidester, T. (1990a) Preliminary results from the evaluation of cockpit resource management training: Performance ratings of flight crews. *Aviation, Space and Environmental Medicine*, 61, 576-579.
- Helmreich, R., Wilhelm, J., Kello, J., Taggart, W. & Butler, R. (1990b) Reinforcing and evaluating crew reseasource management: Evaluator/LOS instructor reference manual. Austin: NASA/UT Technical Manual 90-2.
- Houle, T. (1995) An operational model for the evaluation of Crew Resource Management (CRM) skills in Line Operations Simulation (LOS). In N. McDonald, N. Johnston & R. Fuller (Eds.), *Applications of Psychology to the Aviation System.* Aldershot: Avebury.
- ICAO (1989) Flight Crew Training: CRM & LOFT. Human Factors Digest, 2. Montreal: ICAO.
- JAA (1997) JAR-FCL Part 1 (Aeroplane) Flight Crew Licensing Requirements. Hoofddorp: JAA.
- JAA (1996) JAR-OPS. Flight operations. Hoofddorp: JAA.

Jensen, R. (1995) Pilot Judgment and Crew Resource Management. Aldershot: Avebury.

- Jentsch, F., Sellin-Wolters, S., Bowers, C. & Salas, E. (1995) Crew coordination behaviors as predictors of problem detection and decision making times. In *Proceedings of the Human Factors and Ergonomics Society 39th Annual Meeting*. San Diego: HFES.
- Johnston, N., Fuller, R. & McDonald, N. (1995) (Eds.) Aviation Psychology. Training and Selection. Aldershot: Avebury.

- Kaempf, G. & Klinger, D. (1992) Integrated measurement of crew resource management and technical flying skills. Final technical report, Battelle Memorial Institute, Fairborn, OH. Klein Associates.
- Lanzano, J., Seamster, T. & Edens, E. (1997) The importance of CRM skills in an AQP. In Proceedings of the 9th International Symposium of Aviation Psychology. Ohio: OSU.
- Lauber, J. (1984) Resource management in the cockpit. Air Line Pilot, 53, 20-23.
- Law, J. & Sherman, P. (1995) Do raters agree? Assessing inter-rater agreement in the evaluation of air crew resource management skills. In *Proceedings of the 8th International Symposium of Aviation Psychology*. Ohio: OSU.
- Law, J. & Wilhelm, J. (1995) Ratings of CRM skill markers in domestic and international operations: A first look. In *Proceedings of the 8th International Symposium of Aviation Psychology*, Ohio: OSU.
- McDonald, N., Johnston, N. & Fuller, R. (1995) (Eds.) Applications of Psychology to the Aviation System. Aldershot: Avebury.

McDonnell, L., Jobe, K. & Dismukes, K. (1997) Facilitating LOS Debriefings: A Training Manual. NASA Technical Memorandum 112192. NASA Ames.

