The Assessment of Registrars' Non-technical Skills in the Emergency Department

Dr Lynsey Anne Flowerdew

Imperial College, London

Department of Surgery and Cancer

Submitted for the degree of MD(Res)

July 2011

Abstract

In recent years safety in medicine has been high on the agenda, both for government and for healthcare providers. This thesis starts by describing the evolution of patient safety and then goes on to explore error specifically in the Emergency Department (ED). Focus is drawn to the role of non-technical skills for improving safety.

The initial broad aim of this research was to learn how the ED team could function better to improve patient care. An interview study is developed to investigate how ED staff change their behaviour during periods of high demand and to determine the direction of future research. This study highlights that staff would benefit from increased awareness of the non-technical skills that contribute to effective teamwork and enhanced patient safety. The interviews also reveal the leadership role of the registrar is of particular importance. Therefore, a series of studies are developed to identify and describe the non-technical skills required by Emergency Medicine trainees, with a specific focus on leadership.

The process of developing a provisional assessment tool for assessing non-technical skills in the ED is described. This draws on published literature and curricula as well as considering existing methods of assessment. The assessment tool is revised using re-analysis of staff interviews and a series of preliminary observations in the ED. Content validity of the tool is measured using a survey of expert opinion and this helps to further refine the tool components. An experimental study reveals that whilst adequate levels of inter-rater reliability are achievable, rater accuracy appears to be more problematic. Various sources or rater error are also explored and this leads onto a larger, multicentre observational study investigating use of the tool in the workplace. Further data for reliability is collected and field notes are analysed to provide a detailed description of the non-technical skills displayed by ED registrars.

Findings of the studies are summarised and limitations, applications and further research are discussed.

Acknowledgements

First of all, I would like to thank my husband and daughter for their patience and support during the writing of this thesis. I am especially grateful to my mother, who knows more about non-technical skills than any retired grandmother should.

I would like to express appreciation to my supervisors, Dr Maria Woloshynowych, Dr Ruth Brown and Prof Charles Vincent, for their wisdom and encouragement.

I would also like to take this opportunity to thank Dr Arran Gaunt, Dr Jessica Spedding and Dr Ajay Bhargava for their hard work collecting data for the multi-centre study. Finally, I would like to thank all the registrars who graciously let us observe them and, without whom, this research would not have been possible.

Declaration of Originality

The work included in this thesis in my own work and all else is appropriately referenced.

Contents

Abstract 2

Acknowledgements 3

List of Illustrations 11

List of Tables 12

List of Abbreviations 13

1 Introduction: Safety in the Emergency Department

Chapter overview 14

1.1 Evolution of patient safety 14

1.2 The size of the problem in the emergency department 15

1.3 The error prone ED 16

1.4 Models of error and incident analysis 19

1.5 The challenges of safety research in the ED 20

1.6 Factors contributing to error in the ED 22

2 Introduction: Teamwork and Non-technical skills in the ED

2.1 Defining teamwork, non-technical skills and leadership 25

2.2 Origins of teamwork research in industry and healthcare 27

2.3 Leadership 28

2.3.1 The importance of leadership 28

2.3.2 Leadership in the ED 29

2.4 Improving patient safety in the ED 30

2.4.1 Team training and Crew Resource Management 30

2.4.2 Simulation 33

2.4.3 Other interventions to improve non-technical skills 33

2.5 Assessing non-technical skills 34

2.5.1 Methods of assessing non-technical skills 34

2.5.2 The development of behavioural marker systems in healthcare 35

2.5.3 Evaluation of behavioural marker systems 38

2.5.4. Assessing non-technical skills in EM 39

3 Staff Interview Study

Chapter overview 42
3.1 Background 42
3.2 Aims 43
3.3 Methods 44
3.3.1 Design and participants 44
3.3.2 Materials 44
3.3.3 Procedure 45
3.3.4 Data analysis 45
3.3.5 Ethical Issues 46
3.4 Results 46
3.4.1 Stressors 46
3.4.2 Effects of working under pressure 48
3.4.3 Suggested Interventions and advice 50
3.4.4 Senior versus junior staff 53
3.5 Discussion 54

4 Non-technical skills associated with safety and error in the ED – A scoping review of the literature

Chapter overview 57
4.1 Background 57
4.2 Methods 58
4.2.1 Stage 1 58
4.2.2 Stage 2 60
4.3 Results 61
4.3.1 Stage 1 61
4.3.2 Stage 2 61
4.4 Discussion 70
4.4.1 Limitations 70
4.4.2 Summary 71

5 Development of a provisional observational tool to assess non-technical skills in the ED

Chapter overview 72
5.1 Background 72
5.2 Aims 75
5.3 Methods 77
5.4 Results 79
5.4.1 Review of ED safety literature 79
5.4.2 Review of non-technical skill literature 79
5.4.3 Curricula 81
5.4.4 Roles and Responsibilities 81
5.4.5 Existing Forms of Assessment in the ED 82
5.4.6 Summary of sources 84
5.4.7 Inclusion or exclusion of skills to produce a provisional assessment tool 85
5.4.8 Summary of results 91
5.4.9 The rating scale 93
5.4.10 Evaluation of feasibility 94

6 Revision of the Assessment Tool

Chapter overview 95
6.1 Background 95
6.2 Study 1 ï Recoding Interviews 97
6.2.1 Aims 97
6.2.2 Methods 97
6.2.3 Results 98
6.2.4 Summary of Findings 105
6.3 Study 2 - Field observations 106
6.3.1 Aims 106
6.3.2 Methods 106
6.3.3 Results 107
6.3.4 Summary of findings 114
6.4 Inclusion or exclusion of skills for Version 2 of assessment tool 114

6.5 Discussion 118

7 Validation Questionnaire Study

Chapter overview 121
7.1 Background 121
7.2 Aims 122
7.3 Methods 122
7.3.1 Design 122
7.3.2 Participants and Procedure 122
7.3.3 Materials 123
7.3.5 Data analysis 124
7.4 Results 125
7.4.1 Assessing content validity 126
7.4.2 Identifying major omissions in the tool 127
7.4.3 Exploring staff attitudes to assessment 130
7.5 Discussion 132
7.1 Limitations 133

8 Training EM physicians to use the tool for assessment of non-technical skills

Chapter overview 135
8.1 Background 135
8.2 Aims 135
8.3 Methods 136
8.3.1 Design 136
8.3.2 Participants 137
8.3.3 Materials 137
8.3.4 Procedure 137
8.3.5 Data analysis 138
8.4 Results 139
8.4.1 Inter-rater agreement 139
8.4.2 Accuracy 140

8.4.3 Participant feedback 143
8.5 Discussion 147
8.5.1 Inter-rater agreement and accuracy 147
8.5.2 Feedback from participants 149
8.5.3 Strengths and Limitations 150
8.5.4 Summary 150

9 A Multi-centre Observational Study

Chapter overview 152
9.1 Background 152
9.2 Aims 153
9.3 Methods 154
9.3.1 Design 154
9.3.2 Setting 154
9.3.3 Participants 154
9.3.4 Materials 156
9.3.5 Procedure 156
9.3.6 Ethical Issues 157
9.3.7 Data analysis 158
9.4 Results 159
9.4.1 Evaluation of reliability 159
9.4.2 Observability 164
9.4.3 Description of the range of skills 164
9.4.4 Comparison of hospitals 171
9.5 Discussion 173
9.5.1 Reliability 173
9.5.2 Observability 176
9.5.3 Description of range of skills 177
9.5.4 Comparison of hospitals 177
9.5.5 Strengths and Limitations 178
9.5.6 Summary 179

10 General Discussion

Chapter overview 180

10.1 Research Findings 180

10.2 Limitations and alternative approaches 183

10.3 Future Research 185

10.3.1 Piloting using the tool 185

10.3.2 Use in the resuscitation room 185

10.3.3 Other healthcare professionals 185

10.3.4 Non-technical skills and outcome 186

10.3.5 Other future research 186

10.4 Application of Research Findings 187

10.4.1 Formative assessment of trainees 187

10.4.2 Other assessment applications 187

10.4.3 Trainees as assessors 188

10.4.4 Training 189

10.4.5 Selection of trainees 190

Appendices

Appendix A	Materials for the interview study	208
Appendix B	Materials for the development of a provisional tool	218
Appendix C	Materials used to refine the assessment tool	226
Appendix D	Materials for the validation questionnaire study	229
Appendix E	Materials for the training day study	235
Appendix F	Further data for the multicentre study	236

List of Illustrations

Fig.3.1	Stressors mentioned by the ED team	47
Box 4.1	Non-technical skills associated with error and safety in the ED	61
Fig. 5.1	Millerôs pyramid for assessing clinical competence	74
Fig. 5.2	Summary of steps to develop the assessment tool	76
Box 5.1	Skills identified in the review of ED safety literature	79
Box 5.2	Skill list used for data collection	80
Box 6.1	12 non-technical skills used for coding interviews and observations	97
Fig. 7.1	Assessment form	128
Fig. 9.1	Mean and dispersion of ratings	166
Fig. 9.2	Comparison of mean rankings for each skill	171

List of Tables

Table 2.1	Methods used to derive behavioural marker systems in healthcare	36
Table 2.2	Techniques available to identify non-technical skills	37
Table 4.1	Summary of data extracted from Stages 1 and 2	62
Table 5.1	Summary of various sources used in the development of the provisional tool	84
Table 5.2	Skills included in the provisional tool	92
Table 5.3	Rating scale for the assessment tool	93
Table 6.1	A summary of findings from the interview study	105
Table 6.2	Summary of findings from observations in the ED	114
Table 6.3	Summary of skills supported for inclusion in the revised tool	115
Table 6.4	Categories and skills with associated definitions	117
Table 7.1	Participants in validation survey	125
Table 7.2	Responses to Q2 -Should non-technical skills be explicitly assessed?	130
Table 7.3	Responses to Q3 ï Would registrars like more feedback on these skills?	131
Table 7.4	Responses to Q4 ï Would it be valuable for consultants to actively observe registrars running the department?	132
Table 8.1	Values for variance and inter-rater agreement	140
Table 8.2	Accuracy of ratings and Mean Absolute Deviation from the reference rating	141
Table 8.3	Responses given on the feedback questionnaire	143
Table 9.1	ICCôs for individual skills and overall scale for each pair of observers	160
Table 9.2	Sources of rater error	162
Table 9.3	Observability of each of the 12 skills	164
Table 9.4	Descriptive statistics	165
Table 9.5	Examples of poor and good behaviour observed during the study	167
Table 9.6	Results of Mann-Whitney tests comparing hospitals for the skill Supervision and Feedback	172
Table 9.7	Results of Mann-Whitney tests comparing hospitals for the skill SA Gathering Information	172

List of Abbreviations

ACAT	acute care assessment tool
ANTS	Anaesthetists Non-Technical Skills
ARMC	Academy of Royal Medical Colleges
CDU	Clinical Decision Unit
CEX	Clinical Evaluation Exercises
CEM	College of Emergency Medicine
CVI	Content Validity Index
CRM	Crew/ Crisis Resource Management
DM	decision making
DOPS	direct observation of practical skills
ED	Emergency Department
EM	Emergency Medicine
FAST	Focused Assessment with Sonography for Trauma
FCEM	Fellow of the College of Emergency Medicine
FMEA	Failure Modes and Effects Analysis
ICC	Intra-class coefficient
ICU	Intensive Care Unit
LOSA	Line Operations Safety Audits
MAD	mean absolute deviation
MCEM	Membership of the College of Emergency Medicine
MLCF	Medical Leadership Competency Framework
MSF	multi-source feedback
NASA	National Aeronautics and Space Administration
NOTSS	Non-Technical Skills for Surgeons
OTAS	Observational Teamwork Assessment for Surgery
REC	Research and Ethics Committee
RCA	Root Cause Analysis
SA	situational awareness
SBAR	situation background assessment recommendation
SHO	senior house officer
WBA	workplace based assessment

1 Introduction: Safety in the Emergency Department

Chapter overview

This chapter introduces a brief history of patient safety in healthcare and describes influences from both national government and the study of human factors. Studies looking at critical incidents and adverse events reveal high rates of preventable error in the emergency department (ED) compared to other areas of the hospital. Consideration is given to the unique characteristics of the ED patient, illness and setting that may account for this. The methods and challenges of researching error in the ED are presented and the factors contributing to error are described. In particular, studies reveal that failures in teamwork provide the conditions that allow error to occur in the ED. Focus is directed on the importance of ónon-technical skillsô, including decision-making, communication and assertiveness. The relationship between non-technical skills and safety in healthcare is explored further in the next chapter. Leadership is identified as a vital aspect of teamwork and methods for identifying and assessing non-technical skills are reviewed. The introductory chapters provide the foundation for subsequent studies described in the thesis.

1.1 Evolution of patient safety

Working in the NHS today it is hard to imagine a time when patient safety was an unknown concept. We have always known that patients occasionally receive the wrong medication or that some patients contract infections whilst in hospital. However the focus on patient safety as a distinct set of ideas and initiatives can be traced back less than 20 years. One of the most influential reports was the Harvard Medical Practice Study commissioned by the US Department of Health (Leape, 1991). The study, which reviewed a sample of hospitalisations to assess the number of potentially compensable cases, was designed in response to increasing rates of litigation. The reportôs authors estimated between 44 000 and 98 000 annual US deaths were due to medical error and, for the first time, the true size of the problem was exposed.

In 1996, the Quality in Australian Healthcare Study was published which stated that 16.6% of hospital admissions resulted in an adverse event (considered to be an unintended injury

caused by medical management rather than the disease process). In the UK, Sir Liam Donaldson produced the report óAn Organisation with a Memoryô for the Department of Health (Department of Health, 2000). This document stated that research-based estimates show that adverse events occur in around 10% of hospital admissions and approximately half of these are considered preventable (Vincent, 2001). The report also estimated the cost at £2 billion a year in additional hospital stays alone, without taking into account the wider economic costs. A number of recommendations were set out relating to a variety of issues from staff training to incident reporting. Five years later, the National Audit Office published a report which recognised that some important progress was being made in creating a more open culture in which errors could be reported and discussed (National Audit Office, 2005). However, the authors once again expressed concern that we did not appear to be learning from mistakes or implementing change in a way that prevented similar events from recurring.

In contrast to the procedural and regulatory nature of these reports, the study of human performance introduced a whole new approach to tackling patient safety. This focus on human factors and ergonomics developed largely outside of healthcare in high risk industries such as aviation, nuclear power, and the oil and gas industries. Its roots stem from psychology and engineering and, over the decades, has developed from straightforward cockpit redesign to analysis of complex sociotechnical systems. The study of human performance looks to aspects such as teamwork and the work environment to explain both the sources of error and to find solutions (Helmreich, 1996, Stone, 2004). However, there is a growing realisation that the safety problems of healthcare are difficult and simple application of existing solutions from other industries is not a viable strategy. Patient safety will continue to present challenges to researchers, politicians and healthcare providers for many years to come.

1.2 The size of the problem in the emergency department

A number of major epidemiological studies in a variety of countries have investigated hospital adverse events (Leape, 1991, Wilson, 1996, Thomas, 2000, Vincent, 2001, Baker, 2004). The studies found that adverse events were commonly related to drug treatment and operative complications, such as wound infections. Some adverse events were deemed to be unpredictable and unpreventable, such as an allergic reaction in a patient with no previous

history of medication allergy. Authors also attempted to determine if the adverse event was due to clinical negligence and, therefore, preventable. Examples included failure of administration of medication or diagnostic error. Researchers found that a small proportion of all adverse events stem from care in the ED (1-3%) compared to the operating room (around 40%) (Leape, 1991, Thomas, 2000). However, these studies consistently identified that EDs have the highest proportion of preventable events (50-70%). These studies were based on record reviews of in-patents and, as a high proportion of ED patients are never admitted, it is likely the burden of preventable injury is even higher.

An Australian surveillance study based on both voluntary and mandatory reporting of ED incidents revealed about 97% of the critical incidents were deemed preventable (Vinen, 2000). The team defined critical incidents as ñany unintended event which is inconsistent with routine hospital practice or of the quality of patient care which has had or could have had a demonstrably adverse outcome for a patientò. Therefore, incidents may or may not result in an adverse event. An observational study based in two UK emergency departments found 1-2% of patients who attend the ED suffer a critical incident (Thomas, 2008). A total of 443 critical incidents were detected by review of ED records, direct observation in the ED, review of radiology and laboratory reports and critical incident reporting by staff. Another UK study identified all patients who died within 7 days of admission through the ED in a 4 month period. Two ED consultants then independently reviewed the quality of care they received (Nafsi, 2007). Of the 95 cases investigated, 12% of deaths were determined to be possibly or probably preventable due to failures in the ED. This can be extrapolated to almost 35 potentially preventable deaths per year in a single ED (annual ED attendance of 65 000) and does not include discharged patients so this may be an underestimate. These statistics clearly demonstrate a need to focus on error reduction to improve morbidity and mortality in patients attending the ED.

1.3 The error prone ED

It is important to consider what is unique to emergency care in order to understand why the ED is susceptible to error. Failures in patient care may be attributed to three main factors relating to the patient, the illness and the setting of the ED.

The patient

Typically, the emergency patient is unknown to the doctor and the visit is unscheduled (Croskerry and Sinclair, 2001). Many arrive without a list of their medications and patients often struggle to recount a concise summary of their medical problems. This is compounded by difficulties contacting general practitioners and a lack of access to medical records. Communication may be difficult or impossible due to dementia, delirium, alcohol intoxication, unconsciousness or unknown foreign languages. This means that the emergency physician has to work with high levels of uncertainty. Patients may be disruptive or aggressive due to psychiatric disturbance, drug and alcohol use or may arrive involuntarily accompanied by the police. Such patients may be uncooperative, hostile or purposively deceptive. The ED may be seen as the only available access to medical help for marginalised members of the community such as the homeless and illegal immigrants or may be used as a refuge for victims of domestic violence. These patients may be fearful of divulging information and may pose particular difficulties arranging safe follow up care.

The illness

The type of medical illness or injury encountered in emergency practice is unlimited. Problems are undifferentiated and of varying acuity. Cosby and Croskerry describe high ósignal-to-noise ratioô where many harmless conditions mimic significant illness and many significant conditions present with a harmless appearance (Cosby and Croskerry, 2009). These factors partly explain why diagnostic error is a particular problem in the ED. The nature of acute illness means that it is often necessary to intervene rapidly with incomplete information and this poses increased risk to patients. Many patients in the ED are critically unwell and require interventions that are likely to carry inherent risk. Examples may include thorocostomy and endotracheal intubation. Furthermore, these interventions take place in less controlled or sterile environments than that of the operating theatre or intensive care unit. For many patients, the ED is the first medical point of contact during their illness. This means the signs and symptoms of their disease may not be fully manifested, making diagnosis difficult. It also highlights the vital importance of close monitoring as patients may deteriorate suddenly and unexpectedly.

The setting

Emergency departments are open all day, every day, and whilst ambulances may be diverted occasionally, departments never close. The workload is essentially limitless and emergency departments are frequently overcrowded and understaffed (Fatovich, 2003, Cowan, 2005). An economically efficient approach suggests the ED should run at capacity most of the time. This leaves little room for the inevitable surges that come with local events, disasters, seasonal illnesses or even random variation. Staff in UK emergency departments have had the additional pressure of ensuring patients are either admitted or discharged within four hours of arrival. This time constraint has encouraged rapid, sometimes hasty decisions to be made. Historically, EDs have been staffed by predominantly junior, inexperienced doctors. Although it is now recognised that this needs to change, there is a shortage of experienced emergency physicians in the UK. In many departments, the leadership role is filled by registrars and clinical fellows with little or no training in the complexities of coordinating a busy ED. Over the years, departments have often grown in an ad hoc way, with little attention to design for safety or efficiency (Cosby, 2009, Gosbee, 2004). Furthermore, physicians are expected to manage multiple patients simultaneously, often complicated by constant interruptions and breaks in tasks (Chisholm, 2000). These factors lead to a busy and sometimes chaotic environment that is vulnerable to errors and mistakes.

By considering the unique characteristics of the emergency patient, their illness and the emergency department setting, it is easy to understand why errors occur. However, the situation is further compounded by the hidden nature of error in the ED. This can be partly attributed to the lack of ównershipôof a patient, as many healthcare providers are involved in each patient visit (Cosby and Croskerry, 2003). In addition, fragmented delivery of care relies on adequate feedback mechanisms if clinicians are to learn from their mistakes (Kostopoulou and Shepherd, 2000). There is a distinct lack of feedback in the ED and Croskerry identified a number of reasons for this (Croskerry, 2000). Systems for routinely returning outcome data, such as discharge summaries, to emergency practitioners are uncommon or nonexistent. Emergency physicians are time pressured and this leaves little time to make phone calls or track reports that would reveal diagnostic error and thus facilitate learning. Most errors become apparent after the patient has either been discharged or admitted. There is often no reliable mechanism for general practitioners or ward staff to feedback to the ED so that errors can be highlighted and solutions considered. Transition of

care due to shift work means that even when errors are uncovered in the ED they may be detected by someone unrelated to the incident. Incident reporting systems can be a valuable method of ensuring errors are brought to light and investigated. However, not all incidents are reported and incomplete recordings can make follow-up difficult (Stanhope, 1999, Tighe et al., 2006).

Despite the multitude of reasons for considering the ED a high risk environment, it is not acceptable to simply conclude that error is inevitable in the ED. Fortunately, patient safety is gaining increasing attention and research efforts have started to focus on how to make the ED a safer environment.

1.4 Models of error and incident analysis

In order to design ways of reducing error, it is first necessary to have a full understanding of the causes of error in the ED. Traditionally, error causation has focussed on individual blame. This approach views unsafe acts as the result of forgetfulness, inattention or poor motivation. Subsequently, attempts to reduce error have relied on poster campaigns to increase awareness, staff training, disciplinary action or other efforts aimed at reducing human variability. An alternative approach is to accept that humans are fallible and consider errors as a consequence rather than the cause of a problem. Countermeasures rely on understanding the error provoking properties of the system and developing appropriate defences and safeguards to protect patients. However, these barriers may have weaknesses or defects that Reason likened to the holes in a slice of Swiss cheese (Reason, 2000). The presence of holes in any one slice does not usually produce a negative outcome. However, when the holes in many slices (or defences) line up, the conditions are set for errors to be propagated and result in potential harm to patients. The ôholesô may represent either active failures or latent conditions and most adverse events are due to a combination of both (Reason, 2000). Active failures occur around the time of the incident, with immediate effect, and are usually committed by people in direct contact with the patient. Latent conditions such as understaffing, equipment failure or design deficiencies provide an environment that allows errors to occur. Improving patient safety relies on gaining a deep understanding of

both the human factors associated with errors as well as the making the system as robust as is possible.

A number of methods have been developed for investigating errors and identifying the human and system factors involved. Root cause analysis (RCA) is a retrospective approach to error analysis and was widely applied to investigate large industrial accidents (Adamski, 2000) but has been adapted for use in the NHS (National Patient Safety Agency, 2004). RCA provides a structured and process-focussed framework with which to approach sentinel events. However, the investigation is subject to hindsight bias and may be interpreted that a single, root cause can be determined. The London Protocol describes a process for accident investigation that evaluates the task, team factors, individual factors, patient factors, the work environment and the organisation (Taylor-Adams, 2004). Whilst retrospective, this method emphasises a forward-looking approach to identify what can be learnt from the event to prevent similar incidents happening in the future. Failure modes and effects analysis (FMEA) prospectively examines the likelihood of failure before a process or system is designed (or redesigned) to minimise risk (DeRosier et al., 2002, Duwe et al., 2005). FMEA utilises the multi-disciplinary team to analyse an error-prone process with the aim of preventing major adverse events from occurring. A similar, less time-consuming approach is being tested for widespread use in the NHS and is known as Prospective Hazard Analysis (Ward, 2006). This shows that a variety of types of incident analysis, including both retrospective and prospective methods, are available for clinicians and researchers to develop a deeper understanding of error.

1.5 The challenges of safety research in the ED

The analysis of known incidents is a common method of investigating error in healthcare and has made important contributions to the understanding of error in the ED (Vinen, 2000). However, research has shown that analysis of critical incidents detects only a small fraction of errors and adverse events demonstrable by other means (Stanhope, 1999, Sari, 2007).

An alternative to the analysis of incident reports may include note review but documentation in the ED is notoriously brief and sometimes woefully inadequate. A study was undertaken

to see if a retrospective review of ED documentation could detect incidents registered in an anonymous reporting system. Thirty sets of notes were reviewed independently by two blinded ED physicians and they each missed nearly half of the reported incidents (Shapiro, 2002). This highlights the weaknesses of large studies such as the Harvard Medical Practice study (Leape, 1991) which relied on notes review to detect medical errors. Other difficulties in gathering data result from the lack of standardised coding for presenting complaints making it difficult to identify and study a particular condition. Researchers may have to trail through a pile of notes labelled ónunwell adultô or risk losing relevant case notes. Possibly the most significant barrier to effective research is discovering outcomes of patients when 80% are discharged from the ED. This is a particular problem when considering that some of the most consequential failures in the ED stem from discharging a patient who should have been admitted (Pope, 2000). Identifying and following up these patients may be virtually impossible if they chose to go elsewhere for their subsequent care.

Investigating error in the ED may also involve qualitative research methods looking at staff and patients in the natural setting. However, there are a number of barriers to effective research efforts. Firstly, clinical work in emergency care is complex, characterised by staff managing multiple patients and multiple tasks in parallel. This is in contrast to the operating room, out-patient clinic or primary care setting where single-patient care is the norm. This complexity introduces a whole new set of contributing factors when considering sources of error which is difficult to disentangle (Wears et al., 2010). Secondly, the distributed nature of clinical work over patients, over space and over providers provides practical challenges to observational research. A decision needs to be made whether to follow the clinician or the patient and accept that observations will be fragmented and potentially give an incomplete picture. Moreover, the rapid timescale on which events unfold in the ED makes it difficult for observers to recognise the features of the context that are pertinent to understanding how things went wrong.

Whilst some of these difficulties are inevitable, Wears et al considered strategies for overcoming some of the challenges that research in the ED presents (Wears et al., 2010). For example, the use of multiple observers may help to build a more complete picture than a single observer limited to particular patient, clinician or geographical area. Training observers in slower-paced environments such as the operating theatre or ICU could help researchers develop their observational skills for use in the more rapid pace of the ED setting.

Most importantly, researchers should anticipate these issues and actively search for solutions and alternatives during the planning stage of any ED project.

1.6 Factors contributing to error in the ED

Analysis of these sources of error enables consideration of how to address error in the ED. A UK research group analysed critical incidents in two EDs to identify and classify common órisk typesô (Thomas et al., 2004). Risk is defined as óan avoidable increase in the probability of an adverse outcome for a patientô (Boreham, 2000). They concluded the commonest critical incidents, and also highest level of severity, were failure of patient assessment, treatment and discharge arrangements. Whilst they were able to conclude that these were the riskiest elements of a patientôs care, the study did little to identify the causative factors underlying the incidents. Furthermore, the discussion states that these problems can be addressed by improvements in training which is likely to be an oversimplification of the problem. Subsequently, the research team went on to identify the root causes of critical incidents to determine whether incidents were due to technical, organisational or human failure (Thomas, 2008). These were then further sub-classified and analysis revealed the most common root causes were related to organisational issues outside the ED; internal management issues; human errors relating to knowledge or task verification and execution; and issues relating to patient behaviours. By considering factors external to the clinician, equivalent to Reasonôs latent conditions (Reason, 2000), this study highlights that error in the ED is more complex than simple mistakes made at an individual level. However, compartmentalising error in such a way fails to take into account the multifaceted nature of most ED error. As well as reflecting lack of knowledge, errors related to task execution may indicate inadequate supervision or work overload due to staffing shortages.

Whilst the studies undertaken by Thomas and colleagues provides some insight into error in the ED, analysis of critical incidents is likely to give an incomplete picture of medical error due to deficient reporting (Stanhope, 1999, Sari, 2007). In an attempt to overcome this limitation, a US team undertook a prospective study collecting data on error in the ED by interviewing staff during and at the end of each shift (Fordyce, 2003). Reported errors were categorised as diagnostic studies (22%), administrative procedures (16%), pharmacotherapy

(16%), documentation (13%), communication (12%), environmental (11%) and other (9%). The researchers determined that 98% of errors did not have a significant adverse outcome. Staff were encouraged to complete error report forms that not only described the error but also the circumstances surrounding the error. However, closer inspection of the results suggests that little additional information was given to understand the sources of error. For example, 15 errors involved inappropriate ordering of medication but there is no explanation why this occurred. Indeed the authors allude to this weakness in the study by describing the narratives as ôbriefô and explaining that no further information could be gained because of the use of anonymous reporting.

The previous studies give a good account of the types of error that occur in the ED but fall short of explaining why. A UK study investigated both the active failures and latent conditions associated with critical incidents (Boreham, 2000). Active failures included those found in studies previously described, such as failure or delay in initiating treatment and failure to obtain diagnostic information. The latent conditions that underlay these failures were (i) workloadï resource imbalance, (ii) cognitive errors and (iii) hierarchy. In particular, the authors highlighted the importance of ôsapiential authorityô, where knowledge gained from experience or privileged access to information is recognised in addition to official status. For example, the first person to see the patient may gain crucial information that is not available to a senior colleague. This demonstrates the importance of junior staff being assertive and willing to offer information. Furthermore, senior staff need to be receptive to input from others and involve the team in the decision-making process. This more detailed description of the conditions that allow errors to occur reveals the importance of team interactions.

Cosby led a retrospective study involving analysis of cases referred to an ED physician review committee to identify factors that contributed to ôpatient management problemsô (Cosby, 2008). The authors used this term rather than error as they believed error implies carelessness and focuses on blame, whereas, ôproblemô suggests a more constructive analysis and implies there is something to be solved. Cases were detected during routine morbidity and mortality surveillance and referrals were made by ED staff, admitting services, consultants or quality assurance managers. Over a 15 year period 636 cases were investigated and, as in previous studies, the diagnostic process was the leading phase of work in which problems were identified. The main contributing factors were (i) clinical tasks (e.g.

flawed medical reasoning, lack of knowledge and affective influences), (ii) patient factors (e.g. uncooperative or inability to communicate) and (iii) teamwork factors (e.g. miscommunication, poor team transition, authority gradients and failure to supervise). Again, this study demonstrates the value of teamwork and makes some attempt to identify specific aspects associated with error in the ED.

A comparison of studies described above reveals how elements of teamwork such as hierarchy and failed supervision may contribute to error. The MedTeams Project was set up in the US to investigate more fully the team factors related to error in the ED. Systematic observation in the ED and a series of expert panel meetings were used to produce a model of teamwork behaviours (Risser et al., 1999). The researchers identified 15 main behaviours and 48 teamwork actions that were associated with error. The four actions most commonly associated with error were failures of (i) prioritisation, (ii) asserting oneself, (iii) cross-monitoring actions of team member, and iv) developing a plan. These were implicated as a primary contributor to error in more than 20% of cases. The findings of this study were applied directly to the development of a training course aimed at improving teamwork in the ED (Morey, 2002). This focussed on the actions necessary for both the prevention of error and mitigating the effects of error.

These studies have utilised a variety of approaches to give an overview of the sources of error in the ED and provide an insight into both the types of error and the contributing factors. Teamwork failure has been shown to provide the conditions that allow preventable error to manifest in the ED. This body of research also shows that certain aspects of the patient process are error-prone, in particular diagnosis, which reflects the importance of decision-making in the ED. Other common factors include recognising the impact of authority gradients and the value of good communication as well as accepting the fact that some contributory factors lie outside the ED team, for example external management and patient factors. Therefore attempts to reduce error in the ED by focussing entirely on clinical training and technical skills are likely to have limited success. More attention needs to paid to ónon-technical skillsô, such as decision-making, communication and assertiveness. These skills have traditionally been learnt through trial and error and more targeted interventions may involve encouraging behaviour feedback and non-technical skills training. The next chapter describes how non-technical skills may contribute to safety in healthcare, and more specifically EM, through assessment and training.

2 Introduction: Teamwork and Non-technical skills in the ED

2.1 Defining teamwork, non-technical skills and leadership

Studies described in the previous chapter reveal that teamwork failure is an important contributing factor to error in the ED. Non-technical skills are closely related to teamwork however there appears to be some confusion over the use of these terms and they are often undefined and used ambiguously. Indeed, in a recent review of non-technical skills in Emergency Care, Cooper wrote ñNon-technical skills are predominantly defined as teamwork skillsé this begs the question as to why the term non-technical skills is used in the first place?ò (Cooper, 2010, p14). Whilst teamwork and non-technical skills are linked, the terms should not be used interchangeably. Team skills are one of the team-based competencies (in addition to the knowledge and attitudes) necessary for effective team functioning (Dickenson, 1997). Flin defines non-technical skills as ñthe cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient task performanceò(Flin, 2003, p1) and lists teamwork and leadership as examples. This suggests that ónon-technical skillsô is an over-arching term and teamwork is one component.

The distinction between team skills and non-technical skills is illustrated in the ANTS (Anaesthetists Non-technical Skills) and NOTSS (Non-Technical Skills for Surgeons) assessment tools (Fletcher et al., 2003, Yule et al., 2006). Both contain a teamwork category, which includes elements such as exchanging information, support of others and coordinating team activities. Whilst other skills such as decision-making should ideally utilise team input, these skills do not necessarily contribute to effective team functioning.

Whilst these definitions may appear straightforward, in practice, the terms are often used interchangeably. The MedTeams project was a large study that examined the teamwork skills related to patient safety in the ED (Risser et al., 1999). Many of the óteamwork actionsô listed would appear to be non-technical skills rather than teamwork skills. For example, óapply problem-solving strategiesô is certainly important to ensure optimum team output but does not necessarily contribute to the effectiveness of teamwork.

The situation is further complicated by the fact that skills can be assessed either at an individual level or a team level. It is often considered that assessment of an individual refers to non-technical skills and assessment of the overall team measures teamwork skills. Whilst this may often be the case, it is not a strict rule. It is theoretically possible to rate an individualôs teamwork skills (and the ANTS category of ôTeamworkô obviously allows for this), however, some skills like cooperation are clearly interdependent making individual assessment more difficult. Furthermore, the purpose of assessing teamwork is often different from that of non-technical skills. OTAS (Healey et al., 2004) is the team equivalent of NOTSS. It was developed to assess the overall functioning of the team because the researchers felt that performance of the team as a whole was most critical.

A similar problem exists with the use of the term leadership, which is also a subheading under the umbrella term of non-technical skills (Flin, 2003). Salas identified important aspects of leadership such as directing and coordinating the activities of others, assessing team performance, assigning tasks, motivating team members and organising and executing management plans (Salas, 2004). Therefore good leadership incorporates a number of separate skills such as situational awareness, workload management, teaching/ feedback, team building and problem-solving. In fact it covers almost all of the common non-technical skills. Clearly these skills may also be demonstrated by other team members and this reflects the more recent view that leadership is not confined to those in senior positions. It appears that leadership does not require *different* skills, it is just that there is a different emphasis on the responsibility of those actions. For example, a junior team member may assign tasks from time to time but it is the leaderôs responsibility to manage the overall workload of the team members. When considering the ED, each team member should maintain situational awareness, for example to be aware if there is a particularly long waiting time, but it is ultimately the team leaderôs responsibility to ensure team members are updated. The purpose of a team leader is to ensure someone has an overview of the situation that is more difficult for individual team members to maintain.

This thesis focuses on non-technical skills in the ED and therefore includes skills required for effective and safe team functioning as well as individual tasks. The importance of good leadership is evident in the published literature (see Chapter 4.3) and this is reinforced by the findings of subsequent studies included in the thesis. The definitions provided show how leadership is closely related to non-technical skills and not a separate set of skills. The

assessment tool developed during the course of this MD includes the most relevant ED non-technical skills and incorporates important elements of leadership within the skill definitions.

2.2 Origins of teamwork research in industry and healthcare

Air travel is often quoted as the safest mode of transport but that has not always been the case. Similarly, having a general anaesthetic was once considered risky but now children are routinely anaesthetised for relatively minor procedures because it is considered to be ósafeô. This improved safety record is largely attributed to a shift in focus from individual blame to understanding the human factors contributing to accidents. In the 1970s, a series of aircraft accidents were investigated and 70% were found to be due to human error rather than technical faults or weather. Subsequently, the aviation industry held a conference at NASA which brought together psychologists and airline pilots to consider how to identify and manage the human factors contributing to accidents. Analysis of óblack boxô recordings of communications between pilots and air traffic control revealed failures in leadership, poor team coordination, communication breakdown, lack of assertiveness, inattention and inadequate decision making (Flin, 2008). This led to the development of training courses in non-technical skills known as cockpit resource management. This term was rapidly replaced with crew resource management (CRM) with the acknowledgement that ground and cabin staff also make a vital contribution to safety. CRM is defined as óusing all the available resources ï information, equipment and people ï to achieve safe and efficient flight operationsô (Lauber, 1984). Assessment and training of non-technical skills is now a mandatory part of pilotsô ongoing licensing and trainee pilots are educated in basic human factors knowledge early in their training.

CRM utilises simulators to allow air crews to practise dealing with team and technical problems in a non-threatening training environment. There are clear parallels with the use of simulators in medical training and this comparison was noted by a team of anaesthetists in 1989. This led to the development of a training programme known as Anaesthesia Crisis Resource Management (Howard et al., 1992) and CRM now refers to crisis, rather than crew, resource management in most medical training. This has subsequently spread to other areas of medicine such as surgery and emergency medicine and will be discussed later.

Many emergency physicians in the UK may think that the interest in human factors is a recent development for EM, particularly with a surge in simulation training courses available. However, the MedTeams project and subsequent TeamSTEPPS programme started in the US almost 15 years ago (Risser et al., 1999, Morey, 2002). In the UK, the introduction of non-technical skills training can be partly attributed to overspill from anaesthetics. ED trainees often rotated into anaesthetic jobs where CRM was more widely available. These anaesthetist-centred training programmes were expanded to include the multidisciplinary team and ósecond generationô courses incorporated clinicians from anaesthetics, emergency medicine and surgery (Flin and Maran, 2004).

The focus on clinical leadership has had two major driving forces in recent years. Lord Darziôs report óHigh Quality Care for Allô was published in 2008 after a year-long review and highlighted the importance of clinical leadership for improving quality within the NHS (Darzi, 2008). Although the report focussed on many of the managerial aspects of medical leadership, it reinforced the belief that strong leadership is essential for providing good clinical care. Around the same time, the Academy of Royal Medical Colleges published a well-researched curriculum for clinical leadership (Clough, 2009) that covered both managerial and óshop-floorô leadership skills and competencies. Whilst, influences from government and regulatory bodies have helped raise the profile of clinical leadership, the importance of leadership was already well established in industry and academic medicine.

2.3 Leadership

2.3.1 The importance of leadership

There is a great deal of literature that illustrates the importance of effective leadership in teams (Kendall, 2004, Kozlowski, 1999, Burke, 2006a, Marks, 2000). Salas described leadership as one of the ñBig Fiveò in teamwork as it is a core component of all good teams and has been shown to have the greatest impact on team performance (Salas et al., 2005). Burke et al conducted a meta-analysis of the leadership literature and concluded that when task interdependence is higher, team leadership is more important in achieving efficacious team performance outcomes (Burke, 2006a). This is relevant in hospital medicine, and

especially the ED, where patients are looked after by a number of care providers and each clinician is responsible for the immediate care of several patients at any one time.

