UNIVERSITY OF ABERDEEN

SCPMDE Project: RDNES/991/C

The Identification and Measurement of Anaesthetists' Non-Technical Skills

Review of Behavioural Marker Systems in Anaesthesia

Workpackage 2 Report
Version 1.1

G. Fletcher, R. Flin and P. McGeorge

Industrial Psychology Group, William Guild Building, King's College, ABERDEEN, AB24 2UB. Tel. 01224 273210, Fax 01224 273211

SCPMDE Project: RDNES/991/C

The Identification and Measurement of Anaesthetists' Non-Technical Skills

University of Aberdeen and Scottish Clinical Simulation Centre

Review of Behavioural Marker Systems Anaesthesia

Workpackage 2 Report
Version 1.1

G. Fletcher, R. Flin and P. McGeorge

Industrial Psychology Group
University of Aberdeen

Document History

Version	Date	Comments
Draft 1	29th March, 2000	Internal Aberdeen review
Draft 2	17th April, 2000	Steering Group review
Draft 3	12th May, 2000	Final review
Version 1	15th May, 2000	Release to SCMPDE
Version 1.1	10th July, 2003	Adaptations and for pdf formatting

Contents

1. Introduction **1**

1.1 Background 1
1.2 Overview of main sections of report 2
1.3 Behavioural Markers 2

2. Method **8**

2.1 Introduction 8
2.2 Review of behavioural marker systems and checklists 8
2.3 Informal survey 8

3. Review of Behavioural Markers and Checklists **10**

3.1 Introduction 10
3.2 Description of Behavioural Marker Systems 11
3.3 Behavioural Marker Content Analysis 16

4. Survey of use of Behavioural Markers and Checklists **23**

4.1 Introduction 23
4.2 Use of markers/checklist 23
4.3 Development of systems and scales 24
4.4 Trainer training 25

5. Discussion **26**

6. Conclusion **30**

7. References **31**

Acknowledgements

The authors would like to thank those anaesthetists and researchers who provided valuable information for this report in response to the survey, and through more general communications. Again, the support and advice of members the project Steering Group is gratefully acknowledged.

1. Introduction

1.1 Background

This is the second report from the University of Aberdeen and the Scottish Clinical Simulation Centre's two year research project(1999-2001), funded by the Scottish Council for Postgraduate Medical and Dental Education (SCPMDE), reference: RDNES/991/C, investigating non-technical skills in anaesthetists. It describes WP2, a review of behavioural markers systems currently used in the training and assessment of anaesthetists' non-technical skills.

The objective of the project is to develop a taxonomy of anaesthetists' non-technical skills and identify behavioural markers for these skills that can be used to support training in the simulator. Within this project the term non-technical skills is used to describe *'anaesthetists' attitudes and behaviours in the operating theatre environment not directly related to the use of medical expertise, drugs or equipment'*. This definition includes both social/interpersonal skills and cognitive/mental skills, and thus would cover skills and behaviour relating to communication, team co-ordination, leadership, decision making, situational awareness and workload management. Such aspects of human performance have been the focus of research and training for several years now in aviation, and Crew Resource Management (CRM) training, which addresses them, is now a compulsory component of a European commercial pilot's licence.

The literature reviewed for the WP1 report[1] suggests that there is beginning to be a trend towards training non-technical skills for anaesthetists (and other medical professionals) particularly with the increasing availability of high fidelity simulators. These provide a learning environment where the necessary skills for coping with complex, dynamic and often uncertain tasks can be practised and discussed without any risk to patient safety. Indeed, although the number of serious anaesthetic mishaps is small, of those that have been reported and analysed, as many as 80% (Cooper, Newbower, Long & McPeek, 1978; Williamson, Webb, Sellen, Runciman & Van Der Valt, 1993) have been found to have human factors causes associated, many related to non-technical skills. Examples of these include: drug swaps, distractions, performance in emergency situations, and monitoring failures (more details on studies into critical incidents can be found in the WP1 literature review). The prevalence of human error causes in incidents is not unique to medicine, other industries, e.g. aviation, report similar findings. Traditional approaches to training that provide technical knowledge and practical psychomotor skills do not directly address these problems and so one approach for dealing with this is to acknowledge that human error in medicine is inevitable given the complex nature of the tasks (Helmreich & Merritt, 1998) and to support error management through training in non-technical skills. However, before it is possible to develop such training and evaluate its effectiveness, it is necessary to identify exactly what non-technical skills are important in anaesthesia and what behavioural indicators are observable to demonstrate effective performance.

[1] G. Fletcher, R. Flin and P. McGeorge 'WP1 Report: Review of Human Factors Research in Anaesthesia. Version 1, 10 February 2000'. Interim Report on SCPMDE Research Grant RDNES/991/C.

The aim of WP2 was to identify and analyse existing taxonomies of anaesthetists' non-technical skills. The review of literature for WP1 indicated that non-technical skills checklists and behavioural marker systems do exist in anaesthesia and are being used for training, notably for Anesthesia Crisis Resource Management (ACRM: Gaba, Fish & Howard, 1994) and for Team Oriented Medical Simulation (TOMS; Sexton, Marsch, Helmreich, Betzendoerfer, Kocher, Scheidegger and the TOMS Team, 1997). Some of the marker systems identified are really checklists developed to guide performance and actions during training. In this respect they were not developed as marker systems for behavioural assessment but essentially serve this function by providing a framework for skills development. Other systems could more accurately be described as behavioural marker systems as they were developed to include behavioural examples to give guidance for trainees and assessors. These may also be described as checklists but their focus is on the observers identification of activities rather than the anaesthetists performance of the activity. In this report the term marker system is used as a general description for both, although the individual terms are also used in relation to specific systems. This report gives more detail on how these systems were developed and how they are used during anaesthetist training. However, as this type of training is not as widespread as its counterpart in aviation, this report will also discuss some of the recent advances made in the aviation industry particularly with regard to the use of behavioural markers for assessment of non-technical skills. While the purpose of the anaesthetists' non-technical skills is not to develop a performance assessment tool *per se*, the same principles apply because a necessary element of training is to provide feedback to trainees on their performance. Therefore, some of the issues associated with this are included in this report.

1.2 Overview of main sections of report

This report is divided into six main sections. In this first section, the idea of behavioural markers and their use in training has been introduced; in the second section, the method for the identification and analysis of anaesthetic behavioural markers and checklists is described. Section three describes results of a small survey on the current use of the checklists and behavioural marker systems in anaesthesia and section four gives details of the identified marker systems and the key skills addressed by them. In section five the findings of the workpackage are discussed and in section six a set of the most critical and commonly identified anaesthetist non-technical skills and behaviours is provided for use in subsequent stages of the project.

1.3 Behavioural Markers

Before considering the use of behavioural markers in anaesthesia, it is important to understand what is meant by behavioural makers in the context of non-technical skills training and why they are important. To do this, it is necessary to look to the aviation industry where CRM or non-technical skills training is now a legal requirement in Europe and thus the tools for supporting it are well advanced.

In the literature review CRM was defined as "using all the available resources – information, equipment, and people – to achieve safe and efficient flight operations"

(Lauber, 1984, p20). This is a widely used definition and lists of CRM skills have been variously developed to support the training of this aspect of performance. As described in the literature review, the earliest CRM programmes in aviation tended to focus on pilot interpersonal skills and personality but over the last twenty years CRM training has developed into a broader concept focusing now on the wider team and organisation, and incorporating all aspects of communication, co-ordination, workload management, decision making and situational awareness necessary for effective error management. In addition to the content of courses having changed, the level at which they seek to influence behaviour has also changed. Seamster, Prentiss and Edens (1997) noted that "traditionally CRM training has emphasised the attitude component of crew behaviour" (p500) but now the emphasis has shifted to the behaviour itself. This brings with it a shift towards training skills and not just knowledge, although the knowledge part of CRM is also still taught. This differentiation between knowledge and skills becomes very important to researchers investigating CRM when the issue of evaluation is raised. Following an FAA definition, Seamster *et al.* (1997) describe knowledge as "the information required to develop skills and attitudes" and skill as "the ability to perform an activity or action" (p500). Clearly while the knowledge element of CRM is vital, it is the effective execution of the associated skills that is important for flight safety (e.g. knowing about decision making is different from having good decision making skills). Furthermore, it is CRM skills that are evaluated in the FAA's Advanced Qualification Programme (AQP; Birnbach & Longridge, 1993), which US airlines can voluntarily sign up to as a way of meeting various regulatory requirements. One of the key elements of AQP is evaluation of CRM skills in realistic simulator scenarios during Line Oriented Flight Training (LOFT) and Line Operational Evaluation (LOE). To support this process it has been necessary for airlines to conduct extensive task analyses to identify core CRM skills and their associated behavioural markers.