- Maschke, P., Goeters, K., Hormann, H. & Schiewe, A. (1995) The development of the DLR/ Lufthansa Crew Resource Management traing. In N. Johnston, R. Fuller & N. McDonald (Eds.) Aviation Psychology: Training and Selection. Aldershot: Avebury.
- Naef, W. (1995) Practical application of CRM concepts: Swissair's Human Aspects Development program (HAD). In *Proceedings of the 8th International Symposium of Aviation Psychology*. Ohio: OSU.
- NOTECHS (1997) The Evaluation of Non-Technical Skills of Multi-Pilot Aircrew. Project Report. Amsterdam: NLR.
- NTSB (1991) Annual Review of Aircraft Accident Data: US Carrier Operations Calendar Year 1988, NTSB/ARC-91/01. Washington, DC: NTSB.
- NTSB (1994) Safety Study. A Review of Flight Crew-Involved Major Accidents of US Air Carriers 1978-1990, NTSB/S594/01. Washington, DC: NTSB.
- Oransanu, J. (1997) Stress and aeronautical team decision making. In R. Flin, E. Salas, M. Strub & L. Martin, (Eds.), *Decision Making Under Stress*. Aldershot: Avebury.
- Orasanu, J. (1994) Shared problem models and flight crew performance. In N. Johnston, N. McDonald & R. Fuller (Eds.) Aviation Psychology in Practice. Aldershot: Avebury.
- Pettit, M. & Dunlap, J. (1997) Cockpit leadership and followership skills: Theoretical perspectives and training guidelines. Project report. Washington: FAA.
- Reason, J. (1997) Managing the Risks of Organisational Accidents. Aldershot: Avebury.
- Seamster, T. & Edens, E. (1993) Cognitive modelling of CRM assessment expertise: Identification of the primary assessors. In Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting. San Diego: HFES.
- Seamster, T., Edens, E., McDougall, W. & Hamman, W. (1994) Observable crew behaviors in the development and assessment of Line Operational Evaluations (LOE'S). Study Report DTFA01-93-C-00055. Washington: FAA.
- Seamster, T., Edens, E. & Holt, R. (1995) Scenario event sets and the reliability of CRM assessment. In Proceedings of the 8th International Symposium of Aviation Psychology. Ohio: OSU.
- Seamster, T., Hamman, W. & Edens, E. (1995) Specification of observable behaviors within LOE/LOFT event sets. In *Proceedings of the 8th International Symposium of Aviation Psychology*, Ohio: OSU.
- Seamster, T., Prentiss, F. & Edens, E. (1997) Methods for the analysis of CRM skills. In Proceedings of the 9th International Symposium of Aviation Psychology. Ohio: OSU.
- Taggart, W. (1991) Advanced CRM training for instructors and evaluators. In *Proceedings of the 6th* International Symposium of Aviation Psychology, Ohio: OSU.
- Taggart, W. (1995) The NASA/UT/FAA Line/LOS Checklist: Assessing system safety and crew performance. In *Proceedings of the 8th International Symposium on Aviation Psychology*. Ohio: OSU.
- Wiener, E., Kanki, B. & Helmreich, R. (1993) (Eds.) Cockpit Resource Management. San Diego: Academic Press.
- Weiner, E. & Nagel, D. (1988) (Eds.) Human Factors in Aviation. San Diego: Academic Press.
- Wilhelm, J. (1991) Crew member and instructor evaluations of line oriented flight training. In *Proceedings of the 6th International Symposium of Aviation Psychology*. Ohio: OSU.
- Williams, D., Holt, R. & Boehm-Davis, D. (1997) Training for inter-rater reliability: Baselines and benchmarks. In Proceedings of the 9th International Symposium on Aviation Psychology. Ohio: OSU.

Acknowledgements

The authors would like to thank all those who participated in the study, for giving up their time and providing copies of their course materials and checklists. We are particularly grateful to Captain Glyn David of Bond Helicopters for his advice, Captains Graham Dainty and Chris Cane of Bristow Helicopters for opening their simulator and facilitator training sessions to us; Roseanne Beale of British Airways for travelling to Aberdeen to spend a day with us and Professor Bob Helmreich and his team at University of Texas, Austin for their generous help and hospitality during our visit.

•

•

•

•

.

Non technical Skills (NOTECHS) Framework: Categories and Elements

•

•

•

••••

•

•

•

•

••••

•

•

•

•

••••

•

•

•

•

•

CATEGORIES	ELEMENTS
N. SCANE AND ST	Team building and maintaining
Co-operation	Consideration of others
	Supporting others
	Conflict solving
	Use of authority/assertiveness
Leadership and	Maintaining standards
managerial skills	Planning and co-operation
	Workload management
	System awareness
Situation awareness	Environmental awareness
	Anticipation
	Problem definition/diagnosis
Decision	Option generation
making	Risk assessment/option choice
	Outcome review



48

.