Leadership has also been shown to play an important role in safety (Hofmann, 2004). Some industries, such as oil and gas, are now promoting ósafety leadershipô which incorporates (i) monitoring and reinforcing workersô safe behaviours; (ii) participating in workplace safety activities; (iii) being supportive of safety initiatives; (iv) emphasising safety over productivity (Flin, 2004). This concept is something that should be encouraged in medical leaders to improve patient safety.

In addition to team effectiveness and safety, leadership style has also been associated with staff well being. Studies in the intensive care unit and operating room have shown that leadership behaviour that encourages team members to communicate openly and participate in decision making decreases the likelihood of negative employee outcome, such as burnout, low job satisfaction and poor organisational commitment (Reader, 2007, Edmondson, 2003).

As previously mentioned, leaders should be proficient in an array of non-technical skills such as situational awareness, workload management, team motivation and problem-solving. Leadership can be considered a special means of coordination and it is specifically the feedback and monitoring roles that have been shown to correlate with overall team performance outcomes (Burke, 2006b). In the medical setting, a good leader must not only be a good clinician but should also have a strong sense of team dynamics and understand the impact of personal interactions and communications (Shapiro et al., 2008).

The Academy of Royal Medical Collegesô leadership curriculum has now been integrated into all speciality curricula (this leadership curriculum is reviewed in more detail in Chapter 5). This highlights the increasing importance of leadership skills for doctors in the future.

2.3.2 Leadership in the ED

The importance of leadership has also been clearly demonstrated in the ED (Andersen, 2010, Andersen et al., 2010, Fernandez, 2008, Kaissi, 2003, Lubbert, 2009, Simmons, 2010, Brodie, 2009). Manser reviewed the current literature on teamwork in ódynamic domainsô of

healthcare, where the members of a team are constantly changing, such as operating rooms, intensive care and emergency medicine (Manser, 2009). Effective leadership was identified as one of the main factors associated with high clinical performance.

Most literature related to the ED focuses on leadership in resuscitation teams. Research has shown that resuscitation leaders with a more ôhands-onô approach were associated with teams that displayed less adaptability and coordination, and were subsequently less effective (Cooper, 1999). Studies have also shown that successful resuscitation teams have leaders that exhibit explicit task distribution (Marsch, 2005) and ôdirective leadership behaviourô (Tschan, 2006). However, this direct style of leadership may be dynamic and it is accepted that leadership styles often change depending on the situation. For example, a more directive style may be appropriate when a quick response or decision is required. Within the ED, this may mean that during busy periods, a senior doctor may explicitly state a revised diagnosis after reviewing a juniorôs patient rather than guiding the doctor through the decision making process. A different approach may be taken if more time is available, to improve the educational experience of the junior doctor. Furthermore, emergency teams have been shown to demonstrate ôadaptive leadershipô where the specifics of the leader responsibilities change depending on the skill level and experience of individual team members and the team as a whole (Klein, 2006). This may mean delegating leadership responsibilities to other team members or changing team leader when new team members arrive on the scene.

Leadership has been shown to be important for team effectiveness, safety and staff wellbeing. The literature highlights the vital contribution leadership makes in resuscitation teams and, in particular, the changing styles and roles of ED clinical leadership.

2.4 Improving patient safety in the ED

2.4.1 Team training and Crew Resource Management

It is important to understand exactly what it is about teamwork that promotes safety and, therefore, how training should be approached. Shapiro comments that ñteamwork training is

not a specific fix for any one type of error. Ratheré behaviours capable of increasing system resilienceò (Shapiro et al., 2008). This reflects the idea described by Risser that teamwork actions are a ósafety netô to protect patients from inevitable system and human failings (Risser et al., 1999). Safety research in general has started to focus on how caregivers can avert an adverse outcome by interrupting the propagation of multiple small failures (Wears, 2000).

Central to these ideas is the acceptance that, no matter how well staff are trained and how well systems are designed, errors will inevitably occur. Teamwork actions that may mitigate the effect of these errors include, for example, closed-loop communication. This involves the receiver acknowledging and repeating messages to ensure understanding. This would pick up if an error had been made either by the person giving the message or by the person receiving it. Cross-monitoring team member actions involves team members watching each other for errors and problems, and relies on a culture where this is acceptable and encouraged. The teamôs resilience to error can be strengthened by practising back-up behaviour to support team members who are overloaded and assertively speaking up when a team member feels a patient is at risk. Furthermore, a team approach to decision-making may help to counteract some of the many biases that can influence ED clinicians.

The two training courses mentioned in the previous section, Anaesthesia Crisis Resource Management and TeamSTEPPS, are both based on concepts originating from aviation Crew Resource Management. However, the courses differ in that the anaesthesia programme focuses on training the individual provider whereas TeamSTEPPS trains the teams as a whole. CRM can be seen as a set of three barriers to the occurrence of errors by: (i) preventing errors occurring in the first instance, (ii) if errors do occur, then it provides staff with techniques for identifying errors before they have an operational effect (i.e. an adverse event), and (iii) can mitigate the consequences of errors (Helmreich, 1999). CRM includes training related to resource management, situational awareness, decision making, team working and leadership as well as learning about human error and human performance limitation.

A meta-analysis of aviation literature concluded that CRM has a statistically significant impact on both attitudes and behaviour (O'Connor et al., 2008). Whilst some healthcare based studies have been able to show improved team performance with CRM training, the evidence demonstrating improved patient safety is sparse (Gaba, 1992, McCulloch, 2009,

Taylor et al., 2007, Wright et al., 2009, Schull, 2001, Eppich, 2008). This may be due to the methodological difficulties of measuring improved patient safety or a reduction in error. Other researchers have commented on the lack of a ógold standardô method for measuring performance and evaluating the effectiveness of CRM (Kim et al., 2006, Baker, 2005). This may be why some researchers have simply analysed end of course attitude questionnaires to evaluate the success of training (Reznek et al., 2003, Grogan, 2004). One study looking at clinical training in a simulator versus clinical training combined with CRM, showed no significant difference in non-technical skill performance (Zausig, 2009). This may be because a single day of training is insufficient to produce a statistically significant difference in performance and, indeed, many CRM courses are purposively spread over 2 or 3 days. However, the study revealed that the quality of medical management correlated with the quality of non-technical skills, suggesting that they are inextricably linked. This is consistent with another studyôs findings where the groups with the best medical management of advanced life support also demonstrated more non-technical skills competence (Marsch, 2004). Flin commented that the most significant impact of CRM training on safety may come from the shift in cultural norms (Flin, 2009). For example, widespread training may make it more acceptable and common-place for junior members of the team to speak up about potential risks. Therefore, the effects of this may only become evident over a period of time.

One of the most extensive studies investigating the impact of team training on patient safety was carried out in the ED as part of the MedTeams project (Morey, 2002). The experimental group consisted of 684 physicians, nurses and technicians who undertook a team training programme. Assessments of the team and ED safety outcome measures were made prior to training and at four and eight months after training. They reported significant improvements in the quality of team behaviours and reductions in clinical error rates (30.9% to 4.4%). They also revealed improvements in attitudes to teamwork as well as staff assessments of institutional support. It should be noted that as well as team training, the intervention included introducing organisational change to further encourage teamwork behaviours. It is therefore hard to determine which aspect of the intervention was responsible for the improvements seen. Indeed, some of those involved in the project felt that the most important factor was simply bringing together people from multi-disciplinary backgrounds to train together.

2.4.2 Simulation

One of the most common methods of training non-technical skills is simulation, which has generally utilised CRM principles. Most courses include theoretical background as well as hands-on simulator scenarios and many proponents argue that it should include a multi-disciplinary team (Murray, 2000).

Whilst simulation appears to be promising, a study by Shapiro was unable to demonstrate any significant difference in teamwork behaviours when adding simulator training to a didactic teamwork curriculum (Shapiro, 2004). The study involved four teams (each consisting of five multidisciplinary staff members) which were randomly allocated into either a control group or experimental group. All participants had already received classroom teaching as part of the TeamSTEPPS training programme and the experimental group received an additional 1 day simulation-based training. Each team was subsequently followed for 2 months and assessed using the same observational framework used for the MedTeams project. The intervention group showed a trend towards improvement ($p=0.07$) and the failure to detect a significant difference may be attributed to the small sample size.

Many are now recognising that although simulator training for CRM has significant potential benefits, it neglects some of the other important aspects of an ED physiciansô role, such as coordinating and supervising the wider ED team. Crisis management should also include managing multiple patients and dealing with issues external to the ED (Small et al., 1999). One suggested solution is the use of multiple-patient simulation scenarios which allows for expanded communication scenarios, introduces the concept of triage and enables increased scenario complexity (Kobayashi, 2007). Simulation has also been used to improve non-technical skills with leadership training, cognitive forcing strategies and psychological training on human factors (Shapiro et al., 2008, Bond, 2004, Müller et al., 2007).

2.4.3 Other interventions to improve non-technical skills

Whilst simulation CRM and team training have formed the basis of most training of non-technical skills, there are an array of other methods in use within medicine and industry (Flin, 2008). This may be information based (such as lectures or e-learning), demonstration based

(using actors or real-life video clips) or practice based. In addition to CRM, practice based training may involve small group work, role plays, desktop exercises or full scale exercises such as a site emergency (Crichton, 2001). Croskerry highlights the use of metacognition, which is essentially ñthinking about thinkingò, to develop cognitive forcing strategies to reduce the effects of biases when making clinical decisions (Croskerry, 2003). More imaginative solutions have involved the use of an online ñvirtual EDò for training leadership and CRM principles (Youngblood, 2008). A trial has shown this to be equally as effective as simulation based training and has the advantage of being less expensive.

The evidence presented highlights some of the opportunities and challenges related to improving non-technical skills. However, following the lead of aviation, a more efficient and reliable way of training these transferable skills may be to target medical undergraduates and other trainee healthcare professionals.

2.5 Assessing non-technical skills

2.5.1 Methods of assessing non-technical skills

Many training courses have been developed with the aim of improving non-technical skills in healthcare. However, a commonly held view is that in order to train effectively, you must first be able to measure what you are training. Measuring the progressive development of an individual or group is necessary to evaluate the impact of any training programme (O'Connor et al., 2008). Furthermore, assessment of non-technical skills is necessary to provide specific feedback to trainees and prioritise training needs for an individual or group (Flin, 2008).

One method of assessment is the use of ôbehavioural marker systemsô to structure observations in the workplace or simulated setting. These systems are composed of a number of non-technical skills, grouped into categories, and each skill is linked to explicit behavioural markers which illustrate good and poor practice. This is combined with a rating scale to enable assessors to rate skills based on observable behaviours and provide structured and specific feedback.

In addition to observational rating scales, non-technical skills have been assessed by questionnaires. For example the Situation Awareness Global Assessment Technique (SGAT) involves ófreezingô a simulation and questioning participants about their perception of the situation (Wright, 2004). Other questionnaires used in emergency care include the Communication Competency Questionnaire (Cooper, 2007) and the Adaptive Leadership Behaviour Description Questionnaire (Cooper, 1999) which measures a resuscitation team leadersôcommand and control.

2.5.2 The development of behavioural marker systems in healthcare

A number of rating systems to assess non-technical skills have been developed across an array of medical specialities. Initially, behavioural marker systems for medical teams were simply adapted from existing aviation tools. For example, Gaba et al adapted the aviation LOSA (Line Operations Safety Audits) checklist for use in simulated anaesthetic crises (Gaba, 1998). This was done without any specific research into the similarities or differences between aviation and medicine. LOSA was also used as the basis to develop a behavioural marker system for neonatal resuscitation (Thomas, 2004). Whilst the research team recognised that there were likely to be some generalisable aspects of team behaviour, they understood that it would be inadequate to directly apply the LOSA markers. They modified the aviation tool by observing video recordings of resuscitations as well as using data from focus groups and surveys. Similarly, the Oxford NOTECHS system for surgeons was adapted from the NOTECH aviation tool by a process of task analysis and consultation with a panel of surgeons and human-factors experts (Mishra et al., 2009).

In 2002, a team from Aberdeen University developed a behavioural marker system for anaesthetists called ANTS (Anaesthetistsô Non-Technical Skills) (Flin et al., 2003). Unlike previous systems that were adapted from aviation tools, the team went back to first principles to identify the salient non-technical skills. The process involved:

1. a literature review to extract skills and group common themes to guide subsequent data collection
2. cognitive task analysis interviews with 29 anaesthetists
3. workshop discussions with the research team to produce a taxonomy

4. refining the taxonomy using data from incident reports, attitude surveys and observations.

Shortly afterwards, a framework for surgeons, NOTSS (Non-Technical Skills for Surgeons) was developed (Yule et al., 2006). The stages included:

1. cognitive task analysis interviews with 27 surgeons
2. incorporating the results from a literature review, observations and attitude survey
3. deriving a skills taxonomy using an expert panel on behavioural markers
4. using panels of surgeons to modify the structure of the framework and 16 consultant surgeons to add the exemplar behaviours
5. refining these behaviours at 2 multidisciplinary review meetings

A summary of the methods used by some of the teams to develop behavioural marker systems for assessing non-technical skills is shown in Table 2.1 below. This was used to determine the steps needed to identify the core EM non-technical skills and guide subsequent studies in the thesis.

Table 2.1 Methods used to derive behavioural marker systems in healthcare

Method	NOTSS	ANTS	Oxford NOTECH	Neonatal
Cognitive task analyses (interviews)	•	•		
Attitude survey/s	•	•		•
Literature review	•	•		
Mortality reports	•			
Incident reports		•		
Observations	•	•		25 resus
Expert input/ panel	•		•	
Focus groups				•
Task analysis			•	

It can be seen that most attempts to produce a behavioural marker system have used more than one method of collecting data. This triangulation of results helps to ensure that the final

list of skills is as comprehensive as possible. Flin categorised the techniques used to identify non-technical skills into: analysis of events, questioning techniques and observational techniques (Flin, 2008). A summary of this is shown in Table 2.2 below.

Table 2.2 Techniques available to identify non-technical skills

Type	Techniques
Event-based Examining accident reports to identify patters of behaviour	• accident/near miss analysis • confidential reporting systems analysis
Questioning techniques Soliciting information directly from role-holders of the job under investigation	• interviews • focus groups • questionnaires and surveys
Observational techniques Watching individuals or teams carrying out tasks	• direct • participant • remote (e.g. video recordings)

Flin also mentions training syllabi from aviation regulatory bodies as a source of information regarding non-technical skills. As previously mentioned, the Academy of Royal Medical Colleges developed a leadership curriculum and although this is not a list of observable non-technical skills, the framework gives a comprehensive overview of what skills are expected of a trainee (Clough, 2009). The College of Emergency Medicine curriculum also covers a variety of ógeneric skillsô which includes many speciality-specific non-technical skills and related behaviours (College of Emergency Medicine, 2009).

Whilst most tools are speciality specific, the CATS (Communication and Teamwork Skills) assessment tool covers generic teamwork skills for application in any medical speciality (Frankel, 2007). This is based on the argument that many non-technical and teamwork skills are common to all areas of medicine. Assessment tools have not only been used for physicians and surgeons but have also been developed for scrub nurses (Mitchell, 2009) and medical students (Wright et al., 2009)

2.5.3 Evaluation of Behavioural Marker Systems

It is essential that any behavioural marker system is thoroughly evaluated to ensure it is valid and robust. The ANTS system was evaluated using 50 consultant anaesthetists who had attended a 1 day training session (Fletcher et al., 2003). Participants watched eight videos of scripted anaesthetic situations and rated the non-technical skills of the main anaesthetist in each scenario using the ANTS system. Participants were also given an evaluation questionnaire and the combination of rating forms and questionnaires provided data to assess validity, completeness of the scale, observability, inter-rater reliability, sensitivity, internal consistency, usability and acceptability. The team concluded that the ANTS scale provided adequate validity, reliability and usability based on a one day training course. However, inter-rater reliability (or agreement) ranged from 0.55-0.65, well below the recommended level of 0.7. Also, the evaluation involved scripted scenarios under experimental conditions so extrapolating the results to real clinical scenarios may be difficult. An Australian team attempted to evaluate ANTS using unscripted scenarios of routine anaesthesia and found poor inter-rater reliability and inadequate correlation of scores between participants and experts (Graham et al., 2010). This study highlighted the need to train assessors in basic non-technical skills concepts before attempting specialised training in use of a behavioural marker system.

Similarly, the NOTSS system was evaluated using video tapes of scenarios in a simulated operating room (Yule et al., 2008). Following a short training course of 2.5 hours, 44 surgeons watched 6 scenarios and rated the lead surgeon using the NOTSS scale. Participantsô ratings of the behaviours demonstrated in each scenario were compared to expert ratings (ółaccuracyôô), and assessed for inter-rater reliability and internal consistency. Accuracy for acceptable/ unacceptable behaviour only achieved a minimum level of 60% for all categories and inter-rater reliability was unacceptable for 2 of the 4 categories. This was attributed to insufficient training but the team concluded that the scale was adequately reliable for use. There was no mention of validity tests.

OTAS is an observational tool developed for assessing teams in the operating room (Healey et al., 2004). Unlike ANTS and NOTSS, this system assesses the overall team and rates team behaviours (such as coordination and communication) as well noting completion of tasks. The tool was evaluated using data collected during only 12 elective procedures (Sevdalis,

2009). The research team compared ratings between a single novice observer, who had received minimal training, and two experts, as well as comparing the pair of experts. Although they found strong correlations between the expert/expert ratings, there were significant differences between expert/novice ratings, suggesting that adequate training is vital for the use of OTAS. However, a clear limitation is the small number of raters used, making it difficult to draw firm conclusions from the study.

In 2001, an international group of human factors experts met to discuss the development and use of behavioural marker systems in industry and healthcare (Klampfer, 2001). They published a series of recommendations, including a list of characteristics of a good behavioural marker systems such validity, reliability and usability. However, there was no clear description of how these measures should be evaluated or what criteria should be used. The evaluation of marker systems described above highlights some of the difficulties of gaining adequate levels of reliability from a single observation. In contrast, most workplace based assessments in use in the UK have been evaluated using ôGeneralisability Theoryõ (Wilkinson, 2008). Accepting that reliability will not be achieved using a single encounter, this method utilises a large number of assessors and each trainee is assessed on multiple occasions. This may represent a more feasible approach to the widespread assessment of non-technical skills in medical education. However, as yet, there is no tool that comprehensively assesses non-technical skills in the ED setting.

2.5.4. Assessing non-technical skills in EM

Assessment of non-technical skills in EM has been attempted using a variety of different methods, tailored to specific settings within the ED. The MedTeam project identified essential EM teamwork behaviours by tailoring aviation evaluation criteria to the emergency care setting (Risser et al., 1999). This involved a total of 200 hoursô observation by behavioural scientists followed by a series of expert panel meetings, incorporating the expertise of emergency care physicians, emergency nurses and behavioural scientists. The objective was to produce a research tool (rather than an assessment tool) to help identify areas where errors occur and use this to formulate a training programme. This meant that the final observational framework was longer and more complicated than would be expected for an assessment tool and comprised 18 skills.

Assessment of non-technical skills within EM have invariably centred on the critically ill or injured patient, usually in a simulator environment (Malec, 2007, Holcomb, 2002, Cooper, 2007, Ottestad, 2007, Kim, 2006, Cooper, 2009). For example, Ottestad and colleagues developed a technical and non-technical rating scale for individuals and teams managing septic shock (Ottestad, 2007). Another system focused on assessing the management of critically ill patients using the Ottawa Crisis Resource Management Global Rating Scale (Kim et al., 2006). This assesses performance in the five categories of CRM using a 7-point Likert scale, as well as giving a score for overall performance. The assessment tool was developed by three acute care physicians with experience as resuscitation instructors, using consultation with other physicians and instructors at a national critical care course. No details were given regarding exactly how the tool was developed or the number of participants at the course. The scale was subsequently modification using the Delphi process using only four participants, including the three initial physicians. Inter-rater reliability at a category level, measured using intraclass coefficients, reached only 0.475-0.626, well below the commonly accepted level of 0.7.

The Team Emergency Assessment Measure (TEAM) was developed for use in the resuscitation room. The development process involved simply a review of the literature followed by creation of a draft instrument using an expert clinical team (Cooper et al., 2010). Evaluation of inter-rater reliability and retest reliability was determined to be ófairô at a level of only 0.55 and 0.53 respectively.

As yet, few researchers have stepped out of the resuscitation room environment to consider the skills necessary for teamwork in the wider ED. Here is it important to consider the coordination of a larger team and the problems posed by competing demands from multiple patients. This apparent reluctance to incorporate more characteristic EM scenarios is likely to be due to the perceived difficulties of this type of assessment, such as the wider geographical area that would need to be observed. Furthermore, the necessary empirical evidence related to non-technical skills in the wider ED is scarce and this may reflect the challenges of undertaking studies in such a complex setting (as previously discussed in Chapter 1.5). Assessment of non-technical skills, both in aviation and in healthcare, has largely been restricted to the simulator environment. Simulating a single resuscitation room or operating theatre is clearly more straightforward than attempting to replicate the full ED setting. Whilst

the simulator is a much more controlled setting than the workplace, allowing greater flexibility, it is also artificial and performance does not necessarily reflect true practice.

The subsequent studies included in this thesis describe the development and validation of a tool to assess non-technical skills in the workplace, incorporating the skills necessary during routine work as well as the management of critically ill or injured patients. The studies include a review of the literature, interviews with staff, observations within the ED and a survey of expert opinion. The designs were based on the best methods for developing an observational tool as described above. The decision to focus on the assessment of non-technical skills stems from the findings of an initial interview study conducted with ED staff. A description of this study is given in the following chapter.

3 Staff Interview Study

Chapter overview

The initial, broad aim of this thesis was to consider how the ED team could function better to provide safe and high quality clinical care. In order to set the direction of future studies, an exploratory interview study was undertaken with staff in the ED. The study identified behaviours that staff exhibit under pressure and considered how these could have a positive or negative effect on the team. The findings related to a number of fundamental skills such as leadership, communication and workload management that are vital for team function but not traditionally taught or assessed. Following on from this study, subsequent thesis chapters focus on developing a tool to assess these and other key non-technical skills in the ED.

3.1 Background

Healthcare providers are increasingly aware that it is important to understand the human factors involved in providing good clinical care. The ED, in particular, is a challenging environment in which to provide consistently high quality care for all patients. Understanding the behaviours staff exhibit under pressure is essential for considering improvement strategies yet there is limited research based in the ED.

Non-healthcare research undertaken in experimental conditions has shown that, under pressure, individuals lose team perspective and focus shifts to the individualôs own task (Driskell, 1999). As a result, some of the safety functions of the team may be lost, for example, checking each other for errors and monitoring the workload of team members. Excessive demand has also been associated with reduced explicit communication and poorer team coordination (Moon et al., 2004, Kleinman, 1989). Janis introduced the idea of ógroupthinkô which states that, under stress, a group may make faulty decisions due to conformity and a lower likelihood of seeking information or advice from outside the group (Janis and Mann, 1977). It has been noted that junior flight-crew members are less likely to question the captain under emergency conditions, sometimes deferring to the extent of not offering valuable task information (Foushee and Helmreich, 1988).

Studies on individual surgeons have shown undue levels of stress can lead to impaired judgment, decision making, and communication (Wetzel et al., 2006). One ED study demonstrated that senior medical and nursing staff were susceptible to communication overload at times of increased demand (Spencer et al., 2004). Other ED research has identified communication breakdown during periods of high stress and negative effects on decision making (Xiao, 1996, Coiera, 2002, Schull, 2001). Whilst these studies highlight the negative impact of stress on team function, it is possible that óstressingô a team may lead to enhanced communication and improved team function under certain circumstances. Indeed, some studies have highlighted the positive influences of stress on teamwork (Drach-Zahavy and Freund, 2007) and some people thrive when working in high pressure environments.

Objective measures of overcrowding in the ED do not always accurately reflect staff perceptions (Raj et al., 2006). Furthermore, subjective assessment of excess ED demand does not correlate between staff members, i.e. individual staff members perceive workload and overload differently (Reeder, 2003). This suggests that pressure on the ED team is more complex than simply high patient numbers. Many ED physicians will recognise there are days when demands in the ED are high but these do not necessarily correlate with excessive patient attendances. Therefore, it is also important to understand the factors that contribute to high pressure in the ED.

This was an exploratory study used to identified key stressors for ED staff and investigate positive and negative behaviours associated with working under pressure. Interviews with ED medical and nursing staff were also used to generate ideas for interventions that would improve how the team functions and guide future studies.

3.2 Aims

The aims of this interview study were:

- to identify key stressors for staff working in the ED
- explore the behaviours exhibited by ED staff when working under pressure
- consider interventions that may improve how the team functions

3.3 Methods

3.3.1 Design and participants

A qualitative design was employed using semi-structured interviews comprising a series of open-ended questions (see Appendix A) which allowed staff to speak freely about their experiences in the ED. Data was collected from staff working in the ED of St Maryôs Hospital, London. This is a busy teaching hospital ED with an annual patient attendance of 90 000 per year. Each clinical shift is coordinated by a nurse-in-charge and ED registrar, under the supervision of an ED consultant. A purposive sampling method was employed to recruit both medical and nursing staff of varying seniority. Recruitment continued until no significant new themes emerged from the data (ótheoretical saturationö).

The inclusion criteria for participants were as follows:

- qualified nurse or doctor working in the ED
- able and willing to give informed consent to participate in the study

The exclusion criteria for participants were as follows:

- medical students or student nurses
- staff currently seeking help for work-related stress

3.3.2 Materials

Materials consisted of the following and can be found in Appendix A:

1. Interview questions/ topic guide
2. Participant information sheet
3. Participant consent form
4. Occupational Health advice sheet
5. Coding framework

3.3.3 Procedure

The study was advertised using posters in staff areas of the ED at St Maryôs Hospital seeking volunteers for the study. Staff were also invited to participate by the lead investigator. They were given a verbal explanation of the study and it was emphasised that participation was completely voluntary and confidential. Participants were also given a written participant information sheet that provided information about the interview process as well as directing staff to Occupational Health for any necessary support.

Prior to commencing the interview, participants were asked if they had any questions and were asked to sign a written consent form. Interviews were undertaken by a sole investigator and consisted of open-ended questions on: perceived stressors (ôWhat makes you feel under pressure at work?ô); effects of stress for the individual (ôHave you noticed that you behave differently when youôre under pressure?ô) and perceived effects on other team members (ôHow does the behaviour of others change when they appear to be under pressure?ô); as well as possible interventions (ôDo you have any suggestions how the team could work together more effectively, especially during periods of increased demand?ô). Previous studies with surgeons revealed their reluctance to admit feeling stressed so the use of the word ôstressô was avoided (Arora et al., 2009, Wetzel et al., 2006). Participants were also asked to describe circumstances where high pressure in the ED may have contributed to an error.

3.3.4 Data analysis

Audio-recorded interviews were anonymised and responses were transcribed verbatim. Data were analysed to extract broad themes from the interviews and responses were coded using Nvivo computer programme. Half the interviews were analysed independently by a second member of the research team who had training in human factors research. There were no major disagreements regarding the coding of data. The coded material was subject to member check to reduce investigator bias. This involved showing participants a summary of how their responses had been coded to ensure the true meaning was retained (Lincoln, 1985).

3.3.5 Ethical Issues

Ethics approval was granted by Barnet Enfield and Haringey Research and Ethics Committee and St Maryôs Research and Development Department. Informed written consent was collected and staff were made aware that participation in the study was entirely voluntary and they could withdraw at any time. All data collected from participants were anonymised to maintain confidentiality. Thought was given to what would happen if a clinical incident was disclosed during the course of the interview and it had not been reported or investigated. It was explained on the Participant Information Sheet that only significant incidents would require disclosure to a third party and this would be done confidentially wherever possible. This problem did not arise. As the study looked at factors contributing to pressure on staff, the interviews may have brought to light a particular problem a staff member had coping with stress at work. Therefore, contact was made with the head of Occupational Health who agreed to provide support and staff were offered a leaflet explaining the support services available.

3.4 Results

22 participants took part in the study, consisting of 4 consultants [C1-4], 7 registrars [R1-7], 5 lower grade doctors [L1-5] and 6 nurses [N1-6].

3.4.1 Stressors

When questioned, all staff members recognised that the ED can be a high pressure environment at times and over half of respondents said they worked under pressure most of the time. Respondents were able to describe a number of factors that contributed to this and Figure 3.1 summarises the stressors mentioned by staff during the interviews.

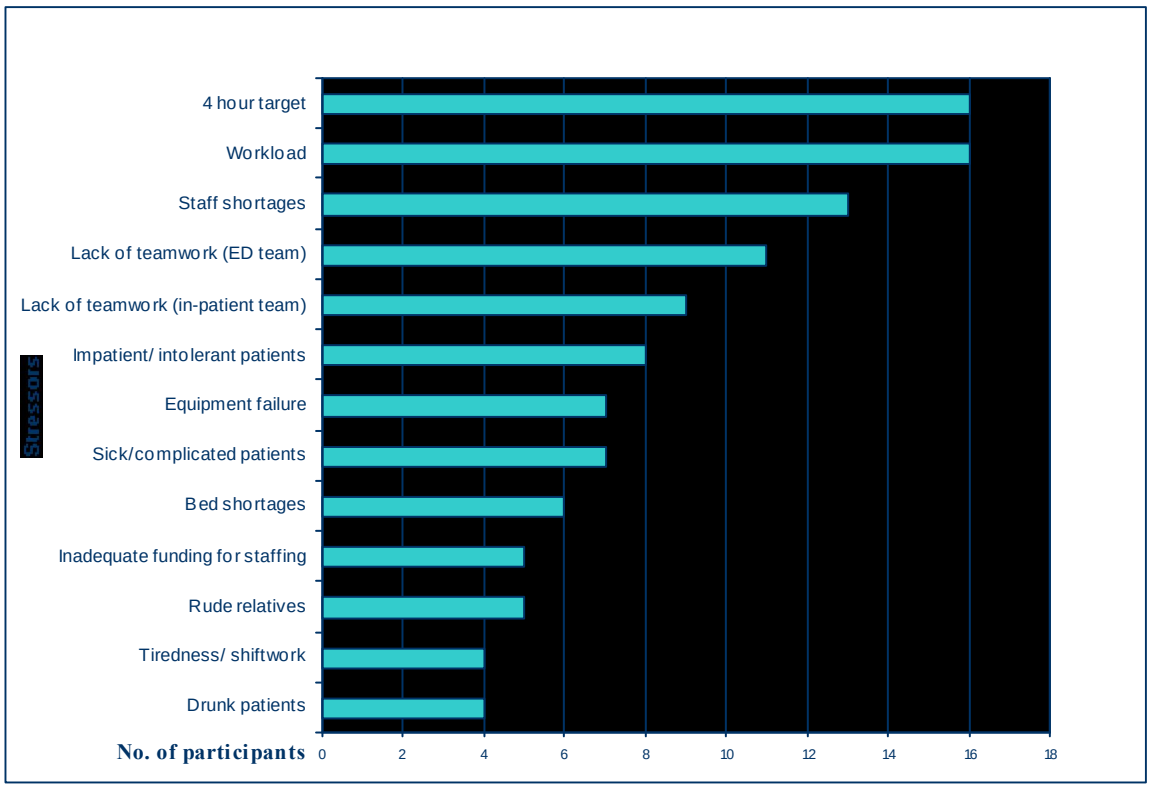


Figure 3.1 Stressors mentioned by the ED team

The most frequently mentioned stressors are linked, i.e. high workload, combined with staff shortages makes the four hour target more difficult to achieve (this was the government target that 98% of ED patients should be admitted or discharged within four hours of arrival). However, other factors may explain why staff feel they work under pressure even when workload and staffing are balanced. Lack of teamwork, both within the ED and with the wider inpatient teams, appeared to be a significant contributor to pressure in the ED. Respondents described ñlack of cooperationò and ñvarious conflicts with other team membersò as contributing to pressure. One staff member said ñI could be fully staffed and have a bad day and I could have less and have a good day. Itôs obviously about how the team works togetherò[N2]. Other stressors centred on the patients and relatives (e.g. rude relatives or drunk patients) and other resource issues (e.g. equipment problems, inadequate funding for staff and bed shortages).

3.4.2 Effects of working under pressure

Staff described a multitude of ways their behaviour changed when working under pressure. This included task related actions such as rationalising investigations (3/22), referring without investigation results (3/22) or abbreviating documentation (6/22). Numerous staff described examples of back-up behaviour (9/22) or delegation of tasks (11/22). However, many responses fell into 4 categories relating to communication, departmental overview, managing staff and managing patients.

1. Communication

In general, junior staff described communicating less, in particular to senior staff, during periods of stress or high demand. They gave reasons such as ñI donôt want to communicate as much because I know that they are stressed and they can get angryò [L4] and ñI try not to speak to senior registrars because I know theyôre really, really busy and I attempt to do things myselfò [L5]. Three junior doctors said that this sometimes meant discharging or referring patients without first speaking to the registrar, which they would have done under normal circumstances. Whilst this may be perceived as a safety concern, one doctor said ñI had a lot of patients where I might have gone to ask but I thought, no, I know that really, thatôs fineò [L3]. This suggests that as long as the doctor knows their capabilities and limits, reduced communication may promote more independent and self-assured working. However, four junior doctors said that they sometimes found senior staff ñaggressiveò and ñunapproachableò and two stated this meant that they were less likely to ask for help.

There were complaints about the rudeness or abruptness of other team members (9/22) as well as staff admitting to being less polite when working under pressure (9/22). Six participants described the importance of speaking calmly and not shouting. For example, one nurse said ñHe stands there with a quiet voice and never raises his voice, which instils confidence in the entire teamò [N1].

Participants also described being more succinct and direct both with other staff members (5/22) and with patients (7/22). This involved more specific instructions, more direct questioning or less detailed explanation. Senior staff acknowledged that some

communication events are shortened, for example, ñIôm probably not giving the SHOs enough time to tell me their storiesò [R5]. However, they increased communication related to coordinating the department. Four participants described more frequent exchanges between the nurse-in-charge and lead clinician to plan and prioritise patient tasks, problem-solve and identify any patients of particular concern. One doctor said, ñI think the communication has to step up a few gears and you have to be on top of everything thatôs going oné and if itôs not that busy then you might not need to talk to the nurse in charge for half an hourò [R1].

2. Departmental overview

During periods of high demand, staff described maintaining an overview of the department by utilising the Patient Tracking System, calling a mini-board round and ñeyeballingò patients in cubicles or the waiting room. Two registrars commented that they were less likely to see their own patients, or lead a resuscitation, to ensure they were free to oversee the department and coordinate the team. Staff also described the importance of requesting regular updates about the state of the whole hospital, including bed availability and staffing issues. These comments reveal how staff in a leadership role maintain situational awareness, ensuring they have a good ómental modelô representing the status of the current task and the surrounding working environment

3. Managing staff

Senior staff described how they monitored team workload to maintain efficient patient flow. For example, they checked how long a doctor had been with a patient or noted when the last ócas cardô had been picked up. Two registrars said they instructed juniors to report back after a specified time and three senior doctors said they were likely to seek out juniors early to discuss a patient plan. Consultants and registrars described explicit allocation of tasks and stipulating a patientôs management plan rather than coaching a junior doctor through it. Senior staff redirected staff according to ability (e.g. surgical background to minors) and monitored juniors to identify those who needed more support. Participants explained how they enlisted the help of speciality doctors, GPs, paediatric ED doctors and consultants to help during periods of high demand. Six participants described exerting authority to get the

team working faster. Whilst senior medical staff changed their leadership style to be more directive under pressure (6/10), conversely, two senior nurses commented on allowing nurses to use their own initiative and encouraging them to use clinical judgement. One said ñtrust yourself, allocate appropriately, let them get on and have confidenceò [N2].

4. Managing patients

Participants described utilising the Clinical Decision Unit (CDU) to clear cubicles and reviewing or re-triaging patients when waiting times were extended. While two registrars said they were less likely to see patients when under pressure, three consultants explained the importance of seeing patients to ñqueue-bustò, especially if a difficult patient was next to be seen. One consultant said ñI tend to have a quick shifty through all the cards in the box, see if there is anyone I can see quicklyò [C2] suggesting they preferentially select certain patients to ensure they donôt get too involved. Another said, ñMy own style is to pick up the card and see the patientí the senior nurse is in charge of how things are organised and my job is to see patientsò [C4].

Whilst it can be seen that staff change their behaviour in a multitude of ways some interviewees said they resisted changing how they worked when under pressure (3/22). However, they all qualified this by saying it was patient care that should not be abbreviated and that some re-prioritisation was necessary. For example, ñI donôt think you should alter your practice because itôs busier, itôs just a different emphasis placed on different areasò.

3.4.3 Suggested Interventions and advice

Whilst four staff mentioned interventions that focussed on increasing resources, such as improved staffing or new equipment, others described suggestions that would improved team function, as follows:

1. Leadership and Team Training

The vital importance of leadership was highlighted in a number of quotes such as ñI think the way the team works under pressure is a factor of how the team leadership works under pressureò [R2] and ñI think it always just goes down to whoôs in chargeò [N6]. The most common suggestion to help the team function better under pressure was improving team skills and in particular leadership skills (7/22) as there was perceived to be a significant gap in training. The lack of preparation for leadership amongst registrars was mentioned by several interviewees. One registrar commented, ñAt the end of my first year I went into a hospital where you were very much expected to lead the department with no real instructioné and I had a terrible yearò [R6]. A senior registrar said, ñThereôs a sense sometimes that this is your shift, crack on with it, youôre in charge and live or dieé itôs quite obvious sometimes when people havenôt got good insight into how to run thingsò [R2]. Similarly, one consultant said, ñI think it is to do with actually sitting down and teaching people how they can manage the departmenté you canôt expect someone just to know how to do itò [C3].

Several participants mentioned specific ways leadership and team training could be addressed such as mentoring, situational teaching and simulation training. Six interviewees mentioned Crew Resource Management or team building exercises. One staff member felt that improved relationships with ward and laboratory staff would ensure everyone worked together for the common good of the patient. Another suggested returning to set teams for night shifts to improve team cohesion. The general consensus appeared to be that if the wider team functioned well then staff could cope with the high demands of working in the ED. However, respondents also recognised that a simple training course was unlikely to provide long-lasting improvements and two people commented that sustained behaviour change requires a change in attitude at an organisational level. Whilst staff were aware of the financial burdens of some interventions, one staff member reflected ñIt doesnôt cost anything to build a teamò [C4].

2. Advertising Breaks

Taking breaks during periods of high pressure were a source of tension. Five respondents commented on other staff members ñdisappearingò off for breaks or mistrust about the length of break taken. Interviewees described how time was wasted looking for staff members who had gone on a break. However, four staff highlighted the importance of having breaks for efficient and healthy working. One respondent suggested using a notice board to inform colleagues of the actual start time of a break, avoiding unnecessary searching or mistrust.

3. Practising back-up behaviour

Twelve (55%) interviewees mentioned that staff not helping out was a source of tension. However, respondents also described examples of back up behaviour (9/22) such as doctors pushing trolleys and nurses taking bloods. Interestingly, senior nurses were more likely to take on extra tasks to support colleagues who were struggling whilst doctorsô reasoning tended to focus on the belief it was simply faster or more efficient.

4. Staying calm

Another common theme that emerged from the data was the importance of staff members remaining calm when under pressure (16/22) and many referred specifically to the team leader (6/22). Staff gave reasons for this such as ñmakes the place feel less busyò, ñreassuring that things are okò and ñgives you confidence they know what they are doingò. Half commented on the perceived negative effects for people who get stressed and some gave examples such as ñaffects decision makingò and ñinefficientò. Whilst participants identified that staff should remain calm, there were few practical suggestions for achieving this except taking some ñtime outò (2/22) or physical de-stressing techniques (1/22). One respondent simply said, ñPeople who panic shouldnôt be in A&Eò [N1].