The main source of guidance for the development of behavioural markers in aviation has been the NASA/UT LINE/LOS Checklist (LLC: Helmreich, Wilhelm, Kello, Taggart & Butler, 1991), which was developed from analysis of accident and incidents and from psychology research. The checklist has been revised several times and is used extensively to evaluate crew CRM skills and performance during simulator sessions and line audits. It is also used as a research tool for investigating human factors aspects of performance. LLC is structured around key dimensions or elements of the CRM components of flight operations and to assist in assessment of the elements, behavioural markers are given as examples of each concept (Helmreich, 1991). The main dimensions for crew performance in this original LLC were: Communications Process and Decision Behaviour, Team Building and Maintenance, Workload Management and Situational Awareness, and Overall Technical Proficiency. In the current version of LLC, version 4.4 (Helmreich, Butler, Taggart & Wilhelm, 1997) there are four main human factors dimensions: Team Management and Crew Communications, Situational Awareness and Decision Making, Automation Management, and Special Situations, plus two others: Technical Proficiency and Overall Observations. The evolution of this checklist indicates that even in aviation, the use of skill lists and behavioural markers is still being developed. Indeed the issue of how to group skills together is still unresolved and Flin and Martin (1998) observed that different CRM skill lists often appear to group items in different ways, which can have a profound influence on how a particular skill is interpreted.

In a paper describing development of a cognitive model of instructor CRM assessment, Seamster and Edens (1993) gave experienced instructors 60 'LOFT concepts', mainly taken from LLC, and asked them to group related concepts. The resulting groups were subjected to cluster analysis which revealed items were clustered initially as for CRM assessment and technical assessment. The CRM assessment items were grouped as cognitive and interpersonal skills. The cognitive skills related to: problem identification, prioritisation, and workload management, and the interpersonal skills related to teamwork through: communication, group climate, and leadership/followership. These results suggested that experts assessing CRM skills actually use specific groupings of skills to assist them in their task. In a subsequent paper, Seamster, Hamman and Edens (1995) highlight the importance of using classification systems to group CRM behaviours as a means of facilitating the assessment process. They note that while the behaviours related to the interpersonal or team skills are directly observable, those related to the cognitive or mental skills are "less observable and often have to be inferred based on observation of interpersonal or technical behaviours" (p664). They believe that making distinctions between the different types of skills makes it easier to develop the behavioural markers for the elements and it also helps instructors with "compiling those elements into a meaningful hierarchy so that their relationship is understandable as well as usable" (Seamster & Edens, 1993, p126).

One of the fundamental features of the LLC is that it addresses crew rather than individual performance, although it does allow for comments to be made on individuals. However, marker systems have also been developed to address non-technical skills in individual pilots. An example of one such system is NOTECHS. The NOTECHS system (1998) was developed in a European research initiative to meet the new legislative requirements for pilots to be evaluated on non-technical skills as part of their flight crew licence. NOTECHS was developed from a review of current marker systems and from psychology research, with the specific purpose of being used for CRM assessment in the simulator. The system divides skills into four main groups: Leadership and Management; Co-operation; Decision Making; and Situational awareness. The first two categories are the social skills and the last two categories are the cognitive skills. Each main skill is divided into further element and behavioural markers. Communication is not included as a separate category because the authors argued that "communication skills are inherent in all four categories and the listed behaviours all involve communication" (NOTECHS, 1998, p62). At the current time, the NOTECHS system is being validated in a large pan-European experiment to ensure that as a measurement system for CRM skills it is reliable, culturally robust and easy to use (O'Connor et al, 2002). Such evaluation is vital when developing behavioural marker system for use in any profession, to be certain that it really does address the important skills and can be used accurately and effectively by trainees, trainers and assessors alike.

Following an extensive review of research and practice in this area, Flin and Martin (2001) identified five key areas that should be considered when developing and using behavioural markers for CRM skills assessment, which apply equally to anaesthesia as to aviation:

1.Unit of assessment

This refers to whose performance is being assessed. In many of the aviation studies they reviewed, performance was assessed at the team level and not at the individual pilot level. This was also found to be the case in anaesthesia studies discussed in the WP1 literature review. The most likely reason for this is the use of aviation team performance checklists as the basis for the anaesthesia ones. However, as with aviation, while performance of the whole team is important, at the training level individuals need to be considered so they can develop the necessary skills to allow them to work in any number of different teams, in different situations and adapt to the conditions as appropriate. Moreover final assessment of pilots' competence in technical skills for licences is at the individual not the team level.

2. *Identification of CRM skills and their associated behavioural markers*

It is this process that is the main objective of the present project. Some of the main skill areas in aviation have been discussed above and those currently being used in anaesthesia are reviewed in the following sections. However, Flin and Martin observed in their study that while the main CRM core skills seem to be similar, they are described differently between airlines and behavioural markers may have been selected to address the activities of particular simulator scenarios. It is not unlikely that a similar phenomenon will be encountered in anaesthesia and, therefore, when developing the taxonomy, the focus should be kept initially at the skill category level and only later move to the behavioural markers. It also seems clear that the skills and behaviours cannot be developed without consideration of factors such as the unit of assessment because good indicators of skilled performance by a whole team may be different to good indicators of skilled performance by an individual within a team context. Perhaps the most important issue when developing the markers is that they relate to behaviour that can actually be observed. In their guidance for Advanced Qualification Programme (AQP) developers, Birnbach and Longridge (1993) state that evaluation of CRM should "focus on clearly defined, observable behaviour rather than inferences about underlying cognitive processes" (p272). However, the cognitive skills are of particular interest in anaesthesia and so, as with aviation, the only way to assess these skills, and for trainees to demonstrate that they are using them, is through their related behaviours. To reduce any possible negative effect of having to make such inferences, the NOTECHS (1998) report recommends basing them on "obvious observable facts and behaviours" For example, while it is difficult to tell that pilot has good situational awareness, observing, or not, that he/she asks particular questions, gathers information, and discusses plans or the status of the aircraft, gives an indication as to the pilot's level of situation awareness.

3.Assessment Method

As there is no one, direct and measurable output of an individual's CRM training (measuring reduction in errors and improved error trapping are difficult as both are confounded by so many other factors), CRM proficiency is usually measured by observer judgement of performance using a scale based on behavioural markers. Flin and Martin (2001) note that while there is extensive psychological research on designing rating scales for performance measurement, there have been very few studies comparing different types of rating scales used in CRM evaluation, particularly for individual pilots. Thus there are a wide range of scales with different numbers of points and different anchors being used. In aviation the usual number of points is four or five but ultimately, when pilot non-technical skills assessment is

linked to licensing, a two point pass/fail system is used. There is perhaps a difference here as to what is necessary to stamp a licence or declare an individual competent and the requirement to provide meaningful feedback from assessment during training. For the latter purpose, it is necessary for both the instructor and the trainee to know the level of performance against a predefined set of criteria so further activities can be tailored appropriately, e.g. a rating of '2' may equate to minimum standard with more training necessary.

4.Rater Reliability and 5.Rater Training

Flin and Martin (2001) report that studies of rater reliability have shown differences not only between different raters (inter-rater reliability) but also within rater (intra-rater reliability) differences for ratings given across whole scenarios or for particular event sets. Therefore, there are two issues that need addressing: how to ensure different raters give equal ratings for the same performance, and how to rate differences in a given behaviour across a scenario. The first of these areas has received considerable attention recently because of its implications for pilot licensing and flight safety (Holt, Boehm-Davis and Beaubien, 2001). It is well acknowledged that rating behaviour is open to a number of systematic errors or biases such as: central tendency, contrast effect, initial impression and latest behaviour (Aiken, 1996; Furnham, 1997). Within aviation, Hamman, Beaubien and Holt (1999) identified the possibility of errors relating to personal interpretation of Standard Operating Procedures (understanding of the marker system contents would also apply here), memory and use of the rating scale. The consequence of such errors could mean proficient pilots do not pass the training course, or more dangerously ineffective pilots do pass the training. Therefore, one of the key requirements when developing a behavioural marker system and rating scale is to evaluate the inter-rater reliability of the scale. Hamman *et al.* (1999, p1214) identified five characteristics that can be assessed when giving feedback to instructors about their performance as LOE evaluators:

- Congruency between each instructor's distribution of judgements and the groups' distribution of judgements
- Systematic differences between the instructor's mean ratings and the groups' mean ratings
- Consistency with which instructors are able to shift their ratings up with good performance or down with poor performance
- Instructors' ability to discriminate between crews of varying levels of performance
- Absolute level of inter-rater agreement on each scale item.

Hamman *et al.* (1999) give feedback on these different aspects of inter-rater reliability during special training courses given to help instructors develop their rating skills and reduce bias. However, they point out that "when presented independently, such feedback can be misleading" (p1214) but that used together they can be used to produce a "multi-faceted profile of their strengths and weaknesses as evaluators" (p1214). The importance of training raters to reduce judgement biases and improve inter-rater reliability is particularly highlighted by Flin and Martin. Therefore, when developing a behavioural marker system, part of ensuring good inter-rater reliability is to develop an effective training programme for users of the system. The training should be designed to ensure that trainers are able to use the behavioural markers

accurately and consistently. It should also include regular follow-up calibration. Obviously this is particularly important if the markers are being used to assess performance for a CRM skills evaluation. The area of trainer training has received much attention in aviation and within anaesthesia is also of great importance, particularly as the role of trainer is often part of a consultant's job, irrespective of aptitude or appropriate training.