-



APPENDIX 1C Non-technical Skills (NOTECHS) Framework: Elements & Behaviours for Category - Situation Awareness

APPENDIX ID Non-technical Skills (NOTECHS) Elements and Behaviours for Category - Decision





••••• • • •

UT/FAA LINE/LOS CHECKLIST, VERSION 4.4

Observer is to complete one form for each flight segment

	1
Airline	-
Date (Mo. Yr.)	-
Observer ID	0
Route	
A/C Type & Series	-
Hrs. Observed	

		D	ls.
LOE	LOFT	Scenario	Line Ob

Demographics	Capt.	1st Off	Engineer
Domicile			
Years of experience - all airlines			
Years in position - this A/C			

26. LAG INU. VI
CB: TCG INU. UI

	K One	DUX
First leg flown together		More than one day flown together
First day flown together		

CREW PERFORMANCE RATING BY PHASE OF FLIGHT

ndard - The demonstrated behavior Outstand	I Standard - The demonstrated behavior Outstand	Observed Standard - The demonstrated behavior Outstand	ns - Observed Standard - The demonstrated behavior Outstand	ectations - Observed Standard - The demonstrated behavior Outstand	Expectations - Observed Standard - The demonstrated behavior Outstand	inimum Expectations - Observed Standard - The demonstrated behavior Outstand	is Minimum Expectations - Observed Standard - The demonstrated behavior Outstand	Minimum Exnectations - Observed Standard - The demonstrated behavior Outstand	performance is Minimum Expectations - Observed Standard - The demonstrated behavior Outstand	I The demonstrated helpavior Observed Standard The demonstrated helpavior Outstand	Contract of the second se
ndard - The demonstrated behavior Outs	I Standard - The demonstrated behavior Outs	Observed Standard - The demonstrated behavior Outs	ns - Observed Standard - The demonstrated behavior Outs	ectations - Observed Standard - The demonstrated behavior Outs	I Expectations - Observed Standard - The demonstrated behavior Outs	inimum Expectations - Observed Standard - The demonstrated behavior Outs	is Minimum Expectations - Observed Standard - The demonstrated behavior Out	Minimum Expectations - Observed Standard - The demonstrated behavior Out	nerformance is Minimum Expectations - Observed Standard - The demonstrated behavior Out	Desved Standard The demonstrated helavior Out	And the second behavior of the demonstration of the
motec and maintaine crew ever	and an	and an and an and a second sec	minimim momotee and maintaine craw ave	and more anisterious line and						TYEE DETTOFTIATION IS A MALINIMUM PAPECIAUUUS - UDSCIVEU STATIUMU - 1110 UCHININA ALCO DETTOFTIATION OF A MALINIMUM PAPECIAUUUS	Observed performance is Minimum Expectations - Observed Standard - The demonstrated behavior Out
	n promotes and maintains crew excep	minimum promotes and maintains crew excep	and minimum promotes and manimum crew creep	meets minimum promotes and maintains crew excep	ce meets minimum promotes and maintains crew excep	rformance meets minimum promotes and maintains crew excep	his performance meets minimum promotes and maintains crew excep	ns. This performance meets minimum promotes and maintains crew excep	bectations. This performance meets minimum promotes and maintains crew excep	ow expectations. This performance meets minimum promotes and maintains crew excep	ly below expectations. This performance meets minimum promotes and maintains crew excep
ctiveness. This is the level of sp	r effectiveness. This is the level of sp	room for effectiveness. This is the level of sp	ample room for effectiveness. This is the level of sp	there is ample room for leffectiveness. This is the level of sp	nts but there is ample room for effectiveness. This is the level of sp	mirements but there is ample room for effectiveness. This is the level of sp	requirements but there is ample room for leffectiveness. This is the level of sp	necessary requirements but there is ample room for effectiveness. This is the level of sp	where necessary requirements but there is ample room for leffectiveness. This is the level of sp	where necessary requirements but there is ample room for leffectiveness. This is the level of sp	instances where necessary requirements but there is ample room for effectiveness. This is the level of sp
formance that should be normally	herformance that should be normally	formance herformance that should be normally	d of nerformance herformance that should be normally	his level of nerformance nerformance that should be normally	ant This level of nerformance herformance that should be normally	purcurents out under is anapte room for performance that should be normally	as improvement This level of nerformance herformance that should be normally	accounted improvement. This level of nerformance herformance that should be normally	where increasedy requirements out mere is ample room for preformance that should be normally	to and eventuates immovement. This level of performance incremence that should be normally	instances where necessary requirements out more is ample room for performance but where that should be normally
formance that should be	e performance that should be	formance performance that should be	I of performance performance that should be	his level of performance performance that should be	ent. This level of performance performance that should be	provement. This level of performance performance that should be	les improvement. This level of performance performance that should be	examples improvement. This level of performance performance that should be	nt, and examples improvement. This level of performance performance that should be	t present, and examples improvement. This level of performance performance that should be	vas not present, and examples improvement. This level of performance performance that should be
formance that sho	e performance that sho	formance performance that sho	I of performance performance that sho	his level of performance performance that sho	ent. This level of performance performance that sho	provement. This level of performance performance that sho	les improvement. This level of performance performance that sho	examples improvement. This level of performance performance that sho	int, and examples improvement. This level of performance performance that sho	t present, and examples improvement. This level of performance performance that sho	vas not present, and examples improvement. This level of performance performance that sho
ctiveness. ormance	r effectiveness.	formance performance	s ample room for effectiveness	there is ample room for effectiveness his level of performance performance	effectiveness ent. This level of performance desired for effective crew	provement. This level of performance performance	requirements but there is ample room for effectiveness improvement. This level of performance performance	necessary requirements but there is ample room for effectiveness examples improvement. This level of performance performance	where necessary requirements but there is ample room for effectiveness nt, and examples improvement. This level of performance performance	t present, and examples improvement. This level of performance performance in the intervences.	instances where necessary requirements but there is ample room for effectiveness vas not present, and examples improvement. This level of performance performance is how and the desired for effective room for here in f
	r effe	room for effe formance pert	s ample room for effe	there is ample room for effe his level of performance perf	the meets minimum pro- nts but there is ample room for effe ent. This level of performance performance	rformance meets minimum pro quirements but there is ample room for effe provement. This level of performance perf	his performance meets minimum pro ary requirements but there is ample room for effe les improvement. This level of performance perf	ns. This performance meets minimum pro- necessary requirements but there is ample room for effe examples improvement. This level of performance perf	pectations. This performance meets minimum pro- where necessary requirements but there is ample room for effe nt, and examples improvement. This level of performance perf	ow expectations. This performance meets minimum pro- ces where necessary requirements but there is ample room for effe t present, and examples improvement. This level of performance perf	ly below expectations. This performance meets minimum pro- instances where necessary requirements but there is ample room for effe- vas not present, and examples improvement. This level of performance perf-