5. Motivating staff

Three participants discussed the motivation to work harder during periods of high demand. Two senior staff members commented that they felt junior doctors had no incentive to work harder, especially when this related to achieving targets. One junior doctor reflected on the lack of positive feedback and another said that there should be rewards for doing well. These comments reflect the importance of setting team goals and ensuring all staff are motivated to achieve these goals. This can be summarised by a respondent who stated that ñstaff will accept stress and working under pressure if there is a sense of fairness from aboveé youôve got to show the troops that you love themò [C4].

3.4.4 Senior versus junior staff

It became clear during the interviews that most senior staff had acquired a number of strategies to deal with overload and high pressure in the department. However, a ónewô registrar, admitted to having very few suggestions for dealing with increased demand, despite an induction programme that included scenario teaching for management issues. Staff said they learned techniques by modelling senior colleagues or had worked it out by trial and error, for example, ñI learned the hard way [that] not escalating things and not involving senior people then led to problemsò [C2]. However, these learning approaches may be inefficient and unreliable and without any explicit training in what is ócorrectô, staff may pick up bad habits. For example, one senior doctor admitted ñputting a nose out of jointò to get something done, which fits with junior staff commenting on ñaggressiveò seniors who are then viewed as unapproachable.

This study also revealed a high level of misunderstanding between different grades of staff. For example, senior doctors may feel obliged to know what is going on with each patient in the department (i.e. situational awareness) but one junior described this as ñchecking upò. Similarly, another junior doctor complained about registrars who ñ*even* go to review a patient in personò although this may indicate good data gathering for decision making. One junior doctor mentioned they would never disagree with a registrarôs clinical judgement ï clearly highlighting the potential danger of authority gradients. These examples reveal that some

junior doctors lack understanding about responsibilities within the ED team and, indeed, may not fully comprehend the purpose of teamwork and its contribution to safety.

3.5 Discussion

This study gives insight into the stressors experienced by ED staff in a London teaching hospital. The two most common stressors, excessive workload and staff shortages, may be considered two sides of the same problem, that is, an imbalance between workload and resources. By definition, stress occurs when a person perceives that demand exceeds available resources (Lazarus, 1984). This imbalance is exacerbated by government targets, where time pressure becomes an issue in addition to clinical need. Studies with surgeons in the operating room identified distractions and interruptions as a common stressor (Arora et al., 2009). However, interruptions were not mentioned by staff in this study, possibly because it is considered normal in the ED (Chisholm, 2000, Woloshynowych et al., 2007). Interestingly, both groups of clinicians interviewed described team factors as a major stressor.

This study goes some way to explaining why there are days the department may not have excessive patient numbers but staff still feel like they are working under pressure. It would appear that teamwork, both within the ED and with in-patient staff, is integral to smooth running of the department. Furthermore, participants identified that team leadership is a principle deciding factor on how well the team functions, which is consistent with the published literature (Shapiro et al., 2008, Cooper, 1999, Manser, 2009). Studies have shown that leaders of resuscitation teams working under pressure often choose a more directive style with explicit task distribution (Tschan, 2006, Marsch, 2005). Comparable results were found in this study. Whilst resuscitation team leaders were less successful when adopting a 'hands on' approach (Cooper, 1999), interviews with consultants indicated they expected team leaders to see patients as well as manage the department during periods of high demand. Previous research has shown that communication overload may be a problem during periods of high demand in the ED (Spencer, 2004). Interview responses suggest ED staff adapt by being more concise or making fewer exchanges between senior and junior staff, though it was also clear that other communication events became more frequent, particularly those necessary for coordination.

Many of the behaviours interviewees described in this study relate to ónon-technical skillsô, such as situational awareness, communication or leadership. For example, maintaining an overview of the department refers to ósituational awarenessô and managing staff and patients would be considered óworkload managementô. Staff often commented how they learnt these behaviours the hard way through trial and error rather than formal training. Participants described the importance of leadership and yet it was clear that the registrars who filled this role had very little preparation or training. There appeared to be no specific instruction for registrars regarding what skills and behaviours were expected when coordinating the department and this had negative implications not only for the individual but for the team as a whole. Other issues raised by staff related directly to patient safety, for example, when junior staff feel unable to ask for help or believe it is unacceptable to question senior colleagues. Interviews revealed how it is important for registrars and other senior staff to remain calm and approachable and highlighted problems caused by authority gradients in the ED. Non-technical skills were shown to be essential for a team working under pressure yet these skills were rarely explicitly discussed or taught. This revealed a fundamental gap in medical education. Clearly trainees need to be given a more structured approach to learning the skills that have implications for both efficiency and safety in the ED.

Subsequent studies described in this report therefore focus on non-technical skills for the ED registrar, as they perform a vital leadership role in the department. The first step towards improving registrarsô non-technical skills is to identify those skills essential for trainees. Further research was necessary as there are no published studies focussing on these skills in a UK setting. The skill set was integrated into an assessment tool to structure feedback, promote self-reflection and target training needs.

There are some limitations of this study that should be considered. Interviews reflect self-reported behaviour, rather than actual behaviour, and this relies on respondents giving accurate and honest answers. This may be limited by incomplete recollection, misunderstanding the question or embarrassment. However, in this study, participants appeared to be relaxed and open and many were willing to admit to less-than-best practice. The sample used in this study was small and represented the views from one London ED therefore it is difficult to generalise the findings to other ED staff. Despite this, interviews are a useful tool to uncover issues that may need further investigation and lead to subsequent research using other methodologies, such as observation (Jensen, 1991).

This exploratory study draws attention to ED traineesô lack of preparation for leadership and reveals the ad hoc way that many staff accrue vital non-technical skills. Medical education should expand to include training related to non-technical skills, in addition to clinical skills, in order to help individuals and teams working under pressure in the ED. Improving non-technical skills, through structured assessment in the ED, is the focus of this thesis. The following chapter describes a review of published literature linking non-technical skills to safety in the ED and provides the foundation for identifying the salient non-technical skills.

4 Non-technical skills associated with safety and error in the ED – A scoping review of the literature

CHAPTER OVERVIEW

This chapter summarises the published empirical evidence linking specific non-technical skills to safety and error in the ED. A scoping review of the literature was undertaken, in two stages, to identify nine key non-technical skills for EM clinicians. The chapter concludes with limitations of the review.

4.1 Background

There have been several literature reviews looking at teamwork and non-technical skills in the ED (Kilner and Sheppard, 2010, Eppich, 2008, Manser, 2009). Whilst some appear promising, on further inspection they offer little evidence specifically related to EM. For example, in a systematic review of teamwork and patient safety in dynamic domains of healthcare, such as the operating room, intensive care and emergency medicine, only 11 of the 60 articles reviewed and referenced involved the ED or resuscitation teams (Manser, 2009). The reviewed papers relevant to EM focussed on team performance or evaluation of simulation training rather than attempting to demonstrate any clear association with error or safety in the ED. In general, EM teamwork literature usually relates to the development or evaluation of simulation-based training courses for improving non-technical skills in resuscitations. Other reviews have focussed on staff attitudes to teamwork and training rather than concentrate on safety aspects (Kilner and Sheppard, 2010). There is no review of the literature that specifically lists the non-technical skills associated with safety and error in the ED. Furthermore, it is essential to consider the skills necessary for all aspects of EM work rather than focus exclusively on critically ill or injured patients.

The aim of this scoping review was to first identify, and then explore in more depth, the non-technical skills specifically linked to safety and error in the ED.

4.2 Methods

A scoping review was carried out in preference to a systematic review of the literature, as systematic reviews are narrow in focus and are guided by specific research questions (Glasziou, 2001). In contrast, a scoping review forms part of the preliminary investigative process to give an overview of the range and nature of existing evidence (Davis, 2009) and guides more focussed lines of research.

This review was conducted in two stages. Stage 1 involved undertaking a search of the literature to produce a list of non-technical skills related to safety and error in the ED. Stage 2 utilised this list of individual skills to guide a further search of the literature and expand the evidence for review. This method was chosen because preliminary searches of the literature identified very few papers focussing specifically on non-technical skills and safety in the ED. Relevant studies were missed because they either (i) focussed on isolated skills, such as communication, or (ii) reported a range of skills but failed to use the general terms ónon-technical skillsô or óteamworkô. Therefore, in order to complete an adequate review of the literature it was necessary to first identify important skills.

4.2.1 Stage 1

Search strategy Medline, EMBASE and PsychINFO databases were searched. This was supplemented by searching Health Information Resources (formerly National Library of Health) and reviewing reference lists of seminal articles and relevant systematic reviews. Keywords and MeSH terms were generated around three main concepts, namely non-technical skills, emergency medicine and error. The search terms were:

Concept 1: ñNon-technicalò or ñteam workò or ñhuman factorò or ñleadershipò (and all variations)

Concept 2: ñEmergency medicineò or ñemergency departmentò or ñaccident and emergencyò or ñemergency physicianò or ñresuscitationò or ñtrauma teamò or ñarrest teamò (and all variations)

Concept 3: ñErrorò or ñsafetyò or ñcritical incidentò or ñadverse eventò (and all variations)

Results from the three searches were combined and limited to papers published between 1988 and November 2010, using human samples, with abstracts available. Sensitivity of the search was checked by ensuring known relevant papers were retrieved by the search, and keywords were modified or added accordingly. Thus, hand-searching informed and was incorporated into the search strategy.

Inclusion and exclusion criteria

All articles retrieved by the search were assessed by two EM clinicians (LF and RB) and consensus was reached following discussion. Eligibility judgements were made according to the following inclusion and exclusion criteria:

Inclusion criteria:

- Empirical study e.g. observational, case note analysis, interview or survey etc
- Adult or paediatric emergency medicine
- The main focus of the paper was related to
    a) identifying non-technical skills related to error or safety in the ED, or
    b) where non-technical skills were identified as a significant contributory factor

Exclusion criteria:

- Related to specific types of error e.g. incorrect radiological or ECG interpretation
- Pre-hospital care or neonatal resuscitation (as this usually takes place in the delivery suite)
- Assessment of an intervention
- Review of the literature or commentary (this is not empirical evidence; however reference lists were checked)

Of 98 articles identified through the Ovid search, 22 abstracts were potentially relevant. Full text copies of the papers were sought and screened according to the selection criteria. Six articles were excluded because they were precursor or follow-on studies of other papers that were included for review. Thirteen articles in total were excluded. Two additional papers were found by manual searching of reference lists. Searches using Health Information Resources did not reveal any new relevant papers. A total of 11 articles were used for data extraction.

Data Extraction

Data on study design, country of origin and sample were recorded. A Grounded Theory approach (Strauss, 1998) was used to analyse data and categorise skills. This is an iterative method where the data is examined to extract common themes, rather than using a predefined theoretical coding framework.

4.2.2 Stage 2

Skills identified from analysis of papers extracted during the Stage 1 review were used to generate new keywords for use in a secondary search. Search terms included the main skill headings identified in Stage 1 plus alternative terms commonly used by authors of included articles, as shown in Table 1. For example, papers that provided data coded into the skill ósassertivenessô also frequently included the terms hierarchy and authority gradients within the text. The choice of terms was decided pragmatically and some apparently useful terms were later excluded because the search retrieved too many irrelevant articles. For example the term ómonitorô appeared in articles related to situational awareness but searches produced excessive numbers of papers related to monitoring devices.

The final list of search terms used to expand and replace Concept 1 were communication, handover, handoff, workload, interruption, anticipate, plan, brief, situational awareness, supervision, feedback, standards, authority, assert, hierarchy, decision-making and all variations. Concepts 2 and 3 remained unchanged. These terms were searched using Medline, EMBASE and PsychINFO databases and relevant papers were identified using the

same inclusion and exclusion criteria described for Stage 1. The secondary search identified an additional 713 papers, 21 of which fulfilled the inclusion criteria.

4.3 Results

4.3.1 Stage 1

The factors identified as being associated with safety and error in the ED were categorised into 9 broad skills (see Box 4.1). Identifying the skill headings was an iterative process, with additional headings added as data analysis progressed.

1. Communication
2. Workload management
3. Anticipation
4. Situational awareness
5. Supervision and feedback
6. Leadership
7. Maintenance of standards
8. Assertiveness
9. Decision-making

Box 4.1 Non-technical skills associated with error and safety in the ED

4.3.2 Stage 2

A total of 32 articles were used to describe the main non-technical skills related to safety in the ED and a summary of the data extracted is shown in Table 4.1.

Table 4.1 Summary of data extracted from Stages 1 and 2

				Non-technical skills identified in the ED literature								
Ref.	**Study design**	**Country**	**Sample**	**Communication**	**Workload Management**	**Anticipation**	**Situational Awareness**	**Teaching Feedback & Supervision**	**Leadership**	**Maintenance of standards**	**Assertiveness**	**Decision making**
Anderson (2010)	Critical Incident Analysis	Denmark	122 critical incidents	Communication			Patient monitoring/ observation		Leadership			
Anderson (2010)	Interview	Denmark	11 ATLS instructors	Closed-loop communication	Avoid task overload				Identifiable leader Hands-off	Maintaining standards & guidelines	Assertion	Team decision making
Apker (2007)	Interviews	USA	12 interviews	Handover/ expectations								
Beach (2003)	Case study	USA	N/A	Team transitions			Team situational awareness					Diagnostic error/ cognitive bias
Boreham (2000)	Observation & critical incident analysis	UK	30 months								Authority gradients	Cognitive errors
Brodie (2009)	FMEA	UK	16 multi-disciplinary interviews						Identifiable leader			
Campbell (2007)	Case study	USA	N/A	Team transition							Authority gradients	Cognitive error
Cosby (2003)	Critical incident analysis	USA	636 cases	Communication Shift handover				Failure to supervise			Authority gradients	Medical reasoning
Eisenberg (2005)	Observation	USA	1 year	Handover/ team transition							Hierarchy	
Elshove-Bolk (2004)	Claims analysis	Netherlands	256 claims					Supervision				Diagnostic error

Famularo (2000)	Critical incident analysis	Italy	4 cases									Cognitive error
Fernandez (2008)	Focus group	USA	ED staff	Closed-loop communication	Back-up behaviour Coordinating activities	Planning	Patient & System monitoring Team cognition	Debrief	Leadership			
Fordyce (2003)	Observational & interview	USA	7 days 346 errors	Communication								Diagnostic error
Hamman (2010)	Scenario Simulation	USA	1 simulation	Handover								
Henneman (2006)	Focus groups	USA	20 ED nurses			Anticipation	Surveillance					
Hicks (2008)	Survey	Canada	84 ED doctors & nurses	Communication	Prioritising Coordinating team actions							
Hoff (2008)	Observations & Interviews	USA	ITU, surgical & EM teams		Workload Interruptions							
Horwitz (2009)	Survey	USA	139 respondents 40 errors	Handover	Workload						Professional conflict	Diagnostic error Incomplete info
Kachalia (2007)	Claims review	USA	122 claims	Handover	Workload			Supervision				Cognitive errors
Kennedy (2009)	Observations & interviews	Canada	88 EM & internal medicine trainees		Workload							Asking for help
Kaissi (2003)	Survey	USA	261 ED, ICU and OR nurses	Communication		Briefing		Debrief	Identifiable leader		Assertiveness Conflict resolution	Team input in decision making
Kilroy (2006)	Critical Incident Analysis	UK	18 EM consultants & trainees					Supervision				

Laxmisan (2007)	Observation & interviews	USA	3 months		Interruptions Multi-tasking							Cognitive error
Lubbert (2009)	Observation	Holland	387 trauma videos				Patient supervision		Identifiable leader	Following protocol		
Magid (2009)	Survey	USA	3562 ED doctors & nurses	Shift handover	Dealing with interruptions							
Redfern (2009)	FMEA	UK	N/A	Shift handover/ team transition								
Risser (1999)	Critical incident analysis	USA	54 critical incidents	Communicate plans Check back communication	Cross-monitor actions of team Prioritise Assign roles		Monitor execution of plan (maintain S.A)				Assert a position Apply 2 challenge rule	Offer/ seek info. for decision making Develop/ execute a plan
Schenkel (2003)	Critical Incident interviews	USA	16 medicine, surgical & obstetric trainees					Supervision Feedback				Diagnostic error
Simmons (2010)	Focus group	USA	19 NICU and ED nurses	Communication	Back up behaviour		Shared mental models		Leadership		Conflict resolution	
Stella (1996)	Critical incident analysis	Australia	112 incidents	Communication				Supervision				Diagnostic error
Wears (2002)	Case report	USA	Single case report		Interruptions	Anticipation	Situational awareness	Supervision				
White (2004)	Review of claims	USA	74 claims	Communication				Supervision		Documentation Failure to follow procedures		Diagnostic error

The nine main skills are now discussed.

1. Communication

Twenty papers commented on the association between communication and error in the ED. One research team interviewed ED staff at 3-4 hourly intervals during a shift and found that 12% of errors involved miscommunication (Fordyce, 2003). An Australian team monitored critical incidents and then asked staff to make suggestions on the reporting form. Improved communication was identified as the most important corrective strategy in 21% of incidents (Stella et al., 1996).

Numerous studies have identified communication failure at shift changeover or transfer of information between providers as a source of error in the ED (Beach, 2003, Cosby, 2003, Apker, 2007, Campbell, 2007, Eisenberg, 2005, Hamman, 2010, Horwitz, 2009, Kachalia, 2007, Magid, 2009, Redfern, 2009). One research team specifically commented on how different expectations between ED and in-patient teams had a negative influence on information flow (Apker, 2007). A case study highlighted how transition between multiple care providers can perpetuate a failure to diagnose the underlying medical illness (Beach, 2003). The authors described how this transition may cause communication breakdown due to insufficient or inaccurate data, mistimed or delayed information, poorly organised data and the insertion of ópseudo-informationôô. A year-long observational study in 2 EDs indentified four communication events in a patientôs journey that were linked to potential adverse outcomes: during triage, ambulance transfer, ED handover and admission (Eisenberg, 2005). The observers also noted barriers to effective communication included (i) hierarchy, where nurses and junior doctors were reluctant to question senior colleagues, (ii) noise in the department and (iii) the absence of somewhere to speak in private.

The potential for closed-loop communication to enhance information exchange and reduce error was identified in three studies (Risser et al., 1999, Fernandez, 2008, Andersen, 2010). This type of communication, where the message is repeated to ensure correct transmission, could potentially reduce errors due to óslipsô (errors in the performance of skill-based behaviours that occur when our attention is diverted).

2. Workload management

Workload management is a complex skill that covers a wide range of behaviours. The review identified a number of aspects of workload management relevant to safety and error, including (i) assigning roles (ii) dealing with interruptions, (iii) prioritising tasks for individual patients and triaging patients, (iv) practicing back-up behaviour, (v) allocating tasks and coordinating team activities and (vi) cross-monitoring actions of team members to manage workload (Fernandez, 2008, Hicks, 2008, Magid, 2009, Risser et al., 1999, Simmons, 2010, Wears, 2002). These elements clearly overlap with other non-technical skills, as dealing with interruptions involves effective communication and appropriate assertiveness, and cross-monitoring actions requires adequate situational awareness. A study using both observations and interviews to explore why trainees may be reluctant to ask for help, concluded that one of the main barriers was heavy workload (Kennedy, 2009) and a study looking at diagnostic errors in the ED identified excessive workload as a contributing factor in 23% cases (Kachalia, 2007). Another study involving observations and interviews concluded that intense workload and excessive interruptions reduced ED physiciansô ability to attend to safety and error (Hoff, 2008).

3. Anticipation

This skill refers to the planning and briefing that is often associated with the anticipated arrival of a critical patient. Preparations may include allocating team roles, identifying available equipment or noting ED capacity (Fernandez, 2008). However, this skill also refers to the ability to react to a change in a patientôs condition, for example, by transferring an unstable patient with resuscitation kit and intubation equipment (Wears, 2002). Anticipation is also important for considering the disposition of a patient. One of the studies included in this review commented on significant delays involving consultation from speciality teams and transfers to the intensive care unit (Magid, 2009). A focus group study identified óanticipationô as one of the main ways ED nurses identified and subsequently interrupted and corrected medical errors (Henneman, 2006).

4. Situational Awareness

Situational awareness involves surveying the environment to pick up cues that may need action, requesting information from others and relaying information to the team. The studies included in this review described a number of elements that are essential for good situational awareness in the ED. These include adequate (i) patient monitoring and observation, (ii) system monitoring (tracking team resources and environmental conditions), (iii) team cognition and shared mental models (which involves sharing information so that each team member has a similar understanding of task requirements and role responsibilities), and (iv) monitoring execution of plans (Andersen et al., 2010, Fernandez, 2008, Lubbert, 2009, Risser et al., 1999, Simmons, 2010, Wears, 2002). óSurveillanceô is also important for identifying and averting medical error (Henneman, 2006).

5. Supervision and Feedback

The importance of supervision in the ED has been highlighted in a number of studies (Kilroy, 2006a, Cosby, 2003, White, 2004, Kachalia, 2007, Stella et al., 1996). Supervision failure has been implicated in up to a third of adverse events (White, 2004) and a third of closed malpractice claims in the USA (Kachalia, 2007). A study of closed claims in the Netherlands found that resident supervision by a consultant was documented in only 15% of medical records and the authors commented on the need for improved supervision of junior doctors (Elshove-Bolk, 2004). A critical incident (interview) study based in the UK identified four key situations where supervision is required: (i) performance of a key practical skill, (ii) arrival at a key diagnosis, (iii) the performance of a key management step, and (iv) support in dealing with a complex referral (Kilroy, 2006a).

Many clinicians are aware that lack of feedback in the ED is a particular concern. Feedback is not only important from within the ED team, it is also vital that ED staff enquire about patient outcome from admitting teams or records to enhance the learning process. A US-based research team interviewed medical, surgical and obstetric residents (not EM residents) about errors involving ED care and found that the majority of errors were not discussed with the ED team (Schenkel, 2003). Other

papers reviewed identified the need for debriefing after critical cases as a method of feedback (Fernandez, 2008, Kaissi, 2003). In particular, a group at the Emergency Medicine Consensus conference specified that debrief should involve discussion of team processes as well as errors and ónear-missesô, rather than focussing on performance and outcome (Fernandez, 2008).

6. Leadership

There is a great deal of evidence, both within and outside medicine, that points to the importance of effective leadership and this is reflected in the fact that seven articles linked leadership skills with safety and error in the ED (Andersen et al., 2010, Andersen, 2010, Fernandez, 2008, Kaissi, 2003, Lubbert, 2009, Simmons, 2010). In particular, the need for a clearly identifiable leader is highlighted (Andersen, 2010, Lubbert, 2009, Kaissi, 2003, Brodie, 2009). A UK research team using FMEA to identify hazards in resuscitation concluded that óa lack of designated leaderô is both highly probable and a severe risk of harm to the patient (Brodie, 2009). One interview study also commented on the conflicting demands put on resuscitation team leaders (Andersen, 2010). The commonly accepted principle that the leader should maintain a hands-off approach is challenged in situations where the leader is considered the most competent in solving technical problems.

7. Maintenance of standards

Three articles included in this review noted the importance of maintaining standards and following protocols for ensuring patient safety (Andersen, 2010, Lubbert, 2009, White, 2004). Studies mainly focussed on teams in the resuscitation room (Lubbert, 2009, Andersen, 2010), however adhering to procedures is clearly important for any management plan. Difficulties may arise in determining whether any deviation from protocol is the result of knowledge gaps or errors in execution. There is also overlap with other skills, as failure to maintain standards may reflect work overload or cognitive errors of decision making. Failing standards covers other practices such as

poor documentation, which was found to be a contributing factor in 20% of closed malpractice cases (White, 2004).

8. Authority & Assertiveness

Skilful use of authority and assertiveness is essential for EM clinicians. This review identified several circumstances where error may occur, including (i) failure to assert a position or corrective action (Risser et al., 1999, Kaissi, 2003, Andersen, 2010), (ii) inadequate conflict resolution (Horwitz, 2009, Simmons, 2010, Kaissi, 2003), and (iii) rigid authority gradients (Eisenberg, 2005, Cosby, 2003, Campbell, 2007, Boreham, 2000). Researchers in an observational study witnessed the effect of junior staff who felt unable to question senior staff (Eisenberg, 2005). A 30 month observational study investigating error in the ED identified the importance of authority derived from experience, unique access to information or being at hand in an emergency, in addition to authority from formal status (Boreham, 2000).

9. Decision-making

The Netherlandsô closed claims review found 48% of cases were due to ñfailure to diagnose properlyò (Elshove-Bolk, 2004) and an interview study in the US found that eight of the sixteen cases discussed involved diagnostic error (Schenkel, 2003). This was contrasted with nursing staff reports which mainly involved errors in delivery of care. A UK study investigating critical incidents identified cognitive errors as one of three primary latent conditions (Boreham, 2000). Cognitive errors include slips and mistakes, cognitive bias (such as confirmation bias, looking for confirming evidence to support a hypothesis, rather than to refute it) and errors of interpretation. A case study highlighted a multitude of ways diagnostic error may occur, including failure to consider an alternative diagnosis, error due to inheriting someone elseôs thinking, errors of prevalence estimation and affective errors (Campbell, 2007). A review of four major clinical incidents in an Italian ED highlighted how overreliance on the use of patterns during cognitive processes may cause errors (Famularo, 2000). Staff in the ED are particularly prone to cognitive error, in part, because of the high frequency of interruptions and the tendency to multi-task (Laxmisan, 2007). This review identifies

the importance of team input in decision making as a key strategy to reduce error in the ED (Andersen, 2010, Kaissi, 2003, Risser et al., 1999).

4.4 Discussion

The aim of this chapter is to give a summary of the main ED non-technical skills literature, drawing specifically on empirical evidence. There are likely to be other articles of benefit, for example those that focus on team performance but are unable to demonstrate a direct association between specific behaviours and safety. Now that the principle skills have been identified further reviews would enable a deeper understanding of the elements required for each individual skill, however this is not within the scope of this thesis.

4.4.1 Limitations

This review shows that searching for relevant articles is difficult as there is no standardised use of terms and ónon-technical skillsô as a phrase has only recently been introduced into medical terminology (Glavin, 2011). This means that authors may not explicitly use the term in their article and, therefore, an indirect method of searching for relevant papers was needed. The coding of data to broad skills relied on interpretation and some behaviours identified in the articles could have been categorised differently. For example, dealing with interruptions was coded into workload management but may also fall within the skill communication. The secondary search could have been further expanded to incorporate other keywords which may have retrieved additional papers. However, this scoping review was designed to give an overview of the literature available rather than attempting to identify all papers on this topic.

The evidence presented in this review is qualitative rather than quantitative, i.e. interviews, questionnaires and observational research. Using the recommended GRADE system (Guyatt, 2008) for assessing the quality of studies (high, moderate,

low, very low) all the articles identified would be rated low or very low. However, qualitative research is useful for identifying and defining the óproblemô and which can then be further explored quantitatively. Whilst it is impossible to quantitatively determine the effect of skills like ónticipationô, it is possible to quantitatively assess the impact of non-technical skills training on error reduction. This has been attempted with limited success, with most studies only demonstrating improved team performance or positive effects on staff attitude (Jankouskas, 2007, McCulloch, 2009, Gaba, 1992, Wright et al., 2009).

4.4.2 Summary

An initial search identified only 11 articles describing how non-technical skills contribute to safety and error in the ED. However, analysis revealed a number of recurring themes that allowed identification of nine essential non-technical skills. These are communication, workload management, anticipation, situational awareness, supervision & feedback, leadership, maintenance of standards, authority & assertiveness and decision-making. A secondary search, using these nine skills and related terms, uncovered a further 21 relevant papers. This enabled a more detailed description of the non-technical skills associated with safety and error in the ED. The following chapter incorporates the findings of this review, together with other sources, to produce a provisional tool for the assessment of ED registrarsô non-technical skills.

5 Development of a provisional observational tool to assess non-technical skills in the ED

CHAPTER OVERVIEW

The previous chapter describes a review of the literature which is the first step to identifying the salient non-technical skills for ED physicians. This chapter starts by describing reasons for choosing to assess ED registrarsô non-technical skills using a behavioural marker system. A summary of the steps used in the development of the tool is given, followed by a detailed description of the methods used in Phase 1 to produce a provisional tool. The reasoning for inclusion and exclusion of each non-technical skill is presented and the rating scale is discussed.

5.1 Background

The initial interview study highlights the important contribution that effective non-technical skills and teamwork make to the smooth running of the ED. Furthermore, good leadership, especially that of the registrar, was seen to be vital for a team working under pressure. Traditionally, the development of UK trainees has been somewhat lacking in this area, with many registrars getting only limited feedback and training in non-technical skills such as leadership and teamwork. Over recent years, many healthcare organisations and governing bodies, such as the General Medical Council, have made a commitment to improving non-technical skills for British doctors. This has been evident with the recent surge in leadership training opportunities, for example medical leadership fellowships available through the Darzi Fellowship programme.

Training related to non-technical skills is increasingly available in the UK. In 2008 the College of Emergency Medicine (CEM) set up a specialist group with the aim of improving non-technical skills through simulation training. However, there is currently no standardised way of measuring non-technical skill performance that is both specific to EM and comprehensive, covering resuscitation and routine work. A

robust method of assessing non-technical skills is necessary to identify training needs, facilitate feedback and measure the impact of any training intervention.

A number of assessment tools exist for assessing non-technical skills in resuscitation scenarios (Kim et al., 2006). However, these tools largely mirror CRM principles that originated in aviation with little attention paid to the unique characteristics of working in the ED. Furthermore, these tools focus entirely on a single critically ill patient and fail to encompass the skills necessary for managing multiple patients or issues related to coordinating the wider department. Indeed, the management of resuscitation patients makes up only a small proportion of an ED physiciansô daily workload and non-technical skills are equally as important for dealing with lower acuity patients.

In surgery and anaesthetics, the most common method of assessing non-technical skills is in the form of observation. Skills are often rated using a behavioural marker system, such as ANTS or NOTSS (see Introduction Chapter 2.5). The advantage of a behavioural marker system over other forms of observational assessment is that it explicitly describes the salient skills and associated behaviours. Such an overt method of assessment is less important when the target is familiar and performance criteria are well known. For example, medical professionals have a well established understanding of what constitutes a good history and examination. In contrast, there is currently no clear consensus among EM physicians which non-technical skills are most relevant in the ED or what constituent behaviours represent these skills in practice.

Observation of non-technical skills may take place either in the workplace or, more commonly, in a simulator environment. The advantage of a workplace based assessment is that it incorporates the complex interactions of authentic situations that cannot be replicated in a simulator. As illustrated in Figure 5.1, Millerôs pyramid of assessing clinical competence shows that actual behaviour (ódoesô) is a higher level than simulator performance (óshows howô) (Miller, 1990).

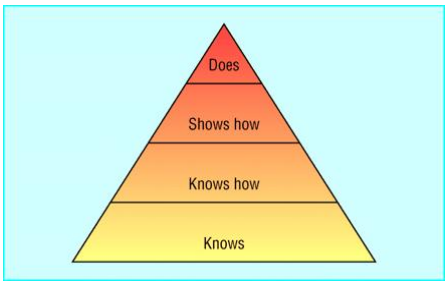


Figure 5.1 Millerôs pyramid for assessing clinical competence

This relies on the assumption that assessments of actual practice are a better reflection of routine performance than assessments done under artificial conditions. Workplace based assessment target the highest level of the pyramid. The most challenging aspect of improving non-technical skills is not the knowledge base but learning how to integrate that into daily practice and the competing demands of working in the ED. Furthermore, workplace based assessments may be the only realistic method of evaluating performance related to higher level tasks such as coordinating the larger team and managing multiple patients. Therefore, workplace based assessments of non-technical skills may be particularly suited for EM trainees.

The tool described in this thesis has been designed for use in formative assessment to: promote reflection, guide future learning, and monitor progress (Cox et al., 2007). Encouraging staff to pause and reflect on their performance of non-technical skills is particularly important in the fast-paced ED.

It is clearly important that all members of the ED team have proficient non-technical skills. However, the interview study revealed that those in a leadership role are of particular importance. The lead physician may be either a consultant or registrar who, in addition to meeting the clinical needs of patients, also takes responsibility for overseeing and coordinating staff and patients on the óshop floorô. In the department involved in the interview study, the leadership role was predominantly taken by the registrar, under the supervision of an ED consultant. Registrars are also one of the primary targets for workplace based assessment, although with revalidation, assessment of consultants may be in the not too distant future. For those reasons, the

tool was developed primarily for the assessment of EM registrars, though it is expected that these skills would be the same as those required for ED consultants.

With this in mind, a series of studies were undertaken to produce an observational tool for the assessment of registrarôs non-technical skills in the ED, with a particular focus on the leadership role.

5.2 Aims

The development of the assessment tool was divided into two phases, each with specific aims. The aims were:

Phase 1: To produce a provisional assessment tool based on published literature to evaluate the feasibility of assessing non-technical skills in the ED.

Prior to undertaking more detailed studies a provisional tool was tested to determine the feasibility of observing the ED registrar. This was needed because the ED registrar works over a large geographical area rather than the limited confines of an operating theatre or resuscitation room. A trial was also necessary to check if a sufficient variety of skills would be observable in a reasonable time frame.

Phase 2: To identify any significant omissions from the tool using staff interviews and field observations and to extract examples of positive and negative behaviour to explain the skills in the context of the ED. A secondary aim was to establish whether skills included in the tool were observable.

The second phase involved more in depth studies to further develop the tool. This ensured a full set of skills was identified and that these skills were linked to observable behaviours. This is described in Chapter 6: Revision of the Observational Tool. The two phases of development are illustrated in Figure 5.2 below.

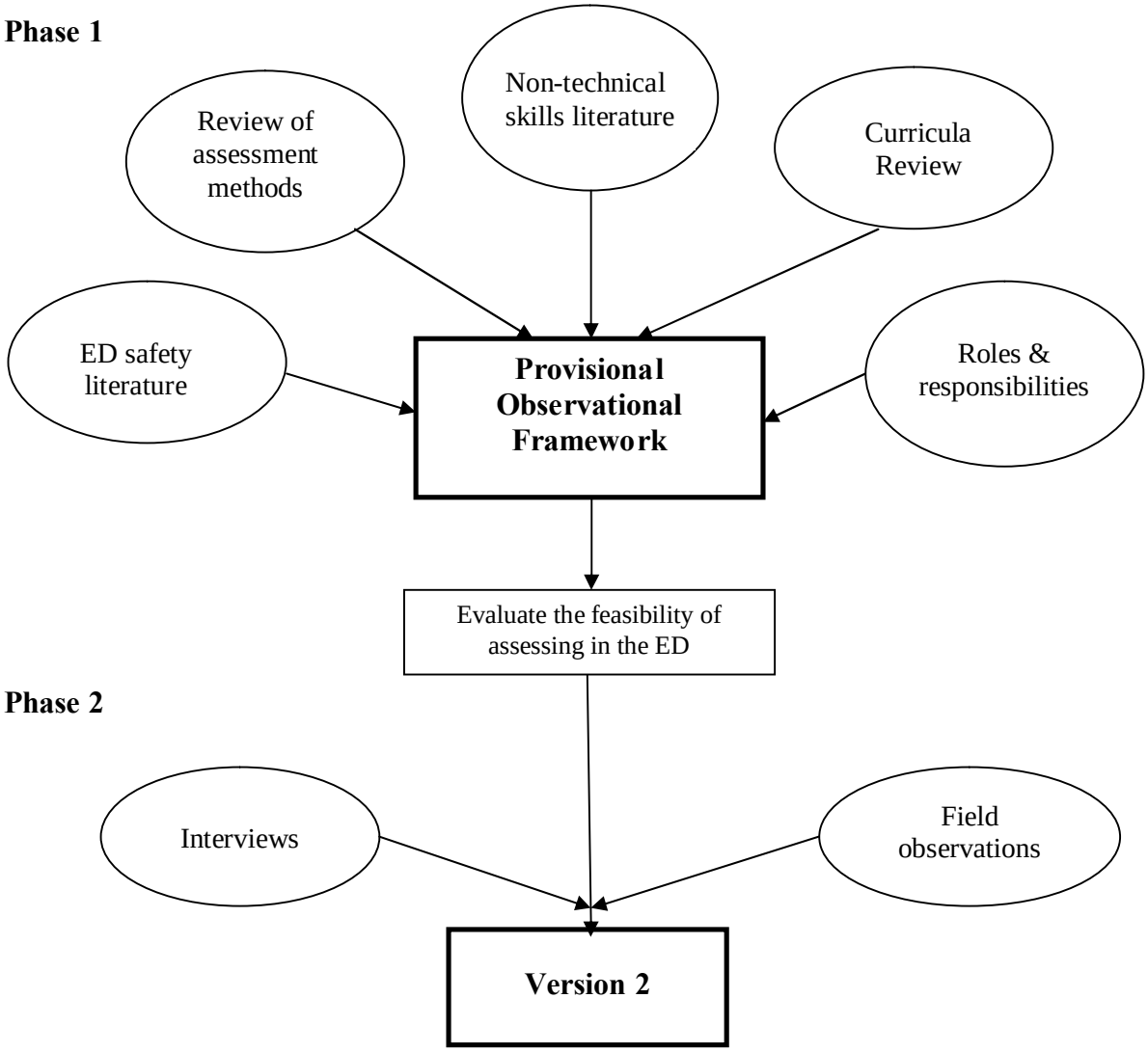


Figure 5.2 Summary of steps to develop the assessment tool

5.3 Methods

The provisional observational tool was based on a comprehensive review of the literature. This included:

1. Examining the ED safety literature

Relevant databases were searched to retrieve articles with the aim of identifying those skills shown to contribute to safety and error in the ED. This is fully described in Chapter 4: A scoping review of the literature.

2. A review of existing tools used to assess non-technical skills in healthcare

A review of the literature was undertaken to identify a list of non-technical skills that had been included in other observational tools. This was based on the understanding that many key non-technical skills are relevant to all medical teams (Baker, 2005, Frankel, 2007). However, it is also accepted that EM provides some unique challenges related to the patient, their illnesses and the ED setting, as previously discussed. Combining the skills taken from the ED safety literature with those identified in other areas of healthcare provided a list of potentially relevant skills that guided subsequent analysis.

Articles were also examined to explore the methods used by other research teams to construct behavioural marker systems for use in healthcare, as discussed in the Introduction Section 2.5. This helped to determine the steps needed to identify the non-technical skills for the ED registrar.

3. Analysis of relevant curricula

The CEM curriculum (College of Emergency Medicine, 2009) describes a set of generic skills that are expected of EM trainees. This covers, for example, communication and documentation. Also, the Academy of Royal Medical Colleges (ARMC) has produced a curriculum for medical leadership (Clough, 2009). Many attitudes and behaviours described in these documents relate to non-technical skills that may be observed in the ED. These curricula were reviewed in detail to determine which competencies relate to non-technical skills and to ensure the final observational tool reflected this.

4. A review of current methods of assessment for EM trainees

The CEM and ARMC curricula not only describe relevant skills, but also state the current methods used to assess each element, for example multi-source feedback. The existing assessment methods were reviewed to ensure that an assessment tool was developed that complemented what was already in place.

5. Consideration of registrarsô roles and responsibilities

The assessment tool needed to cover the primary roles and responsibilities of an ED registrar. These are described in an induction booklet given to new middle grades at St Maryôs Hospital (see Appendix B) and this was used to help decide which non-technical skills were most important.

Primary data analysis from each of the above sources was undertaken and presented to the research team, comprising three EM physicians and two human factors experts. This was necessary to reduce the effect of investigator bias. Subsequently, each skill from the list was considered individually and discussed for inclusion in the assessment tool.