The other aspect of reliability relates to intra-rater reliability and the problems encountered with average performance across a training or assessment scenario. The TARGETS (Targeted Acceptable Responses to Generated Events or Tasks) judging methodology (Fowlkes, Lane, Salas, Franz and Oser, 1994) has been developed for US military helicopter team training and specifies particular behaviours for each core skill in the individual parts of a scenario. These "targeted acceptable responses" can be for individuals or the team and are scored based on their presence or absence. While this approach produces an acceptable level of rater agreement and removes many of the problems of applying general markers to different situations, it also significantly increases the effort associated with developing each scenario. An alternative to this, as used in the LLC, is to divide the scenario into meaningful events based on phase of flight, e.g. take-off, climb, cruise. To use this approach in anaesthesia it would be first be necessary to identify the different phases of the operation with clear start and end points as these phases are determined less obviously by checklists and procedures than they are in aviation.

While there is clearly a move towards the formal evaluation of non-technical skills in aviation, the development of skills taxonomies and behavioural marker systems is also necessary for purposes other than evaluation. Within training programmes, a certain level of assessment is necessary simply to measure individuals' progress on a course and to evaluate its training effectiveness. Training for medical knowledge and practical skills can be more easily assessed, as performance usually relates to identifying cues for diagnosis, selecting treatment, applying solutions and carrying out medical procedures. There are often specific activities that can be scored or measured in an objective way e.g. time to treatment or administration of correct drug that can be used as measures of performance during simulator training. It is clear if an individual has performed the activities correctly or not and so feedback can be given as appropriate. However, if there is no valid measure for the skills being addressed by the course, such feedback and discussion of performance is more difficult and is open to instructor subjectivity. This is particularly so for non-technical skills where there may not be direct, objective measures. Therefore, it is important to develop a framework of non-technical skills to allow core skills to be addressed in the training course (simulator scenarios, notes, etc). These same skills can provide the basis for selecting behavioural markers that can be used to assess performance formally and informally and to structure discussion during de-briefing. Identifying the areas to be addressed during de-briefing is very important as it allows feedback to be structured and guidance given against unambiguous criteria. Obviously there will always remain an element of interpretation surrounding non-technical behaviour but with consideration of the issues discussed in this section, it is hoped that the non-technical skills framework and behavioural markers developed in this project, once fully validated, will provide a tool that can be used with confidence for all aspects of anaesthetist training and assessment.

2. Method

2.1 Introduction

Following on from the findings of the literature review, which showed that various human factors training programmes have been developed for anaesthetists, this workpackage sought to collect more information on the current use of marker systems for non-technical skills in anaesthetist training. Two related activities were undertaken to achieve this: analysis of the available markers to identify common skills and themes and a survey to collect additional behavioural marker systems and discover how they are used.

2.2 Review of behavioural marker systems and checklists

Available checklists and behavioural marker systems were collected from the literature, from Internet sites for anaesthesia departments, and from an informal survey of simulator centres. Based on the findings of the literature review, the criteria for selecting these were kept quite broad, one of the reasons being that similar skill sets appeared to be being used in a number of different ways by different anaesthesia departments. In total, six markers systems or checklists were identified, five through WP1 and one as a result of the survey.

Each set of behavioural markers or checklist was copied into an Excel spreadsheet. Using the different levels of description contained in each marker system, e.g. category or behaviours, the items in each checklist were sorted according to common terms and themes. The different skills or behaviours were sorted into groups at equivalent levels of abstraction, thus at one level, skills like ‘communication’ and ‘leadership’ formed themes. As most systems did not give labels to the different levels, there was an element of interpretation in this process. As the purpose of this analysis was simply to identify those non-technical skills commonly addressed in training, and thus assumed to be important, the sorting process was conduced by one researcher supported by discussions with colleagues.

2.3 Informal survey

Based on the literature reviewed in WP1, it was apparent that a number of medical centres were using lists of key skills and/or associated checklists for both anaesthetist training and research. However, from the reviewed papers it was not always possible to establish exactly how the training was being conducted, how the skill lists were used and for what purpose. Therefore, it was decided to conduct an informal survey of selected individuals to find out more about their use of behavioural markers in anaesthesia training programmes. The reasons for the targeted informal nature of the survey were pragmatic. The number of anaesthesia departments providing non-technical skills training (either ACRM or TOMS) seems to be limited to those owning anaesthetic simulators. Thus a full scale survey of all anaesthetic departments in the UK and abroad would have been very time consuming given the expected outcome that behavioural marker systems are not commonly used in anaesthetist training.

Departments and research groups with simulators are fairly easy to track down through the Internet as most are keen supporters of anaesthetist education and training and are happy to share their experiences with others in the field. While this does mean that the survey may not have collected information from the widest possible population, the survey could be kept quick, simple and informal allowing contacts to be made and further discussion pursued as necessary.

Participants for the survey were selected on the basis of being involved in anaesthetist simulator training. Key figures identified from their research publications and other individuals listed on the University of Rochester (USA) web page[2] of Anesthesia Simulator Users were approached. Individuals from 13 anaesthetic departments and simulation centres were contacted and sent questions in an e-mail communication in February, 2000. These departments were located in: UK (1), USA (6), Germany (1), Switzerland (1), Netherlands (1), Denmark (1), Canada (1) and Australia (1). In several cases contact had already been made with the person and the questions were supplemental to on-going communications. The questions used in the survey were adapted from the semi-structured interview format used by Flin and Martin (1998) to investigate the use of behavioural markers in aviation CRM courses. The survey was introduced as being part of the non-technical skills research project with the aim of collecting information about the development and use of behavioural marker systems in anaesthetist training. The questions asked are shown below. As in the aviation study, in addition to answering the questions, respondents were asked to send a copy of their checklist or marker system. As the questions were e-mailed to individual anaesthetists, responses were received the same way and further questions asked if necessary. The option was also given for the respondents to ask questions about the study either electronically or on the telephone.

1. Do you use a behavioural marker/performance indicator system for anaesthetist training in the simulator?
2. What type of training course do you use it for? (e.g. ACRM, crew resource management)
3. How long have you used the marker system for?
4. How was this system developed? (e.g. developed internally, derived ACRM)
5. If the system was developed externally, what, if any, changes did you make to adapt it for use in your hospital?
6. How do you use the system during training? (e.g. during briefing/de-briefing, for assessment)
7. Does your marker system have a performance rating scale?
8. If so, what is the scale and why was it selected?
9. How are the trainers trained to use the scale?

Eight responses were received: USA (4), Canada (1), Australia (1), Germany (1) and Switzerland (1). The results from the survey questions were reviewed for content to provide information about current practice in anaesthetist training. While the survey provided information on the use of behavioural markers for anaesthetist training and assessment, it only produced one new marker system for the analysis. The other markers systems had previously been identified from the literature review.

[2] http://www.anes.rochester.edu/simulate/simusers

3. Review of Behavioural Markers and Checklists

3.1 Introduction

The following section gives details of the sample marker systems obtained and identified as currently being used for simulator-based training and research into performance assessment of anaesthetists. The six marker systems were collected from a variety of sources: published literature (n=3), communication with the developers (n=1), the Internet (n=1) and the survey (n=1, but also available in the literature). Information on the use of five of these systems was collected from the informal survey and is discussed in greater detail in Section 4 Below. The lists reviewed include both team skills marker systems and checklists used in training. The checklists, which are essentially lists of activity to be completed, are included as they are used to provide a framework for the training and thus address key non-technical skills. Therefore, the review is not a necessarily a comparison of like with like and is not intended to be a comprehensive analysis. However, its utility lies in helping to identify the skills and behaviours commonly considered important for the effective performance of anaesthetists.

The six marker systems that have been described and analysed in this workpackage are:

1. Crisis Management Behavioural Performance Markers
2. Kommunikations Status Team Performance Indicators
3. Operating Room Checklist
4. Anesthesia Crisis Resource Management (ACRM) Principle Markers
5. Categories of behaviour defining aptitude for Iowa anesthesia residents
6. Emergency Team Co-ordination Course Teamwork Behaviour Matrix

In section 3.2 the contents of each marker system is described including its origins and any rating scale used. The order of the presentation is related to their common origins. Systems 1-3 are based on the NASA/UT Line/LOS Checklists and team skill markers (Helmreich *et al.*, 1991), system 4 is the ACRM checklist (Howard *et al.*, 1992), and system 5-6 were developed independently for selection purposes and emergency medicine training respectively. System 7, the NOTECHS (1998) system, is an aviation marker system currently being validated for use as a non-technical skills assessment tool that has been included to provide an additional reference point. Indeed, it shows how the anaesthesia marker systems remain very similar to their aviation-based ancestors. To assist in the description of the marker systems, three labels have been added to describe the different levels of each checklist based on the review of their contents and construction. These descriptions are those used in NOTECHS: category, element and behaviour. Category describes the highest level of skills (e.g. co-operation), element describes the middle level of skills (e.g. supporting others) with several of these usually being grouped into a category. The lowest level is the exemplar behaviour (e.g. offers assistance). This is effectively the indicator or marker behaviour, which can be positive or negative. Figure 1 below shows how the three levels are related. The term ‘item’ is used as a generic description for skills and behaviours at any level in a checklist.