The following performance markers are specific behaviors that serve as indicators of how effectively resource management is being practiced. They are not intended to be exhaustive lists of behaviors that should be seen, but rather as exemplars of behaviors associated with more and less effective crew resource management. It is not expected that all behaviors will be seen for every phase of a specific flight. When performance is rated either as (4) or (1), please describe the causes for the <u>specific</u> rating in the COMMENTS section. Be concise. General comments are to be made at the end of this form.

Team Management & Crew Communications	Pre - Depart	T/O & Climb	Cruise	Des/Appr Landing	SPECIFIC COMMENTS
 Team concept and environment for open communications established and/or maintained, e.g., crewmembers listen with patience, do not interrupt or "talk over", do not rush through the briefing, make eye contact as appropriate. 					
 Briefings are operationally thorough, interesting, and address crew coordination and planning for potential problems. Expectations are set for how possible deviations from normal operations are to be handled, e.g., rejected T/O, engine failure after lift-off, go-around at destination. 					
 Cabin crew is included as part of team in briefings, as appropriate, and guidelines are established for coordination between flight deck and cabin. Passengers are briefed and updated as needed, i.e., delays, weather, etc. 					
 Group climate is appropriate to operational situation, e.g., presence of social conversation at appropriate times. Crew ensures that non-operational factors such as social interaction do not interfere with necessary tasks. 					