5.4 Results

5.4.1 Review of ED safety literature

Analysis of articles retrieved in the original scoping review of literature (described in Chapter 4) was based on themes emerging from the data rather than following a predefined coding framework. The skills identified in the scoping review are shown in Box 5.1.

1. Communication
2. Workload management
3. Anticipation
4. Situational awareness
5. Supervision and feedback
6. Leadership
7. Maintenance of standards
8. Assertiveness
9. Decision-making

Box 5.1: Skills identified in the review of ED safety literature

5.4.2 Review of non-technical skill literature

A second literature review identified an array of behaviours and skills that had been included in other healthcare observational frameworks. The skill list was derived from published work in a variety of different medical specialities including anaesthetics, surgery, neonatal resuscitation and emergency medicine. In total, thirteen observational systems were reviewed (Malec, 2007, Frankel, 2007, Risser et al., 1999, Yule et al., 2006, Fletcher et al., 2003, Mishra et al., 2009, Thomas, 2004, Healey et al., 2004, Cooper, 2007, Ottestad, 2007, Kim, 2006, Holcomb, 2002, Mitchell, 2009).

The observational tools reviewed included both team and individual tools. This was because few researchers have focussed on individualsô non-technical skills and many of the team-based tools describe behaviours that are also considered to be non-technical skills. For example, the MedTeams project was primarily interested in ćteamwork actionsô but the list essentially encompasses a wide range of non-technical skills such as decision making, situational awareness and communication. However, some teamwork behaviours, such as cooperation, may be very difficult to assess individually, and were therefore excluded from the list.

In addition to taxonomies focussing on teamwork and non-technical skills, leadership ratings scales were reviewed (Cooper, 2007) as well as some scales that assessed both technical and non-technical skills (Ottestad, 2007, Holcomb, 2002).

The skills identified in this review of non-technical skills literature were condensed into 13 main skills and are shown in Box 5.2 below (see Appendix B for more detail). Those identified as being associated with error and safety specifically in the ED (shown in Box 5.1) clearly fit into this skill list derived from other behavioural marker systems. Therefore, subsequent data collection used these 13 non-technical skills.

1. Maintaining standards
2. Decision making
3. Workload management
4. Communication
5. Conflict resolution
6. Team building
7. Leadership
8. Support of others
9. Teaching and feedback
10. Authority and assertiveness
11. Situational awareness
12. Coordinating team members
13. Assessing capabilities/ supervision

Box 5.2: Skill list used for data collection

5.4.3 Curricula

In 2009, The CEM curriculum described a set of generic skills that are expected of EM trainees although this is being revised so that the skills are more fully integrated into the curriculum. A thorough review was undertaken to ensure that as many skills as possible from the curriculum were represented in the observational tool. For example the curriculum highlights the importance of accurate and legible documentation which would be covered by the non-technical skill ôMaintenance of Standardsô. The curriculum also mentions delegation and seeking help which is included in ôWorkload Managementô. A full comparison can be found in Appendix B.

The Medical Leadership Competency Framework (MLCF) was developed by the ARMC in collaboration with the NHS Institute for Innovation and Improvement (Clough, 2009). It was based on a review of the literature on medical leadership and a comparative analysis of leadership competency frameworks, as well as consultation with a variety of professional and lay groups. Whilst much of the curriculum focuses on leadership in a managerial role (such as planning service improvement) it also describes the more day-to-day clinical leadership required to support effective teamwork. Many attitudes and behaviours described relate to non-technical skills that may be observed in the ED. For example, the section ôDeveloping Networksô describes how individuals should ñinvite a wide opinion from patients, carers and the wider health teamò. This is reflected in the skill ôDecision Making- option generationô. A full list of leadership competencies that are related to teamwork and non-technical skills can be found in Appendix B.

5.4.4 Roles and Responsibilities

Newly appointed ED middle grades working at St Maryôs Hospital are given an induction booklet that lists what is expected of the lead, bleep-holder registrar and this helped to identify the non-technical skills related to the practical coordination of the department. An attempt was made to retrieve equivalent documents from other EDs within the region but comments from trainees suggested that many departments fail to

explicitly list what is expected of their middle grades. An example of the responsibilities described in the booklet is ñOversee the workload in Majors and ensure a decision is made within 30 minutes of the patient being seenò. This shows that Workload Management is an important skill and highlights the responsibility of registrars for efficient decision making in the department. A full comparison can be found in Appendix B.

5.4.5 Existing Forms of Assessment in the ED

The existing assessment methods mentioned in the CEM and ARMC curricula were reviewed to ensure the observational tool complements what is already in place. The CEM website describes the methods of assessment currently in use. These are:

- Clinical methods ï include observed clinical care during working time, mini Clinical Evaluation Exercises (mini CEX), direct observation of practical skills (DOPS), acute care assessment tool (ACAT), case based discussion and video or observed operating within a team (for example the resuscitation room).
- Office based methods ï such as an audit of cases and review of complaints and clinical incidents.
- Teaching session based assessmentsï this covers a variety of methods such as topic presentations and scenario teaching
- Peer review - Mini PAT (peer assessment tool) also known as Multi-source Feedback (MSF) or 360° feedback. It involves multi-professional feedback and covers the five domains of Good Medical Practice: 1) good clinical care; 2) maintaining good medical practice; 3) relationships with patients; 4) teaching, training & appraising and 5) assessing and working with colleagues (GMC, 2006)
- Formal examinations ï written (MCEM and FCEM short answer questions and multiple choice), OSCE and clinical exams.

Many of the assessment methods focus on clinical skills. Whilst coordinating the department and supervising junior staff is an important role of registrars, especially during out-of-hours, this is not reflected in how a trainee is assessed. MSF allows for specific feedback of some non-technical skills (see Appendix B). Each statement is rated on a scale of 1 to 5. Unlike some peer assessment forms, the CEM form does not give space for additional supporting information or examples so it may be difficult to reflect on a particularly low or high rating. Even if there was additional space for free text, there is no guarantee that respondents would use it or that concrete examples would be given. However, an observational tool for assessing non-technical skills allows immediate and specific feedback to the trainee.

An alternative method for assessing non-technical skills currently in use is ñobserved clinical care during working timeò. However, without a framework to structure feedback, assessment may focus on patient-centred aspects of care rather than associated teamwork or non-technical skills. This is also likely to result in wide-variations in how a trainee may be assessed and depends on the assessorsô personal knowledge of non-technical skills. Furthermore, personal experience has shown that there are significant differences between departments in how often trainees are observed in their clinical duties. It should also be noted that a traineeôs teaching ability is expected to be assessed. This is based largely based on classroom teaching rather than spontaneous, ad hoc óshop floorô teaching, although the FCEM exam now incorporates some teaching stations.

The ACAT-EM is a version of the Acute Care Assessment Tool revised specifically for use in the ED (see Appendix B for further details). It covers a number of important non-technical skills such as workload management and team building but falls short of comprehensively assessing all the primary non-technical skills for EM trainees. Furthermore, it has been translated from an assessment tool designed by a group of clinicians on the Royal College of Physicians curricula committee and used to assess competence of trainees managing an acute medical take (Johnson et al., 2009) without any empirical grounding in EM. Whilst ACAT-EM provides an adequate temporary solution in the absence of any other observational tool to evaluate EM non-technical skills, it lacks the detail needed to assess the complex role of an EM trainee.

This review of existing methods of assessment reveals that although some EM non-technical skills are assessed, the current tools have a number of shortfalls and fail to comprehensively assess these important safety skills.

5.4.6 Summary of sources

Table 5.1 below summarises the results from a review of the published literature as well as curricula, roles and responsibilities and existing methods of assessment. This reflects both the practical aspects of service provision, as described in the roles and responsibilities document, as well as the more team-orientated skills, such as team building, and teaching and coaching covered in the curricula.

Table 5.1 Summary of various sources used in the development of the provisional tool

Non-technical Skill	ED safety literature	CEM curriculum	ARMC curriculum	R&R	MSF	ACAT-EM
Maintaining standards	•	•	•	•	•	•
Decision making	•	•	•	•	•	•
Workload management	•	•	•	•		•
Exchanging information	•	•	•	•	•	•
Conflict resolution		•	•			
Team building		•	•		•	•
Leadership	•	•	•	•		•
Support of others		•	•			
Teaching and feedback	•	•	•		•	
Authority and assertiveness	•		•		•	
Situational awareness	•			•		
Coordinating team members				•	•	
Supervision/ Assessing capabilities	•			•		•

ARMC = Academy of Royal Medical Colleges
R&R = roles and responsibilities
MSF = multisource feedback
ACAT-EM = Acute Care Assessment Tool

5.4.7 Inclusion or exclusion of skills to produce a provisional assessment tool

Following this review of the literature, curricula and methods of assessment, each of the thirteen skills listed in Box 5.2 were considered for inclusion in the provisional assessment tool. The reasoning for inclusion or exclusion of each skill from the framework is given below.

1. Maintenance of standards

As in all areas of medicine, this is an important skill in the ED and significantly contributes to patient safety. However, maintenance of standards is particularly relevant in UK Emergency Departments because of the occasional conflicting priorities of government targets and clinical need. A registrar requires good judgement to ensure targets are met whilst patients receive optimal care. This skill is also used to ensure clinical guidelines are followed and safety standards are met, for example good hand hygiene and legible patient notes, as mentioned in both the CEM and ARMC curricula. The definition covers maintaining standards within the traineeôs own work as well as ensuring compliance within the team. Some tools explicitly mention óCross-monitor team member actionsô or óEngage in error-correction actionsô. This highlights the general principle that it is the responsibility of all team members, not just those in a supervisory role, to be vigilant for errors made by others and intervene if necessary.

2. Decision making

Decision making was a common skill found in the literature and is the foundation of all clinical work in the ED, from arriving at a diagnosis to planning and executing a patient management plan. This also includes aiding a junior doctor in their decision making, problem-solving during periods of departmental overcrowding or other issues that may arise during the course of a shift, for example, a doctor calling in sick or a malfunctioning computer system. The inclusion of this skill in the provisional

framework is supported by the literature (e.g. high prevalence of diagnostic error in the ED) as well as the curricula, and roles and responsibilities of a registrar. Analysis of other assessment tools highlighted the complexity of skills such as decision making and situational awareness. In two behavioural marker systems (Fletcher et al., 2003, Yule et al., 2006) these skills were each divided into 3 or 4 subcomponents which were separately assessed. For example, in the ANTS behavioural marker system, decision making was divided into óidentifying optionsô, óbalancing risks and selecting optionsô and óre-evaluatingô. After discussion with the research team, it was agreed that dividing decision-making (and situational awareness) into sub-skills was also appropriate for EM. These were option generation, selecting and communicating options and outcome review.

a) Option generation: An ED registrar must ensure there is adequate information available before a decision is made regarding a patientsô management. This may involve accessing a number of sources such as hospital notes or speaking to relatives and the GP. Addressing departmental issues may require gathering information from the nurse-in-charge, bed manager or other relevant staff.

b) Selecting and communicating options: When helping a junior doctor, the registrar may either coach the doctor through the decision-making process or, if working under pressure, may take a more directive approach. An important aspect of this skill is verbalising the risks of various options and explaining any changes to a doctorôs management plan. Explanation of decisions is also vital for understanding within the nursing team and plays an important safety role by highlighting errors in the decision-making process.

c) Outcome review: Reviewing the outcome of a decision is vital to take into account any new information that has come to light, consider the impact of treatment and to ensure a decision has been acted on appropriately.

3. Workload Management

This skill is equivalent to ôIdentifying and Managing Resourcesô as described in other behavioural marker systems. One of the major responsibilities of the lead registrar is workload management and this is an essential skill for ensuring patients are seen in a reasonable time frame. Workload management involves managing their own, as well as othersô workload to avoid periods of both over and under-activity. Junior doctors may need encouragement to work more efficiently but registrars must also ensure that doctors do not get overwhelmed by seeing too many patients at one time. This may also require coordinating the team so that individual areas of the department are staffed according to need. Demonstration of this skill may involve the lead registrar delegating tasks to a colleague, knowing when to ask for help and understanding when a situation requires escalation to senior colleagues. Furthermore, prioritising tasks when working under pressure is essential for workload management. The importance of workload management is highlighted by the fact that it is included in the curricula, the roles and responsibilities as well as published literature.

4. Communication

The safety literature shows that effective communication is vital for safe practice in the ED and is covered in both the CEM and ARMC curricula. However, all other skills rely on effective communication so it may be hard to isolate this as a separate skill. Indeed, the ANTS tool does not contain communication as a separate element for this reason (Fletcher et al., 2003). This skill was excluded from the provisional tool.

5. Conflict Solving

This skill has some overlap with ôAuthority and Assertivenessô but refers specifically to disagreements that require resolution. This may include, for example, coming to an agreement with a speciality doctor who initially refused a referral. The skill requires the doctor to choose an appropriate time and place to discuss the problem, i.e. away

from patient areas, and resolve the issue whilst remaining calm. Unfortunately, it is fairly common to have minor disagreements with speciality doctors or radiology personnel so this is an important skill for the ED registrar. The CEM and ARMC recognise the value of this difficult skill with inclusion in both curricula.

6. Team Building

A strong team is vital in the ED and the registrar forms a pivotal role in maintaining this. This may be in the form of motivating the team during times of stress or saying ćthank youô at the end of a difficult shift. Many junior doctors find the ED a challenging place to work with difficult shift patterns and this skill is essential for maintaining a good team spirit. It is also vital that the registrar is perceived as being approachable by staff who need to ask a question or need help. Team building is considered an important competency in both the CEM and ARMC curricula.

7. Leadership

As discussed in the introduction to teamwork and non-technical skills (see Section 2.1.1), leadership is comprised a number of separate skills that are assessed individually, for example, workload management or assessing capabilities. Failure to establish a leader, particularly within a resuscitation team, has been highlighted as a significant contributor to error in the ED. However, within the department, leadership may be allocated on a rota system or the leader may be identified by carrying a bleep. If the department does not adequately assign a lead physician for the shift, this could be considered a failure of the system rather than the fault of an individual. For these reasons, it was decided not to include leadership as a separate skill in the provisional assessment tool.

8. Support of Others

Both the ANTS and NOTTS observational frameworks included this skill and similar definitions were used: óproviding physical, cognitive and emotional help to team membersô (Fletcher et al., 2003, Yule et al., 2006). The definitions of ósupport of othersô also referred to the importance of debriefing and giving feedback or constructive criticism, which is an explicit competency in the curricula. The ED is a very demanding environment in which to work so learning to support others is a key skill. Providing back-up behaviour to support colleagues is important for the smooth running of the team. The registrar needs to provide encouragement as well as feedback and coaching for junior staff. Therefore, this skill was included in the initial ED observational tool.

9. Teaching

Whilst this is an important skill, it may not be observed frequently enough to warrant being included in the provisional tool. Formal teaching, such as small group work or lectures, will not be observed.

10. Authority & Assertiveness

The published literature clearly highlights that this is an essential skill in the ED (Risser et al., 1999, Kaissi, 2003, Horwitz, 2009, Simmons, 2010, Eisenberg, 2005, Cosby, 2003, Campbell, 2007, Boreham, 2000). Whilst óCoping with Pressureô (a skill included in other behavioural marker systems (Yule et al., 2006)) was also considered to be important, it was felt that there were unlikely to be sufficient observable behaviours to warrant making it an individual element. Furthermore, closer inspection of the NOTSS descriptors for óCoping with Pressureô reveals it overlaps significantly with óAuthority and Assertivenessô, for example, ñadopting a suitably forceful mannerò and ñtakes responsibility for the patientò. Both skills refer to the ability of the doctor to take control and speak up when necessary. Whilst there may appear to be some overlap between óConflict Solvingô and óAuthority and

Assertivenessô, the latter includes situations where assertiveness is required in the absence of conflict, such as pointing out an error to a senior or ensuring tasks get done in a timely manner. This skill was included in the provisional assessment tool.

11. Situational Awareness

Situational awareness involves surveying the surroundings in order to pick up cues that need action. This seems to be a ubiquitous skill in the behavioural marker systems reviewed and it is also clearly an essential skill for coordinating the ED. The lead registrar must be aware of waiting times and staffing levels in all areas of the department and, if necessary, allocate staff accordingly. It is also important to have an awareness of the state of the hospital in general and other issues that might affect the smooth running of the department, such as a potentially disruptive or sick patient. Checking the Patient Tracking System is one way of maintaining situational awareness, as well as listening to colleaguesô communications and observing patients in the department.

Some tools divide situational awareness into three elements. This reflects the fact that situational awareness requires not only surveying the environment for cues, but also acting on those cues, relaying information to the rest of the team and planning ahead to prevent problems in the future. Clearly it is possible to be competent at picking up cues but still fail to ensure the rest of the team is updated. Therefore it was decided that the provisional framework would be based on three separate elements: a) Gathering Information; b) Understanding and Communicating, and c) Anticipating. The wording was chosen carefully, for example, ôGathering Informationô was used, as in ANTS and NOTSS, instead of the term ôNoticeô which is used in the Oxford NOTECH. The former suggests a more active seeking of information rather than passively noticing cues in the environment. Though it is not explicitly mentioned in either curriculum, nor explicitly assessed, this skill is clearly vital for smooth running in the ED and is an important element of CRM training and was therefore included in the tool.

12. Coordinating team members

This skill is common in operating room or resuscitation room taxonomies, where several staff members routinely work around a single patient. However, most coordination of the wider ED team is covered in ôWorkload Managementô and ôSituational Awareness ï Understanding and Communicatingô. It is not a specified competency in either the CEM or ARMC curriculum nor is there strong evidence relating it to safety in the ED literature. Therefore, this skill was not included in the provisional tool.

13. Assessing Capabilities

ôAssessing capabilitiesô also referred to as ôUnderstanding team needsô, is a vital skill for ED registrars. Registrars often have to supervise new locum doctors and it is essential that they adapt the level of supervision according to ability. Similarly, different junior doctors require different levels of supervision, with some doctors able to work relatively autonomously and others requiring more regular input. It may also be necessary to consider the experience of a speciality doctor who is receiving a referral as some are relatively inexperienced with limited specialty knowledge. The importance of supervision is stressed in the ED literature. In addition, this skill refers to the doctorôs ability to be aware of or sensitive to a team memberôs fatigue and stress levels. This is essential in such a demanding environment and may be demonstrated, for example, by suggesting a break.

5.4.8 Summary of results

From previous published work and initial observations in the ED it was clear that leadership was difficult to isolate as a separate skill. Salas (2004) specified important aspects of leadership including directing and coordinating the activities of others, assessing team performance, assigning tasks, motivating team members and organising and executing management plans (Salas, 2004). Therefore good leadership incorporates a number of non-technical skills such as situational awareness, workload

management, teaching/ feedback, team building and decision-making. Consequently, leadership was not coded as a separate skill. Instead, elements of leadership behaviour were incorporated into the definitions and behavioural markers of each individual skill.

As described above, Decision-making and Situational Awareness were each divided into 3 subcomponents to enable separate assessment. Therefore, the provisional tool comprised 13 separate elements or skills. The skills excluded from the provisional tool were: communication, leadership, teaching & feedback and coordinating team members. The categories and skills of the provisional observational framework are shown in Table 5.2 below.

Table 5.2 Skills included in the provisional tool

Category	Skill	Definition
Leadership & Management	**Maintenance of Standards**	Includes achieving 4 hour targets as well as clinical and safety standards. Incorporates own work as well as ensuring others meet appropriate standards.
	Assessing Capabilities	Being alert to factors that may limit team membersô ability to perform, including level of experience, stress and fatigue. Aware of skill level and expertise of all team members including nursing staff, speciality doctors etc.
	Workload Management	Includes managing own and others workload, avoiding both over and underactivity. (Reallocating staff due to long wait is considered in óProblem Solvingô or preventing build up of patients in óAnticipatingô; Back-up behaviour is covered in óSupport of Othersô)
	Authority & Assertiveness	Maintains ones position until convinced otherwise. Overlaps with óConflict Solvingô but also includes situations where assertiveness is required in the absence of conflict, eg pointing out error to senior, ensuring tasks get done in a timely manner.
Teamwork & Cooperation	**Team Building**	Traits and behaviours that promote team cohesion and produce an environment that is conducive to efficient teamwork. Appreciate contribution made by support staff such as porters and receptionists.
	Support of Others	Providing physical, cognitive and emotional support to others. Includes back-up behaviour to support the ED team as well as specialty doctors. Also involves providing personal feedback, debfriefing and teaching to facilitate learning.
	Conflict Solving	Suggests conflict solutions. Involves choosing an appropriate time and place to discuss disagreements. May overlap with óAuthority & Assertivenessô.
Decision making	**Option Generation**	Using all resources (written and verbal) to gather information and generate appropriate options for a given problem or task. (NB óSituation Awarenessô involves **identifying** problems). Includes consulting GP, carers, ambulance sheet and inpatient notes as well as understanding where to find hospital/ national guidelines. Also involves surveying whole department and using multiple sources of information when considering reallocating staff.
	Selecting & Communicating Options	Considering risks of various options and discussing this with the team. Involves clearly stating decision and explaining reasons, if necessary. Differs from Sit Awareness óUnderstanding & Communicatingô in that action is expected.
	Outcome Review	Once a decision has been made, reviews suitability in light of new information or change in circumstances. Confirms task has been done.
Situational Awareness	**Gathering Information**	Surveying the environment to pick up cues that may need action as well as requesting reports from others.
	Understanding & Communicating	Cross-checking information to ensure it is reliable. Communicating situation to the team to ensure a shared mental model and keep team óin the pictureô rather than expecting action.
	Anticipating	Anticipating potential issues with staffing or cubicle availability in the department.

5.4.9 The rating scale

Following discussion within the research team, a nine-point rating scale was chosen for the assessment tool. This was divided into Unacceptable (1-3), Acceptable (3-6) and Exemplary (7-9) as shown in Table 5.3

Table 5.3 Rating scale for the assessment tool

Unacceptable Standard Several examples of poor behaviour or behaviour that directly compromises patient safety			**Acceptable Standard** Performance was of a satisfactory standard with mostly good behaviour observed. Standard expected of a competent trainee.			**Exemplary Standard** Performance was of a consistently high standard. A model for other team members.		
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

A number of factors were considered. For example, there are published studies which confirms the theoretical argument that course scales with fewer points may be less reliable than fine scales (Jenkins, 1977). However, error may increase if scale points exceed a raterôs capacity to discriminate and Miller estimated this to be ñseven plus or minus twoò (Miller, 1956). More recently, a study comparing mini-CEX assessments using a 5-point scale and 9-point scale found similar inter-rater reliability results for the two scales (Cook, 2009). However, the 9-point scale appeared to provide more accurate scores for identifying unsatisfactory, satisfactory, and superior scripted levels of competence. Discussion within the research team also highlighted the advantage of using a scale that allowed discrimination between trainees that had marginally poor performance and those that required more serious remedial action.

5.4.10 Evaluation of feasibility

The provisional tool was tested to determine the feasibility of assessing non-technical skills in the ED. This involved conducting four observational episodes, each following the lead registrar for the shift and using the tool to asses their non-technical skills. This determined that a range of skills were observable within a 45-60 minute

period. During each session, a minimum of 10 out of the 13 individual skills were observed. Whilst it was challenging following the registrar around the various sections of the department (i.e., majors, minors and the resuscitation room), observing the registrar was deemed to be feasible. ED registrars found the experience acceptable and commented that they thought it was a good idea to get feedback on how well they were running the department. Based on these findings, subsequent studies were undertaken during Phase 2 to further develop the assessment tool and this is discussed in the following chapter.

6 Revision of the Assessment Tool

CHAPTER OVERVIEW

This chapter describes the use of staff interviews and field observations to refine the provisional assessment tool described in the previous chapter and to produce a second version of the tool. Incorporating the results of these studies ensured that the skills and behaviours included in the tool were not only important in theory but also in practice. A working version of the tool (Version 2) is presented and the chapter concludes with limitations of these studies and associated issues.

6.1 Background

Once the feasibility of assessing non-technical skills in the ED had been established, the provisional assessment tool was revised and adapted based on data gained from a further two studies described below. This was necessary as the provisional tool was an entirely theoretical framework amassed from the literature and it was essential to determine what was important in UK practice. A two-stage approach to first identify skills and then refine the list is recommended and has been used by other research teams (Flin, 2008, Fletcher et al., 2003, Yule et al., 2006). The two subsequent studies were also used to extract examples of good and poor behaviour that would help to illustrate the skills.

Flin (2008) mentions three methods for identifying non-technical skills: event based, questioning techniques and observational techniques - see Table 2.2. She also recommends that a combination of methods is used as no single technique provides a full picture of the relevant non-technical skills. The use of event based methods was incorporated by using the analysis of critical incidents and adverse events found in the literature review.

Questioning techniques includes interviews such as the critical decision method (Klein, 1989). This involves asking relevant staff to describe a critical incident they remember and then probing deeper to reveal key judgements, cues and errors. Whilst

this may elicit important information, it has some disadvantages. For example, interviewees may be unable to recall a relevant incident and a study revealed interviewees may find it difficult to clearly describe an incident where events become blurred due to high workload or pressure (Crandall, 2006). This may certainly be the case for incidents occurring in the ED. However, a review of the transcripts from the initial interview study described in Chapter 3 revealed that numerous non-technical skills were mentioned by ED staff. Interviewees were asked to describe behaviour that had a positive or negative effect on how the team functions and this often involved describing relevant non-technical skills such as assertiveness and communication. Staff were also questioned about the circumstances of an error or poor patient care, although not in the detail usually used for the critical decision method. Therefore, reanalysis of the interview transcripts was undertaken to identify key EM non-technical skills.

Observation may either involve direct observation, where the observer watches how tasks are performed and records behaviour in the workplace, or remote observation, e.g. from video tape (Stanton, 2005). As previously mentioned, observation in the ED may be particularly challenging and a decision has to be made whether to follow the patient, a staff member or remain in one geographical area. Whilst video taping has been used in the ED with some success (Michaelson, 1997, Cooper, 1999), this has been predominantly used for the study of resuscitation teams. The assessment tool under development was aimed for use in the whole ED including the resuscitation room and videoing such a large area would have bee impractical. Therefore, direct observation following the lead registrar was chosen.

The assessment tool was designed to summarise the non-technical skills required for a UK emergency medicine trainee, with a particular focus on those skills required when leading the shift. As discussed in the previous chapter, leadership was not observed as a single skill but comprised a number of skills. Therefore, leadership was not coded separately and subsequent data analysis was based on 12 skills rather than the original 13 skills shown in Box 5.2.

6.2 Study 1 – Recoding Interviews

6.2.1 Aims

The aims were:

1. To determine if there were any skills that ED staff reported in the interviews which could be added to the provisional tool.
2. To identify behavioural markers linked to non-technical skills in the context of the ED to further illustrate the skill definitions.

6.2.2 Methods

The interview transcripts of the 22 participants described in Chapter 3 were recoded using Nvivo to identify non-technical skills used by staff in the ED. The coding framework was based on 12 main non-technical skills, shown in Box 6.1, that had been used for previous collection of data (see Appendix C for coding framework).

1. Maintaining standards
2. Decision making
3. Workload management
4. Communication
5. Conflict resolution
6. Team building
7. Support of others
8. Teaching and feedback
9. Authority and assertiveness
10. Situational awareness
11. Coordinating team members
12. Assessing capabilities/ supervision

Box 6.1 12 non-technical skills used for coding interviews & observations

6.2.3 Results

The main findings from the interview study are given below. Each of the 12 skills are summarised and quotes from participants are given. This is followed by a decision whether this supports inclusion in the revised assessment tool.

1. Maintenance of Standards

Eleven (50%) participants mentioned the importance of maintaining patient safety and delivering quality care. This was usually stated in reference to the competing pressures of the four hour target and delivering safe and appropriate patient care. Six interviewees commented on need to keep accurate and contemporaneous notes. Other comments involved maintaining standards with respect to drug administration and reviewing x-rays (e.g. inadequate x-rays, not checking the name etc). One registrar described an incident on the Clinical Decision Unit and recalled that the protocol which states patients should be seen by a senior doctor prior to transfer had not been followed.

Decision:	Supports inclusion

2. Decision Making

a) Option Generation

Participantsô description of decision making generally centred on either patient management or overload in the department. One of the registrarôs primary leadership responsibilities is to supervise and assist junior doctors in their decision making. One senior doctor commented that it is good practice to see the patient in person if asked a question by a junior doctor. Staff also commented on the use of teamwork to make decisions, for example, ñYou can admit to your juniors, actually I donôt know, what are your thoughts?ò[R1] and ñA fresh pair of eyes can change everythingé because when Iôm absolutely exhausted I donôt actually think appropriatelyò [N4]. When the department is overcrowded with long waiting times, staff solutions included utilising

speciality doctorsô goodwill to help, contacting bed managers and consultants, borrowing a doctor from the paediatric ED or enlisting the help of the GP working in the department.

b) Selecting and communicating options

Nursing staff described the difficulties of doctors who donôt communicate plans. For example, ñtheyôll clerk the patient, and theyôll put the notes in the trolley and not actually inform you of what the plan of action is and Iôll look through the notes an hour later and see theyôve written the plan, óAdmitôò[N6] and ñthey write up the prescription but they donôt tell anyone so it doesnôt get doneò[N5]. Two junior doctors mentioned the importance of giving reasons and explanations when asking someone to do something or disagreeing with a management plan. Five doctors commented that they felt senior staff interfered or changed management plans unnecessarily or without an adequate explanation. Staff also commented on the importance of the registrar being decisive.

c) Outcome review

Outcome review was occasionally mentioned. Three senior team members described telling staff to report back after seeing a patient; and one reported that they often checked that a nurse had initiated treatment.

Decision:	Supports inclusion

3. Workload Management

This was the most commonly described non-technical skill. Eleven staff described examples of delegation and nine participants explained how they practiced back-up behaviour. Staff agreed that when working under pressure it was more important to ñblur the boundariesò and help each other out. However, some junior staff felt that people should ñjust do their jobò. For example, ñI can dip urine but a nurse canôt see a patientò[L5]. This differing opinion appeared to be a cause of tension between staff. Three senior staff also described the importance of escalating when overloaded. Both junior and senior staff highlighted the importance of seniors helping colleagues to

boost team morale and to be seen ñleading from the frontò. Prioritisation and multi-tasking were mentioned as important skills. Workload management was also associated with errors, for example, ñé youôre involved with the coroner on the phone or something like that then all of a sudden two or three hours later you go round there [Clinical Decision Unit] and you find this person moribund in bed and you think what the hell happened to the last two or three hours. You havenôt made the time to go and see them..ò [R2]. Doctors described positive and negative examples where registrars tried to get juniors working faster. For example, one junior doctor said, ñI would have thought he would say, óoh is there a problem with this patientô, instead of just, ówhy are you still seeing this patient?ô ò[L2]. Senior staff described how, during busy times, they would allocated patients to SHOs or press for decisions to be made to get the patients moving. Both senior and junior staff thought this was a good thing if done the right way. For example, ñI think heôs great, heôs like, letôs sort this out, letôs get this done. He really marshals everyoneò[L5].

Decision:	Supports inclusion

4. Communication

Eighteen respondents referred to the value of communication for effective teamwork. Six of the 22 staff described the importance of being clear and concise and senior staff often commented they were more prescriptive and directive when speaking to team members. Miscommunication was cited as a contributor to error. For example, ñHe said, this patient needs Tazocin, so that patient got Tazocin, the one he was talking to. Unfortunately, he meant the other patientò[N6]. Another respondent described how poor handover between the ED team and medical in-patient team may have contributed to the death of a patient.

Decision:	Supports inclusion

5. Conflict Solving

Twelve respondents commented on conflict with staff members external to the ED team, either ward staff or speciality doctors. For example, ñRefusal to accept the referral, and there were examples of it during this morningsô ward round, is always a source of conflictò [R5], and ñlike a senior surgeon who comes down and tries to take overé Itôs balancing what youôre supposed to be doing with what their expectations are of what you should be doingé those kind of things do cause a bit of frictionò [R6].

Decision:	Supports inclusion

6. Team Building

This was mentioned by almost all interviewees as vital for smooth working in the ED. Staff described the negative effects of rude or unfriendly staff - ñwhen people lose their manners and maybe even raising voices, itôs probably the most counterproductive thing because it has a totally demoralising effect on the teamò[N3]. A sense of humour and having a positive attitude were important attributes for someone in charge to motivate and encourage staff. Many people stated that back-up behaviour not only improved efficiency but also had a strong influence on team morale. People mentioned the importance of having good relationships with speciality doctors, ward staff and ambulance crews, ñthey can make or break youò[N1].

Decision:	Supports inclusion

7. Support of Others

Although this was coded separately, most references related to back-up behaviour which was covered in ôWorkload Managementô. Two junior doctors commented on the importance of having the support of registrars or consultants when experiencing difficulty with a speciality doctor or radiologist. Cognitive support is included in Decision Making and no-one mentioned emotional support.

Decision:	No strong evidence to support inclusion

8. Teaching and feedback

Four staff members mentioned taking the opportunity to teach during periods of inactivity or low pressure. Four respondents described the importance of appropriate and fair feedback. One junior doctor said, ñI think it was fair of her to say, you should have referred him earlier, but she was quite difficult about it and rude, and in front of everyoneò [L1].

Decision:	No strong evidence to support inclusion

9. Authority and Assertiveness

Junior doctors (Foundation level or ST1) commented on the way that senior colleagues ñharassedò or ñgot intimidatingò when the department was busy and said it would have been more helpful if they were given help instead of just ñpressuredò. One said that this use of authority made them go slower; ñWe were much more inefficient because we were trying to think of what heôs going to ask us rather than how to finish quicklyò [L4]. Another junior doctor said that they knew of colleagues who would not ask certain senior staff a question because they found them unapproachable, which is clearly a safety issue. Interestingly, three senior physicians admitted they were sometimes ñwilling to put a nose out of jointò [R2] or ñget stroppyò [C1] to get something done but did not appear to appreciate the negative effects this may have on the team. One senior nurse said, ñI canôt be like that every day, or I could be but I wouldnôt get the job done. Itôs not as effectiveò [N2].

The important issue of authority gradients was highlighted when a junior was asked if they ever disagreed clinically with a registrar. He said, ñNo, theyôre the boss. I always do what they tell me to do.ò [L5]. Five interviewees mentioned that some senior staff appeared to ñlose controlò when under pressure. Six participants also commented on the positive effect of having someone calm and confident in charge. Four interviewees described errors because staff did not assert themselves. For example, a junior nurse felt there may have been a child protection issue and mentioned it to medical staff but did not speak up when a senior doctor discharged the

child. This highlights the need to balance being over and under assertive and also to remain calm under pressure.

Decision:	Supports inclusion

10. Situational Awareness

a) Gathering information

Registrars use a number of sources for gathering information about the state of the department. Two reported that the Patient Tracking System is vital for monitoring and the Nurse-in-charge is useful, not only for keeping the registrar updated on patients, but also for keeping an eye on the doctors. Two senior colleagues also describing ñeye-ballingò patients or casting an eye over patients waiting in cubicles and the waiting room to see if there was anyone that looked unwell, especially when waiting times are prolonged. Gathering information may also involve speaking to the bed manager or the nurse in charge of CDU for assessing the bed state. Although many ED staff may not be familiar with the term ósituational awarenessô these responses show that they are continually practicing this skill throughout a shift. Board rounds were mentioned by many staff as an important method of keeping up-to-date. Board rounds happen at specified times during the day and involve medical staff and the nurse-in-charge reviewing the management plan of each patient. Registrars also spoke to junior doctors individually to gather updates although one doctor perceived this as ñchecking upò. Clinical aspects of situational awareness, such as patient monitoring, were not discussed.

b) Understanding and communicating options

The concept of communicating situational awareness generally focussed on whether junior staff are aware of how busy the department is, for example, ñI honestly donôt believe that your junior members of the team necessarily have either the insight or the capability to decide for themselves itôs a busy day so weôre going to work differentlyò[N3]. In fact one junior doctor said they actively avoided keeping an eye on the wait times and said ñI try not to look at how many patients there areé If thereôs ten patients on the board in majors, itôs not going to make any differenceé Iôm

still going to see the patient the same amountò[L2]. In contrast, another doctor said, ñThey say, óDo you know that thereôs a two hour, three hour wait?ô I know, of course yeahò[L5]. This shows that while it is vital to communicate situational awareness, a culture must exist where this is perceived as supportive rather than as harassment.

c) Anticipation

The skill óAnticipationô usually involved planning ahead to book a bed for a possible admission. Staff also mentioned being proactive and anticipating problems. One staff member described a successful registrar - ñtheyôre more active probably, rather than waiting for doctors to come up to them with problems, they are seeking them out before they become big problems or theyôre seeking out people who they know need more supportò[N1].

Decision:	Supports inclusion

11. Coordinating the Team

Fifteen respondents commented on the importance of senior staff being able to coordinate the team effectively. This mainly involved moving staff between areas in the department (e.g. from Majors to Minors) and organising breaks.

Decision:	Supports inclusion

12. Assessing Capabilities

This was generally referred to when describing allocation of nurses rather than doctors. Two doctors mentioned the importance of adapting your level of supervision according to the juniorôs capabilities.

Decision:	No strong evidence to support inclusion

6.2.4 Summary of Findings

The table below shows a summary of the findings from recoding of the interview study. It states the number of participants who mentioned the importance of the skill and whether there was strong or weak evidence to support inclusion in the revised assessment tool.

Table 6.1 Summary of findings from the interview study

Skill	No. of participants who mentioned skill	Interviews support inclusion	Weak evidence to support inclusion
1. Maintenance of standards	11	•	
2. Decision making	20	•	
3. Workload management	23	•	
4. Communication	18	•	
5. Conflict resolution	12	•	
6. Team building	22	•	
7. Support of others	2		•
8. Teaching and feedback	8		•
9. Authority and Assertiveness	17	•	
10. Situational awareness	21	•	
11. Coordinating the team	15	•	
12. Assessing capabilities	2		•

6.3 Study 2 - Field observations

6.3.1 Aims

The aims of the study were:

1. To ensure the skills in the assessment tool were observable and sufficiently common to warrant inclusion.
2. To extract examples of positive and negative behaviour that would be used to further illustrate the skill definitions.

6.3.2 Methods

Design and Setting

Direct observations of the lead ED registrar were conducted. Observations were undertaken in the departments of St Maryôs Hospital, Paddington, and Charing Cross Hospital, London.

Procedure

Participants were notified, via email, of the general purpose of the study and were given the opportunity to be excluded if they wished. They were told that neither the individual participant, nor the consultants, would be given any feedback about performance. Verbal consent was sought prior to each observation episode. Verbal consent was also sought from the nurse-in-charge for each shift and other staff members were informed as much as was practical. The observations were all carried out by the lead investigator and involved following the coordinating (or ôbleep-holderô) registrar for approximately an hour. Extensive field notes were recorded as the lead registrar was observed going about their usual clinical duties. The notes provided details of observed behaviour and communication events as well as documenting any perceived errors or examples of substandard practice. Participants

were informally questioned about how they felt being observed and whether they thought it changed their practice. The field notes were then typed and analysed using the same coding framework as that used for the interview study above (see Appendix C).

6.3.3 Results

A total of 20 observational episodes were completed across the two sites (10 in each department) lasting between 45 and 70 minutes. Thirteen registrars were observed during the course of the study and this included both training grade doctors and non-training grade, such as clinical fellows. A sample of field notes taken is given in Appendix C.

Each skill will be described individually and followed by a summary statement on whether this supports inclusion in the revised assessment tool.