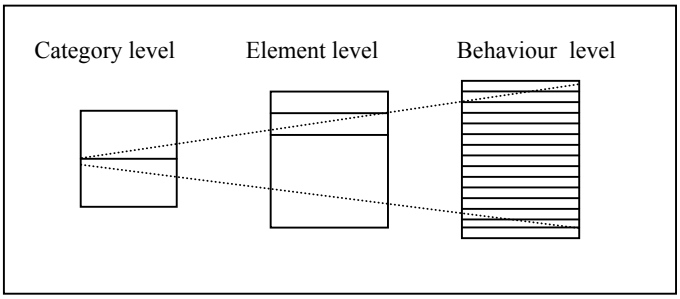

Figure 1. Relationship between levels of skill

Section 3.3 details the results of the content analysis on the six marker systems. It contains lists showing the comparable items and the key skills identified. It also contains a general description of the marker systems outlining some of the differences and similarities between them and briefly highlighting some of the difficulties associated with the analysis process. Noticeable by its absence in the following list of checklists and behavioural markers is any system provided by the Royal College of Anaesthetists for consultant anaesthetists involved in training in the UK. Such a checklist can be found in the Royal College of Anaesthetists' booklet on Specialist Training in Anaesthesia, Supervision and Assessment (1994) and is an 'end of module assessment exercise' where performance in four main areas: clinical skills, knowledge, postgraduate activities and attitudes, is assessed on a behaviourally anchored three point rating scale. The skills addressed are very broad and could probably be applied to any area of medicine. While covering important aspects of performance, they do not appear to have been developed specifically with the anaesthetic task or environment in mind. Also, while this checklist exists, it is not used and has been superseded by other checklists which are now themselves under revision. As a result it has not been included in this analysis.

3.2 Description of Behavioural Marker Systems

<u>System 1.</u>

Title	Crisis Management Behavioural Performance Markers
Source	Gaba, Howard, Flanagan, Smith, Fish and Botney (1998) Researchers based at Stanford University Medical School and Palo Alto Veterans Affairs Medical Centre
Origin	Adapted from aviation team performance markers in the NASA/UT Line/LOS Checklist (1995)
Use	Used as a behavioural performance marker system in research investigating performance assessment for simulator sessions in ACRM training. Not actually used in normal training but being investigated as a measurement tool, particularly inter-rater reliability. Used in conjunction with a technical assessment system covering clinical actions.
Contents	12 elements (including two for overall effectiveness) and 42 behaviours. Total number of items 54.

Rating Scale	Uses a five point rating scale, ratings are given for two phases of the operation: 0 – not observed (also have to select rating from 1-5) 1 – poor performance 2 – minimally acceptable 3 – standard performance 4 – good performance 5 – excellent performance
Other	This system was derived from the aviation checklist for the specific purpose of investigating behavioural performance in crisis management. It is still under development and has not been used for assessment during training although this may be considered in the future. Evaluation of inter-rater reliability of the system found that it was fairly high and that reducing the scale from five to four might improve this further. The greatest problem was found with averaging performance over time as it was found to vary greatly even when the scenarios were divided into two phases.

System 2.

Title	KOMSTAT Team Achievement Indicators
Source	Document published on the University of Basel-Kantonspital, Department of Anaesthesia Internet site[3]: Kommunikations-Status (KOMSTAT), Operationssal-Beobachtungen, Ver. 2.0, 07/97 . Translated from German.
Origin	Developed from the University of Texas Operating Room Checklist (see below)
Use	Used as a operating room team performance checklist during observation of Team Oriented Medical Simulation/Human Aspects Development medical training. Used to rate all aspects of the teams' performance during simulator scenarios (not just specific groups or skills) and to give feedback to the team immediately after the procedure.
Contents	3 categories, 15 elements, 39 behaviours. Total number of items 57.
Rating Scale	Uses a four point assessment scale for different phases of the operation and considering the anaesthetic and surgical team perspectives and the interface between the two groups. The rating scale anchors are: 1. Insufficient – achievements are clearly behind expectations. Behaviours seem to be not motivated enough, more irritated and dissociated. Could lead to a reduction of outcome. 2. Minimal – achievements are still acceptable, but should be better. Behaviour is still reasonable but barely sufficient. As a standard is not satisfactory. 3. Standard – this quality and level of achievements and behaviour should the norm. It corresponds to the daily routine. 4. Outstanding – this achievement represents outstanding abilities.

[3] Swiss anaesthesia Server: http://www.medana.unibas.ch
Also, Human Aspects Development Medical training: http://hadmedical.ch

Other	The focus on the assessment with this marker system is the whole team (anaesthetic or surgical). The checklist has also been used to observe behaviour in real operations. The element level items are all listed on a single scoring page grouped in the categories, the behaviour descriptions and the rating scale are on separate pages.

System 3.

Title	Operating Room Checklist (ORCL) Version IV
Source	Operating Room Checklist, American version (1995), obtained from University of Texas
Origin	Developed from the aviation Line/LOS checklist and Line Operational Safety Audit
Use	As System 2. Used for observation of operating room teams during Team Oriented Medical Simulations and real operations.
Contents	A two level checklist, the items could be described as categories and elements or, as specific examples contained within items in the lower level, elements and behaviour. This latter division of items is used in this review because the level of detail of the low level items makes it difficult to isolate specific skill elements. 6 elements and 19 behaviours. Total number of items 24. Considers non-technical performance and also includes overall observation of technical performance, team performance and adherence to guidelines.
Rating Scale	Rating scale is four point anchored scale, again used at different phases of the operation with markings for the anaesthetic and surgical teams and the interface. Specific comments are requested for ratings of 1 or 4. 1. Poor – observed performance is significantly below expectations. This includes instances where necessary behaviour was not present, and examples of inappropriate behaviour that was detrimental to mission effectiveness. 2. Minimum expectations – observed performance meets minimum requirements but there is ample room for improvement. This level of performance is less that desired for effective team operations. 3. Standard – this demonstrated behaviour promotes and maintains team effectiveness. This is the level of performance that should be normally occurring in operations 4. Outstanding – performance represents exceptional skill in the application of specific behaviours, and serves as a model for teamwork – truly noteworthy and effective.
Other	This checklist is effectively the same as the one described in System 2. But is probably an earlier version and thus there are some differences.

System 4.

Title	Anesthesia Crisis Resource Management (ACRM) Principle Markers
Source	Howard, Gaba, Fish, Yang and Sarnquist (1992) Researchers based at Stanford University Medical School and Palo Alto Veterans Affairs Medical Centre
Origin	Developed originally from aviation research CRM training and NASA/UT team skills to provide a basis for ACRM training.
Use	The markers/checklist forms the basis for the ACRM course. The items are discussed during the knowledge part of the course and during debriefing. Therefore it is considered to be a checklist for the key skills and behaviours.
Contents	13 elements (principle markers) and 20 behaviours. Not all elements have associated behaviours as some items at the element level are themselves behaviours. One of the element level items is overall crisis management. Total number of items 33.
Rating Scale	While the source document for the checklist shows a scales of 1-5 next to each principle marker, the paper does not discuss if or how this is used. It could be for performance assessment. Or could even be for trainees to evaluate the course in this area.
Other	This checklist forms the basis for ACRM training world-wide. An example of this is from a simulation centre in Australia. In this instance, the checklist has been adapted slightly to include higher level categories related to the phases of a crisis (prevention, planning, response and recovery). The prevention category is new and is described by an element level item 'understanding of precipitating and augmenting factors'. The rating scale on the checklist has been reduced to three points but is not really used in training.

System 5.

Title	Categories of behaviour defining aptitude for Iowa anesthesia residents
Source	Altmaier, From, Pearson, Gorbatenko-Roth and Ugolini (1997) Researchers based in the Division of Psychological and Quantitative Foundations and in the Department of Anesthesia at the College of Medicine, University of Iowa.
Origin	Developed from critical incident interviews and grounded theory analysis. Similar behavioural markers were developed in other fields of medicine (radiology, obstetrics) by the same team and the system also builds on earlier medical work investigating non-cognitive skills.
Use	The system was not developed for training but for selection of anaesthesia residents. However, as it was developed as a behavioural marker system it is included. No additional information is given on the effectiveness of this system for selection or if it was ever used subsequently for performance evaluation purposes.

Content	6 elements and 18 behaviours (items in this list could also considered as 6 categories and 18 elements). For each of the lowest level items there is also an associated example of effective and ineffective behaviour with it in the form of a critical incident. Total number of items, excluding critical incident examples, 24.
Rating Scale	None is given
Other	-

System 6.