COMMENTS Des/Appr Landing Cruise T/O & Climb Pre-depart of automation is reduced or disengaged, or automation is effectively used to reduce well in advance who is flying the leg, establish responsibility for non-routine types Crewmembers demonstrate high levels of vigilance in both high and low workload e.g., effective Operational decisions are clearly stated to other crewmembers and acknowledged, crew members. Adequate time is provided for completion of tasks, e.g., establish Crewmembers speak up, and state their information with appropriate persistence, Captain coordinates flightdeck activities to establish proper balance between command authority and crew member participation, and acts decisively when the 13. Guidelines are established/followed for the operation of automated systems, i.e., when systems will be disabled, programming actions that must be verbalized and Automated systems are used at appropriate levels, i.e., when programming demands could reduce situational awareness and create work overloads, the level Secondary operational tasks are prioritized so as to allow sufficient resources for dealing effectively with primary flight duties, e.g., dealing with passenger needs, conditions, e.g., active monitoring, scanning, cross-checking, attending to radio inquiry about uncertainty of clearance limits, clarification of confusing/unclear Workload and task distribution is clearly communicated and acknowledged by PF and PNF duties and responsibilities with regard to automated systems are Crewmembers periodically review and verify the status of aircraft automated Crew plans for sufficient time prior to maneuvers for programming of Flight Crew prepares for expected or contingency situations including approaches, Crewmembers verbalize and acknowledge entries and changes to automated until there is some clear resolution and decision, e.g., effective advocacy & and include cabin crew and others when appropriate, e.g., good cross-talk systems, e.g., optimum cruise level, correct profile for active runway. Crewmembers ask questions regarding crew actions and decisions, Team Management & Crew Communications (Cont.) calls, switch settings, altitude callouts, crossing restrictions established/followed, e.g., FMS entry and cross-checking. Situational Awareness & Decision Making assertion: "I'm uncomfortable with..., Let's.. between pilots, everyone "on same page. weather, etc. e.g., stays "ahead of curve". crew meals, company communications. **Automation Management** Management Computer. systems parameters. of communications. ATC instructions. situation requires. acknowledged. workload. 10. 11. 12. 14. 15. 16. 17. 18. 5. 2. 6 9 ŝ

. • • • • • • • • • • • • . • .

 Positive and negative performance feedback is given a made a positive learning experience for the whole crew objective, based on observable behavior, and given contakeoffs and/or landings. Performance feedback is accepted objectively and non-takeoffs and/or landings. When conflicts arise, the crew remains focused on the hand. Crewmembers listen actively to ideas and opinit when wrong, conflict issues are identified and resolved activity, caffeine management, walking through the care activity, caffeine management, walking through the care activity. 		Landing	
 Performance feedback is accepted objectively and non- 21. When conflicts arise, the crew remains focused on the hand. Crewmembers listen actively to ideas and opinit when wrong, conflict issues are identified and resolved 22. Crewmembers recognize fatigue and take specific stepi alertness, e.g., use of fatigue countermeasures such as activity, caffeine management, walking through the ca			
 When conflicts arise, the crew remains focused on the hand. Crewmembers listen actively to ideas and opinit when wrong, conflict issues are identified and resolved Crewmembers recognize fatigue and take specific step alertness, e.g., use of fatigue countermeasures such as activity, caffeine management, walking through the ca 			
22. Crewmembers recognize fatigue and take specific step alertness, e.g., use of fatigue countermeasures such as activity, caffeine management, walking through the ca			
23. Crew actions avoid the creation of self-imposed workly late descents due to lack of situational awareness/plant			
24. Crewmembers recognize and report work overloads in "I'm getting loaded up here; Can you take over"			
25. When appropriate, crewmembers take the initiative and knowledge and experience, i.e., new: crewmembers, ro			

Te	chnical Proficiency	Pre - Depart	T/O & Climb	Cruise	Des/Appr Landing	COMMENTS
26.	Sterile cockpit SOP used in an appropriate manner, e.g., keep social conversation during taxi and below 10,000 ft, to a minimum, avoid distractions during these phases.			XXX		
27.	Altitude and terrain awareness, e.g., verbalize and acknowledge altitude clearances and changes, make proper callouts, demonstrate awareness of significant terrain.					
28	Checklist compliance , e.g., use proper normal and abnormal checklists in the prescribed manner, do not read from memory, look at items being read, complete checklist and call out completion.					