1. Maintaining Standards

This was a commonly observed skill with clear implications for safety. For example, ñYou canôt send that patient home until we get the U&E blood results backò and ñYou canôt see down to T1 on this c-spine x-ray. Itôs inadequateò. The registrar was also seen regularly referring doctors to clinical guidelines (or checking themselves) to ensure the correct patient management plan was followed. Maintenance of standards included observing doctors washing hands between patients. Some examples of poor standards included illegible handwriting, not putting a patientôs name on an ECG and acknowledging but not acting on a patient monitor alarm in an adjacent cubicle.

Decision:	Supports inclusion

2. Decision Making

a) Option generation

Registrars were observed encouraging the use of multiple sources of information before deciding on a patientôs management plan. This included obtaining hospital notes, old x-rays, previous ECGs, contacting the patientôs GP and accessing internet resources, such as Toxbase. Registrars also took advice from consultants, other registrars, radiologists, physiotherapists and occupational therapists. Some registrars were particularly skilful at asking the necessary probing questions of junior team members to get information in an efficient way. In addition to patient management, examples of Option Generation included calling the surgical registrar to come down to help with an excessive wait for the surgeons and liaising with the nurse in charge over cubicle availability problems.

b) Selecting and communicating options

Registrars were routinely seen making decisions and following up with reasoning for the decision. Staff generally appeared to be decisive although decisions were occasionally deferred until the ward round. Registrars were very directive at times, for example, telling a junior doctor exactly what they needed to do for a patient and what to say when referring to the surgeons. There was also evidence of decisions not being communicated effectively. For example, forgetting to tell the nurse in charge about an admission and so subsequently the hospital bed was not booked.

c) Outcome review

This was usually observed when a registrar asked a junior about a patient they had previously discussed to check if they had been successfully referred or to enquire whether a patientôs condition had improved. Other examples included asking a member of the inpatient team the outcome of a patient for their own personal interest or learning.

Decision:	Supports inclusion

3. Workload Management

This was a very common skill and often involved the registrar monitoring team workload and improving efficiency. For example, ñif youôre waiting for bloods can you go and see Mr T for me?ò It was sometimes difficult to separate skills, for example, ñHave you finished seeing Mrs B yet?ò could be seen as both situational awareness and workload management. Examples of delegation (ñCan you do these bloods?ò) and back-up behaviour (answers phone for nurse in charge or gets own medication) were also commonly observed. There was one example of ineffective workload management when the department was very busy and an inexperienced registrar said ñWe need to get some movementò to an empty room. Whilst this was probably óthinking aloudô, it was not followed up by any particular action in an attempt to solve the problem. There were a few times when registrars were seen multi-tasking (e.g. writing in patientsô notes whilst listening to someone) but it is difficult to assess if this is good or bad behaviour, i.e. efficiency versus safety. Also, prioritising is an important skill for registrars; however, this is not easily observable as it is largely a cognitive process.

Decision:	Supports inclusion

4. Communication

Although communication was a skill excluded from the provisional framework, the field observations revealed important episodes of communication that would not necessarily be coded elsewhere, for example, speaking to a radiologist to request an investigation or making a referral to specialty doctors. This looked at the quality of communication events, such as concise handover, correct use of terminology, speaking clearly and loudly enough to be heard. Many other skills in the assessment tool also rely on communication, for example, Situational Awareness ï Understanding & Communicating and Decision making ï Selecting & Communicating Options. However, these skills generally assess whether or not the event occurred (i.e. did the doctor communicate the need to book an inpatient bed) rather than the quality of the communication. Observations identified episodes of miscommunication (e.g. unclear

instructions) and examples where good communication avoided an error (e.g. closed-loop communication or use of clarification). For example, when a doctor gave an incorrectly dosed verbal drug order, the nurse repeated it back and the doctor subsequently noticed the error. On another occasion, a nurse asked for clarification when a doctor asked for a non-standard dose of a drug and the doctor explained that the patient had already received a small dose prior to arrival in the ED.

Decision:	Supports inclusion

5. Conflict Solving

This was rarely observed. Some episodes that may be considered conflict solving were captured under Authority and Assertiveness (see below).

Decision:	No strong evidence to support inclusion

6. Team building

Team building was not observed as a specific set of behaviours but rather the overall approach of the doctor. Some doctors clearly had a very good relationship not only with the ED team but also with specialty doctors, paramedics, etc. It was noted that some registrars were particularly approachable and friendly to junior doctors who were requesting advice or help. Other examples included staff saying thank you and the use of humour. Obviously, this may be strongly influenced by the Hawthorne Effect as staff may behave differently because they are being observed.

Decision:	Borderline evidence to support inclusion

7. Support of Others

This element was included in both the ANTS and NOTSS taxonomy but in this observational study examples of this skill were mainly coded elsewhere. For example, back-up behaviour was assessed in ôWorkload Managementô. Other behaviours such as providing reassurance and thanks were covered in ôTeam Buildingô.

Decision:	No strong evidence to support inclusion

8. Teaching and feedback

Only eleven of the 20 observations had episodes of teaching, which were aimed at either junior doctors, medical students or nurses. These were mainly brief encounters, for example, explaining an abnormality on an ECG. No feedback was observed (positive or negative).

Decision:	Borderline evidence to support inclusion

9. Authority and Assertiveness

There were many situations where registrars displayed the skill Authority and Assertiveness. In general this was appropriately used. For example, to get something done in a timely manner or to impress on staff the urgency of the situation. However, there was one episode where a junior doctor appeared upset about the way she had been spoken to. Several situations were observed where the use of authority appeared to have safety implications, for example, a registrar reiterating to the nurse in charge that a cubicle was required urgently for a patient with a cardiac history and ongoing chest pain. Assertiveness by staff who needed to deal firmly with interruptions was also observed.

Decision:	Supports inclusion

10. Situational Awareness

a) Gathering information

Much of the lead registrarsô time was spent on this skill. This included using the Patient Tracking System to keep an eye on various areas in department or looking at the triage notes of individual patients. Registrars also displayed situational awareness by running through the list of Majorôs patients with the nurse in charge or asking a doctor for an update on their patient. Registrars were seen checking that all the doctors had turned up for the shift or simply overhearing other team membersô conversations. One example involved listening to the paramedic handover to the nurse-in-charge to pick up if a sick patient had been brought in. This œeavesdroppingô was a particularly noticeable behaviour in some registrars, often the more senior doctors.

b) Communicating

This commonly involved keeping the nurse in charge updated on patient activity or informing the nurse looking after an individual patient what was happening. There were also examples where staff failed to keep team members updated on the situation. This differed from Decision making ï Selecting and communicating where plans were communicated because action was required rather than simply to keep team members in the picture.

c) Anticipating

This skill was evident when staff averted the ED 4 hour breach by anticipating delays, for example, when needing a specialty review or a side room. Registrars also planned ahead to ensure cubicle availability by moving patients early to the CDU. Staff were seen trying to contact a particular team before the end of the day when they suspected a patient may need a review.

Decision:	Supports inclusion

11. Coordinating the Team

Most examples of coordinating staff involved moving staff between majors and minors. However, this was almost exclusively observed in St Maryôs ED as the ED at Charing Cross Hospital has a completely separate and independently staffed Urgent Care Centre. Whilst coordinating staff is an important and potentially difficult skill for those departments that require movement of staff between areas, this could be assessed in óDecision Makingô (i.e. if there is a particularly long wait in one area of the department). Coordinating staff was also observed when organising breaks.

Decision:	No strong evidence to support inclusion

12. Assessing Capabilities

Only three clear episodes of assessing capabilities were observed. These involved checking a junior was confident or competent doing a procedure. Several times the registrar was heard to say ñAre you all right?ò However, it is hard to say if this is assessing capabilities, situational awareness, support of others or team building. It is likely that the registrar was constantly using this skill, for example, to adapt the level of supervision according to the ability of the junior doctor. However, this is predominantly a cognitive process and, therefore, not easily observable.

Decision:	No strong evidence to support inclusion

6.3.4 Summary of findings

The table below shows a summary of findings from observations in the ED and indicates whether the study provided strong, borderline or weak evidence for inclusion in the revised assessment tool.

Table 6.2 Summary of findings from observations in the ED

Skill	Strong evidence from observations	Weak evidence from observations	Borderline evidence
1. Maintenance of standards	•		
2. Decision making	•		
3. Workload management	•		
4. Communication	•		
5. Conflict resolution		•	
6. Team building			•
7. Support of others		•	
8. Teaching and feedback			•
9. Authority and assertiveness	•		
10. Situational awareness	•		
11. Coordinating the team		•	
12. Assessing capabilities		•	

6.4 Inclusion or exclusion of skills for Version 2 of the assessment tool

The final decision whether to include the skill in the revised tool was determined by the research team reaching a consensus. The research group consisted of 1 ED consultant, 2 ED registrars with a background in patient safety and 2 human factors experts. The group met on two occasions and using the data gathered from the two studies, in addition to evidence from published literature, each skill was again

discussed and reconsidered for inclusion. Factors taken into consideration included: (i) the weight of empirical evidence available linking the skill to safety in general and more specifically to EM, (iii) whether the skill was adequately assessed using current methods of assessment, and (iii) how observable the skill was (both frequency of observation and the degree hidden cognitive processes are involved). A summary of the two studies which shows which skills are supported for inclusion in the revised framework is shown in Table 6.3.

Table 6.3 Summary of skills supported for inclusion in the revised tool

Skill	Interview	Observations
Maintenance of standards	•	•
Decision making	•	•
Workload management	•	•
Exchanging information	•	•
Conflict resolution	•	
Team building	•	○
Support of others		
Teaching and feedback		○
Authority & assertiveness	•	•
Situation awareness	•	•
Coordinating the team	•	
Assessing capabilities/ supervision		

- • Strong evidence for inclusion
- ○ Borderline evidence for inclusion

Reasons for excluding skills included (i) the skill was observed infrequently and there was no strong evidence for inclusion from other sources, i.e. literature or interview study (for example, there was insufficient evidence for Conflict Resolution as an independent skill so it was included in the definition of óAuthority and Assertivenessô), or (ii) the behaviours observed or mentioned in the interviews already fell within the definition of another skill group (for example, evidence of

ôCoordinating the teamô involved behaviours that were coded in ôWorkload Managementô).

Two further skills were considered for exclusion after analysing results from the observational study; ôTeaching and Feedbackô and ôAssessing capabilities/ Supervisionô. Whilst brief episodes of teaching were observed, feedback (positive or negative) to team members was rarely seen during the observations. The literature describes the vital importance of feedback for improving individualsô performance and the distinct lack of this observed behaviour may actually suggest the need for its inclusion. Furthermore, the importance of óshop floorô apprenticeship teaching has been well established and registrars rarely receive feedback on this type of teaching skill. Supervision and assessing capabilities involves predominantly cognitive processes so this skill was difficult to observe. However, there is a strong association between failed supervision and error in the ED so these skills were merged into a single element ôSupervision and Feedbackô, and teaching was encompassed in the skill definition. ôQuality of Communicationô was added in order to assess important communication events such as handover that would be difficult to rate elsewhere.

Once the skills set had been decided coded interview transcripts and field notes were reviewed to determine the behavioural markers linked to each skill included in the assessment tool. Samples were extracted by the lead researcher and subsequently discussed and agreed by the research team. The final list of skills and corresponding definitions are shown in Table 6.4 below.

Table 6.4 Categories and skills with associated definitions

Category and Skill	Skill Definition
MANAGEMENT AND SUPERVISION	
Maintenance of Standards	Subscribes to clinical and safety standards as well as considering performance targets. Monitors compliance.
Workload Management	Manages own and others' workload to avoid both under and over-activity. Includes prioritising, delegating, asking for help and offering assistance.
Supervision and Feedback	Assesses capabilities and identifies knowledge gaps. Provides opportunities for teaching and constructive feedback.
TEAMWORK & COOPERATION	
Team Building	Provides motivation and support for the team. Appears friendly and approachable.
Quality of Communication	Gives verbal and written information concisely and effectively. Listens, acknowledges receipt of information and clarifies when necessary.
Authority & Assertiveness	Behaves in an appropriately forceful manner and speaks up when necessary. Resolves conflict effectively and remains calm when under pressure.
DECISION MAKING [3]	
Option Generation	Uses all resources (written and verbal) to gather information and generate appropriate options for a given problem or task. Involves team members in the decision making process.
Selecting & Communicating Options	Considers risks of various options and discusses this with the team. Involves clearly stating decisions and explaining reasons, if necessary.
Outcome Review	Once a decision has been made, reviews suitability in light of new information or change in circumstances and considers new options. Confirms tasks have been done.
SITUATIONAL AWARENESS[3]	
Gathering Information	Surveys the environment to pick up cues that may need action as well as requesting reports from others.
Anticipating	Anticipates potential problems such as patient clinical deterioration or departmental issues and discusses contingencies.
Informing the Team	Cross-checks information to ensure it is reliable. Communicates situation to keep team 'in the picture' rather than just expecting action.

[3]Decision Making and Situational Awareness have been divided into three separate sub-skills

6.5 Discussion

Staff interviews and a series of preliminary observations provide a rich source of data for determining the salient non-technical skills for EM physicians. These studies reveal the importance of using triangulation of data from a number of sources to produce a complete list of observable skills. For example, whilst conflict resolution featured strongly in the interview study, it was observed rarely in practice. Similarly, the published literature gives weight to the importance of providing feedback but this rarely observed and including this in the tool may promote more effective use of feedback in the ED.

It is important to consider the limitations of any behavioural marker system, including that: (i) not all aspects of performance will be captured in limited periods of observation; (ii) some behaviours may be important but are observed infrequently, e.g. conflict resolution; and (iii) the system relies on human observers who have their own limitations, e.g. bias, distraction, overload (Flin, 2008). This assessment tool is based on evidence from a number of different sources which helps to counteract some of the limitations of each research method alone. For example, the interview study measures peoplesô attitudes and self-reported behaviour which may be different from their actual behaviour. Observational studies look at behaviour in a real life setting but single observations may be unrepresentative of a participantôs normal behaviour.

The interview topic guide was not designed to specifically identify non-technical skills and this may have affected the responses given. However, it is not practical to simply ask interviewees which non-technical skills they feel are most important as many staff were unfamiliar with the concepts involved. Therefore, an indirect method of questioning was required and asking questions related to staff interactions under pressure is a suitable alternative. The interviews revealed that only 12 of the 23 respondents could recall a suitable incident or error, and the details surrounding the event were often sparse, so the critical decision method may not have been a viable option. Furthermore, the critical decision method focuses entirely on safety aspects and makes no attempt to identify skills that are related to efficiency, which is an integral part of the definition of non-technical skills.

Obviously any observational research should consider the potential impact of the Hawthorne effect. This is the tendency for research participants to change their behaviour as a result of their awareness of being studied. There is some evidence that the Hawthorne effect may be minimal or even absent in healthcare research (Schnelle, 2006, Hohenhaus et al., 2008, Niebuhr et al., 1981), possibly because staff are highly focussed on the task and are likely to be too busy to be distracted by observers (Niebuhr et al., 1981). Even so, whilst the effect may have influenced staff to have performed better (or possibly worse) it is unlikely this resulted in a different range of skills being demonstrated.

Many observational studies utilise experts in human factors and although they have training in observational methods most have limited clinical expertise. The advantage of this study was that the sole investigator had knowledge of observational techniques as well as substantial experience as an ED physician. This was vital to understand the complex workings of the ED. Also the registrars appeared to be relaxed and comfortable being observed by ósone of their ownô which made the observations less threatening and intrusive. However, there is an ongoing debate as to whether medical or human factors experts make the most appropriate observers (Carthey, 2003). Some studies use both to gain a richer dataset e.g. surgical studies (Healey et al., 2004).

The interview study and observations were carried out in two EDs within the central London area and it could be argued that this may not be representative of other EDs. However, the two departments used in the studies were very different from each other. For example, one contained both Majors and Minors areas whilst the other had no Minors and instead had a completely separate Urgent Care Centre. Also, one department had a significantly higher consultant presence on the shop floor during the observational periods. Both of these factors are likely to influence the roles and responsibilities of the registrar. Furthermore, subsequent validation studies have incorporated the views and observations of clinicians from across the country.

Triangulation of data from the literature review, staff interviews and field observations enabled the development of an assessment tool to assess ED registrarsô non-technical skills. The next step was to undertake a thorough evaluation of the tool to determine validity, reliability, and acceptability. The following chapters describe a

series of evaluative studies, starting with assessment of content validity, using a survey of expert opinion.

7 Validation Questionnaire

CHAPTER OVERVIEW

This chapter describes the evaluation of the assessment toolôs content validity using a survey of EM consultants, trainees and senior nurses. The findings of the study enabled revision of the tool to ensure a full range of non-technical skills and behaviours were encompassed. The survey explored staff attitudes to assessment of non-technical skills in the workplace and the chapter concludes with consideration of limitations of the study.

7.1 Background

Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose (Haynes, 1995). Demonstrating content validity is an essential element in the development of an assessment tool (Eignor, 2001). Quantitative measures of content validity traditionally use a panel of ósubject matter expertsô to rate items for relevance (Lawshe, 1975). This enables identification of possible irrelevant items on an assessment tool. Grant and Davis suggest that in addition to this type of óitem contentô evaluation, panel experts should also comment on óitem styleô to ensure item wording in clear and unambiguous, and ócomprehensivenessô to ensure all dimensions of the content domain are included (Grant and Davis, 1997). When experts do not agree or when the panel identifies missing domain areas, the instrument should be revised (Lynn, 1986).

The initial interview study with EM staff highlighted that non-technical skills such as communication and assertiveness are important for ED registrars. The safety literature also supports the need for a focus on these skills. This was the basis for deciding to develop a tool for the assessment of non-technical skills in the ED. However, it is also important to seek the opinion of ED staff to determine if they feel structured observational assessment is necessary and to ask trainees if they would like feedback about their leadership and non-technical skills. Also, previous studies were

limited to two London EDôs and this may not be representative of other departments in the country so surveying the opinion of a wider audience was necessary. Therefore, this study was designed to assess the toolôs content validity and comprehensiveness, as well as survey staff attitudes towards this type of assessment.

7.2 Aims

The aims of the study were threefold:

- To quantitatively assess the content validity of exemplar behaviours used in the assessment tool
- To identify any significant omissions from the list of non-technical skills
- To explore Emergency staff attitudes regarding observational assessment of ED registrarsô non-technical skills.

7.3 Methods

7.3.1 Design

A survey design, in the form of a questionnaire, was chosen as this enables collection of data from a large, geographically dispersed sample of the target population. The initial studies used to inform the development of the assessment tool were based in two London teaching hospitals and it was important to survey the opinions of staff from a variety of departments and backgrounds.

7.3.2 Participants and Procedure

A purposive sampling method was used to survey a minimum of 100 Emergency staff from around the UK. There is no agreed number of experts needed to establish content validity (Polit, 2007). However, a larger sample reduces the effect of agreement due to chance which is discussed further in ôData Analysisô. Furthermore,

Emergency Departments, and the role of the ED registrar, differ widely across the UK so a large sample ensured this variation was taken into account. Larger samples also reduce bias introduced by the selection process. The participants included EM consultants and registrars and senior emergency nurses. Although the participants were not experts in non-technical skills they could be considered ósubject matter expertsô in that they are capable of determining what skills and behaviours are essential for a competent ED registrar.

Selection of participants to take part in the study involved the lead investigator travelling to various training days and academic meetings attended by EM physicians in the south of England. Researchers have commented that it is beneficial to select participants from different geographic locations (Grant, 1990). Therefore, the validation survey was also presented at a patient safety study day held at the College of Emergency Medicine and attended by EM physicians and nurses from around the UK.

It was presumed that whilst many participants would be able to identify the behaviours and skills requisite for a competent ED registrar, many would be unfamiliar with the term ñnon-technical skillsò. As one of the aims was to ensure that there were no significant omissions from the list of non-technical skills, the survey was distributed to groups after a short presentation focussing on non-technical skills and safety. Researchers have commented that providing participants with a conceptual basis for the instrument is essential for accurate content review (Grant and Davis, 1997). Questionnaires were completed anonymously and subsequently analysed by the research team.

7.3.3 Materials

The materials comprised of a 3 page questionnaire (see Appendix D). Participants were asked to read the skill descriptions and exemplar behaviours and rate each behaviour according to how important they are for an ED registrar, on a scale of 1-5 (see Appendix D). Only examples of good behaviour were used in the questionnaire; 3 behaviours for each of the 12 skills (total 36 behaviours). This was chosen because

the assessment tool lists a total of 72 good and poor behaviours and this would have been too time-consuming for staff to complete. Participants were also encouraged to annotate the list of skills and behaviours with any comments or feedback, such as ambiguous or unsuitable wording.

The second part of the questionnaire invited respondents to give suggestions for skills or behaviours that they believed were important and had not been listed. Three further questions were used to gauge opinion about the value of direct observation for the assessment of registrarsô non-technical skills in the ED. Responses were free text, rather than yes/no, to encourage participants to add additional comments.

7.3.5 Data analysis

Content validity was measured by calculating the Content Validity Index (CVI) for each item on the questionnaire. This is the number of experts giving a rating of 4 or 5 (ôVery importantô or ôEssentialô) divided by the total number of experts ï i.e. the proportion in agreement about relevance (Martuza, 1977). The CVI of the overall questionnaire was calculated by taking the mean of all 36 item CVIôs.

This method was chosen after considering a number of factors. Firstly, CVI is simple to compute and interpreting the result of óproportion in agreement about relevanceô is intuitive. It is also important to consider the probability of chance agreement. Some researchers have suggested using Fleiss Kappa (i.e. multi-rater kappa coefficient) which corrects for chance (Wynd, 2003). However, this index calculates the extent to which experts reach *any* type of agreement, not the extent to which there is consensus about *relevance* of an item (i.e. the participants could all agree an item is irrelevant, giving a high kappa value, but that would not indicate that an item should remain in the scale). CVI does not correct for chance but as the number of respondents increases, the probability of chance agreement reduces. For example, with a sample of 9 experts, the probability that raters would all agree on relevance of an item is 0.002, or 2 in 1000 (Polit, 2007). Based on this, it was decided to use CVI and sample as large a population as possible. There is no consensus on what CVI value is acceptable for establishing content validity of an item and it depends on the size of the

sample (a smaller sample requires a higher CVI value). However, Lynnôs widely cited guidelines state a CVI of 0.75 would be considered ñexcellentò with 16 experts (Lynn, 1986). This translates to 75% of participants rating the item óvery importantô or éssentialô and was chosen as the cut-off for the purposes of this study.

The free text questions were coded according to themes that emerged from the data rather than a predetermined coding framework. Calculations of CVI and coding were completed by the lead researcher. The research team, comprising of 3 EM physicians and one human factors expert, reviewed the results to determine how problem items should be revised.

7.4 Results

In total, 148 questionnaires were completed and comprised EM consultants, registrar trainees (SpR/ ST 4-5), non-training senior clinical fellows and senior nurses (see Table 6.1 below).

Table 7.1 Participants in validation survey

Job	No.	Percent
EM consultant	65	44
EM trainee	53	36
EM middle grade	7	5
EM senior nurse	23	15
Total	148	100

7.4.1 Assessing content validity

Of the 36 behaviours included in the survey, nine items failed to achieve a CVI Ó0.75 and were therefore revised (see Appendix D). Also two behaviours had significant missing data (9 and 7 missing values compared to 1-4 missing for all other

statements) so these statements were also reviewed. The overall CVI for the questionnaire was 0.81.

The highest scoring statements, both with a CVI of 0.98, were ôRemains calm under pressureô and ôExplains importance of ensuring sick patient is stable rather than moving patient for organisational reasonsô. The lowest scoring items were ôPractices back-up behaviour when nurse is busy e.g. does a set of observationsô (CVI 0.55) and ôUpdates team about waiting timeô (CVI 0.52).

Participants were encouraged to annotate the questionnaire and these comments were used to help revise problem items (either low CVI or significant missing data). For example, three participants noted that the ôAnticipatingô behaviours were considered to be more of a nursing role than physician. Two participants commented that the term ôreprimandô (ôReprimands doctor for illegible notes...ô) was too severe and this was one of the lowest scoring behaviours. One participant remarked that it is not always appropriate to disagree with senior colleagues (ôStates differing opinion to senior colleague regarding patient management planô) although it may be essential if patient safety is at risk. This potential contradiction may have explained the low response rate on this item. Other comments related to the widespread use of alcohol gel (ôWashes hands between patientsô) and the lack of Minors and Majors areas in many departments now (these terms were used in ôAnticipatingô behaviour statements). Two participants noted that working faster does not always maintain safety and quality (ôMotivates team to work faster during periods of high workloadô) and this may explain the low CVI.

7.4.2 Identifying major omissions in the tool

Participants were asked to add any relevant non-technical skill or behaviour not listed in the questionnaire and a total of 101 suggestions were made (some respondents listed more than one skill or behaviour). The comments were reviewed to determine if these represented any major omission in the tool. The research team concluded that all the relevant statements could be categorised into the existing 12 non-technical skills. For example, 7 participants mentioned the importance of trainees taking breaks

as well as ensuring other team members take a break. This was incorporated into the skill ôWorkload Managementô and did not reflect the omission of a major skill.

Participant suggestions were used to improve or replace low scoring behavioural statements. These included comments on clear handover (2 participants), team leadership in resuscitations (n=3), handling interruptions (n=2), being aware of own limitations or admitting they donôt know (n=2), ensuring prescribed treatment is given in an appropriate time frame (n=1), communicating using SBAR (n=1), and identifying a junior doctor in difficulty (n=1).

Other comments reflected good practice but were not considered non-technical skills. For example, seven participants mentioned empathy and four wrote compassion (also humility, honesty, integrity and sense of humour). Three respondents made comments about keeping patients and relatives informed which would not be considered a non-technical skill, although involving patients and carers in decision-making was included in one of the behavioural statements. Some comments were disregarded because they were too vague, for example, ópromotes team workô or órisk managementô. Others were discounted because there are not readily observable because the behaviour is infrequent (e.g. two participants mentioned debriefing after unsuccessful resuscitation) or involved cognitive processes (e.g. óability to reflect on own practiceô). The majority of other comments related to activities already listed in the exemplar behaviours (e.g. óawareness of when stressedô = óremains calm under pressureô and ómaturity in taking advice from nursing staffô = ólistens to team memberôs inputô) or skill definitions (e.g. conflict resolution and prioritising). A copy of the final version of the assessment tool is shown in Figure 7.1 below.

Figure 7.1 Assessment form Assessor Trainee Date

	Element		Rating	Observations
Management & Supervision	**Maintenance of Standards**	Subscribes to clinical and safety standards as well as considering performance targets. Monitors compliance.		
	Workload Management	Manages own and others' workload to avoid both under and over-activity. Includes prioritising, delegating, asking for help and offering assistance.		
	Supervision and Feedback	Assesses capabilities and identifies knowledge gaps. Provides opportunities for teaching and constructive feedback.		
Teamwork & Cooperation	**Team Building**	Provides motivation and support for the team. Appears friendly and approachable.		
	Quality of Communication	Gives verbal and written information concisely and effectively. Listens, acknowledges receipt of information and clarifies when necessary.		
	Authority & Assertiveness	Behaves in an appropriately forceful manner and speaks up when necessary. Resolves conflict effectively and remains calm when under pressure.		
Decision-Making	**Option Generation**	Uses all resources (written and verbal) to gather information and generate appropriate options for a given problem or task. Involves team members in the decision making process.		
	Selecting & Communicating Options	Considers risks of various options and discusses this with the team. Involves clearly stating decisions and explaining reasons, if necessary.		
	Outcome Review	Once a decision has been made, reviews suitability in light of new information or change in circumstances and considers new options. Confirms tasks have been done.		
Situational Awareness	**Gathering Information**	Surveys the environment to pick up cues that may need action as well as requesting reports from others.		
	Anticipating	Anticipates potential issues such as staffing or cubicle availability in the department and discusses contingencies.		
	Informing the Team	Cross-checks information to ensure it is reliable. Communicates situation to keep team 'in the picture' rather than just expecting action.		

		Examples of good behaviour	Example of poor behaviour
Management & Supervision	**Maintenance of Standards**	• Notices doctorôs illegible notes and explains the value of good note keeping • Ensures clinical guidelines are followed and appropriate pro forma is completed • Explains importance of ensuring sick patient is stable prior to transfer	• Fails to write contemporaneous notes • Does not wash hands (or use alcohol gel) after reviewing patient • Fails to adhere to clinical safety procedures
	Workload Management	• Sees a doctor has spent a long time with a patient and ascertains the reason • Deals with interruptions efficiently • Ensures both themselves and other team members take appropriate breaks	• Fails to act when a junior is overloaded and patient care is compromised • Focuses on one particular patient and loses control of the department • Fails to escalate appropriately when overloaded
	Supervision & Feedback	• Gives constructive criticism to team member • Takes the opportunity to teach whilst reviewing patient with junior doctor • Gives positive feedback to junior doctor who has made a difficult diagnosis	• Criticises a colleague in front of the team • Does not adequately supervise junior doctor with a sick patient • Fails to ask if junior doctor is confident doing a practical procedure unsupervised
Teamwork & Cooperation	**Team Building**	• Even when busy, reacts positively to a junior doctor asking for help • Says thank you at end of a difficult shift • Motivates team, especially during stressful periods	• Harasses team members rather than giving assistance or advice • Speaks abruptly to colleague who asks for help • Impolite when speaking to nursing staff
	Quality of Communication	• Gives an accurate and succinct handover of the department • Ensures important message is heard correctly • Gives clear referral to specialty doctor with reason for admission (e.g. SBAR)	• Uses unfamiliar abbreviations that require clarification • Repeatedly interrupts doctor who is presenting a patientôs history • Gives ambiguous instructions
	Authority & Assertiveness	• Uses appropriate degree of assertiveness when specialty refuses referral • Willing to speak up to senior staff when concerned • Remains calm under pressure	• Fails to persevere when specialty refuses appropriate referral • Shouts instructions to staff members when under pressure • Appears panicked and stressed
Decision making	**Option Generation**	• Goes to see patient to get more information when junior is unclear about history • Encourages team membersô input • Seeks help when unsure	• Does not look at previous ED notes/ old ECGs when necessary • Fails to listen to team members input for patient management • Fails to ensure all relevant information is available when advising referral
	Selecting & Communicating Options	• Verbalises consideration of risk when sending home patient • Discusses the contribution of false positive and false negative test results • Decisive when giving advice to junior doctors	• Uses CDU to avoid making treatment decisions • Alters junior doctorôs treatment plan without explanation • Forgets to notify nurse-in-charge of admission
	Outcome Review	• Reviews impact of treatment given to acutely sick patient • Follows up with doctor to see if preliminary plan needs revising • Ensures priority treatment has been given to patient	• Fails to establish referral outcome of complicated patient • Sticks rigidly to plan despite availability of new information • Fails to check that delegated task has been done
Situational Awareness	**Gathering Information**	• Uses Patient Tracking System appropriately to monitor state of the department • ŒEyeballsôpatients during long wait times to identify anyone who looks unwell • Notices doctor has not turned up for shift	• Fails to notice that patient is about to breach and no plan has been made • Ignores patient alarm alerting deterioration of vital signs • Fails to notice that CDU is full when arranging new transfers
	Anticipating	• Identifies busy triage area and anticipates increased demand • Discusses contingencies with nurse-in-charge during periods of overcrowding • Prepares trauma team for arrival of emergency patient	• Fails to anticipate and prepare for difficulties or complications during a practical procedure • Fails to ensure that breaks are planned to maintain safe staffing levels • Fails to anticipate and plan for clinical deterioration during patient transfer
	Informing the Team	• Updates team about new issues such as bed availability or staff shortages • Keeps nurse-in-charge up to date with plans for patients • Communicates a change in patient status to relevant specialty doctors	• Notices the long wait but fails to check the rest of the team is aware • Fails to inform team members when going on a break

7.4.3 Exploring staff attitudes to assessment

The last three questions on the questionnaire were used to evaluate the views of ED staff regarding the assessment of non-technical skills in the workplace using a structured observational tool. Overall, the response was very positive. One participant said, ñGood summary of shop floor work for registrars. Excellent workò. Another said, ñReally like your points, very insightful because team work is the essence of ED medicineò.

The first question asked whether non-technical skills should be explicitly and specifically assessed during training (see Table 6.2 below).

Table 7.2 Responses to Q1 -Should non-technical skills be explicitly assessed?

	Frequency	**Percent**
yes	77	52
no	10	7
difficult to assess	5	3
not sure	4	3
missing	52	35
Total	148	100

Approximately a third of participants did not respond (16 consultants, 19 trainees and 17 nurses). Of those who did respond, over 80% agreed registrars should be assessed and only 10% disagreed (2 consultants and 8 trainees). There were a number of strongly positive responses such as ñWe should have done this a long time agoò and ñAbsolutely, otherwise leadership skills will never be developedò. The reason given for opposing this type of assessment was invariably related to the fact that they believed informal and overall assessment, for example using ñsurreptitiousò observation, was more appropriate. Five respondents commented that this type of assessment would be difficult and two participants thought theses skills should be assessed but wanted assurance the assessment tool was ñvalidò and ñrobustò. Four people highlighted the benefits this type of assessment would bring by encouraging

feedback and reflective practice. Another two commented on the need for continuous assessment rather than one-off observation and two participants mentioned that assessment needs to be done in parallel with training in non-technical skills and leadership.

The second question asked whether registrars would like more feedback from consultants on these skills and the responses are summarised in Table 6.3 below. Of those trainees and middle grades who responded (n=41) 88% said they would like more feedback and only 10% (n=4) said they did not. Six respondents commented on the overlap with current MSF assessments, although one mentioned that MSF would be more useful if it included specific comments as evidence. Two people noted that junior doctors and nursing staff would also be a useful source of feedback.

Table 7.3 Responses to Q2 ï Would registrars like more feedback on these skills?

	Frequency	**Percent**
yes	36	24
no	4	3
not applicable	88	60
perhaps	1	1
missing	19	13
Total	148	100

The majority of people (93%) who answered question 3 said that they believe it would be beneficial for consultants to actively observe trainees running the department (missing data had a similar distribution to the first question) as shown in Table 7.4 below. One trainee simply said, ñDefinitely, how else can you learn or be properly assessed?ò In contrast, one trainee wrote ñAll of this is obvious and I donôt quite see the point. We will train our SpRôs like our bosses train usò.

Table 7.4 Responses to Q3 ï Would it be valuable for consultants to actively observe registrars running the department?

	Frequency	Percent
yes	91	61
no	7	5
missing	50	34
Total	148	100

Four respondents suggested registrars should observe and feedback on consultant leadership and an additional three consultants said they would like more feedback on their skills. Five trainees expressed concern about some consultantsô non-technical skills or said they would only like feedback from select consultants. One said, ñIt depends on the consultants... Perhaps education and training for the trainers?ò Eleven ED staff said that informal observation of trainees already happens in their department, although four of those commented that standardising and structuring the process may be beneficial. One participant said ñI would hope that the majority of ED consultants do this on a daily basis ï though the feedback is just as important and often not done as often as it should beò. Interestingly, one trainee said ñAbsolutely! But majority of the time we are neither observed nor supervisedò.

Three participants commented on the possible Hawthorne effect that may result from being observed by a consultant and one suggested the charge nurse at night may be better placed. Eight respondents expressed concern about the time required for this type of assessment, including one who said ñYes, without a doubt ï just desperately want the time to do soò.

7.5 Discussion

Three quarters of the exemplar behaviours in the questionnaire were deemed relevant and one quarter required revision. Participants gave an extensive range of suggestions and comments to help revise and improve the behavioural statements. Feedback from

respondents also highlighted the concerns some staff have regarding observational assessment of non-technical skills in the ED. For example, there were a number of comments about the time required for this type of assessment and concerns regarding bias due to the Hawthorne effect. Trainees and consultants need to be convinced of the benefit of a new tool over existing methods of assessment and evaluation. Furthermore, assessment should be implemented in parallel with training efforts. Assessors may need to be trained, not only in using the tool, but also in non-technical skill principles in order to be seen as credible. These comments are important to understand any barriers to wider implementation of the tool.

Despite these reservations, most staff recognised the potential benefits of assessment. Although many consultants already observe their registrarsô non-technical skills, this assessment often lacked structure and would inevitably vary considerably between departments and consultants. Furthermore, it was suggested that although trainees may be having their performance judged, they are not always given adequate feedback in order to encourage learning and improvement. There also seems to be considerable variation in the perceived level of observation and introducing an assessment tool may help to standardise practice.

7.1 Limitations

One of the primary limitations of this study is that not all exemplar behaviours from the assessment tool were included in the questionnaire. This was the trade-off for choosing to sample a large number of ED staff rather than a smaller group who were assembled specifically for the purpose of validating the tool. Indeed, the high proportion of missing data for the free text questions at the end of the survey may imply that the questionnaire was too long. Furthermore, many poor behaviours are simply the absence or opposite of good behaviour and, indeed, some observational tools only give exemplar behaviours (Hull, 2011; Frankel, 2007). Other studies have been designed to validate the tool in its entirety to help overcome this issue.

There may have been some bias introduced by the sampling method used as the participants were largely a self-selected group who attended meetings related to

academic medicine and patient safety and may not represent the views of the wider EM community. However, for validation purposes, the participants are expected to be ósubject matter expertsô and this cohort is more likely to have a background knowledge of the issues related to assessment of non-technical skills. Whilst nurses may appear to be underrepresented in the sample, the priority was to survey consultants and registrars who will be the eventual assessors and subjects.

One weakness of using CVI for analysing content validity is that it converts data into dichotomous relevant/ non-relevant and therefore does not consider all the variation in expert responses. However, data based on a 5 point scale is more susceptible to outliers. For example, a couple of respondents rated a substantial proportion of items low relevance and this is more likely to reflect attitudes towards the assessment tool rather than the items themselves. It is also important to note that content validity degrades with time as definitions and constructs evolve (Haynes, 1995). For example, ten years ago, it may have been appropriate to include statements related to Majors and Minors areas but as the ED structure changed, this has become outdated.

Overall, this study achieved Grant and Davisô recommendations for establishing content validity by assessing item content, item style and comprehensiveness. Subsequent studies will further evaluate the assessment tool and examine the issues related to training assessors.

8 Training EM physicians to use the tool for assessment of non-technical skills

CHAPTER OVERVIEW

The study described in this chapter was designed for the purpose of evaluating the assessment tool in an experimental setting using consultant and trainee EM physicians. Feedback gained from participants was analysed to further examine concerns and barriers to wider implementation of a workplace based tool for assessing non-technical skills. Sources of rater error are explored and limitations of the study are discussed.

8.1 Background

It is clear that a behavioural marker system needs to be evaluated before it is used for assessment or research (Klampfer, 2001). The ôutility indexô described by Cees van der Vleuten is a framework for assessment tool design and evaluation (van der Vleuten, 1996). It has five components (i) reliability, (ii) validity, (iii) educational impact, (iv) cost efficiency and (v) acceptability. Whilst evaluation efforts often focus on reliability and validity, it is vital that a tool is considered acceptable to those who will be using it. Surveying opinion, and addressing any concerns raised, is important to aid future implementation and ensure sustainable use. The purpose of this study was to assess inter-rater reliability and acceptability, using data gathered in an experimental setting.

8.2 Aims

The specific aims of this study were to:

- To assess inter-rater agreement (reliability) in an experimental setting
- To assess rater accuracy
- To obtain feedback on use of the tool from a wide range of participants

8.3 Methods

8.3.1 Design

An experimental method was used whereby trained participants watched a series of films and rated the principal EM physiciansô non-technical skills using the assessment tool. This method had been used by other research teams to evaluate a behavioural marker system (Fletcher et al., 2003, Yule et al., 2008).