Title	Teamwork Behaviour Matrix
Source	Small, Wuerz, Simon, Shapiro, Conn and Setnik (1999) Researchers based at various Harvard Medical School and various Boston Hospitals and at MedTeams, Dynamic Research Corporation, Boston.
Origin	Developed for an emergency medicine physicians Emergency Team Co-ordination Course (ETCC) at the Center for Medical Simulation in Boston following the principles of ACRM and based on US Army aviation training.
Use	Team Behaviour Matrix is used to support scenario generation, direction and debriefings. Not used for performance assessment although the behaviours could be observed
Content	5 categories (labelled team dimensions), 18 elements (labelled primary descriptors) and 40 behaviours (labelled teamwork actions by core team). Total number of items 63.
Rating Scale	No rating scale is used
Other	This marker system was developed for Emergency Medicine physicians not anaesthetists but the parallels in work environment and tasks are similar, except obviously all cases are emergencies.

System 7.

Title	NOTECHS
Source	NOTECHS Final Report (1998) Developed for EC through collaborative research between NLR (Netherlands), IMASSA (France), DLR (Germany) and University of Aberdeen
Origin	Developed from review of other aviation CRM checklists and behavioural marker systems and theoretical analysis
Use	Currently being validated for use as a non-technical/CRM skills assessment tool in European airlines and possibly as an Acceptable Means of Compliance for European Joint Aviation Requirements.
Content	Nominally divided into cognitive and social skills. 4 categories, 15 elements and 40 behaviours (these are markers for good behaviours, poor behaviours are also given in the NOTECHS report). Total number of items 59..

Rating Scale	Rating scale being evaluated has five points: Very poor – endangers flight safety directly Poor – in other conditions could endanger flight safety Acceptable – does not endanger flight safety but needs improvement Good – enhances flight safety Very Good – optimally enhances flight safety and could serve as an example for other pilots.
Other	This is included as a typical example of an aviation marker system.

3.3 Behavioural Marker Content Analysis

As not all the checklist and marker systems are constructed in the same way or for the same purpose, direct comparison of the number and type of skills/behaviours included is difficult. Of the six anaesthetic or medical systems listed above (NOTECHS is only included for reference), four had two levels and two had three levels. Where the lists only have two levels, sometimes they appeared to be at the category and element level and sometimes at the element and behaviour level. Obviously this is a fairly arbitrary distinction and labelling has only really be added to assist in the content analysis. The items in the systems could be reviewed to identify common items at any of the levels. The average number of items per checklist was 42 (range from 24-63). Systems with two levels of description had an average of 34 items (range 24-54) and systems with three levels has an average of 60 items. Looking at all six systems, the highest level of each list, the category level, contains the smallest number of items, from 3-5 (6); the middle element levels contain from 12-18 items; and the lowest levels contain from 18-42 behaviours. This is except behavioural marker system 4, which contains 18 elements or behaviours, plus an example of both effective and ineffective behaviour for each in the form of a critical incident example. Only two systems had their own labels attached to the different skill levels (e.g. category, element, behaviour): systems 4 and 5. Most systems showed the element and behaviour items listed using the highest level item as the heading.

Contained in several of the systems are elements and behaviours relating to overall performance and technical performance. Also of note in the ACRM based checklists is the inclusion of several items not really describing either specific behaviours or skills. Examples of this include ‘initial crisis management’ or ‘calls for help’. These are clearly very important actions though not perhaps at the same level of abstraction as other skills in the list such as ‘attention allocation’ and ‘leadership’. Again this reflects the fact that the lists were developed to provide a framework for ACRM training and not to be used specifically as behavioural marker systems. Thus the checklist items are listed in sequential order for responding to a crisis; the stages have not been labelled in the original but as mentioned others have added these as categories. Other systems have the items broken down into similar types of skills and behaviours, e.g. team concerns, situational awareness and decision making. Often, it may be that the same skill is included at different levels in the different systems. For example, ‘vigilance’ is at the behaviour level in system 4 but at the element level in system 1, with behavioural markers including ‘monitoring’ and ‘cross-checks all sources of information’.

However, in spite of the differences in level and grouping of skills, the content analysis shows that there are common skills between the different checklists. This is to be expected for several systems as they are different variations of the same checklist or were derived originally from the aviation team skills markers. Even comparing the anaesthetic and medical marker systems and checklists to the aviation system shows that many of the skills are the same. One of the main exceptions are skills and behaviours related to teaching. This is a requirement important for consultant and trainee anaesthetists that is not paralleled in aviation. Other differences seem to exist at the behaviour level of the checklists and systems. This follows on from WP1 which suggested that while some basic non-technical skills may be found to be common between the aviation and anaesthetic domain, the behavioural indicators of effective performance may be different.

Items included in the category level from the systems identified are shown below. They are taken from System 2: KOMSTAT Team Performance Indicators and System 6: ETCC Teamwork Behaviour Matrix, there are the only systems that could be sorted at this level.

- Team Concerns [2]
- Improve team skills [6]
- Maintain team structures and climate [6]

- *(Decision finding process /)* Communication [2]
- Communicate with the team [6]

- Decision finding process *(/Communication)* [2]
- Apply problem-solving strategies [6]

- Management of work situation [2]
- Execute plans and manage workload [6]

The results of the analysis to identify common themes at the category level are shown in Table 1.

Category Level	
Social Skills	Communication
	Team organisation, co-ordination and working
Cognitive Skills	Planning and managing the work
	Decision making and problem solving

Table 1. Anaesthetists' Non-Technical Skills identified from existing marker systems – category level.

As a demonstration of how the same skill can be classified in different ways, in system 2, communication is considered more as a cognitive skill (with decision making) and in system 6 it is considered as a social skill (with reference to the team).

This cognitive-social skill distinction is not made explicitly in the anaesthesia marker systems but items do appear to be grouped along those lines. NOTECHS does make the distinction between cognitive and social skills but as communication is considered necessary for all non-technical skills, it is not included as a separate item in the system at all. In Table 1 it has been included as it was explicitly mentioned in both lists analysed. It is currently listed as a social skills, but whether it should really is regarded as a cognitive skill, a social skill or both is something that will need to be investigated later in the project. Obviously at this stage the items at category level could have been grouped in a number of ways as there were only two marker systems to work with.

Therefore, for the purposes of identifying key skills in anaesthesia, perhaps the most useful level for analysis of these marker systems is the middle or element level. Sorting at the lowest, behaviour level results in a large number of items, many of which are different indicators of the same skill type. It is also clear that the lowest level behaviours could be grouped together in a number of different ways. By sorting and comparing middle level skills and behaviours, it is possible to obtain a good idea of the skill areas that are considered to be important.

The clusters of items shown below are the result of sorting skills at the middle (or sometime top) level according to common words and meanings. As the items have been mixed up, the number in square brackets indicates marker system or checklist from which each item originated.

- Briefing [2]
- Orientation [1]
- Planning [4]
- Preparation/Planning/Anticipation [1]
- Preparedness [5]
- Prepares for post-event care [4]
- Conduct situational planning [6]

- Distribution of work [2]
- Distribution of workload [4]
- Workload Distribution [1]
- Manage team resources and workload [6]

- Attention allocation [4]
- Priorities [2]
- Prioritise tasks [6]
- Anticipation [2]
- Attention/survey [2]

- Recognition of events [4]
- Vigilance [1]
- Data monitoring [5]
- Cross-monitor [6]

- Decisions [2]

- Situational Awareness & Decision Making [3]
- Apply decision making method [6]
- Maintain situation awareness [6]
- Engage in error-correction actions [6]

- Assessment & re-evaluation of patient status [4]
- Re-evaluation [1]
- Conduct secondary triage [6]

- Communication [4]
- Communications [1]
- Interpersonal Skills [5]
- Communications, Co-ordination and Leadership [3]
- Use standards of communication [6]

- Information/Question [2]
- Inquiry/Advocacy/Assertion [1]
- Reception [2]
- Persistence [2]
- Offer information to team [6]
- Request information from team [6]

- Feedback [1]
- Feedback [2]

- Leadership [4]
- Leadership/Followership [1]
- Establish leadership [6]

- Group Climate [1]
- Team care/climate within the group [2]
- Team forming [2]
- Team leadership or management [2]
- Engage in informal team improvement strategies [6]
- Engage in formal team improvement strategies [6]
- Team Building [3]
- Organise team [6]
- Cultivate the team [6]
- Special Situations [3]
- Resolve conflicts [6]

- Response to teaching [5]
- Teaching [2]
- Consultation and Instruction [3]

- Calls for help [4]
- Getting help [2]
- Emergency declared early, not late [4]

- Initial crisis management [4]
- Emergency situations [5]
- Implement plan [6]

- Medical knowledge relevant to event [4]
- Technical Skills [5]

- Overall ANESTHESIA Crew effectiveness [1]
- Overall crisis management [4]
- Overall HOT-SEAT Person effectiveness [1]
- Overall Observations [3]

If a single term is now used for each of the 16 skill element groups listed above a single set of key non-technical skill areas can be identified. These results are shown below in Table 2. The labels of cognitive, social and other have been added to indicate the type of skill apparently being addressed and an example of a behavioural marker associated with items in that skill element group.