Overall Crew Rating	Il technical proficiency	Il crew effectiveness

Overall Rating	her systems events that occur during flight.		ronment, e.g., WX, ATC, Traffic, MEL's,	ent on conditions affecting flight ->.
nsiderations	erity of abnormals and other sys	ited 1=low to 4=high.	mplexity of operating environme	1 1=low to 4=high.Comment on

In those cases where the actions of a particular crewmember may be particularly significant to the outcome of the observed behavior, enter the relevant item number from the LLC, the crew position involved, the rating assigned, check one or more phases of flight, and include supporting

1		5.5 Store 1.5			 	
Comments						
Flight	Des/Appr Landing			yment		
se(s) of	Cruise			ght sec		
te Phas	T/O & Climb			this fli		
Indica	Pre- Depart			nts on		
	Rating			omme		
	Crew Position			tional C		
				5	1.1.5	

APPENDIX 3

SEMI-STRUCTURED INTERVIEW SCHEDULE

QUESTIONS FOR CRM STUDY

Introduction

C

We are conducting a study of CRM and in particular, behavioural markers used in CRM training and assessment. We would like to ask you some questions about your behavioural marker schemes.

- 1. Have you introduced behavioural markers into your CRM course?
- 2. When did you introduce behavioural markers into your CRM course?
- 3. What behavioural markers do you use on your CRM courses?
- 4. What was the basis of your behavioural markers system? (eg. LLC?)
- 5. Where did this group draw the behavioural markers from? (Prompt what elements did they base their behavioural markers on?)
- 6. Why are these elements important to good CRM?
- 7. For each element:

How do you teach this element to students on a CRM course?

8. For each element:

What behaviours do you look for in LOFT to show that CRM students have learned the principles you are teaching them?

- 9. What feedback do you give to discuss these behaviours with your CRM students?
- 10. Do you consider behavioural (social) skills and technical skills during a single LOFT session?
- 11. How do you train/help your trainers to look for behaviours which indicate good CRM?
- 12. How do you train/help your trainers to use behavioural markers?
- 13. How does your organisation show its support for your use of behavioural markers?
- 14. How do your CRM students react to your use of behavioural markers?
- 15. What do CRM students think of you using behavioural markers to indicate CRM behaviour changes?



APPENDIX 4



Department of Psychology

University of Aberdeen

CAA CRM Behavioural Markers Questionnaire

Introduction

This questionnaire asks about the behavioural markers (crew performance indicators) system, for non-technical/ CRM skills, that is in place in YOUR AIRLINE. We have filled in answers to as many of the items as we can. Please could you complete the other items as fully as possible and return the whole to us. All information you include will be de-identified before it is included in any of our reports or publications.

General information about the marker system YOUR AIRLINE uses

- 1.1 Name of behavioural markers (BM) system ____
- 1.2 Who developed the BM system?
- 1.3 If YOUR AIRLINE developed the system itself what year was development begun?
- 1.4 What was the research basis (e.g. Helmreich's NASA/UT work, LCII4)?

1.5 When was the system first used in YOUR AIRLINE? _____

1.6 How was the initial concept of the BM system introduced?

During CRM courses During other courses Through posters Through LOFT sessions As wallet note cards Other, please specify

 1

1.7 What changes have you made to your BM system since it was introduced?

Details of behavioural marker system

2.1 List the markers in your BM system, or enclose a copy of your BM sheet.

•

•

.

•

•••••

2.2 Does it have a rating scale?

±

Yes... No...

2.3 How many points are on the scale? 2 3 4 5 6 other ____

2.4 What are the labels for the points e.g. poor, standard, excellent?

- 2.5 Why was this type of scale chosen?
- 2.6 Does the system consider technical skills with non technical skills?