A grant was awarded by the CEM to fund film production costs and to cover the expenses of the training day. Draft scripts were written by the lead investigator with subsequent modifications made by other members of the research team (two EM physicians and one human factors expert) during a series of meetings. The scenarios were chosen to portray a range of registrar types and each film depicted examples of all 12 skills with varying levels of performance. Filming took place over three days, using a combination of volunteer medical staff and professional actors. The final product was three short films lasting between 5 and 7 minutes. Each film was rated individually by a team of experts in non-technical skills which included 4 EM physicians, 1 anaesthetist with a background in non-technical skills and 1 human factors expert. Subsequently, the group met to discuss scores and the team reached a consensus on the ócorrectô score for each skill (the reference rating). These reference ratings were used to analyse the accuracy of participantsô assessments.

Participants attended a 1 day course (see Appendix E for the programme) and training was provided by 3 EM physicians who were experienced in non-technical skills assessment and debriefing. The course consisted of (i) an introduction to the importance of leadership and non-technical skills for EM; (ii) background on the development of the assessment tool, instruction on how to use the tool and explanation of skill definitions and the rating scale; (iii) practice using the tool using two filmed scenarios that were intermittently paused to allow participants to discuss the skills displayed; (iv) a test scenario which provided data for subsequent analysis; and (vi) a session on debriefing. Approximately 3 hours were dedicated specifically to training participants and practice using the tool.

8.3.2 Participants

Participants were invited to the training day via email distributed to members of the CEM throughout the UK. This included EM consultants, staff grade doctors and trainees.

8.3.3 Materials

Materials consisted of the following and can be found in Appendix E:

- Assessment sheets (shown in Figure 7.1) including definitions, exemplar behaviours and the rating scale
- DVD of 3 filmed scenarios
- Answer sheets for the 3 films listing the behaviours exhibited for each skill and the reference rating
- Feedback questionnaires (see Appendix E). This included both yes/no tick boxes and free text questions related to participant demographics, ease of use of the tool, feedback on the training process, strengths and weaknesses of the tool and whether the participant would be willing to use the tool in practice

8.3.4 Procedure

Attendees received training in non-technical skills and use of the tool as described above. During the test scenario, participants were asked to take notes of observed behaviours and then time was given at the end of the film to fill out the assessment form and give a rating for each skill. Participants were instructed to rate the scenario individually without consultation with others. Assessment forms were collected for subsequent analysis of rater agreement and accuracy. At the end of the training day, participants were asked to complete the feedback questionnaire.

8.3.5 Data Analysis

Data analysis proceeded in accordance with the 3 principal aims of the study.

1. Inter-rater agreement:

Within-group inter-rater agreement (equivalent to inter-rater reliability) was calculated for each skill using the formula

$$r_{wg} = 1 - (S_x^2 / \delta_E^2)$$

where $S_x^2 =$ observed variance and $\delta_E^2 =$ the variance that would be expected if all judgements were due exclusively to random error. This is calculated using

$$\delta_E^2 = (A^2-1)/12$$

where A = the number of alternatives in the response scale, which is this case is 9 (i.e. a 9-point scale). This statistic has been used to estimate inter-rater agreement for other behavioural marker systems in medicine (Fletcher et al., 2003) and is an appropriate method for designs in which one set of judges rate a single target, i.e. one registrar actor (James et al., 1984). An alternative method involves calculating intra-class correlations (Shrout, 1979) for each skill but this requires a repeated measure design and this was not practical within the time constraints. Whilst there is no agreed level that is considered acceptable reliability, many researchers use the criterion >0.7 (Yule et al., 2008, Williams, 1997).

2. Inter-rater accuracy

Inter-rater accuracy was measured by calculating the percentage of ratings correct to within 1 point of the reference rating. Percentages were also calculated using a condensed scale of unacceptable, acceptable and exemplary for comparison. Mean absolute deviation (MAD) was used to determine the degree of error for each skill (i.e. to estimate the ósizeô of the error in addition to how often ratings were incorrectly rated).

3. Feedback

Questionnaires collected from each participant were analysed using basic descriptive statistics and content review. In particular, forms were analysed to determine if any common themes emerged relating to weaknesses of the tool, assessor training and acceptability.

Internal consistency was not measured as the scenarios were fully scripted, thereby confounding any estimation of correlation between skills within a category.

8.4 Results

A total of 43 EM physicians attended from around the UK, including 25 trainees and 18 consultants. The mean number of yearsô experience in EM was 8 years (range 1-28 years). All attendees were unfamiliar with the assessment tool.

8.4.1 Inter-rater agreement

The values for inter-rater agreement varied from 0.56 for Maintenance of Standards to 0.88 for Quality of Communication as shown in Table 8.1. All skills, except Maintenance of Standards, exceeded the 0.7 cut-off for acceptable agreement. Variance was least for the Quality of Communication and largest Maintenance of Standards.

Table 8.1 Values for variance and inter-rater agreement

	Variance S_x^2	**Inter-rater agreement** r_{wg}
Maintenance standards	2.902	0.56
Workload management	1.733	0.74
Supervision & Feedback	.951	0.86
Team Building	1.301	0.80
Quality of communication	.787	0.88
Authority & Assertiveness	1.395	0.79
D.M. Option Generation	1.070	0.84
D.M. Selecting & Comm.	1.850	0.72
D.M. Outcome Review	1.725	0.74
S.A. Gathering Info	1.625	0.76
S.A. Anticipating	1.651	0.75
S.A. Informing the Team	1.691	0.75

D.M. = decision making
S.A. = situational awareness

8.4.2 Accuracy

Table 8.2 shows the percentage accuracy of ratings for each skill using the 9 point scale (correct within 1 point of the reference rating) and with a condensed 3 point scale of unacceptable, acceptable and exemplary. Accuracy within 1 point ranged from 14% for Maintenance of Standards to 84% for Team Building. Condensing the scores into unacceptable, acceptable and exemplary did not improve accuracy results.

Table 8.2 Accuracy of ratings and Mean Absolute Deviation from the reference rating

	Missing (*significant missing data)	**Reference Rating**	**Accuracy** (Within 1 point) %	**Accuracy** (Unacceptable/ Acceptable/ Exemplary) %	**Mean Absolute Deviation**
Maintenance of Standards	1	2	14	14	1.27
Workload Management	-	3	37	14	1.06
Supervision & Feedback	1	8	63	63	0.78
Team Building	-	7	84	65	0.92
Quality of communication	-	5	58	58	0.72
Authority & Assertiveness	1	3	19	5	0.98
D.M. Option Generation	2*	8	40	40	0.82
D.M. Selecting & Comm.	4*	5	58	58	1.10
D.M. Outcome Review	5*	7	37	19	1.02
S.A. Gathering Info	3*	5	63	63	1.09
S.A. Anticipating	3*	4	51	67	1.09
S.A. Informing the Team	1	3	44	19	1.08

There were four skills that experts deemed the actor registrar had performed to an unacceptable level, i.e. reference rating ÒBÓ (highlighted in grey in Table 8.2). The table clearly shows that these skills had the poorest level of accuracy and suggests possible bias due to leniency. This was further investigated by reviewing assessment forms and it became apparent that some error was due to participantsô reluctance to rate unacceptable even though poor behaviour was documented. For example, one participant identified two negative behaviours in Workload Management ñCould have delegated sharps boxò and ñDoesnôt ask for consultantôs helpò and then rates the skill 7 (exemplary). Similarly, one participant documented in the skill Authority and Assertiveness ñSpoke to surgeons but gives in easilyò and gives a rating of 7 (exemplary).

There appeared to be disagreement between observers as to whether the actor registrar had been appropriately assertive with the surgical registrar who refused to make

arrangements to deal with the long wait for surgical patients. Eleven participants did not document anything related to this conversation. Of those that did comment, approximately half (56%) thought the registrar acted with appropriate assertiveness whilst the other half (44%) commented that he should not have conceded. This issue also came up during discussions between the experts when deciding on the reference rating. Although some initially thought the actor was appropriately assertive, two EM consultants said that it would have been more appropriate to escalate the problem to the duty consultant. On further reflection, everyone agreed that that would have been the ideal course of action. This example highlights the disagreement regarding what is the ócorrectô behaviour for a registrar. However, it also suggests that some differences may be resolved by discussion and highlights the importance of ensuring an adequate debrief follows any assessment.

It can be seen that the skill Maintenance of Standards had the greatest error rate and also had the poorest inter-rater agreement (see Table 8.1 and Table 8.2). This was partly due to the way the scenario was scripted. In the film, the actor registrar is seen signing a drug chart and notices a potential overdose of metoprolol but fails to act on it. This skill was, therefore, given a reference rating of 2 indicating potential risk of direct patient harm. Review of the assessment forms revealed that most participants failed to notice this behaviour, or at least failed to document it on their scoring sheet. Only five participants noted the error; 1 incorrectly categorised it in Outcome Review, 2 participants correctly categorised it but still gave a rating of 7 (exemplary) and a further 2 noted the error and gave scores of 1 and 2 (unacceptable). This clearly highlights some sources of rater error: (i) failure to identify a behaviour, (ii) incorrect categorisation, and (iii) leniency.

Five of the 12 skills were given ratings with a Mean Absolute Deviation of less than 1 point and another six had a MAD just over 1 point (1.0 ï 1.1). Maintenance of Standards had a much larger MAD (1.27) which is likely to be due to the reasons discussed above.

8.4.3 Participant feedback

All participants completed feedback questionnaires at the end of the training day. A summary of yes/no tick box responses is given in Table 8.3.

Table 8.3 Responses given on the feedback questionnaire

Question	Response N (%)		
	Yes	**No**	**Don't know/ didn't answer**
Q3: Did you find the assessment tool easy to use on the filmed scenarios?	23 (53)	20 (47)	-
Q4: Do you feel you have had sufficient training to begin assessing trainees in your own department?	26 (60)	15 (35)	2 (5)
Q5: Do you think you would be able to explain to others in your department how to use the tool?	37 (86)	5 (12)	1 (2)
Q9: Consultants only ï Would you use the tool for assessing trainees in your own department	16 (89)	2 (11)	-
Q10: Trainees only ï Would you assess other colleagues using the assessment tool?	25 (100)	0	-
Q11: Trainees only ï Would you like someone to use the tool to observe and assess your leadership and non-technical skills?	24 (96)	1 (4)	-
Q12: Has todayôs session on identifying and assessing leadership and non-technical skills made you reflect on your own practice?	42 (98)	1 (2)	-

Usability

Questions related to usability of the tool highlighted a number of potential issues. In response to Question 3, just under half of participants said they did not find the assessment tool easy to use on the filmed scenarios. Seventeen respondents said it was simply because they needed more practice or were unfamiliar with the tool. This was also reflected in responses to the question, ñDo you feel you had sufficient training to begin assessing trainees in your own department?ò to which 35% or respondents said ónoô. Six people said they would like copies of the DVD for practice

and two said they would like written notes. Three participants said it was difficult to observe and concentrate on recording examples.

Responses to questions also revealed some participants found the tool complicated to use. Seven people commented on the apparent overlap between domains or found that behaviour seemed to fit into more than one category. One participant wrote, ñThere is a lot of overlap with some sections, for example, [Selecting and] Communicating Options and Informing the Team. The difference in these is quite subtleò. Three respondents qualified this point by saying they thought it would become easier as they became more familiar with the tool. Five participants said they thought there were too many categories and three said that the tool could be improved by simplifying the structure. For example, one respondent wrote, ñThough slightly different, ócommunicationô is pictured several times and situational awareness could be one single headingò. Four people suggested a revised rating scale may be better, for example, ñIs such an extensive rating scale required? [Is] acceptable vs. unacceptable vs. exemplary sufficient?ò Eighty six percent of participants said they would be able to explain how to use the tool but again.

Comparison to existing methods of assessment

Other questions related to existing methods of assessing leadership and non-technical skills and the perceived advantages and disadvantages of this kind of assessment. When asked how this tool differed from existing observation of registrarsô leadership and non-technical skills, 16 participants did not answer (or stated ñdonôt knowò or ñnot applicableò) and 10 said that assessment was not done in their department. One Year 7 SpR wrote, ñNot formally done. Random observations are shared, not usually positive onesò. Two trainees commented on difficulty getting consultants to delegate leadership responsibility to a registrar. For example, ñItôs not done. Iôve sometimes tried to start the process but the consultant ends up taking over shop floor managementò.

Nine participants said they thought the tool was more structured and explicit than current methods of assessment. Participants commented that existing observations

were more informal and ad hoc (4/43) and less comprehensive (3/43). Nine respondents noted that the tool made observations more objective and more accurate. Comments included, ñIt helps structure how you assess and... would help eliminate biased opinionò and ñBetter observation ï currently a lot of opinion/ comment givenò. Eight people noted that the tool allowed more identification of a broader range of behaviours and an array of competencies.

Respondents also said that time devoted specifically to observations would be beneficial rather than ñjust snap shotsò (2/43). Three other participants noted the benefits of direct observation with immediate feedback, for example, ñItôs done in real time, not retrospectively with biased memory. Feedback is more useful fresh in the mind of observer and traineeò.

Four respondents said they could appreciate the potential benefits for the observer, for example ñUseful for assessor too... may change your own/departmental practicesò. Four participants wrote positive statements regarding the self-reflection and feedback process, for example, ñGood to get feedback. How you are perceived is often very different from what you thinkò.

A number of disadvantages of this observational tool compared to existing methods of assessment were also noted. In particular, 11 participants commented on the possible Hawthorne Effect and comments included, ña lot of people do badly when under the spotlightò and ñyou do things to fool your consultantò. Eight respondents thought the assessment process would be intimidating or stressful and 14 expressed concern that the tool appeared to be time consuming and labour intensive. Six participants thought the assessment was subjective or vulnerable to bias. For example, one wrote ñThere can be some bias if the assessor has a preconceived opinionò and another wrote, ñmost raters will end up rating >5ò. Two people commented on the probability of the assessor getting interrupted by other staff or getting involved in patient care. Two participants were concerned that feedback may not be done appropriately and could be seen as confrontational or overly negative, especially if undertaken by someone unskilled in debriefing.

Future implementation

Sixteen of the 18 consultants (89%) said that they would use the tool for assessing trainees in their department. One wrote, ñI am convinced of the need to guide leadership development. It is unacceptable to assume that people will just pick it up. Structure empowers processò. In contrast, another wrote, ñNot convinced. Possible cloning of stronger trainers and their ideologyò.

All 23 trainees said they would assess colleagues using the assessment tool and only one registrar said they would not be willing to be observed and assessed. The trainee wrote, ñNo ï I wouldnôt like it but think it would be usefulò. Registrars identified the benefits of observing and assessing colleagues, especially more senior doctors. One trainee commented, ñI would be particularly keen to use this tool to assess and learn from consultant leadership skillsò and another wrote, ñI think it would be really valuable for both partiesò. Two trainees commented it may be difficult to give feedback to colleagues and one wrote, ñAlthough I might feel a little awkward applying it to colleagues of equal grade or higher, I imagine they would be really receptiveò.

Registrars could clearly see the importance of undergoing direct observational assessment of their skills. One wrote, ñIt would give feedback on things no-one usually sees and would make me focus on leadership much more. Great preparation for consultant lifeò and another wrote, ñI think it would be an uncomfortable experience but probably extremely usefulò.

Reflection

The final question asked whether the training day had encouraged participants to reflect on their own behaviour and all but one participant answered óyesô. A whole range of comments were made such as ñI may be abrupt when under stressò, ñrealising the effect of taking on too much and micro-managementò, ñI donôt use as many opportunities to teach as I shouldò, ñbeware of non-verbal behaviourò and ñreflect on friendliness vs. authority and assertivenessò.

It was clear from the feedback that training and discussion that focussed on the assessment of others was also beneficial for the assessor. One participant also commented, ñIôm thinking of reading the óGood Practiceô sheet (i.e. the list of exemplar behaviours that accompany the skill definitions) which you gave me...everydayò.

8.5 Discussion

8.5.1 Inter-rater agreement and accuracy

Values for inter-rater agreement varied from between 0.56 to 0.88. This compares favourably with estimations made for the ANTS tool which varied from 0.55 to 0.67 at the element level (Fletcher et al., 2003). All skills, except Maintenance of Standards, achieved a level of inter-rater reliability in excess of the recommended level of 0.7. The accuracy of participantsô ratings varied significantly but was particularly poor for skills with a reference rating in the óunacceptableô range, suggested possible bias due to leniency. The effect of leniency has been documented for other workplace based assessments (Hawkins, 2010). Response bias, such as leniency, contributes to systematic variance and serves to inflate the estimate for inter-rater reliability (Guion, 1965). Error due to response bias is not limited to the methods used in this study. This systematic error may explain why the results showed good inter-rater reliability but less favourable results for accuracy. Accuracy was poorer than that reported by Fletcher et al. (2003). However, the ANTS system uses a 5 point rating scale so accuracy to within 1 point of the reference rating is considerably easier to achieve.

In addition to a response bias, this study identified a number of other sources of error to account for variation in rater agreement and explain the accuracy results. Review of assessment sheets highlighted issues with (i) disagreement between observers regarding what constitutes good and poor behaviour, (ii) failure to identify a behaviour, and (iii) incorrect categorisation. Similar sources of error were reported by

Australian researchers when evaluating training for a behavioural marker system (Graham et al., 2010).

It became clear that assessors were unfamiliar with some of the concepts related to non-technical skills let alone the tool. It is not surprising that EM physicians disagree on what behaviour is acceptable given that there is limited training available on non-technical skills in the UK. Some standards have been set, for example, defining good hand washing practice or demonstrating the benefits of communicating using the SBAR approach (Haig, 2006). However, the boundaries of good and poor behaviour for many other non-technical skills are yet to be determined and variation in safety beliefs is well documented (Palumbo, 2009). In contrast, most physicians are familiar with the standards required for adequate history taking or physical examination. Whilst opportunities were given for the participant audience to discuss positive and negative behaviours illustrated in the filmed scenarios, this could have played a larger part of the training programme if time allowed. Now that preliminary studies have been completed, further efforts to train assessors using the assessment tool should include more time devoted to non-technical skillsô principles and the relationship to safety. Indeed, recommendations state that formal training is required in applicable aspects of non-technical skills over a minimum two days (Klampfer, 2001).

No amount of training can completely eliminate variation in opinion, for example, regarding what constitutes appropriate assertiveness or micromanagement. If assessment of non-technical skills were to become more widespread, the debate and discussion generated would help to raise the profile of these issues. More research is needed into the positive and negative effects of certain behaviours before a definitive answer can be given. This issue also highlights the importance of ensuring an adequate debrief session follows every assessment. The debrief gives trainees and assessors the opportunity to discuss the positive and negative aspects of observed behaviour and ideally resolve any disagreement.

Participantsô failure to identify certain behaviour was also a source of error. This is probably due to attention overload and the problem may have been exacerbated by using artificial scenarios with fewer natural pauses than normally occur in real life. Indeed responses on the feedback questionnaires revealed that participants found it

difficult to observe and simultaneously record examples. However, less attention will be diverted to thinking about what skill is being observed as assessors become more familiar with the tool, allowing them to concentrate fully on the observations.

A review of the assessment sheets also identified that participants had incorrectly categorised behaviours. Feedback highlighted that some attendees found the tool complicated and noted apparent overlap of domains. Evaluators of the ANTS tool also commented on ñconcerns about complexityò and they explained that this is likely to be due to lack of familiarity with the system (Fletcher et al., 2003). This was also the case in our study as a number of participants commented that categorisation was likely to get easier with practise. Indeed, 35% of attendees felt they had not had enough training which is not surprising given only 3 hours of dedicated training was provided. Another sign that participants may have struggled to categorise behaviour is shown in the frequency of missing data. Whilst many skills had either 0 or 1 missing data, some skills had much more missing data (2-5 missing ratings). These were skills with less intuitive definitions and participants may be less familiar identifying the relevant behaviours. In addition to further training, apparent overlap between domains could be addressed by extending the list of good and poor behaviours to incorporate a wider selection of examples. Whilst incorrect categorisation has a negative effect on reliability and accuracy results, it is less of a problem in practice. Trainees would still gain educational benefit from the assessment process as long as the debrief adequately covers good and poor observed behaviours.

8.5.2 Feedback from participants

Feedback from participants highlighted a number of concerns such as bias due to the Hawthorne Effect or lack of objectivity. Other practical issues were raised as some participants thought the process was too time-consuming and others commented on the likelihood of interruptions from other staff during an assessment. Despite this, the majority of consultants and trainees thought that this type of assessment was worthwhile and would be willing to try it.

This study also highlighted potential benefits for the observer as well as the subject and it is worth considering the advantages of using trainees to assess other trainees. There is likely to be less influence from the Hawthorne Effect and learning through the observation of others is fundamental to the apprenticeship model of teaching. The benefits are emphasised by the fact that almost all attendees said that the training session led them to reflect on their own behaviour. This is particularly relevant as one of the main criticisms of behavioural marker systems is the time-consuming nature of assessor training (Graham et al., 2010). Also, some would argue that trainees have more available time than consultants for performing assessments. Furthermore, registrars are ócconsultants in trainingôso it is essential that they are taught and develop skills in assessment and feedback. However, it is also important to consider that some trainees may feel uncomfortable giving or receiving feedback from another trainee, especially if the assessor is a junior colleague.

8.5.3 Strengths and Limitations

This study provides important data for evaluating a tool to assess registrar non-technical skills. Feedback was sought from both consultants, who may act as assessors, as well as trainees, who may be the subject of assessment the process. However, this study used scripted scenarios in an experimental setting which are clearly very different from real-life conditions in the ED. Ideally, extra time would be dedicated to training, although other teams have used 8 hours of training without achieving adequate reliability and accuracy (Graham et al., 2010). Additional time would also enable data to be gathered from several scenarios, allowing more detailed analysis of variance. Despite these limitations, this study provides encouraging results and suggests that further studies are warranted in the natural ED setting.

8.5.4 Summary

Good levels of inter-rater agreement were achieved in this study. However, 3 hoursô training provides inadequate results for accuracy and ratings may be subject to bias and other sources of error. Furthermore, feedback reveals participants would like

more extensive training and opportunities to practice in order to feel confident using the tool for assessment. Responses regarding acceptability are encouraging and there is evidence that using trainees as assessors may be worth exploring further. Despite limitations due to the experimental design and limited assessor training, this study shows that a larger study based in the ED is justified and this further explored in the following chapter.

9 A Multi-centre Observational Study

CHAPTER OVERVIEW

The previous chapter describes evaluation of the assessment tool under experimental conditions using taped scenarios. The study presented in this chapter was designed to evaluate the tool in the natural ED setting, focussing specifically on reliability and the observabiliy of non-technical skills in the workplace. A summary of the various observed behaviours linked to each non-technical skill is also given and sources of rater error are proposed. The chapter concludes with strengths and limitations of the study as well as consideration for future research.

9.1 Background

The utility of any assessment tool relies on a number of factors such as reliability, validity and acceptability. However, reliability (i.e., the ability to discriminate between doctors in a consistent and replicable manner) is of key importance, as an assessment cannot be viewed as valid (i.e. to be measuring what it is thought to be measuring) unless it is reliable (Streiner, 1995).

Reliability is the ratio between subject variance (which is what we want to measure) and subject + error variance (where what we want to measure is obscured by measurement error). Error variance comes from a variety of sources such as (i) inter-observer variation, e.g. due to different observers consistently marking high or low and observersô differing view of what constitutes good performance, (ii) intra-observer variation, e.g. where observerôs markings vary for no apparent reason (e.g. a ógood dayô or a óbad dayô), and (iii) case specificity, which describes a traineeôs variation in performance from one challenge to another (Crossley et al., 2002). Whilst some error variance in inevitable, it is important to both quantify and understand sources of rating error in a real life setting.

In addition to enabling evaluation of an assessment tool, observation in the workplace is a rich source of data for exploring ED registrarsô non-technical skills. A review of the literature reveals a lack of research focussing on non-technical skills in UK Emergency Departments (see Chapter 4). Studies have either focussed on the skills related to safety in a resuscitation situation (Brodie, 2009, Cowan, 2006) or have focussed on individual skills such as communication (Redfern, 2009) or supervision (Kilroy, 2006a). One observational study designed to identify sources of risk in a UK ED identified some safety issues related to human factors, such as authority gradients and cognitive errors, but failed to give a complete description of the contribution from non-technical skills. Understanding which skills are performed well and which tend to be problematic will also have applications for the training of nontechnical skills.

The two broad objectives for this study were (i) to evaluate a tool for assessing EM non-technical skills, and (ii) to gain a deeper understanding of the array of non-technical skills displayed by registrars in UK emergency departments.

9.2 Aims

Specifically, the aims of the study were:

1. To evaluate inter-rater reliability and test-retest reliability of the assessmentt tool
2. To determine the observability of each of the 12 skills in the tool
3. To describe the range of non-technical skills displayed by ED registrars
4. To determine if observed non-technical skills vary between the departments included in the study

9.3 Methods

9.3.1 Design

This was a multi-centre observational study using direct observation in the workplace. The aim was to collect approximately 100 observational episodes using trained observers in four EDs.

9.3.2 Setting

Data was collected from four NHS Emergency Departments comprising two teaching hospitals and two district general hospitals in London and the south of England.

9.3.3 Participants

Participants were EM doctors, trainee registrars or non-training grade equivalent, working on the middle grade rota. Participants were selected at random depending on who was working on the days the observer was available to undertake the observations.

The inclusion criteria for participants were as follows:

- EM registrars currently working in one of the four specified EDs

The exclusion criteria for participants were as follows:

- locum doctors
- any staff member that had refused to give consent for the study

Observers

Four observers (LF, AG, JS and AB) collected the data. All observers were EM trainees from the South East Thames region who had volunteered to collect data for the study. They completed a half day training course which included an overview of the assessment tool as well as detailed descriptions of the individual skills. A series of scripted filmed scenarios (described in Chapter 8) were used to further illustrate the skills and gave observers an opportunity to practice using the rating scale. These, and other hypothetical scenarios, were discussed in detail to ensure observers fully understood how to identify and categories skills. Practice ratings were discussed to aid calibration and ethical issues, such as intervening to avert or disclose a clinical error, were covered.

The first two workplace based observations for each observer were considered training exercises and novice observers undertook the assessments paired with the lead investigator. After each observation, both observers completed the assessment form and any discrepancies were discussed. This gave new observers an opportunity to ask any additional, practical questions.

Study sites were departments that the observers were either currently working in or had recently left. This was for practical reasons (i.e. more convenient to undertake observations) and also because it more closely reflected how the tool would eventually be used (i.e. physicians rating other physicians who were known to them). Whilst, it is acknowledged that the effect of previous working interactions and may introduce some bias when rating a colleagueôs skills, this bias is unavoidable when using any workplace assessment. It is possible that some registrars may feel uncomfortable being observed by a colleague and they were given the opportunity to be excluded from the study.

On completion of the study, observers were asked to give feedback on using tool and their experiences during the study.

9.3.4 Materials

Materials consisted of the following and can be found in Appendix F:

1. Observational assessment forms ï as shown in Figure 7.1
2. Participant Information Sheets
3. Consent forms

9.3.5 Procedure

Ethical approval was granted from North West London REC1 and site-specific approval was sought from each of the four study sites.

Prior to commencing observations, ED registrars and middle grades in each of the participating EDs were notified in writing via email of the purpose of the study. They were also sent a copy of the Participant Information Sheet and consent form and were given the opportunity to be excluded or ask any clarifying questions. Written consent was sought from each participant prior to commencing the first observation and verbal consent was sought prior to each subsequent observation. A brief explanation of the study was given to the nurse-in-charge for the shift (if they were not already aware) and as many other staff members working in the department as possible.

The assessment involved following the registrar for a 45-70 minute period and recording detailed field notes on observed behaviours. Ideally, the lead or ôbleep holderô registrar was observed but some departments do not allocate a lead registrar so any of the available registrars were included. For each observational period, the observerôs name, the time, date and hospital were also recorded. At the end of the observation period, the investigator used the field notes to populate the assessment form and give a rating for each skill based on documented behaviours. Participants were offered informal, individual feedback by observers. Approximately one third of all assessments used a pair of observers (LF and one other) to obtain inter-rater reliability data. When more than one observer was used, each observer was blinded from the other observerôs ratings. After each paired assessment, ratings were compared and disagreements in ratings were discussed. For single observer

assessments, the observer filled in the assessment sheet and this was reviewed by a second member of the research team to detect any possible errors (e.g. a behaviour that appeared to be incorrectly categorised) or bias (e.g. high or low scores that did not appear to fit with the evidence documented). This was then discussed with the observer and notes were taken.

9.3.6 Ethical Issues

Informed written consent, and subsequently verbal consent, was obtained from all participants prior to undertaking observations. Consent could be withdrawn at any time, for example, if participants felt it was not appropriate to conduct observations for that particular day. All data collected from participants were anonymised using a unique Participant Identification Number and participants were assured that no individual feedback about performance would be given to consultants or managers.

It was anticipated that during the course of the observations a clinical error or risk may be identified by an observer. The decision whether or not to intervene was made at the time by the observers who are all experienced ED physicians. Whilst the observer aimed to be as unobtrusive as possible, patient safety was always considered the priority. Any incident was immediately brought to the attention of the participant being observed. If the clinical error was significant, the participant was encouraged to report the incident in the usual manner. Whilst minor errors were identified and were dealt with immediately, no significant clinical incidents occurred during this study.

9.3.7 Data analysis

1. Evaluation of reliability

a) Inter-rater reliability

Inter-rater reliability refers to the degree of agreement between two or more observers assessing the same encounter and was evaluated by calculating the intra-class coefficient (ICC). A cut off of Ó0.7 was chosen for this study and a full description of the reasoning for the choice of this test and cut-off is given in Appendix F.

ICCôs were calculated for each skill and for the overall scale and this was repeated for each pair of observers. Also, ICCôs were calculated for the mean scores across the whole of the instrument, which is how other workplace based assessments (WBAs) have been evaluated in the UK (Wilkinson, 2008). Assessment sheets and notes of discussions between observers were also analysed to explore possible sources of error.

b) Test-retest reliability

Test-retest reliability measures the consistency of an assessment tool when administered on different occasions to the same sample (i.e. stability of scores over time or cases). This was calculated using Spearmarnôs rho on data from the first two assessments of participants who were observed more than once and was calculated at both the skill and overall scale level. Normally, a correlation of more that 0.7 is considered acceptable (Murphy et al., 2009).

2. Observability

Percentages (i.e. the number of assessments where the skill was observed at least once divided by the total number of assessments) were used to determine the observability of each of the 12 skills. There is no agreed level a skill should be observed to warrant

inclusion in a tool. However, it is logical to expect a skill to observed more often than not (i.e. >50%) (Abell, 2009).

3. **Description of the range of skills**

Each of the individual skills were analysed to determine the average rating and dispersion of ratings. This was presented as the mean rating for each skill and 2 standard deviations from the mean. Additional analysis involved reviewing all of the assessment forms to extract examples of good and poor behaviour exhibited by registrars.

4. **Comparison of hospitals**

The Kruskal-Wallis Test was used to compare hospitals to determine if any significant difference existed in skill level. *Post hoc* procedures involved using Mann-Whitney tests to determine which pair of hospitals were significantly different for a particular skill. Bonferroni corrections were used to reduce the risk of Type 1 errors (i.e. false positives).

9.4 Results

A total of 96 assessments were completed and 43 registrars were observed.

9.4.1 Evaluation of reliability

Table 9.1 below summarises the ICC calculations for the individual skills as well as the overall scale for each pair of observers.

Table 9.1 ICCôs for individual skills and overall scale for each pair of observers (to 2 decimals)

Skill	AG-LF	95% CI		JS-LF	95% CI		AB-LF	98% CI	
Maintenance of standards	0.74	0.09	0.95	0.47	-0.16	0.87	0.71	0.01	0.95
Workload management	0.78	0.23	0.96	0.56	-0.09	0.91	0.63	-0.11	0.94
Supervision & feedback	0.83	0.29	0.97	0.12	-0.49	0.74	-0.25	-0.98	0.66
Team building	0.64	-0.12	0.92	0.69	-0.08	0.94	0.71	-0.10	0.96
Quality of comm.	0.62	-0.25	0.93	0.55	-0.13	0.90	0.12	-0.88	0.82
Authority & assertiveness	-0.11	-0.66	0.62	0.46	-0.52	0.89	-0.16	-0.73	0.65
DM Option generation	0.80	0.29	0.45	0.53	-0.26	0.90	-0.10	-0.44	0.59
DM Selecting & communicating	0.57	-0.27	0.91	0.35	-0.35	0.83	-0.14	-0.68	0.66
DM Outcome review	0.09	-0.43	0.71	0.10	-0.37	0.70	0.77	0.12	0.21
SA Gathering information	0.69	-0.12	0.94	0.48	-0.48	0.89	0.13	-0.17	0.69
SA Anticipating	-0.21	-0.04	0.64	0.75	0.06	0.95	0.68	-0.19	0.95
SA Informing the team	0.39	-0.25	0.85	0	-0.90	0.73	0.47	-0.49	0.91
Using mean scores	**0.82**	**0.30**	**0.97**	**0.70**	**0.03**	**0.94**	**0.59**	**-0.13**	**0.93**
Overall scale	**0.58**	**0.41**	**0.70**	**0.53**	**0.35**	**0.67**	**0.42**	**0.21**	**0.59**

The confidence intervals for ICCs of individual skills are very large, due to the small sample size, making it difficult to make any meaningful interpretations. However, looking at the values across the three pairs of observers it is possible to identify skills that had particularly low ICCs and those that were generally high. For example,

Authority and Assertiveness had low ICCs for all three pairs (-0.16 to 0.46) as did Informing the Team (0 to 0.47). Similarly, Workload Management had generally high ICCs (0.56 to 0.78) as did Team Building (0.64 to 0.71).

At a scale level, none of the three pairs of observes achieved the desired level of 0.7 for inter-rater reliability (0.58, 0.53 and 0.42). However, ICCs calculated using mean scores for each assessment were more acceptable (0.82, 0.70 and 0.59).

Eighteen participants were observed more once and data from the first two observations were used to calculate test retest reliability. Spearmanôs rho calculation using all 12 individual skills was 0.26 and using mean scores was 0.70.

Possible sources of error

A review of assessment sheets and discussion notes identified the following possible sources of rater error and are summarised in Table 9.2 below. Further examples of errors extracted from assessment sheets are given in Appendix F.

Table 9.2 Sources of rater error

Error	Example
Simple mistakes made with categorisation of behaviours	One observer documented ñThroughout the hour, the registrar didnôt check on the status of the department or ask juniors how they were getting on with patients.ò This was placed in Workload Management. A second investigator reviewed the assessment sheet and questioned whether this would be better in SA ï Gathering Information and the original observer, on reflection, agreed.
Overlapping definitions of a skill	This meant that some behaviours were erroneously double coded so that the trainee was unfairly penalised for a single behaviour or their ratings were inappropriately boosted. For example, ñAdvises SHO to use seizure proformaò increased the rating in both Maintenance of Standards and Supervision and Feedback and should have only been coded once.
Misunderstandings of the definition of a skill	This most commonly occurred with the skill Quality of Communication as many other skills relied on communication. However, observers were instructed to code behaviour in Quality of Communication only when it did not fit in another, more specific skill. For example, ñSought advice from ENT registrarò and ñlistened to nurses input about management of patientò could have been better coded into DM ï Option Generation.
Concealed cognitive processes and reasoning	Registrars sometimes appeared to be ignoring other staff members. It is difficult to ascertain where this is rudeness and therefore categorised in Team Building or simply not paying attention and therefore relates to Situational Awareness.

Failure to rate objectively	This was picked up when a second observer reviewed assessment sheets and the documented behaviour did not appear to match the rating given. For example, ñEnsured SHO considered disposition of patientò in Anticipation scored 7. The second investigator thought this was generous for a single positive behaviour. After reflection, the original observer agreed.
Missed behaviours	When two observers were used, it became clear that sometimes one of the observers simply missed behaviour that the other observer had seen. This was either because the observer happened to be looking somewhere else, lack of attention or because noise in the department meant that one observer was unable to hear some communication.
Observed behaviour not recorded/judged	During one observation, the police were seen walking into a cubicle where the registrar and patient were having a consultation. One observer commented on the lack of authority exerted by the registrar for the police unnecessarily interrupting, with no consideration for whether the patient was undressed etc. The second observer had seen the same incident but had not thought to document it and agreed that it should have been noted.
Disagreement between observers regarding what constitutes good and poor behaviour	A registrar reviewed the closure of a simple head wound even though the experienced nurse said it was unnecessary. One observer said this gave the appearance of mistrust and displayed poor Team Building whilst the other observer noted it as good Outcome Review. After discussion, both observers agreed it was an unnecessary action it should have been documented as poor behaviour.
Observer not familiar with the department	One observer gave a registrar a poor score for Team Building because he could tell that the registrar had offended a nurse he knew well whereas the other observer, who was a visitor to the department, did not notice.

9.4.2 Observability

The observability of each skill is shown in Table 9.3. Whilst all skills were observed more than 50% of the time, four were observed considerably less frequently than the others. These were Supervision and Feedback (77%), Authority and Assertiveness (68%), Outcome Review (76%) and Anticipation (70%).

Table 9.3 Observability of each of the 12 skills

(* observed less frequently)

Skill	**% Observed**
Maintenance of standards	100
Workload management	99
Supervision & feedback	77*
Team building	100
Quality of comm.	97
Authority & assertiveness	68*
DM Option generation	95
DM Selecting & communicating	88
DM Outcome review	76*
SA Gathering information	88
SA Anticipating	70*
SA Informing the team	90

9.4.3 Description of the range of skills

Descriptive statistics for ratings given for each skill are shown in Table 9.4. Only one skill had a mean of less than five, which was Maintenance of Standards (mean 4.8). Other skills with relatively low means included Supervision and Feedback, Authority and Assertiveness, DM Selecting & Communicating and Anticipating. The highest mean was calculated for Team Building (mean 6.0). No skill was given a rating of 1 or 9. A series of bar charts showing the frequency of ratings for each skill is displayed in Appendix F.

Table 9.4 Descriptive statistics

	Minimum	Maximum	Mean	Std. Deviation
Maintenance of standards	2	8	4.8	1.27
Workload management	3	8	5.4	1.32
Supervision & feedback	3	8	5.1	1.10
Team building	3	8	6.0	1.09
Quality of comm.	3	7	5.6	1.00
Authority & assertiveness	3	7	5.3	0.88
DM Option generation	3	8	5.8	1.17
DM Selecting & comm.	3	8	5.3	1.00
DM Outcome review	3	8	5.5	0.99
SA Gathering information	3	8	5.7	1.36
SA Anticipating	3	7	5.2	1.00
SA Informing the team	4	8	5.6	0.96

Figure 9.1 below illustrates the dispersion of ratings for each skill. Skills with a larger range of ratings were Maintenance of Standards, Workload Management and SA Gathering Information, whilst skills with relatively low distribution of ratings included Authority & Assertiveness, SA Outcome Review and SA Informing the team.