Obviously these lists could be distilled further and any overlap removed but for the present report they provide a description of the most commonly identified non-technical skills from the reviewed behavioural marker systems and checklists. Certainly, several of the items in the 'Other skills' list could be considered to be already described by other elements in the cognitive or social skill lists. 'Declaring emergency' and 'initial crisis management' are somewhat ambiguous as skills anyway, and perhaps relate more to actions that need to be carried out because they seem to relate to a number of different skills. They are also very clearly related to dealing with crisis events rather than describing generic non-technical skills necessary for all types of situation. Furthermore, if the taxonomy being developed relates only to non-technical skills as the technical/clinical aspects of performance are being considered in other ways, then elements relating to that could be dropped. The only danger of doing this is that the important links between the technical and non-technical aspects of performance could be lost. None-the-less these skills are not central to the taxonomy of non-technical skills. The overall effectiveness item could also be omitted as this relates to a sum of all previously described skills and perhaps technical performance. It is assumed that effective performance on the other non-technical skill elements would result in overall effectiveness, and thus this is not a middle level element but an item resulting from the combination of all skills. It could be considered to be akin to the NOTECHS (1998) pass/fail assessment, which is based on the overall acceptability or unacceptability of performance with regards to safety.

Element Level	
Social Skills	Communication style - crew members acknowledge communication and verify ambiguous communication
	Information sharing - offer information to support planning and decision making
	Feedback - feedback addresses positive and negative aspects of performance
	Leadership - clearly define role of leadership
	Team building and climate - sensitivity for other team members (recognising stress, states of being overworked or tired, potential conflict)
Cognitive Skills	Preparation and planning - crew and team members are made aware of important plans
	Workload management - balance workload within the team
	Prioritisation and allocation of attention - priorities are clearly established and obeyed
	Monitoring and recognition - cross monitor actions of team
	Situation awareness and decision making - engage team in decision making process
	Re-evaluation - re-evaluates the results and side effects of interventions or actions
Other	Teaching - teaching takes place, when opportunity arises
	Declaring emergency - help can be requested if needed or given spontaneously
	Initial Crisis management - initiates life support measures
	Technical and clinical skills - correct use of equipment Overall effectiveness - *specific markers not given*

Table 2. Anaesthetists' Non-Technical Skills identified from existing marker systems – element level with example behavioural markers.

While the above items have been included for completeness, examination of the lists of markers derived in this analysis clearly shows that not all items are included in each list and several items in one group are from the same behavioural marker system. This is inevitable as not all lists contain the same number of items and often lists group similar items together, e.g. system 3 contains communication, co-ordination and leadership in a single element. Where some lists are missing certain elements such as leadership, it may be that behaviours associated with this category/element are described under another heading or may only be made explicit at the lowest level. Even marker systems developed from the same source show these differences, which highlight the fact that these skills are very difficult to categorise. While the debate on where to locate communication as a skill in behavioural marker lists continues, in this review of the non-technical skills at element level, it has been included as a social skill. The reason for this is that in the behavioural marker systems and skill lists reviewed, communication is exemplified with social or interpersonal type behaviours.

Another interesting element with regards to number of lists containing it, is decision making and situational awareness. There are only three systems that explicitly mention these skills, although skills relating to decision making and situation awareness, e.g. recognition of events and vigilance, are included. Possible reasons for this could be these skills relate primarily to cognitive processes and so are difficult to observe, or that they have not been considered to be so important in anaesthesia, or that having not been addressed so strongly in the earlier aviation CRM checklists, which really focused on teamworking, they have not yet been added to the anaesthesia marker systems. That they are included in some of the more recent checklists suggests the latter reason is perhaps the most likely.

4. Survey of use of Behavioural Markers and Checklists

4.1 Introduction

This section briefly describes the results obtained from the seven responses (USA x 4; Canada x 1; Australia x 1; Germany x 1; Switzerland x 1) to the informal survey. In five cases, the marker systems described in the survey have been reviewed in the previous section. The responses relating to these marker systems have been indicated. Given the actual use of behavioural markers, it was not always possible for respondents to answer all the questions, so several respondents simply gave a brief description of current practice with their simulator. While this resulted in less specific data, it provided valuable information into the current use of marker systems in anaesthesia.

4.2 Use of markers/checklist

The first three questions in the survey asked about the use of behavioural marker or performance indicator systems for anaesthetist training in the simulator, the type of training course it is used for and length of time using behavioural markers. The sixth question asked about how the system was used in training. The responses to these questions are combined below.

All the eight respondents were involved with running simulator training courses. Three of these were Anesthesia Crisis Resource Management (ACRM) training courses, one course was a comprehensive human factors training course developed in conjunction with an airline for all operating theatre personnel; one course was being developed for evaluating critical competence; and two respondents did not specify the course type although from the published research studies, one is known to involve training for crisis scenarios. Another simulation centre is developing an evaluation instrument for resuscitation team training. Where information was available about the content of the training courses it suggested that there was usually both a knowledge component and a skills development component.

Only two respondents actually described using behavioural marker systems for training. One is a system developed from the NASA/UT LLC that is used for assessing performance of the whole operating theatre team: KOMSTAT Team Achievement Indications (System 2) and also the Operating Room Checklist (System 3). These training courses were started in 1993 although it is not clear how long the marker system has been used for (the version shown later in this report is dated 1997). The marker system is used by observers during the simulator scenarios and then feedback is given immediately after the procedure. The second marker system uses a standardised points system to evaluate performance of standardised scenarios with a view to evaluating clinical competence. However, this marker system is still being evaluated and is not yet being used in training. As a copy of the marker system could not be obtained, it is not clear which non-technical skills are included or exactly how the scoring is carried out. Another respondent will also be involved in this evaluation study, and in addition they are currently revising their own rating system for resuscitation team training that is being developed as part of a research project. A

fourth institution has a behavioural marker system, developed from the NASA/UT LLC in 1997-8, Crisis Management Behavioural Performance markers (System 1), that is being used for research purposes to investigate behavioural performance assessment. While this system is not used in training apart from for research, it shares many common items with the ACRM principle marker checklist that is used during training, particularly to emphasise key issues during de-briefing (System 4). Two other institutions also run ACRM training and so while they do not use a behavioural marker system they may have used the ACRM skills as a framework during training. One of these simulator centres had also been involved in the development of emergency team co-ordination courses (ETCC) and for this the Teamwork Behaviour Matrix (System 6) is used to develop scenarios and structure debriefing. Given the similarities between the work domains and the type of course, this behavioural marker system is included in the content analysis in the next section. Neither of the two remaining institutions use behavioural markers although one does use technical markers.

With only one out of eight respondents actually using behavioural markers in simulator training (two if the emergency room team training is included) at the present time, the results indicate that use of behavioural markers in training is not yet common practice. However, the tone of the comments suggests that this is something that would be considered in the future once a suitable system has been validated. The development of measurement instruments by three other institutions supports this. Although once developed, these systems might still only be used by some institutions for research into anaesthetist skills and training and not for formal assessment of performance.

4.3 Development of systems and scales

Essentially three marker systems are being used or developed in anaesthesia. Two of these have been developed from the NASA/UT Line/LOS Checklist, albeit independently, and use anchored scales for rating performance. The KOMSTAT Team Performance Indicators currently being used for observing whole team performance during training was adapted with input from University of Texas to take account of behaviours not observable in the operating room or that have different significance in theatre than in the cockpit. The Crisis Management Behavioural Performance Markers that are being developed for use as a research tool were derived from the LLC following extensive changes. This system has been used in conjunction with a scenario specific rating system for technical performance. The marker system being developed for assessing clinical competence includes checklist items with weighted scores and negative scores for missing items. This is currently under evaluation. It is not clear how this system was developed apart from information about the study, a joint effort of several simulator centres. As explicit mention is not made of an originating source it is assumed that the marker system and scale have been developed from scratch.

4.4 Trainer training

An important issue regarding the use of behavioural markers in aviation is that of trainer, training and how to ensure different trainers use behavioural marker systems in the same way, in terms of consistency, sensitivity and congruency with other users (Hamman, Beaubien & Holt, 1999). There were three responses to the question asking about how trainers were trained to use the behavioural performance rating scales: one from the hospital actually using behavioural markers in training and one each from the institutions currently developing marker systems. Trainers using the KOMSTAT Team Performance Indicators are calibrated initially by scoring video tapes and are then re-calibrated every two months through group observation of the same operation. For the Crisis Management Behavioural Performance Markers, training involves watching and rating four scenarios separately and then discussing the scores. However, from their research the developers have found that achieving good inter-rater agreement on specific items is less of a problem than that of how to produce an aggregate score for behaviour that is inconsistent within a scenario. The institution designing the standardised points system for evaluating clinical competence, is currently assessing its inter-rater reliability but believes that specific training should not be necessary to use the system. This suggests that it must address very clear and unambiguous behaviours or perhaps adherence to specific protocols or guidelines.