Yes... No... ±

Training Instructors/ Examiners to use the BM system

3.1 What training do instructors/examiners receive to use the BM system?

3.2 How long are the courses for:

Examiners	
Instructors	

3.3 Do instructors/ examiners follow a protocol when using the BM system?

Yes... No... ±

Status of BM system

0

0

•

•

•

•

•

.

0

•

.

0

0

0

•

4.1 What information is given to pilots to explain how the BM system will be used?

4.2 What is the BM system used for within YOUR AIRLINE (please tick as many as are relevant)?

Ab-initio training		
Recurrent training		
LOFT sessions		
LOE assessment		
Line/ licence checks		
Other, please specify		

4.3 If the BM system is used during pilot training, please give brief details below.

4.4 If your BM system is used during checks, could a pilot fail this check based on their CRM skills?

Yes	No	±

4.5 Are pilots reassessed if they fail the check?

Yes... No... ±

•

•••••

•

•

•

• • • • • • •

4.6 When are they reassessed?

4.7 What feedback do pilots receive about their ratings on the BM scale?

4.8 Following feedback, do pilots have an option for further training or practice? Yes... No... ±

4.9 How is the output or result from using the BM system filed? e.g. does a report go onto a pilot's personal file?

4.10 Are there any further consequences not dealt with above?

Pilots' views

5.1 What feedback have you received from pilots who have taken part in courses or assessments where YOUR AIRLINE's BM system is used?

5.2 What is the level of acceptance of the BM system by pilots? (Please tick one)

High acceptance	
Acceptance	
Cautious acceptance	
Rejection	

5.3 Are you aware of any problems perceived by your pilots with YOUR AIRLINE'S BM system?

Yes... No... ±

5.4 If yes, what are they?

•

0

0

•

•

•

•

•

0

•

•

0

0

•

•

•

•

Future developments

6.1 How do you think BM will develop and be used in YOUR AIRLINE in the future?

Any other comment

7.1 Add any further comments you would like to make on behalf of YOUR AIRLINE.

Thank you for your assistance.



APPENDIX 5

LIST OF CONTRIBUTORS TO THE SURVEYS

•

0

0

0

•

•

•

•

•

.

•

•

•

•

•

•

•

0

•

•

•

.

•

0

•

0

Aer Lingus* Aerospace Crew Research Group, University of Texas (Helmreich) * Air Atlantique Air Canada* Air France* ALPA American Airlines* Ansett Australia* Atlantic Coast* BALPA **Bond Helicopters** British World Airlines **Braathens SAFE** Bristow Helicopters* Britannia* British Airways* British International Helicopters* Carlow International* Cathay Pacific* Continental* Delta* DLR FAA Iberia* ICAO **IFALPA IMASSA** KLM* LMQ (CRM Trainers) Lufthansa* Magec Monarch NASA Ames NLR NOTECHS* Northwest* RAF* Scandinavian Air Services (SAS)* Shell Aircraft US Naval Airwarfare Training Systems Center* Virgin Atlantic Airways

* Indicates organizations who provided copies of their Behavioural Markers Lists


Example Behaviours for Cognitive Elements of Categories:

Decision Making and Workload Management

•

•

•

•

•

0

•

•

•

•

•

•

•

•

•

•

•

•

•

•

•

cognitive	behavioural indicators	other examples of
element		behavioural indicators
attention	answer questions	attentive listening verify automation verbalise entries and changes
information seeking	liaison with outside resources	encourage participation
monitoring	review systems status with crew	changes openly noted ensure crew knows plans
decision making	brief on intent bottom line established	state handling of deviations respond to safety concerns decision stated and acknowledged
inquiry	ask opinions	
situation awareness	check weather reports} check traffic reports} check fuel status} verify environmental constraints	debrief events review a/c and maintenance log
vigilance	acknowledge understanding	crew members state recommendations
time management	performance in accord with plans	program automation beforehand
risk assessment	discuss discrepancies	share doubts identify traps report safety concerns discuss to evaluate options
workload management	specify duties	report overload
preparation	fully brief issues	share experience state intentions set automation guidelines
prioritisation	communicate priorities	bottom line established
planning	ensure crew knows plans	duties established