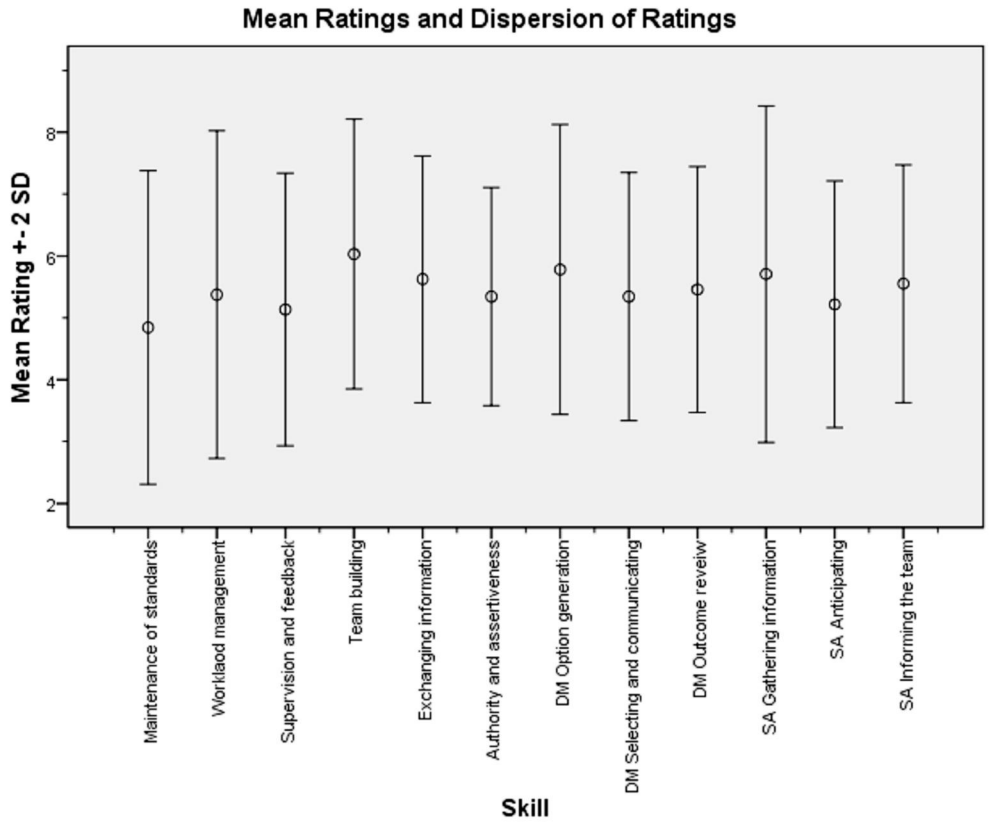


Figure 9.1 Mean and dispersion of ratings

All assessment sheets were reviewed to extract a selection of examples of good and poor behaviours and a sample of these are given in Table 9.5 below. These are only a representative sample of the many examples documented by observers during the course of the study. They were chosen to highlight the main behaviours associated with each of the 12 non-technical skills listed in the assessment tool. A more detailed description of the behaviours observed is given in Appendix F.

Table 9.5 Examples of good and poor behaviour observed during the study

Skills	Examples of good behaviour	Examples of poor behaviour
Maintenance of Standards	Checked list of medication for possible interactions when asked to write up a drug chart Insisted an ECG is repeated because of incorrect lead placement	A patient was transferred to CDU with no documentation despite an acute vertebral fracture being detected on x-ray. Allowed a serious self harm patient (overdose and hanging) to go to the toilet unsupervised without a risk assessment.
Workload management	Prioritised a sickle cell patient with abnormal observations Put post-it notes on cards of patients that needed review Encouraged specialty doctor to make decisions early whilst offering assistance	Bleeped an anaesthetist to assist with a fitting patient and then, while the patient was still fitting, left to sort out another patient. Told a nurse he was not available to see an arrhythmia patient, whilst seeing a non-urgent head injury patient, and made no attempt to help find an alternative doctor.
Supervision and Feedback	Used the phrase ñWhat needs ruling out?ò rather than simply providing the answer when discussing a patientôs management with a junior Gave feedback to ambulance crew about patient they had bought in earlier Spoke to Plastics when referral made by ED SHO is refused	Pointed out a subtle pneumothorax on xray without giving junior doctor the opportunity to spot it SHO was clearly not confident suturing pre-tibial laceration but registrar did not offer help so SHO found another senior doctor Asked a medical student to assist with a log roll without checking they were capable Seen altering a junior doctorsô prescription at the nurseôs request but failed to notify and educate the relevant doctor.

Team Building	Provided support for a nurse who had apparently mismanaged a patient Praised an SHO for picking up a ureteric stone on CT	Appeared dismissive of a junior doctor asking for help Rolled his eyes when a nurse said they were not trained to do cannulation.
Quality of Comm.	Used very specific questions like ñWhatôs your concern?ò when giving advice to junior doctor Reiterated the need to give nebuliser via air with oxygen via nasal cannulae Encouraged the use of check-back communication when giving verbal order to a nurse	Heard saying ñCan you keep an eye on resus?ò to another registrar when going on a break without giving any handover of the patients Failed to use appropriate medical terminology giving a poor description of facial wound when discussing over telephone with Max-Fax doctor. Failed to listen to SHO presentation and had to clarify numerous points that had already been mentioned Said ñletôs get a bag of fluidò during an emergency without specifying which fluid or who was expected to get it
Authority & Assertiveness	Said firmly to surgical registrar ñYou need to see this patientò Reiterated psychiatric symptoms when psychiatric liaison nurse implies symptoms are all organic Registrar only acted as a scribe for a trauma call but pointed out lack of timely review of observations to consultant Was appropriately assertive when SHO was late returning from teaching	SHO appeared intimidated when presenting case to a registrar who questioned aggressively Failed to assert herself with radiologist and unsuccessful getting an out-of-hours CT abdomen for a trauma patient who was subsequently found to have a splenic rupture Confrontational with anaesthetist who was reluctant to intubate a fitting patient without waiting to review impact of treatment

DM Option Generation	Looked at x-rays taken on recent admission and identified missed dislocated hip on an elderly lady Asked about patientôs age and status when asked to prescribe IV fluids Reviewed facial injuries in person when SHO was vague about history Briefly reviewed patient with severe PR bleed before delegating to SHO	Did not get collateral history from witness of possible seizure Seen apparently ignoring an experienced nurse who said a patientôs leg wound smelled of pseudomonas Nurse said patient was unwell and registrar prescribed IV fluids without reviewing patient in person Requested a FAST scan for a patient who had fallen backwards from standing with no abdominal trauma
DM Selecting and Comm.	Asked nurse to insert catheter and explains indication as not initially obvious Gave clear plan to SHO and explained under what circumstances the patient was safe to go home The registrar was unsure so he stated his concerns to the resuscitation team ñWe still donôt know why sheôs short of breathò	Nurse suggests IV 5% dextrose and registrar said ñNo, normal salineò but no explanation why Hesitant dealing with a resuscitation patient and SHO says, ñI donôt know, youôre the reg.ò No discussion of risk or indications for return when registrar suggested a new SHO send home a pregnant PV bleed patient
DM Outcome Review	Checked resuscitation patientôs bloods were en route to the laboratory Reviewed pain of patient who had been prescribed analgesia in the queue to be seen Checked SHO has been successful making referral	Failed to assess the impact of morphine for patient with severe abdominal pain Reviewed patient presenting with status epilepticus but with significant delay Did not check if SHO was successful referring patient to vascular team when already discussed they may be reluctant to accept

SA Gathering Information	Overheard charge nurse and nurse discuss a patient with low blood pressure and offered to help Noticed asthmas patient using own inhaler in the department and made an assessment Had a quick look at a trauma patient who was waiting to be seen	Failed to notice that no BP has been checked for >25 minutes in a patient with significant trauma Heard a tannoy requesting a doctor to the resuscitation room and the registrar failed to make any enquiries Got caught up with a single patient and failed to notice the long wait in the department
SA Anticipating	Prepared to move a patient to Majors to create a free cubicle in the resuscitation room in case a new priority patient arrived Identified early a patient who would need social input to aid discharge Considered the possible deterioration of a warfarin patient with a PR bleed and asks for a transfer to the resuscitation room	No consideration of potential deterioration in a trauma patient with a significant mechanism of injury but stable observations ï later found to have multiple fractures and internal injuries Nurse had to ask if she could move a drunk patient to CDU to clear a cubicle
SA Informing the Team	Told CDU nurses about anticipated transfers Told nurse to inform ward nurses of patientôs psychiatric issues Registrar was seen speaking out a patientôs vital signs during the course of a resuscitation	At the board round, failed to handover that occupational therapy services were unavailable Did not inform team that she was going to the CT scanner with a patient Asked an SHO to see a patient with PR bleed but failed to inform him that the patient was on warfarin

9.4.4 Comparison of hospitals

The Kruskall-Wallis test was used to determine if there was any significant difference in skill level between the four hospitals. This calculation produces mean rankings for each hospital and is shown in Figure 9.2.

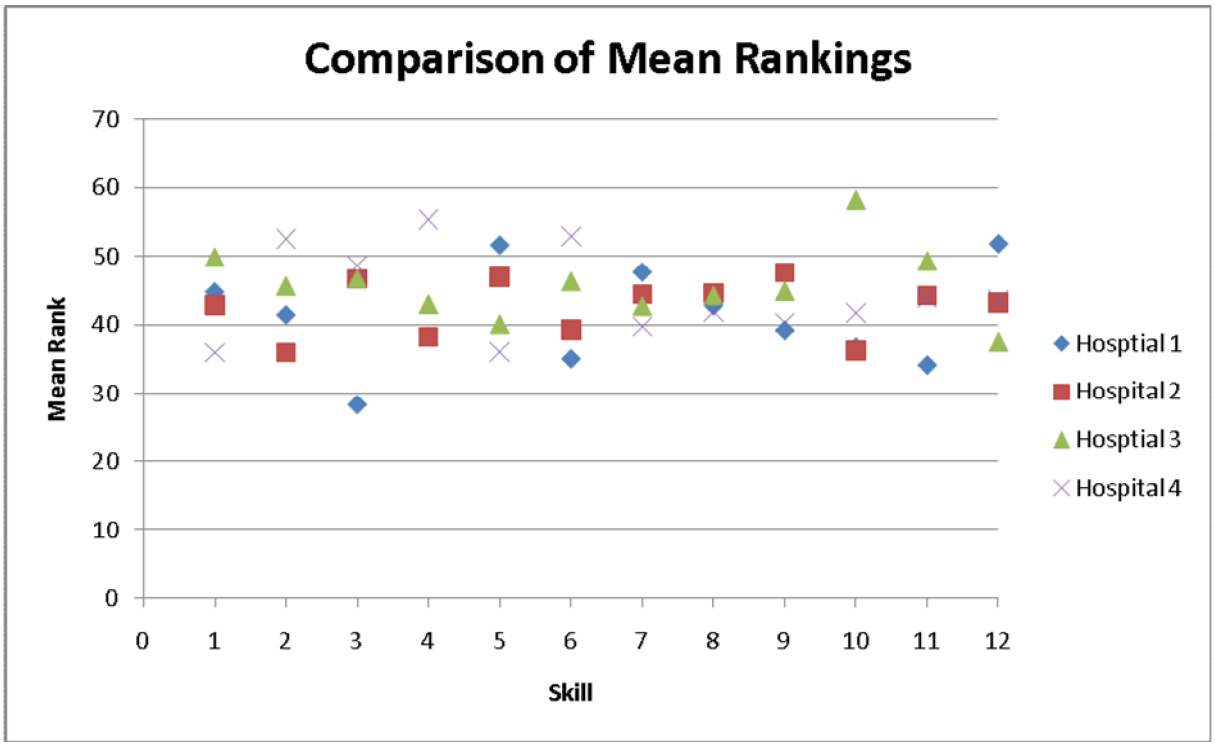


Figure 9.2 Comparison of mean rankings for each skill

A summary of the results of the Kruskal-Wallis test is shown in Appendix F. Two of the 12 skills, Supervision & Feedback and SA Gathering Information, had significantly different distributions of ratings across the four hospitals (p=0.037 and 0.007 respectively). There was no significant difference in performance for any other skills.

Figure 9.2 shows that the mean ranking of Hospital 1 is clearly below the other hospitals for Skill 3 (Supervision and Feedback). Mann-Whitney tests were conducted to compare Supervision and Feedback rankings for Hospital 1 with each of the other three hospitals (i.e. 3 tests). Using a Bonferroni correction, this adjusted the significance level from <0.05 to <0.0167 (i.e. 0.05/3). Table 9.6 shows that the significance level for all three tests is less than 0.0167. Therefore, Hospital 1

participants performed significantly poorer at the skill Supervision and Feedback compared with the other three hospitals.

Table 9.6 Results of Mann-Whitney tests comparing hospitals for the skill Supervision and Feedback

Comparison of Hospitals	**Mann-Whitney (1 tailed test)**
Hospital 1 & 2	0.007
Hospital 1 & 3	0.003
Hospital 1 & 4	0.008

Discussion with observers and review of assessment sheets did not reveal any clear reason for this significant difference in Supervision and Feedback. Although one possibility raised related to the fact that Hospital 1 has the highest level of óshop-floorô consultant presence so registrars may have relinquished teaching and supervision to the consultant. Observers, however, thought that there were still opportunities for teaching that registrars were missing.

Figure 9.2 shows that Hospital 3 had a mean ranking clearly above the other three hospitals for Skill 10 (SA Gathering Information). Mann-Whitney tests were conducted comparing Hospital 3 with the other three hospitals and the same Bonferroni correction was used. Table 9.7 shows that the significance level for all three tests is less than 0.0167. Therefore, Hospital 3 performed significantly better at the skill SA Gathering Information compared to the other three hospitals.

Table 9.7 Results of Mann-Whitney tests comparing hospitals for the skill SA Gathering Information

Comparison of Hospitals	**MannWhitney (1 tailed test)**
Hospital 3 & 1	0.011
Hospital 3 & 2	0.000
Hospital 3 & 4	0.012

In this case, observers could clearly identify differences in registrarsô behaviour at Hospital 3 compared with the other hospitals. In Hospital 3, registrars took on a more formal leadership role, under the supervision of consultants. There was a noticeable increase in Situational Awareness behaviours related to organisational issues such as waiting times, staffing levels and even listening to other peopleôs conversations (e.g. nurses speaking about a sick patient) as they appeared to take on more personal responsibility for the coordination of the department. Although not statistically significant, observers noted that registrars in Hospital 2 performed particularly poorly in this skill. There was no identifiable lead doctor (registrar or consultant) and the nurse in charge took on the main leadership and coordinating role.

9.5 Discussion

9.5.1 Reliability

Inter-rater reliability did not achieve the desired level of 0.7 at the scale level for any of the paired observers. ICC values ranged from 0.42 to 0.58. When traineesô mean scores were used, ICC values were more acceptable and ranged between 0.59 and 0.82. This suggests observers may reliably agree based on overall performance of non-technical skills. Average scores were also used by Graham et al when evaluating the ANTS assessment tool because it eliminates error due to misclassification (Graham et al., 2010).

The reliability results compare favourably with other research undertaken in the UK. In a large study designed to look at implementing WBAs in medical specialities across the UK the authors commented that the mini-CEX was subject to ñsignificant assessor errorò (Wilkinson, 2008). They also used mean scores across the whole of each instrument (mini-CEX, DOPS and MSF). Similarly, poor inter-rater reliability has been documented for the CEX using single encounters (Kroboth et al., 1992). Due to the small sample size, ICCôs calculated at the skill level had very large confidence intervals making meaningful interpretation difficult. However, Workload Management and Team Building had mainly high ICCôs, suggesting good agreement,

whereas Authority and Assertiveness and SA Informing the Team generally had low ICCôs, suggesting poor agreement. These results show areas where the tool may need improvement and highlight skills that need additional attention when assessors are trained to use the tool.

Test retest reliability was not adequate when analysis used 12 individual skills (0.26) but showed good correlation when using mean scores (0.70). For comparison, test retest analysis based on a single encounter, and using averaged scores, for the mini-CEX was calculated to be only 0.23 (Norcini et al., 1995).

Categorising rater error

A number of sources of error were identified to explain variation in ratings between observers which can be categorised into those relating this toolôs specific design, those relating to the training of observers and errors associated with any behavioural marker system. Problems with overlapping skill definitions, or skills that were not clearly understood, are weakness of the tool that could be improved with revision or extension of the definitions and exemplar behaviours. However, if a behaviour is noted but incorrectly categorised the trainee will still gain educational benefit from the assessment process as long as an adequate debrief and discussion takes place.

Training in the use of the tool is also likely to reduce errors due to misunderstandings of skill definitions. Observers may benefit from receiving more advanced training on sources of bias and cognitive error, such as leniency/ severity and the halo/horn effect, to make ratings more objective. Disagreements between observers regarding what constitutes good and poor behaviour is likely to improve as EM physicians receive more training in non-technical skills in general.

With any behavioural marker system, the observer is vulnerable to making simple mistakes, may miss some behaviours or may fail to comment on some observed behaviours. This is particularly so with a tool used in the ED where observers are subject to high volumes of information and may suffer observer fatigue. Another weakness of any behavioural marker systems is the difficulty interpreting skills that

rely on cognitive processes. However, this can be mitigated by good debriefing, where registrars have an opportunity to discuss the reasoning behind observed behaviour. Errors that occurred because the observer was not familiar with the workings of the department are a consequence of the research process and would not happen during a true WBA.

Generalisability Theory to assess WBAs

Behavioural marker systems developed in other areas of medicine such as anaesthetics and surgery have focussed on measuring of inter-rater reliability of an isolated event, as used in this study (Fletcher et al., 2003, Yule et al., 2008, Sevdalis, 2009). This is necessary for high-stakes assessments that occur on a single occasion, which is the case for licensing UK pilots. However, the reliability of most WBAs used in UK healthcare has been based on Generalisability Theory (Wilkinson, 2008, Murphy et al., 2009). This uses a multiple analysis of variance to indicate the magnitude of errors from various specified sources, such as inter-observer variation, intra-observer variation and case specificity, which describes a traineeôs variation in performance from one challenge to another (Crossley et al., 2002). Unlike classic reliability theory which measures these sources of error one by one, as shown in this study, Generalisability Theory allows simultaneous estimation of the components of variance. This is achieved by multiple sampling using a number of different raters on a number of different occasions.

However, this method requires large number of assessors and each trainee is assessed on multiple occasions and is therefore time-consuming and expensive. Furthermore, the complexity of Generalisability Theory is widely acknowledged and there is limited experience amongst statisticians. Reliability based on a single encounter has been shown to be inadequate for a number of commonly used WBAs (Murphy et al., 2009, Wilkinson, 2008, Norcini et al., 1995). All studies have concluded that a number of assessments, which can be calculated using Generalisability Theory, are required to achieve adequate reliability. For example, at least eight different assessors observing at least two encounters each are recommended for the mini-CEX (Wilkinson, 2008).

Problems with reliability are acknowledged in the document ôDeveloping and maintaining an assessment system ï a PMETB guide to good practiceô (PMETB, 2007) in which authors accept that WBAs may not achieve adequate reliability compared to other assessments. This document also warns against placing too much emphasis on numerical scores and highlights the importance of narrative information. For these reasons, documenting specific example of observed behaviour and discussing this in the debrief session is considered vital for the assessment of ED traineesô non-technical skills. Overall, the results of this multicentre study compare favourably with other WBAs and suggest further research, using Generalisability Theory, may be warranted.

9.5.2 Observability

Across all 96 assessments, frequency of observation for each of the 12 skills ranged from 68% for Authority & Assertiveness to 100% for Maintenance of Standards. The skills observed less frequently were Supervision & Feedback, Authority & Assertiveness, Outcome Review and Anticipation. There are likely to be a number of reasons for this. For example, Anticipation is likely to be a highly cognitive process and therefore is not easily observed. Observation of the skill Supervision & Feedback is likely to vary depending on whether the registrar is taking on a leadership role or whether the doctor is working more independently. In practice, assessment of this skill can be targeted, if necessary, depending on what role the doctor is fulfilling during that particular shift. The skills Outcome Review, and Authority & Assertiveness may be only required under certain circumstances and may not be captured in a single encounter. The frequency of ónot observedô is likely to be overestimated. For example, it is possible that there were missed occasions where Anticipation should have been demonstrated by the registrar but was not. However, it is much harder for assessors to recognise the absence of behaviour (unless it leads to a negative outcome), especially during assessments that commonly have very high information load. Despite this, all skills were observed with sufficient frequency such that multiple encounter assessments, as previously recommended, is likely to allow adequate assessment of the 12 skills.

9.5.3 Description of range of skills

This observational study has provided a unique and detailed view of the non-technical skills displayed by ED registrars in the UK. Traditionally most research on these skills in the ED has centred on the single, critically ill patient. The examples described in this study give a broader view of skills required. Indeed, registrars were observed infrequently in the resuscitation room. The results outline examples of good and poor behaviour that would be particularly beneficial to newly appointed registrars. Whilst there were clearly many other examples observed during the study, only a proportion could be listed.

The results describing average ratings for each skill may have application for training non-technical skills, especially when combined with qualitative analysis of field notes. For example, Maintenance of Standards had the lowest mean rating (mean 4.8) and review of field notes highlighted recurrent issues related to hand hygiene, safety of sharps and documentation. Another skill with a relatively low average rating was Supervision and Feedback and observers frequently commented on missed opportunities for teaching. Similarly, DM Selecting and Communicating Options scored relatively low and field notes suggested that registrars often failed to adequately explain reasoning for decisions or discuss risks within the team. Training courses aimed at improving non-technical skills need to explicitly address these issues. Results showing dispersion are helpful to demonstrate which skills are likely to have higher discriminatory value for separating poorly performing trainees from those performing well. If only limited time is available to perform an assessment, focussing on skills with greater dispersion, such as Maintenance of standards and Workload management, may be most useful.

9.5.4 Comparison of hospitals

Comparing hospitals on a skill level allows departments to develop training programmes focussed at the strengths and weaknesses of their own trainees. This data has been made available to the clinical leads of EDs participating in the study for dissemination to other staff. Ratings for two skills were found to vary significantly

between the four departments. The skill Supervision & Feedback was rated significantly lower in the department that had the strongest consultant ‘shop-floor’ presence. Situational Awareness (Gathering Information) was rated significantly higher is the department where registrars took on a more formal leadership role. This gives some insight into the potential effects of shifting leadership roles from the ED registrar to the consultant and is an issue that requires further research.

9.5.5 Strengths and Limitations

Unlike the evaluation of some other behavioural marker systems developed for use in medicine (Fletcher et al., 2003, Yule et al., 2008), this study uses live scenarios rather than scripted videos viewed in a controlled setting. The results of this study are therefore more easily translated into a workplace environment. Furthermore, the setting included a variety of departments, both district general hospitals and teaching hospitals, where the level of consultant presence, and therefore the role of the registrar, varied considerably.

The use of registrar observers may be criticised as, to date, WBAs for registrars are exclusively undertaken by consultants or associate specialists. However, the ability to judge non-technical skills is not limited to consultants and this study reveals some benefits of trainees rating colleagues of a similar level. For example, one of the disadvantages of any WBA is the Hawthorne effect and this is likely to be lower when observed by someone of a similar grade. Also, observers found the process of observing and assessing an educational experience in itself and this is further explored in Chapter 10.

The numbers used in the study are small compared to some other studies used to evaluate tools for use in the workplace (Eisenberg, 2005, Wilkinson, 2008, Murphy et al., 2009). As a result, confidence intervals were large and it is difficult to make any concrete conclusions. However, reliability results were encouraging and suggest that a larger evaluation study, using Generalisability Theory, is warranted.

9.5.6 Summary

This multicentre observational study has demonstrated inter-rater and test retest reliability that compares favourably with other WBAs. A number of sources of error have been identified that relate to the skill definitions, training issues as well as problems that are encountered with any behavioural marker system. Observability of each skill was adequate and could be improved by targeting observations to focus on particular skills of interest. Examples of good and poor behaviour give a broad description of non-technical skills required by ED registrars working in departments in the UK. Comparison of hospitals allowed identification of issues related to consultant supervision and leadership roles that require further research. Also, a larger evaluative to assess the tool using Generalisability Theory is warranted. The following chapter summarises the main research findings included in this thesis and considers the practical application of these findings.

10 General Discussion

CHAPTER OVERVIEW

This concluding chapter reviews the main research findings of each of the studies included in the thesis. This is followed by consideration of the limitations of these studies and reflection on alternative approaches that could have been taken. Suggestions for future research are discussed and the practical applications of the research findings are presented.

10.1 Research Findings

The initial objective of this thesis was to consider ways the ED team could function better in order to provide consistently high quality care for patients. An initial, exploratory interview study with 22 staff identified the main stressors in the ED. These included the ó4 hourô target, excess workload, staff shortages and lack of teamwork, both within the ED and with in-patient staff. Participants described the impact of high pressure on communication practices, departmental overview and the management of staff and patients. The study revealed potential safety issues, for example, when junior staff feel unable to ask for help or believe it is unacceptable to question senior colleagues. Suggested interventions related to leadership and teamwork training, advertising staff breaks, efforts to help staff remain calm under pressure and addressing team motivation. Many of the respondentsô comments related to ónon-technical skillsô such as assertiveness, communication and situational awareness. In particular the study highlighted the importance of clinical leadership and the lack of preparation for this role. Misunderstandings between senior and junior staff and the ad hoc way that staff accrue these skills further indicated the need to focus an intervention that would help raise the profile of non-technical skills in EM. Whilst all team members in the ED clearly require sound non-technical skills, focus was drawn in particular to trainee registrars because of their vital leadership role and the apparent lack of leadership training and guidance at this level.

In order to improve non-technical skills in the ED, it was first necessary to identify the salient non-technical skills. An initial literature search identified only 11 articles describing how non-technical skills contribute to safety and error in the ED. However, analysis revealed a number of recurring themes that allowed identification of nine essential non-technical skills. These are communication, workload management, anticipation, situational awareness, supervision & feedback, leadership, maintenance of standards, authority & assertiveness and decision-making. A secondary search, using these nine skills and related terms, uncovered a further 21 relevant papers to enable a comprehensive review of the literature. This provided an evidence-base for the development of a tool to assess registrarsô non-technical skills in the ED.

Triangulation of data from a number of sources was used to produce a behavioural marker system for use in formative assessment of EM non-technical skills. In addition to a review of the literature, curricula and existing methods of assessment were examined and data were synthesised to produce a provisional tool. This was subsequently revised after reanalysis of staff interviews and a series of preliminary observations in the ED. It became clear that although leadership in the ED was vital, this was not an isolated skill but instead consisted of a number of skills such as situational awareness and decision making. Therefore, leadership was not assessed separately and instead each skill definition was devised to incorporate relevant leadership elements. Also, situational awareness and decision making were complex skills that could not be easily assessed under a single heading so these skills were divided into sub-skills to be assessed separately. An assessment tool comprising 12 skills was produced for use in further evaluative studies.

Content validity of the tool was assessed using a survey of 148 emergency medicine staff. The overall CVI for the questionnaire was 0.81. Nine of the 36 behavioural marker statements evaluated failed to achieve the recommended CVI>0.75 and two had significant missing data. Qualitative analysis of responses enabled revision of problem items to produce a revised list of behavioural markers. The tool was judged not to have any major omissions of non-technical skills. Responses from participants revealed the need to provide more structured and consistent assessment of traineesô non-technical skills in the ED.

A training course to teach EM physicians to use the tool achieved good levels of inter-rater reliability under experimental conditions (11 of the 12 skills reached values >0.7). However, 3 hoursô training provided wide variation in accuracy results for individual skills ï between 14% and 84% correct to within 1 point of the expertsô agreed rating. Analysis of assessment sheets revealed ratings may be subject to leniency bias and other sources of error. Furthermore, feedback from participants highlighted the desire for more extensive training and opportunities to practice in order to feel confident using the tool for assessment. Responses regarding acceptability were positive and the training day encouraged participants to reflect on their own non-technical skills. Feedback from participants in this study and the validation survey highlighted that one of the main barriers to wider implementation of the tool is concerns regarding how time consuming the process would be in practice.

Evaluation of inter-rater and test retest reliability in the multicentre study produced values that compare favourably with other workplace based assessments. Observability of each skill was adequate and could be improved by targeting observations to focus on particular skills of interest. The study produced a vast amount of qualitative data related to non-technical skills in the ED that described good and poor behaviour displayed by UK registrars. Comparison of hospitals allowed identification of potential issues related to consultant supervision and leadership roles that may have accounted for the significant difference in skill level displayed in some departments.

Data from the EM physician training day and multicentre study provided important insight into sources of rating error for the tool. These included rater bias (e.g. failure to rate objectively based on observed behaviour), training issues (e.g. misunderstandings related to skill definitions or apparent overlapping definitions), simple mistakes (e.g. observer saw a behaviour but failed to comment) as well as disagreement between observers regarding what constitutes good and poor behaviour and problems that may be encountered with any observational assessment (e.g. concealed cognitive processes).

The research findings described in this thesis provide an important foundation for developing non-technical skills in the ED which will help the team to provide effective clinical care under challenging conditions.

10.2 Limitations and alternative approaches

Whilst a great deal of important data were generated during the course of these studies, it is important to bear in mind the limitations of the methods chosen. Some shortcomings have already been mentioned in previous chapters but here alternative approaches are considered.

The primary objective of the staff interview study was to explore ways in which the ED team could function better under pressure. This was integral to providing direction for future studies. The transcripts were subsequently used to gain an understanding of the range of non-technical skills that were applicable to staff working in the ED. Whilst this was considered to be an effective use of this rich source of data, the original topic guide was not designed specifically to illicit examples of non-technical skills. An alternative would have been to conduct cognitive task analysis interviews (Flanagan, 1954). This involves asking participants to recount a specific memorable incident and this is then further discussed and probed using a predefined interview procedure. Although this may have produced additional insight into the role of non-technical skills in the ED, this type of interview is usually undertaken by trained psychologists and would have required specific training and practice that was not possible within the limited time constraints of an MD. Furthermore, interviews revealed that few staff could recall an error with sufficient detail to undertake cognitive task analysis.

The literature review may not have encompassed the entire evidence-base for non-technical skills related to safety in the ED. If extra time and resources were available, a more extensive secondary search, using additional search terms, may have retrieved other useful papers. This could have been combined with a quality assessment of papers to exclude studies that were of a poor standard. Whilst this may have

produced a more complete summary of the literature, it is unlikely a completely different set of skills would have been identified.

The initial studies used to develop the assessment tool were all based in two London teaching hospitals, therefore, subsequent studies needed to encompass the wider EM community. This was achieved by surveying the opinion of 148 ED staff from around the country using opportunities at various training days. However, as participants were not gathered for the sole purpose of validating the tool, the questionnaire was of a limited length and not all behavioural marker statements were included. An alternative would have been to gather a smaller focus group to evaluate the tool in full. This may have provided a more in depth assessment of the skills and behaviours but is more susceptible to sampling bias and it would have been costly to reimburse travel from around the UK.

Evaluation of the training day for EM physicians revealed that participants felt further training was required and this was reflected in the inadequate accuracy results for this study. Alternatively, one or two days of training could have been given which would have provided a more complete picture of what could be achieved with extensive training. This would also have allowed more than one scenario to be assessed for data collection, enabling averaging of results over several tests. However, in practice, fewer people may have been willing to attend a training day focussed entirely on a new method of assessment.

The sole use of registrar observers in the multicentre study may be criticised as it is not representative of how most workplace based assessments are currently undertaken. Ideally, a combination of consultant and registrar observers would have been used but there would be obvious difficulties recruiting consultant researchers to such a time-consuming study.

Although these points highlight a number of alternative, and potentially better, approaches to the studies described in this thesis, most suggestions are prohibitively costly or time-consuming and could, therefore, not have been implemented.

10.3 Future Research

10.3.1 Piloting the tool

Firstly and most importantly, a larger evaluative study of the tool is required to more fully assess issues such as rating error and determine the number of assessments required to gain adequate levels of reliability. As previously discussed, a study design based on Generalisability Theory is likely to be the best approach. Such a study would require large-scale training of assessors in order to pilot use of the tool within a deanery, for example. The preliminary evaluative studies included in this thesis provide good supporting evidence that this type of resource-demanding study is warranted.

10.3.2 Use in the resuscitation room

It is also worthwhile considering whether the assessment tool could be modified and validated for use in the resuscitation room. This would expand application of the tool into a simulation setting. Although studies included in this thesis encompassed all aspects of the ED registrarsô role, which included resuscitation, an observational study focussing more specifically on the management of critically ill patients would be beneficial. For example, the twelve non-technical skills included in the assessment tool could be used to structure observation of resuscitation scenarios and expand the list of exemplar behaviours.

10.3.3 Other healthcare professionals

Future research should also encompass other health care professionals working in the ED such as nursing staff and junior doctors as well as ED consultants. These groups form a vital part of the ED team and assessment of their non-technical skills is clearly also essential. Indeed, assessment of consultantsô non-technical skills may conceivably form part of Revalidation in the near future. Some studies used in this

thesis, such as the interviews and literature review, would be applicable for other heath care groups. However, further observational studies would be required to identify specific behaviours relevant to each specific group.

10.3.4 Non-technical skills and outcome

Now that a tool has been established to assess non-technical skills in the ED, opportunities are available to develop a deeper understanding of how non-technical skills contribute to outcome. This is most challenging as there is no easy outcome measure in the ED. Whilst such studies would be complex, possible outcome measures could include 4 hour breaches, staff attitude or observation of errors. For example, one suggestion would be to assess individualsô non-technical skills using the tool and use a second, blinded observer to concomitantly assess errors and technical faults. Another study could correlate staff attitudes (related to safety or morale, for example) to level of non-technical skill performance within the team. Also, the tool could be used to structure observations and investigate the differences between expert and novice ED physicians, to aid and inform training.

10.3.5 Other future research

Individual studies included in this thesis have also inspired ideas for other areas of future research. For example, the interview study revealed the need to explore training initiatives focussed on helping staff to recognise and deal with stress in the ED. Also, the literature review emphasised the need for a series of further reviews, each focussing on a specific skill, in order to gain a full understanding of how these skills contribute to error in the ED. It is vital that future research is undertaken to continue to raise the profile of non-technical skills and highlight the important contribution these skills make to patient safety.

10.4 Application of Research Findings

In isolation, the results of each of the studies have a number of applications within the ED community and have been discussed within the relevant chapter. For example, the interview study highlighted simple practical interventions such as advertising staff breaks and improving team relations with ward staff. The identification of key EM non-technical skills in earlier studies could help focus ideas for safety interventions, such as developing cognitive aids to assist decision making or introducing structured handover proformas to improve communication. Explicitly listing key skills and associated behaviours, based on empirical evidence, will help trainees better understand their role in the department. More directly, the assessment of non-technical skills in the ED has far reaching implications in terms of education and training as well as clinical application.

10.4.1 Formative assessment of trainees

The most important and immediate application of the research findings is the formative assessment of ED registrarsô non-technical skills. Formative assessment is important as it promotes reflection, guides future learning, and facilitates monitoring of progress (Cox et al., 2007). Reflection is particularly important as it is likely many ED staff do not take time out their busy day to stop and consider their performance of non-technical skills. Furthermore, there is evidence that physicians have only a limited ability to self-assess competence (Davis et al., 2006) and this is likely to be exaggerated for interpersonal skills such as team building and communication. Formative assessment is integral to medical education as it reinforces the motivation to learn and can encourage professional development (Friedman Ben-David, 2000).

10.4.2 Other assessment applications

The tool was developed with a specific use in mind ï i.e. the observational assessment of ED registrarsô non-technical skills in the workplace over a 1 hour period.

However, alternative applications should also be explored. Indeed, after the training day, some attendees provided feedback suggesting other uses they had tried. For example, one consultant suggested observing a trainee for a shorter time period and targeting specific events such as observing a registrar as they review a patient plan with a more junior doctor (this could assess decision making, maintenance of standards and teaching & feedback).

Another possibility would be to train senior nurses to assess registrarsô non-technical skills as they are often well placed to evaluate performance, especially on night shifts when consultants may be absent. Currently, most assessment of ED traineesô non-technical skills in the workplace is informally undertaken by consultants as they work alongside trainees, using intermittent observation. Whist the disadvantages of this form of assessment were identified and discussed in Chapter 8, ad hoc observation could be improved by providing structure using this assessment tool. This would allow more detailed feedback on specific skills and potentially encourage evaluation of a broader range of behaviours.

10.4.3 Trainees as assessors

The studies included in this thesis have also led to consideration of a novel approach for the assessment of ED trainees. It is widely accepted that ED consultants are extremely busy having to cope with a combination of clinical duties and management of the department as well as educational commitments. This research highlighted the potential to use registrars to assess other registrars.

In the early stages of the observational research, I began to reflect on my own behaviour and it became clear that observing others may have benefits for the assessor. Respondents at the training day also described how the process of learning to use the tool helped them to reflect and identify behaviours that may need improvement or modification.

Subsequently, registrar observers were used in the multi-centre study and they provided positive feedback related to the benefits of observing colleaguesô non-

technical skills. Two of the three volunteer observers recognised clear advantages of using registrars to assess and provide feedback on non-technical skills in the ED. In particular, they appreciated the need to prepare for consultant life when it would be necessary to undertake assessments and acknowledged that training and preparation for this role was lacking. They noted the difficulty of giving ónegativeô feedback, especially related to interpersonal skills, as they had no real previous experience in this. The third observer was less convinced of the benefits of assessing colleagues but commented on the potential for senior registrars to observe and assess more junior registrars. All three observers described how taking part in the study had led them to reflect on their own behaviour. For example, taking more opportunities to teach, learning to delegate more and acknowledging there are more diplomatic ways to get your point across. This resulted from observing poor behaviour as well as aspiring to model examples of good practice. Observers also acknowledged that there may need to be a culture change, where it is more acceptable to comment on the performance of colleagues of a similar grade. However, with the imminent introduction of Revalidation in the UK, EM physicians are likely to become more familiar with peer evaluation.

10.4.4 Training

Primarily, this research was undertaken to produce an assessment tool that encompasses EM non-technical skills. However, the identification of core non-technical skills and associated behaviours has essentially produced an EM non-technical skills curriculum. This will enable the development of tailor-made training programmes focussing on skills specifically relevant to EM rather than borrowing and translating CRM principles from aviation. Furthermore, CRM and other current training initiatives tend to concentrate on the management of a single critically ill patient and the research included in this thesis covers a wider range of both emergency and routine work, as well as leadership roles.

Alternative training applications include using the tool to structure discussion in a teaching session using the filmed scenarios presented in Chapter 8 as a resource. The non-technical skill taxonomy described in this thesis could also help staff to learn

from critical incidents or adverse events. This could involve reviewing the list of skills to determine if failure of a skill could have contributed to the incident or where better skill performance may have averted an error.

The assessment tool could also be used for research to evaluate the impact of training or other intervention, by measuring the progressive development of individuals.

10.4.5 Selection of trainees

The identification of key non-technical skills for emergency physicians may have implications for the selection of trainees. Following the lead of the aviation industry, other specialities have considered broadening candidate selection criteria to include non-technical skills, such as problem solving and communication (Paice et al., 2010, Patterson, 2005). Indeed, the use of a selection centre testing non-technical skills for recruitment into anaesthesia training has been evaluated with positive results (Gale, 2010). The future selection of ED trainees may benefit from assessment of leadership and other non-technical skills to ensure the most appropriate applicants are chosen.

The variety of assessment and training applications described emphasise the relevance of the research included in this thesis. Whilst further evaluative studies are needed, development of this behavioural marker system provides an important step towards improving non-technical skills and safety in the ED.

References

ABELL, N., KOCH, G.G. 2009. *Developing and validating rapid assessment instruments,* New York, Oxford University Press.

ADAMSKI, P., BRESSLER, H., CHRISTENSEN, D., CROTEAU, R., MCKENNA, C., NORRIS, E., PEW, C., SMITH, L. 2000. *What Every Hospital Should know About Sentinel Events,* Oakbrook, Il, Joint Commission on Accreditation of Healthcare Organizations.

ANDERSEN, P. O., JENSEN, M.K., LIPPERT, A., ØSTERGAARD, D. 2010. Identifying non-technical skills and barriers for improvement of teamwork in cardiac arrest teams. *Resuscitation,* 81**,** 695-702.