5. Discussion

The aim of WP2 was to obtain and review existing behavioural observation scales, marker systems and checklists used in operating theatres and specifically by anaesthetists. What is clear from the results of this investigation is that such scales are not used extensively either in training or assessment. While simulator training is on the increase in anaesthesia, and seeks to address non-technical aspects of performance through human factors theory and scenario-based crisis management training, it is not compulsory. Courses of this type, for example Gaba's ACRM or the Swiss TOMS training programme have been found by participants to be very beneficial although full evaluation of their effectiveness is only beginning to be conducted. The reason for this, ironically, being that before performance can be measured, it is first necessary to develop an effective assessment system and these are still being developed. None-the-less, recognition of the importance of human factors in anaesthesia and medicine is encouraging designers of training courses to focus more on these areas and as a result new systems for measuring non-technical or behavioural performance are being developed.

Two main approaches were taken to this investigation into behavioural marker systems being used in anaesthesia. The first was to take non-technical skill checklists and behavioural marker systems identified in the WP1 literature review and sort them to identify common skills and behaviours. The second, was to discover how these, and any other systems that were available and could be added to the analysis, are used through an informal survey amongst a small sample of simulator centres. From these two approaches, six marker systems/checklists were identified and sorted. Eight responses were received from the survey, showing that only two establishments were using a behavioural marker system for training at the current time. Another establishment has a system used for research, which is being revised, and uses a checklist as a framework for training, and two other groups are evaluating newly developed systems at the present time. One of the marker systems identified was developed as a selection tool and is not known to have been used in training. It was included as it was specifically developed to address non-cognitive skills in anaesthesia residents. Several of the existing systems have been adapted from behavioural marker systems used for CRM training in aviation. Here, CRM marker systems have been rigorously developed and validated for use as assessment tools, and within Europe may be used for evaluating pilot competence for licensing requirements. As the aviation systems form the basis for several of the medical systems, several of the issues relating to the development of behavioural marker systems from this domain have been discussed in this report. To provide some structure to this discussion, it is divided into the five areas of considerations identified by Flin and Martin (2001) and described in the introduction.

1. Unit of assessment
Selecting the unit of assessment in aviation is about deciding whether to focus on an individual performance or the teams' performance. In the USA the focus is on the whole crew, whereas in Europe the focus is on the individual pilot. The consequence of this for anaesthesia is that several of the aviation marker systems, most notably NASA/UT LLC, have been developed to consider team performance. Therefore, many of the behavioural markers focus on team related behaviours and some of the

cognitive skills that can be difficult to observe are inferred through explicit team behaviour, e.g. verbal communication. This assessment of the whole team is also used in anaesthesia and medical team training. The KOMSTAT Team Performance Indicators and the ORCL, which are versions of the same tool, are used for observation of the whole operating team performance. The Crisis Management Behavioural Performance Indicators are also used to team performance but there scope to comment on the 'hotseat' anaesthetist and the anaesthesia crew.

There are two possible problems that arise with scoring the whole team's performance: i) It is difficult to give individual anaesthetists feedback on their own skills and development if they are always being considered in relation to the wider team. This can be particularly so if not all members of the team perform to the same level. ii) There is an assumption that the anaesthetist will be working in a team similar to that on the training course. In an aircraft cockpit there are always a minimum of two pilots this is not necessarily true for anaesthetists. While the anaesthetist will always be working with in the context of the operating theatre team and so needs to be aware of and communicate with the surgeons, nurses, other relevant medical staff and the patient, it may not the case that there is a large anaesthetic team. Often anaesthetists work alone and so, unless there is a training situation, or they need to request assistance from a colleague, many of the non-technical skills needed will be related to planning, maintaining situation awareness and decision making on an individual basis. This means that the behavioural markers used in anaesthesia may not be totally effective if all are based on team related behaviours. Clearly these team skills are very important (and they were what the LLC was designed to investigate) but the cognitive skills are also very important.

Not only have several of the marker systems used in anaesthesia training been developed from aviation courses, they are mostly also used for crisis management/resource management training. In these situations, it is perhaps more likely that the team is enlarged, and communication within and between the medical disciplines becomes even more important than normal. If efforts are to be made to enhance non-technical performance in normal situations as advocated by Helmreich's (2000) error management approach to CRM, then it is necessary to focus on the individuals' skills and address these in simulator training and assessment. It may be that the skills themselves do not alter much from the team to the individual, e.g. planning, decision making, communication seem to be important in all marker systems reviewed but the development of behavioural markers needs to take this into consideration – several of the marker systems already distinguish between leader and team member behaviours.

2. Identification of CRM skills and associated behavioural markers

From the six marker systems identified and reviewed, three were derived from the NASA/UT LLC (two being different versions of the same checklist), the ACRM Principle Markers have their origins in CRM training and the ETCC Teamwork Performance Matrix was based on army helicopter crew co-ordination training. Only the behaviour categories developed for selection at Iowa University have no foundations in the aviation industry. What this means is that there have been few attempts to develop marker systems for anaesthetists' non-technical skills based on systematic analysis of the task requirements. This will be the aim of the next stage in this project, but because such a task is so complex, it is necessary to start with some

previously identified skills that can be used to structure interview questions and provide a reference point for discussion. However, rather than selecting just one marker system for this purpose, the six identified systems were reviewed and the items sorted to produce a set of common items. Two lists of skills were produced, at two different level of detail. Ideally, the lower level items should map directly on to the higher level items but given the limited number of lists in the analysis, the differences in their structure and usage, this linkage was not made. Indeed because the same skill could be included at different levels within a system there was some repetition of items in the two derived lists.

To help provide some structure to the review and sorting of the anaesthesia marker systems, the classification of skills into category/element/behaviour used in NOTECHS and other aviation marker systems were added to the list items. However, this did not really add much to the analysis other than to provide labels for the different item levels. While the medical marker systems are constructed on a hierarchical basis, beyond grouping certain skills together, or behaviours relating to those skills, there was relatively little classification into skill types, e.g. social and cognitive. The most likely reason for this is that most of the marker systems were not developed to be used in that way and were not developed following systematic breakdown of skills into categories, elements and behaviours. Indeed, one difficulty of describing items as certain skill categories or elements arose because not all items can actually be classified as skills. Several of the element level items sorted were not related to skills, i.e. the ability to perform an activity, but to the activity itself, e.g. 'calls for help', 'implement plans', 'prepares for post event care', 'emergency situation'. These items relate to important skills, e.g. 'situational self awareness' or 'use of authority/assertiveness', 'preparation and planning', and 'problem recognition', but are not really skills in themselves. Such items were not excluded from the analysis because that could have resulted in the potential loss of vital information but rather, they were grouped with other items describing equivalent skills. The anaesthesia marker systems were not reviewed at the behavioural marker level because sorting them at this level was found to be open to considerable interpretation, and behaviours could be taken to have different meanings depending on the skill they were describing.

What is interesting is that sorting the anaesthesia marker systems to identify common themes and skills results in skill lists similar to those used in aviation, e.g. NOTECHS. There are two possible reasons for this: i) the lists were derived from those developed in the aviation industry and have not actually had to be significantly altered for use in medicine; ii) the non-technical skills important in anaesthesia are similar to those recognised as being necessary in aviation. As one of the checklists was developed specifically for use in the medical domain, it is possible that both of these statements are partly true. However, until interviews are carried out with experienced anaesthetists in the next workpackage, it will not be possible to say for certain how much the similarity between aviation non-technical skills and anaesthetist non-technical skills is due to cross-feed between the domain and how much is due to similarities in skill requirements.

3. Assessment method

While assessment of non-technical skills is not compulsory in anaesthesia, some attempts have been made by researchers and simulator trainers to develop rating scales to allow for the assessment of skills. Indeed, as previously described, several systems are currently under development, evaluation and revision. Following on from their predecessors in aviation, such rating systems are based on observer judgement of behavioural markers during simulator training sessions. The rating systems seem to use four or five point scales, with descriptors for the level of proficiency associated with each, and against which performance is scored. As with aviation CRM, a necessary requirement for use of observer judgement is that the relevant behaviour can be observed either directly, or indirectly through other behaviours. This is easy for the social skills as they mainly relate to team working skills or interaction with other. Many of the observable behaviours related to cognitive skills also rely on communication. For example, while it is difficult to observe decision making, the collection of information or discussion of options indicate that the process is being followed. Unfortunately, if an anaesthetist is, for the most part working alone, the appropriate social skills will still be observable but some of the cognitive skills may be difficult to detect. In the marker systems reviewed, the strongly cognitive skills appear to be less well represented in the skill lists, perhaps for this very reason. What is clear from the analysis, is that, while markers do need to be based on, or represented by, observable behaviours, the inclusion of skills in the taxonomy should not be based on whether or not it is easy to select these markers.

In terms of the rating scales that could be used with the anaesthesia behavioural markers, as the purpose of the assessment is to provide training feedback not formal evaluation, scales with four or five points seem to be used widely in aviation and also in anesthesia. Perhaps the main limiting factors regarding the rating scale is the level of reliability that can be achieved by multiple raters using the scale to assess the same performance. This issue is addressed briefly below.

4. Reliability and 5. Rater Training

The issue of intra-rater and inter-rater reliability have received some, but not extensive, consideration with regard to the use of behavioural markers in anaesthesia. Perhaps the main reason for this is that performance assessment is not very common and thus the problem has not been widely encountered. Several of the systems identified are used only as checklists and not for performance assessment and thus the reliability of ratings against the items is not an issue. Of the marker systems/checklists identified in this report, those under development and revision are all being evaluated to ensure their reliability. There are clearly difficulties associated with observer judgements of behaviour but there are also ways to help reduce levels of bias. Providing training for the trainers and ensuring they are calibrated to a predefined standard when using the rating scale is important. Where marker systems are being used in anaesthesia this has been recognised and trainers are checked to ensure they are using the scales accurately. For this project, it will be important to keep track of their findings, and during later stages when the taxonomy and marker system are being validated it may be possible to benefit from the experience of others in the field and also from the lessons learned from the aviation project JARTEL

6. Conclusion

From the results of this review of existing behavioural marker systems used in anaesthesia, it is apparent that this is still a new area of training practice. There is no one agreed system used or one common process for developing identifying skills and markers. Indeed much of the source material for the anaesthesia behavioural markers has been derived from aviation CRM training and assessment programmes, specifically the NASA/UT Line/LOS Checklist, in various versions, and is mainly used for training and not assessment.

To obtain a representative set of skills considered to be important for the non-technical aspects of anaesthetists' performance, six systems obtained from developers and the literature were sorted to identify common skills at both the category level, where possible, and the element level (see section 3.2). While there is no real linkage between two resulting of skill lists, it is possible to see how the element skills could be groups into smaller higher levels sets of skills as in the category level.

However, as a workpackage output, it is particularly the skills at the element level that will be used to form the basis of the interviews to be conducted in the next stage of the project (WP3). This list of currently identified anaesthetist non-technical skills, at the element level is shown in Table 3 below. This table is as Table 2 (p21) but without the example behaviours. The WP3 interviews will be conducted with experienced anaesthetists with the aim of investigating the extent to which these skills are considered to be important non-technical skills for anaesthetists, and to identify any skills that need to be added or removed.

Element level	
Social Skills	Communication
	Information sharing
	Feedback
	Leadership
	Team building and climate
Cognitive Skills	Preparation and planning
	Workload management
	Prioritisation and allocation of attention
	Monitoring and recognition
	Situation awareness and decision making
	Re-evaluation
Other	Teaching
	Declaring emergency
	Initial Crisis management
	Technical and clinical skills
	Overall effectiveness

Table 3. Anaesthetists' Non-technical Skills from existing marker systems – element level.

7. References

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

Altmaier, E.M., From, R.P., Pearson, K.S., Gorbatenko-Roth, K.G. & Ugolini, K.A. (1997). A Prospective Study To Select and Evaluate Anesthesiology Residents: Phase 1, The Critical Incident Technique. *Journal of Clinical Anesthesia, 9, 629-636.*

Birnbach, R. A. & Longridge, T. M. (1993). The Regulatory Perspective in E.L. Weiner, B.G. Kanki and R.L. Helmreich (Eds.), *Cockpit Resource Management*. London: Academic Press.

Cooper, J. B., Newbower, R. S., Long, C. D., & McPeek, B. (1978). Preventable Anesthesia Mishaps: A Study of Human Factors. *Anesthesiology, 49*, 399-406.

Flin, R. & Martin, L. (2001) Behavioural Markers for Crew Resource Management: A survey of current practice. *International Journal of Aviation Psychology,* 11 (1), 95-118

Flin, R. & Martin, L. (1998). *Behavioural Markers for Crew Resource Management.* CAA Paper 98005. London: Civil Aviation Authority.

Fowlkes, J.E., Lane, N.E., Salas, E., Franz, T. and Oser, R. (1994). Improving the measurement of team performance: the TARGETs methodology. *Military Psychology, Vol. 6 (1), 47-61.*

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

Gaba, D. M., Fish, S. K., & Howard, S. K. (1994). *Crisis management in anesthesiology*: New York: Churchill Livingstone.

Gaba, D. M., Howard, S. K., Flanagan, B., Smith, B. E., Fish, K. J., & Botney, R. (1998). Assessment of Clinical Performance during Simulated Crises using both Technical and Behavioral Ratings. *Anesthesiology, 89*, 123-147.

Hamman, W.R., Beaubien, J.M. & Holt, R.W. (1999) Evaluating instructor/evaluator inter-rater reliability from performance database information. Proceedings of Tenth International Symposium on Aviation Psychology. Columbus, OH: Ohio State University

Hansberger, J.T., Holt, R.W. & Boehm-Davis D. (1999) Instructor/evaluator evaluations of ACRM effectiveness. Proceedings of Tenth International Symposium on Aviation Psychology. Columbus, OH: Ohio State University

Helmreich, R.L. (2000). On error management: lessons from aviation. *British Medical Journal, 320, 781-785 (18 March).*

Helmreich, R.L. (1991) Strategies for the Study of Flightcrew Behaviour. Proceedings of Sixth International Symposium on Aviation Psychology. Columbus, OH: Ohio State University

Helmreich, R. L., Butler, R. E., Taggart, W. R. & Wilhelm, J. A. (1997). The NASA/University of Texas/FAA Line/LOS Checklist: A behavioral marker-based checklist for CRM skills assessment. (Version 4.4 dated 1.2.97) Austin: NASA/University of Texas/FAA Research Group.

Helmreich, R.L. and Merritt, A.C. (1998*) Culture at Work in Aviation and Medicine*. Aldershot, Hants: Ashgate.

Helmreich, R. L., Schaefer, H.-G., & Sexton, J.B. (1995). Operating Room Checklist (Aerospace Crew Resource Project Technical Report 95-4). Austin, Texas: University of Texas.

Helmreich, R. L., Wilhelm, J. A., Kello, J., Taggart, W. R. & Butler, R. E. (1991) Reinforcing and evaluating crew resource management: Evaluator/LOS instructor reference manual. Austin: NASA/UT Technical Manual 90-2.

Holt, R.W., Boehm-Davis, D.A. & Beaubien, J.M. (2001). Evaluating Resource Management Training. In E. Salas, C. Bowers and E. Edens (Eds.) *Applying Resource Management in Organisations: A guide for training professionals.* Mahwah, NJ: Lawrence Erlbaum Associates.

Howard, S. K., Gaba, D. M., Fish, K. J., Yang, G. S., & Sarnquist, F. H. (1992). Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation, Space, and Environmental Medicine, 63*(9), 763-770.

Lauber, J. (1984) Resource Management in the Cockpit. *Air Line pilot, 53, 20-23*.

Leape, L. L. (1994). Error in medicine. *Journal of the American Medical Association, 272(33)*, 1851-1868.

NOTECHS (1998). Final Report: The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements. NLR-CR-98443. Amsterdam: NLR.

O'Connor, P., Hörmann, H-J., Flin, R., Lodge, M., Goeters, K-M. & the JARTEL Group. (2002) Developing a method for evaluating Crew Resource Management skills: A European perspective. *International Journal of Aviation Psychology, 12 (3), 263-85*

Royal College of Anaesthetists (1994) Specialist Training in Anaesthesia, Supervision and Assessment. London: Royal College of Anaesthetists

Seamster, T.L. and Edens, E.S. (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.L., Hamman, W.R. and Edens, E.S. (1995) Specification of observable behaviours within LOE/LOFT event sets. In *Proceedings of 8th International Symposium on Aviation Psychology*. Columbus: Ohio State University.

Seamster, T.L., Prentiss, F.A. and Edens, E.S. (1997) Methods for the analysis of CRM skills. In *Proceedings of 9th International Symposium on Aviation Psychology*. Columbus: Ohio State University.

Sexton, B., Marsch, S., Helmreich, R. L., Betzendoerfer, D., Kocher, T., & Scheidegger, D. (1997). Participant evaluation of Team Oriented Medical Simulation, *Proceedings of the Second Conference on Simulators in Anaesthesiology Education* (pp. 107-108). New York: Plenum.

Small, S.D., Wuerz, R.C., Simon, R., Shapiro, N., Conn, A. and Setnik, G. (1999) Demonstration of high-fidelity simulation team training for emergency medicine. *Academic Emergency Medicine, Vol. 6 (4), 312-323.*

Williamson, J. A., Webb, R. K., Sellen, A., Runciman, W. B., & Van Der Walt, J. H. (1993). Human failure: an analysis of 2000 incident reports. *Anaesthesia and Intensive Care, 21*(5), 678-683.