ANDERSEN, P. O., MAALØE, R. & ANDERSEN, H. B. 2010. Critical incidents related to cardiac arrests reported to the Danish Patient Safety Database. *Resuscitation,* 81**,** 312-316.

APKER, J., MALLAK, L.A., GIBSON, S.C. 2007. Communicating in the "Gray Zone": Perceptions about Emergency Physician-hospitalist Handoffs and Patient Safety. *Academic Emergency Medicine,* 14**,** 884.

ARORA, S., SEVDALIS, N., NESTEL, D., TIERNEY, T., WOLOSHYNOWYCH, M. & KNEEBONE, R. 2009. Managing intraoperative stress: what do surgeons want from a crisis training program? *The American Journal of Surgery,* 197**,** 537-543.

BAKER, D. P., GUSTAFSON, S., BEAUBIEN, J., SALAS, E., BARACH, P. 2005. *Medical Teamwork and Patient Safety:The Evidence-based Relation*. Agency for Healthcare Research and Quality. Available: http://www.ahrq.gov/qual/medteam/ [Accessed 14th January 2010].

BAKER, G. R., NORTON, P.G., FLINTOFT, V. ET AL 2004. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. *Canadian Medical Association. Journal,* 170**,** 1678.

BEACH, C., CROSKERRY, P., SHAPIRO, M. 2003. Profiles in patient safety: Emergency care transitions. *Academic Emergency Medicine,* 10**,** 364-7.

BOND, W. F. 2004. Using simulation to instruct emergency medicine residents in cognitive forcing strategies. *Academic Medicine,* 79**,** 438.

BOREHAM, N. C., SHEA, C.E., MACKWAY-JONES, K. 2000. Clinical risk and collective competence in the hospital emergency department in the UK. *Social Science & Medicine,* 51**,** 83-91.

BRODIE, A., KINROSS, J., BAILEY, M., AGGARWAL, R. AND VINCENT, C. 2009. Using Failure Mode and Effects Analysis to identify hazards within resuscitation. *International Journal of Risk & Safety in Medicine,* 21**,** 201-215.

BURKE, C. S. 2006a. What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly,* 17**,** 288-307.

CAMPBELL, S. G. 2007. Profiles in patient safety: A "perfect storm" in the emergency department. *Academic Emergency Medicine,* 14**,** 743.

CARTHEY, J. 2003. The role of structured observational research in health care. *Quality & Safety in Health Care,* 12**,** ii13-16.

CHISHOLM, C. D., COLLISON E.K., NELSON, D.R., CORDELL, W.H. 2000. Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine,* 7**,** 1239-1243.

CLOUGH, C., LEWIS, P., THOMAS, L., WADE, W. 2009. Medical Leadership Curriculum. Academy of Royal Medical Colleges.

COIERA, E. W., JAYASURYIA, R.A., HARDY, J., BANAAN, A., THORPE, M.E.C 2002. Communication loads on clinical staff in the emergency department. *Medical Journal of Australia,* 176**,** 415-418.

COLLEGE OF EMERGENCY MEDICINE Generic Skills Curriculum [Online]. College of Emergency Medicine. Available: http://www.collemergencymed.ac.uk/asp/curriculumall.asp?id=7 [Accessed 1st Sept 2009].

COOK, D. A. 2009. Does scale length matter? A comparison of nine- versus five-point rating scales for the mini-CEX. *Advances in Health Sciences Education,* 14**,** 655.

COOPER, S., CANT, R., PORTER, J., SELLICK, K., SOMERS, G., KINSMAN, L. & NESTEL, D. 2010. Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation,* 81**,** 446-452.

COOPER, S., ENDACOTT, R & CANT, R 2010. Measuring non-technical skills in medical emergency care: a review of assessment measures. *Open Access Emergency Medicine,* 2**,** 7-16.

COOPER, S., O'CONNELL, J., JENKIN, A., BULGAR, S. 2007. Collaborative practices in unscheduled emergency care; role and impact of the emergency care practitioner (ECP). *Emergency Medicine Journal,* 24**,** 630-633.

COOPER, S., WAKELAM, A. 1999. Leadership of resuscitation teams; "Lighthouse Leadership". *Resuscitation,* 42**,** 27-45.

COSBY, K., CROSKERRY, P. 2009. The Nature of Emergency Medicine. *In:* CROSKERRY, P., COSBY, K., SCHENKEL, S.M. & WEARS, R.L. (ed.) *Patient Safety in Emergency Medicine.* Philadelphia, PA: Lippincott, Williams & Wilkins.

COSBY, K. S. 2003. A framework for classifying factors that contribute to error in the emergency department. *Annals of Emergency Medicine,* 42**,** 815-823.

COSBY, K. S. & CROSKERRY, P. 2003. Patient Safety: A Curriculum for Teaching Patient Safety in Emergency Medicine. *Academic Emergency Medicine,* 10**,** 69-78.

COSBY, K. S., ROBERTS, R., PALIVOS, L, ROSS, C ET AL. 2008. Characteristics of patient care management problems identified in emergency department morbidity and mortality investigations during 15 years. *Annals of Emergency Medicine,* 51**,** 251-261.

COWAN, J., HASLAM, J. 2006. Resuscitation: Turning guidelines into practice. *Clinical Governance,* 11**,** 160-165.

COWAN, R. M. 2005. Clinical review: Emergency department overcrowding and the potential impact on the critically ill. *Critical Care,* 9**,** 291.

COX, M., IRBY, D. M. & EPSTEIN, R. M. 2007. Assessment in Medical Education. *New England Journal of Medicine,* 356**,** 387-396.

CRANDALL, B., KLEIN, G.A., HOFFMAN, R.R. 2006. *Working Minds. A Practitioner's Guide to Cognitive Task Analysis,* Cambridge, MA, MIT Press.

CRICHTON, M. 2001. Training for decision making during emergencies. *Psihološka Obzorja/Horizons of Psychology,* 10**,** 7.

CROSKERRY, P. 2000. The feedback sanction. *Academic Emergency Medicine,* 7**,** 1232.

CROSKERRY, P. 2003. Cognitive forcing strategies in clinical decisionmaking. *Annals of Emergency Medicine,* 41**,** 110.

CROSKERRY, P. & SINCLAIR, D. 2001. Emergency medicine: A practice prone to error? *CJEM. Canadian Journal of Emergency Medical Care,* 3**,** 271-276.

CROSSLEY, J., DAVIES, H., HUMPHRIS, G. & JOLLY, B. 2002. Generalisability: a key to unlock professional assessment. *Medical Education,* 36**,** 972-978.

DARZI, A. 2008. High Quality Care for All - NHS Next Stage Review Final Report. Department of Health.

DAVIS, D. A., MAZMANIAN, P. E., FORDIS, M., VAN HARRISON, R., THORPE, K. E. & PERRIER, L. 2006. Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. *JAMA: The Journal of the American Medical Association,* 296**,** 1094-1102.

DAVIS, K., DREY, N., GOULD, D. 2009. What are scoping studies? A review of the nursing literature. *International Journal of Nursing Studies,* 46**,** 1386-1400.

DEPARTMENT OF HEALTH. 2000. An organisation with a memory: Report of an expert group on learning from adverse events in the NHS. London: Department of Health.

DEROSIER, J., STALHANDSKE, E., BAGIAN, J. P. & NUDELL, T. 2002. Using Health Care Failure Mode and Effect Analysis: The VA National Center for Patient Safety's Prospective Risk Analysis System. *Joint Commission Journal on Quality and Patient Safety,* 28**,** 248-267.

DICKENSON, T. L. A. M., R.M. 1997. A conceptual framework for team measurement. *In:* BRANNICK, M. T., SALAS E. AND PRINCE, C. (ed.) *Team Performance Measurement: Theory, Methods and Applications.* Mahway NJ: Erlbaum.

DRACH-ZAHAVY, A. & FREUND, A. 2007. Team effectiveness under stress: a structural contingency approach. *Journal of Organizational Behavior,* 28**,** 423-450.

DRISKELL, J. E., SALAS, E., JOHNSTON, J. 1999. Does stress lead to a loss of team perspective? *Group Dynamics,* 3**,** 291.

DUWE, B., FUCHS, B. D. & HANSEN-FLASCHEN, J. 2005. Failure mode and effects analysis application to critical care medicine. *Critical Care Clinics,* 21**,** 21-30.

EDMONDSON, A. 2003. Speaking up in the operating room: how team leaders promote learning in interdisciplinary action teams. *Journal of Management Studies,* 40**,** 1419-52.

EIGNOR, D. R. 2001. Standards for the Development and Use of Tests: The Standards for Educaitonal and Psychological Testing. *European Journal of Psychological Assessment,* 17**,** 157-163.

EISENBERG, E. M., MURPHY, A. G., SUTCLIFFE, K., WEARS, R. L., SCHENKEL, S., PERRY, S., VANDERHOEF, M. 2005. Communication in Emergency Medicine: Implications for Patient Safety. *Communication Monographs,* 72**,** 390 - 413.

ELSHOVE-BOLK, J., SIMONS, M., CREMERS, J. ET AL 2004. A description of emergency department-related malpractice claims in The Netherlands: closed claims study 1993-2001. *European Journal of Emergency Medicine,* 11**,** 247.

EPPICH, W. J. 2008. Team training: Implications for emergency and critical care pediatrics. *Current Opinion in Pediatrics,* 20**,** 255.

FAMULARO, G. 2000. Clinical errors in emergency medicine: experience at the emergency department of an Italian teaching hospital. *Academic Emergency Medicine,* 7**,** 1278.

FATOVICH, D. M. 2003. Entry overload, emergency department overcrowding, and ambulance bypass. *Quality & Safety in Health Care,* 20**,** 406.

FERNANDEZ, R., KOZLOWSKI, S.W.J., SHAPIRO, M.J. ET AL 2008. Toward a definition of teamwork in emergency medicine. *Academic Emergency Medicine,* 15**,** 1104.

FLANAGAN, J. 1954. The critical incident technique. *Psychological Bulletin,* 51(4)**,** 327-58.

FLETCHER, G., FLIN, R., MCGEORGE, P., GLAVIN, R., MARAN, N. & PATEY, R. 2003. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *Bitish. Journal of Anaesthetics,* 90**,** 580-588.

FLIN, R. 2009. Improving patient safety through training in non-technical skills. *BMJ,* 339**,** b3595.

FLIN, R., FLETCHER, G., MCGEORGE, P., SUTHERLAND, A. & PATEY, R. 2003. Anaesthetists' attitudes to teamwork and safety. *Anaesthesia,* 58**,** 233-242.

FLIN, R. & MARAN, N. 2004. Identifying and training non-technical skills for teams in acute medicine. *Quality & Safety in Health Care,* 13**,** i80-84.

FLIN, R., MARTIN, L., GOETERS, K., HOERMANN, J., AMALBERTI, R., VALOT, C. AND NIJHUIS, H 2003. Development of the NOTECHS (Non-Technical Skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety,* 3**,** 95-117.

FLIN, R., O'CONNOR, R., CRICHTON, M. 2008. *Safety at the Sharp End: A guide to non-technical skills,* Farnham, Ashgate.

FLIN, R., YULE, S. 2004. Leadership for safety: Industrial experience. *Quality & Safety in Health Care,* 13**,** i45-i51.

FORDYCE, J., BLANK, F.S.J., PEKOW, P. ET AL 2003. Errors in a busy emergency department. *Annals of Emergency Medicine,* 42**,** 324-333.

FOUSHEE, H. & HELMREICH, R. L. 1988. Group interaction and flight crew performance. *In:* WIENER, E. L. & NAGEL, D. C. (eds.) *Human factors in aviation.* San Diego, CA: Academic Press.

FRANKEL, A., GARDNER, R., MAYNARD, L., KELLY, A. 2007. Using the Communication and Teamwork Skills (CATS) Assessment to measure health care team performance. *The Joint Commission Journal on Quality and Patient Safety,* 33**,** 549.

FRIEDMAN BEN-DAVID, M. 2000. The role of assessment in expanding professional horizons. *Medical Teacher,* 22**,** 472-477.

GABA, D. M. 1992. Improving anesthesiologists' performance by simulating reality. *Anesthesiology,* 76**,** 491-494.

GABA, D. M. 1998. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. *Anesthesiology,* 89**,** 8.

GALE, T. C. E., ROBERTS, M.J., SICE, P.J., LANGTON, J.A., PATTERSON, F.C. ET AL. 2010. Predictive validity of a selection centre testing non-technical skills for recruitment to training in anaesthesia. *British Journal of Anaesthesia,* 105**,** 603-609.

GENERAL MEDICAL COUNCIL. 2006. *Good Medical Practice.* [Online]. Available: http://www.gmc-uk.org/static/documents/content/GMP_0910.pdf [Accessed 17.01.2011].

GLASZIOU, P., IRWI, L., BAIN, C., COLDITZ, G. 2001. *Systematic Reviews in Health Care: A Practical Guide,* Cambridge, Cambridge University Press.

GLAVIN, R. J. 2011. Skills, Training, and Education. *Simulaiton in Healthcare,* 6**,** 4-7.

GOSBEE, J. W. 2004. Conclusion: you need human factors engineering expertise to see design hazards that are hiding in "plain sight"! *The Joint Commission Journal on Quality and Patient Safety,* 30**,** 696-700.

GRAHAM, J., HOCKING, G. & GILES, E. 2010. Anaesthesia Non-Technical Skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? *British Journal of Anaesthesia,* 104**,** 440-445.

GRANT, J. 1990. A methodology for validating nursing diagnoses. *Advances in Nursing Science,* 12**,** 65.

GRANT, J. S. & DAVIS, L. L. 1997. Selection and use of content experts for instrument development. *Research in nursing & health,* 20**,** 269-274.

GROGAN, E. L. 2004. The impact of aviation-based teamwork training on the attitudes of health-care professionals. *Journal of the American College of Surgeons,* 199**,** 843.

GUION, R. M. 1965. *Psychometric Methods,* New York, McGraw-Hill.

GUYATT, G. H., OXMAN, A.D., VISIT, G.E., KUNZ, R, FALCK-YTTER, Y., ALONSO-COELLO, P., SCHUNEMANN, H.J. 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ,* 336**,** 924.

HAIG, K. M., SUTTON, S., WHITTINGTON, J. 2006. SBAR: A Shared Mental Model for Improving Communication Between Clinicians. *Joint Commission Journal on Quality and Patient Safety,* 32**,** 167-175.

HAMMAN, W. R., BEAUDIN-SEILER, B.M., BEAUBIEN, J.M., GULLICKSON, A.M. ET AL 2010. Using simulation to identify and resolve threats to patient safety. *The American Journal of Managed Care,* 16**,** e145.

HAWKINS, R. E. 2010. Constructing a validity argument for the mini-Clinical Evaluation Exercise: a review of the research. *Academic Medicine,* 85**,** 1453.

HAYNES, S. N., RICHARD, D.C.S, KUBANY E.S. 1995. Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment,* 7**,** 238-247.

HEALEY, A. N., UNDRE, S. & VINCENT, C. A. 2004. Developing observational measures of performance in surgical teams. *Quality & Safety in Health Care,* 13**,** i33-40.

HELMREICH, R. L., DAVIES, J.M. 1996. Human factors in the operating room: interpersonal determinants of safety, efficiency and morale. *Bailliere's Clinical Anaesthesiology,* 10**,** 277-295.

HELMREICH, R. L., ET AL 1999. The evolution of crew resource managment training in commercial aviation. *International Journal of Aviation Psychology,* 9**,** 19-32.

HENNEMAN, E. A. 2006. Strategies used by nurses to recover medical errors in an academic emergency department setting. *Applied Nursing Research,* 19**,** 70.

HICKS, C. M., BANDIERA, G.W., DENNY, C.J. ET AL 2008. Building a simulation-based crisis resource management course for emergency medicine, phase 1: Results from an interdisciplinary needs assessment survey. *Academic Emergency Medicine,* 15**,** 1136.

HOFF, T. J. 2008. How work context shapes physician approach to safety and error. *Quality Management in Health Care,* 17**,** 140.

HOFMANN, D., MORGESON, F 2004. The role of leadership in safety. *In:* BARLING, J. A. F., M. (ed.) *The Psychology of Workplace Safety.* Washington: APA Books.

HOHENHAUS, S. M., POWELL, S. & HASKINS, R. 2008. A Practical Approach to Observation of the Emergency Care Setting. *Journal of Emergency Nursing,* 34**,** 142-144.

HOLCOMB, J. B., DUMIRE, R.D., CROMMETT, J.W., STAMATERIS, C.E. ET AL 2002. Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *The Journal of Trauma,* 52**,** 1078.

HORWITZ, L. I., MEREDITH, T., SCHUUR, J.D. ET AL 2009. Dropping the Baton: A Qualitative Analysis of Failures During the Transition From Emergency Department to Inpatient Care. *Annals of Emergency Medicine,* 53**,** 701.
HOWARD, S. K., GABA, D. M., FISH, K. J., YANG, G. & SARNQUIST, F. H. 1992. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. *Aviation Space and Environmental Medicine,* 63**,** 763-70.

HULL, L., ARORA, S., KASSAB, E., KNEEBONE, R., SEVDALIS, N. (2011). Observational Teamwork Assessment for Surgery: Content Validation and Tool Refinement. Journal of the American College of Surgeons, 212(2), 234-243, e235.

JAMES, L. R., DEMAREE, R. G. & WOLF, G. 1984. Estimating within-group interrater reliability with and without response bias. *Journal of Applied Psychology,* 69**,** 85-98.

JANIS, I. & MANN, L. 1977. *Decision-making: A psychological analysis of conflict, choice and committment.*, New York: Free Press.

JANKOUSKAS, T., BUSH, M.C., MURRAY, B. ET AL 2007. Crisis resource management: evaluating outcomes of a multidisciplinary team. *Simulation in Healthcare,* 2**,** 96-101.

JENKINS, G. D., & TABER, T. D. 1977. A Monte Carlo study of factors affecting three indices of composite scale reliability. *The Journal of Applied Psychology,* 62**,** 392-398.

JENSEN, K., JANKOWSKI, N.W. 1991. *Handbook of Qualitative Methodologies for Mass Communication Research,* London, Routledge.

JOHNSON, G., WADE, W., BARRETT, J. & JONES, M. 2009. The Acute Care Assessment Tool: a new assessment in acute medicine. *The Clinical Teacher,* 6**,** 105-109.

KACHALIA, A., GANDHI, T.K., PUOPOLO, A.L. ET AL 2007. Missed and Delayed Diagnoses in the Emergency Department: A Study of Closed Malpractice Claims From 4 Liability Insurers. *Annals of Emergency Medicine,* 49**,** 196.

KAISSI, A., JOHNSON, T., KIRSHBAUM, M.S. 2003. Measuring teamwork and patient safety attitudes of high-risk areas. *Nursing Economics,* 21**,** 211.

KENDALL, D., SALAS, E. 2004. Measuring team performance: review of current methods and consideration of fuuture needs. *Advances in Human Performance and Cognitive Engineering Research,* 5**,** 307-326.

KENNEDY, T. J. T., REGEHR, G., BAKER, G.R. ET AL. 2009. It's a cultural expectation...' the pressure on medical trainees to work independently in clinical practice. *Medical Education,* 43**,** 645.

KILNER, E. & SHEPPARD, L. A. 2010. The role of teamwork and communication in the emergency department: A systematic review. *International Emergency Nursing,* 18**,** 127-137.

KILROY, D. A. 2006a. Clinical supervision in the emergency department: a critical incident study. *Emergency Medicine Journal* 23**,** 105-8.

KILROY, D. A. 2006b. The Foundation Programme and the emergency department: a review of the curriculum and experience of a UK pilot. *Emergency Medicine Journal,* 23**,** 167.

KIM, J., NEILIPOVITZ, D., CARDINAL, P., CHIU, M. & CLINCH, J. 2006. A pilot study using high-fidelity simulation to formally evaluate performance in the resuscitation of critically ill patients: The University of Ottawa Critical Care Medicine, High-Fidelity Simulation, and Crisis Resource Management I Study. *Critical Care Medicine,* 34**,** 2167-2174.

KLAMPFER, B., ET AL 2001. Enhancing Performance in High Risk Environments: Recommendations for use of behavioural marker systems. Zurich: GIHRE.

KLEIN, G., CALDERWOOD, R., MACGREGOR, D. 1989. Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man and Cybernetics,* 19**,** 462-472.

KLEIN, K. J., ZIEGERT, J.C., KNIGHT, A.R., XIAO, Y. 2006. Dynamic delegation: hierarchical, shared and deindividualised leadership in extreme action teams. *Administrative Science Quarterly,* 51**,** 590-621.

KLEINMAN, D. L. A. S., D. 1989. Team performance assessment in distributed decision-making. *Symposium on Interactive Networked Simulation for Training.* Orlando, FL: Institute for Simulation and Training, University of Central Florida.

KOBAYASHI, L. 2007. Multiple Encounter Simulation for High-acuity Multipatient Environment Training. *Academic Emergency Medicine,* 14**,** 1141.

KOSTOPOULOU, O. & SHEPHERD, A. 2000. Fragmentation of treatment and the potential for human error in neonatal intensive care. *Topics in Health Information Management,* 20**,** 78-92.

KOZLOWSKI, S. W., GULLY, S.M., NASON, E.R., SMITH, E.M. 1999. *Developing adaptive teams: a theory of compilation and performance across levels and time,* San Francisco, Jossey-Bass.

KROBOTH, F., HANUSA, B., PARKER, S., COULEHAN, J., KAPOOR, W., BROWN, F., KARPF, M. & LEVEY, G. 1992. The inter-rater reliability and internal consistency of a clinical evaluation exercise. *Journal of General Internal Medicine,* 7**,** 174-179.

LAUBER, J. 1984. Resource management in the cockpit. *Air Line Pilot,* 53**,** 20-23.

LAWSHE, C. H. 1975. A quantative approach to content validity. *Personnel Psychology,* 28**,** 563-575.

LAXMISAN, A., HAKIMZADAA, F., SAYAN, O.R ET AL 2007. The multitasking clinician: Decision-making and cognitive demand during and after team handoffs in emergency care. *International Journal of Medical Informatics,* 76**,** 801-811.

LAZARUS, R. S., FOLKMAN, S. 1984. *Stress, Appraisal, and Coping,* New York, NY, Springer.

LEAPE, L. L., BRENNAN, T.A., LAIRD, N., LAWTHERS, A.G., LOCALIO, A.R., BARNES, B.A., HEBERT, L., NEWHOUSE, J.P., WEILER, P.C., HIATT, H. 1991. The Nature of Adverse Events in Hospitalized Patients. *New England Journal of Medicine,* 324**,** 377-384.

LINCOLN, Y. S. 1985. Case reporting, Member checking and auditing. *Naturalistic Inquiry.* Newbury Park, CA: Sage.

LUBBERT, P. H. W., KAASSCHIETER, E.G., HOORNTJE, L.E. ET AL. 2009. Video registration of trauma team performance in the emergency department: The results of a 2-year analysis in a level 1 trauma center. *The Journal of Trauma,* 67**,** 1412-20.

LYNN, M. R. 1986. Determination and quantification of content validity. *Nursing Research,* 35**,** 382.

MAGID, D. J., SULLIVAN, A.F., CLEARY, P.D. ET AL 2009. The Safety of Emergency Care Systems: Results of a Survey of Clinicians in 65 US Emergency Departments. *Annals of Emergency Medicine,* 53**,** 715.

MALEC, J., TORSHER, L.C., DUNN, W.F., WIEGMANN, D.A., ARNOLD, J.J., BROWN, D., PHATAK, V. 2007. The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simulation in Healthcare,* 2**,** 4.

MANSER, T. 2009. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. *Acta Anaesthesiologica Scandinavica,* 53**,** 143-51.

MARKS, M. A. 2000. Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology,* 85**,** 971.

MARSCH, S. C., MULLER, C., MARQUARDT, K., CONRAD, G., TSCHAN, F., HUNZIKER, P.R. 2004. Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation,* 60**,** 51-6.

MARSCH, S. C., TSCHAN, F., SEMMER, N., SPYCHIGER, M., BREUER, M., HUNZIKER, P.R. 2005. Performance of first responders in simuated cardiac arrests. *Critical Care Medicine,* 33**,** 963-7.

MARTUZA, V. R. 1977. *Applying norm-referenced and criterion-referenced measurement in education,* Boston, Allyn and Bacon.

MCCULLOCH, P., MISHRA, A., DALE, T ET AL 2009. The effects of aviation-style non-technical skills training on technical performance and outcome in the operating theatre. *Quality & Safety in Health Care,* 18**,** 109-115.

MICHAELSON, M., LEVI, L. 1997. Videotaping in the admitting area: a most useful tool for quality improvement of the trauma care. *European Journal of Emergency Medicine,* 4**,** 94-96.

MILLER, G. A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 63**,** 81-97.

MILLER, G. E. 1990. The assessment of clinical skills/ competence/ performance. *Academic Medicine,* S63-7.

MISHRA, A., CATCHPOLE, K. & MCCULLOCH, P. 2009. The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality & Safety in Health Care,* 18**,** 104-108.

MITCHELL, L., FLIN, R. 2009. *Scrub Practitioners' List of Intra-operative Non-technical Skills (SPLINTS),* Farnham, Ashgate.

MOON, H., HOLLENBECK, J. R., HUMPHREY, S. E., ILGEN, D. R., WEST, B., ELLIS, A. P. J. & PORTER, C. O. L. H. 2004. Asymmetric adaptability: Dynamic team structures as one-way streets. *Academy of Management Journal,* 47, 681-695.

MOREY, J., SIMON, R., JAY, G.D., WEARS, R.L., MARY SALISBURY, KIMBERLY A DUKES, SCOTT D BERNS, 2002. Error Reduction and Performance Improvement in the Emergency Department through Formal Teamwork Training: Evaluation Results of the MedTeams Project. *Health Services Research,* 37**,** 1553-1581.

MÜLLER, M. P., HÄNSEL, M., STEHR, S. N., FICHTNER, A., WEBER, S., HARDT, F., BERGMANN, B. & KOCH, T. 2007. Six steps from head to hand: A simulator based transfer oriented psychological training to improve patient safety. *Resuscitation,* 73**,** 137-143.

MURPHY, D., BRUCE, D., MERCER, S. & EVA, K. 2009. The reliability of workplace-based assessment in postgraduate medical education and training: a national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education,* 14**,** 219-232.

MURRAY, W. B., FOSTER, P.A. 2000. Crisis Resource Management among strangers: Principles of organising a multidisciplinary group for Crisis Resource Managment. *Journal of Clinical Anaesthesia,* 12**,** 633-638.

NAFSI, T. 2007. Audit of deaths less than a week after admission through an emergency department: how accurate was the ED diagnosis and were any deaths preventable? *Emergency Medicine Journal : EMJ,* 24**,** 691.

NATIONAL AUDIT OFFICE. 2005. A Safer Place for Patients: Learning to improve patient safety. National Audit Office.

NATIONAL PATIENT SAFETY AGENCY. 2004. Root Cause Analysis Toolkit. [Online]. Available: https://report.npsa.nhs.uk/rcatoolkit/course/iindex.htm. [Accessed 17.1.2011].

NIEBUHR, R. E., MANZ, C. C. & DAVIS, K. R., JR. 1981. Using Videotape Technology: Innovations in Behavioral Research. *Journal of Management,* 7**,** 43-54.

NORCINI, J. J., BLANK, L. L., ARNOLD, G. K. & KIMBALL, H. R. 1995. The Mini-CEX (Clinical Evaluation Exercise): A Preliminary Investigation. *Annals of Internal Medicine,* 123**,** 795-799.

O'CONNOR, P., CAMPBELL, J., NEWON, J., MELTON, J., SALAS, E. & WILSON, K. A. 2008. Crew Resource Management Training Effectiveness: A Meta-Analysis and Some Critical Needs. *The International Journal of Aviation Psychology,* 18**,** 353-368.

OTTESTAD, E., BOULET, J., LIGHTHALL, G. 2007. Evaluating the management of septic shock using patient simulation. *Critical Care Medicine,* 35**,** 769-775.

PAICE, A., AGGARWAL, R. & DARZI, A. 2010. Safety in Surgery: Is Selection the Missing Link? *World Journal of Surgery,* 34**,** 1993-2000.

PALUMBO, A., GRAHAM, J. 2009. Controversies in anaesthesia safety: implications for the teaching and assessment of trainees. *Anaesthesia & Intensive Care Journal,* 37**,** 659.

PATTERSON, F., FERGUSON, E., NORFOLK, T. 2005. A new selection system to recruit general practice registrars: preliminary findings from a validation study. *BMJ,* 330**,** 711-714.

PMETB. 2007. *Developing and maintaining an assessment system – a PMETB guide to good practice* [Online]. Available: http://www.gmc-uk.org/Assessment_good_practice_v0207.pdf_31385949.pdf [Accessed 17.01.2011].

POLIT, D. F. 2007. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health,* 30**,** 459.

POPE, J. H., AUFDERHEIDE, T.P., RUTHAZER, R., WOOLARD, R.H., FELDMAN, J.A., BESHANSKY, J.R., GRIFFITH, J.L., SELKER, H.P. 2000. Missed Diagnoses of Acute Cardiac Ischemia in the Emergency Department. *New England Journal of Medicine,* 342**,** 1163-1170.

RAJ, K., BAKER, K., BRIERLEY, S. & MURRAY, D. 2006. National Emergency Department Overcrowding Study tool is not useful in an Australian emergency department. *Emergency Medicine Australasia,* 18**,** 282-288.

READER, T. W., FLIN R., MEARNS, K., CUTHBERTSON, B.H. 2007. Interdisciplinary communication in the intensive care unit. *British Journal of Anaesthesia,* 98**,** 347-52.

REASON, J. 2000. Human error: models and management. *BMJ,* 320**,** 768-770.

REDFERN, E., BROWN, R. AND VINCENT, C. 2009. Identifying vulnerabilities in communication in the emergency department. *Emergency Medicine Journal,* 26**,** 653.
REEDER, T. J., BURLESON, D.L. & GARRISON, H.G. 2003. The Overcrowded Emergency Department: A Comparison of Staff Perceptions. *Academic Emergency Medicine,* 10**,** 1059-1064.

REZNEK, M., SMITH-COGGINS, R., HOWARD, S., KIRAN, K., HARTER, P., SOWB, Y., GABA, D. & KRUMMEL, T. 2003. Emergency Medicine Crisis Resource Management (EMCRM): Pilot Study of a Simulation-based Crisis Management Course for Emergency Medicine. *Academic Emergency Medicine,* 10**,** 386-389.

RISSER, D. T., RICE, M. M., SALISBURY, M. L., SIMON, R., JAY, G. D. & BERNS, S. D. 1999. The Potential for Improved Teamwork to Reduce Medical Errors in the Emergency Department. *Annals of Emergency Medicine,* 34**,** 373-383.

SALAS, E., BURKE, C.S. AND STAGL, K.C. 2004. Developing teams and team leaders: Strategies and principles. *In:* DAY, D., ZACCARO, S., HALPIN, S.M. (ed.) *Leader Development for Transforming Organizations. Growing Leaders for Tomorrow.* Mahwah, NJ: Lawrence Erlbaum.

SALAS, E., SIMS, D. E. & BURKE, C. S. 2005. Is there a "Big Five" in Teamwork? *Small Group Research,* 36**,** 555-599.

SARI, A. B., SHELDON, T.A., CRACKNELL, A., TURNBULL, A. 2007. Sensitivity of routine system for reporting patient safety incidents in an NHS hospital: retrospective patient case note review. *BMJ,* 334**,** 79.

SCHENKEL, S. M., KHARE, R.K., ROSENTHAL, M.M. ET AL 2003. Resident Perceptions of Medical Errors in the Emergency Department. *Academic Emergency Medicine,* 10**,** 1318.

SCHNELLE, J. F., OUSLANDER J.G. AND SIMMONS, S.F. 2006. Direct observations of nursing home care quality: Does care change when observed? *Journal of American Medical Directors Association,* 7**,** 541-544.

SCHULL, M. J., FERRIS, L.E., TU, J.V., HUX, J.E. 2001. Problems for clinical judgement: 3. Thinking clearly in an emergency. *CMAJ,* 164**,** 1170.

SEVDALIS, N. 2009. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. *Annals of Surgery,* 249**,** 1047.

SHAPIRO, M. J. 2004. Simulation based teamwork training for emergency department staff: Does it improve clinical team performance when added to an existing didactic teamwork curriculum? *Quality & Safety in Health Care,* 13**,** 417.

SHAPIRO, M. J., GARDNER, R., GODWIN, S. A., JAY, G. D., LINDQUIST, D. G., SALISBURY, M. L., SALAS, E., SHAPIRO, M. J., GARDNER, R., GODWIN, S.

A., JAY, G. D., LINDQUIST, D. G., SALISBURY, M. L. & SALAS, E. 2008. Defining team performance for simulation-based training: methodology, metrics, and opportunities for emergency medicine. *Academic Emergency Medicine,* 15**,** 1088-97.

SHAPIRO, M. J., SIMMONS, W.N., WRIGHT, R.O., LINAKIS, J., TRESPALACIOS, F, JAY, G.D. 2002. Can medical errors and near misses reported to an emergency department anonymous error reporting system be identified through chart review? *Academic Emergency Medicine,* 9, 401.

SHROUT, P. E., FLEISS, J.L. 1979. Intraclass correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin,* 86**,** 420-428.

SIMMONS, D., SHERWOOD, G. 2010. Neonatal intensive care unit and emergency department nurses' descriptions of working together: building team relationships to improve safety. *Critical Care Nursing Clinics of North America,* 22**,** 253.

SMALL, S. D., WUERZ, R. C., SIMON, R., SHAPIRO, N., CONN, A. & SETNIK, G. 1999. Demonstration of high-fidelity simulation team training for emergency medicine. *Academic Emergency Medicine,* 6**,** 312-23.

SPENCER, R., COIERA, E. & LOGAN, P. 2004. Variation in communication loads on clinical staff in the emergency department. *Annals of Emergency Medicine,* 44**,** 268-273.

STANHOPE, N., CROWLEY-MURPHY, M., VINCENT, C., O'CONNOR, A.M., TAYLOR-ADAMS, S.E. 1999. An analysis of adverse incident reporting. *Journal of Evaluation in Clinical Practice,* 5**,** 5-12.

STANTON, N. A., SALMON, P.M., WALKER, G.H., BABER, C., JENKINS, D.P. 2005. *Human Factors Methods. A Practical Guide for Engineering and Design,* Aldershot, Ashgate.

STELLA, D., HENDRIE, J., SMYTHE, J. & GRAHAM, I. 1996. Critical incident monitoring is a useful quality improvement tool for the emergency department. *Emergency Medicine,* 8**,** 215-219.

STONE, R., MCCLOY, R. 2004. Ergonomics in medicine and surgery. *BMJ,* 328**,** 1115-1118.

STRAUSS, A. L. 1998. *Basics of Qualitative Research : Techniques and Procedures for Developing Grounded Theory*.

STREINER, D., NORMAN, G. 1995. *Health Measurement Scales. A Practical Guide to their Development and Use,* New York, Oxford Univerisy Press.

TAYLOR-ADAMS, S. 2004. Systems analysis of clinical incidents: the London protocol. *Clinical Risk,* 10**,** 211.

TAYLOR, C. R., HEPWORTH, J. T., BUERHAUS, P. I., DITTUS, R. & SPEROFF, T. 2007. Effect of crew resource management on diabetes care and patient outcomes in an inner-city primary care clinic. *Quality and Safety in Health Care,* 16**,** 244-247.

THOMAS, E. J. 2000. Incidence and types of adverse events and negligent care in Utah and Colorado. *Medical Care,* 38**,** 261.

THOMAS, E. J., SEXTON, J.B., HELMREICH, R.L 2004. Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *Quality & Safety in Health Care,* 13**,** i57.

THOMAS, M. 2008. Incidence and causes of critical incidents in emergency departments: a comparison and root cause analysis. *Emergency Medicine Journal : EMJ,* 25**,** 346.

THOMAS, M., MORTON, R. & MACKWAY-JONES, K. 2004. Identifying and comparing risks in emergency medicine. *Emergency Medicine Journal,* 21**,** 469-472.

TIGHE, C. M., WOLOSHYNOWYCH, M., BROWN, R., WEARS, B. & VINCENT, C. 2006. Incident reporting in one UK accident and emergency department. *Accident and Emergency Nursing,* 14**,** 27-37.

TSCHAN, F., SEMMER, N., GAUTSCHI, D., HUNZIKER, P.R. SPYCHIGER, M., MARSCH, S. 2006. Leading to recovery: group performance and coordinative activities in medical emergency driven teams. *Human Performance***,** 277-304.

VAN DER VLEUTEN, C. P. M. 1996. The assessment of professional competence: developments, research and practical implications. *Advances in Health Sciences Education,* 1**,** 41.

VINCENT, C., NEALE, G., WOLOSHYNOWYCH, M. 2001. Adverse events in British hospitals: preliminary retrospective record review. *BMJ,* 322**,** 517-519.

VINEN, J. 2000. Incident monitoring in emergency departments: An Australian model. *Academic Emergency Medicine,* 7**,** 1290.

WARD, J., CLARKSON, P.J., JEFFCOTT, S. 2006. *Prospective Hazard Analysis* [Online]. Cambridge: University of Cambridge. Available: http://www-edc.eng.cam.ac.uk/research/healthcaredesign/hd4/hazardanalysis/hazardanalysis.pdf [Accessed 11th Feb 2011].

WEARS, R. L. 2000. Beyond error. *Academic Emergency Medicine,* 7**,** 1175-6.

WEARS, R. L., PERRY, S.J. 2002. Human factors and ergonomics in the emergency department. *Annals of Emergency Medicine,* 40**,** 206.

WEARS, R. L., WOLOSHYNOWYCH, M., BROWN, R. & VINCENT, C. A. 2010. Reflective analysis of safety research in the hospital accident & emergency departments. *Applied Ergonomics,* 41**,** 695-700.

WETZEL, C. M., KNEEBONE, R. L., WOLOSHYNOWYCH, M., NESTEL, D., MOORTHY, K., KIDD, J. & DARZI, A. 2006. The effects of stress on surgical performance. *The American Journal of Surgery,* 191**,** 5-10.

WHITE, A. A., WRIGHT, S.W., BLANCO, R. ET AL 2004. Cause-and-effect analysis of risk management files to assess patient care in the emergency department. *Academic Emergency Medicine,* 11**,** 1035-41.

WILKINSON, J. R. 2008. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical Education,* 42**,** 364.

WILLIAMS, D. D., HOLT, R.W., BOEHM-DAVIS, D.A. 1997. Training for inter-rater reliability: baselines and benchmarks. *Proceedings of the 9th International Symposium on Aviation Psychology.* Columbus, Ohio.

WILSON, R. M., RUNCIMAN, W.B., GIBBERD, R.W. ET AL 1996. Quality in Australian Health Care Study. *Medical journal of Australia,* 164**,** 754.

WOLOSHYNOWYCH, M., DAVIS, R., BROWN, R. & VINCENT, C. 2007. Communication Patterns in a UK Emergency Department. *Annals of Emergency Medicine,* 50**,** 407-413.

WRIGHT, M. C. 2004. Objective measures of situation awareness in a simulated medical environment. *Quality & Safety in Health Care,* 13**,** i65.

WRIGHT, M. C., PHILLIPS-BUTE, B. G., PETRUSA, E. R., GRIFFIN, K. L., HOBBS, G. W., TAEKMAN, J. M., WRIGHT, M. C., PHILLIPS-BUTE, B. G., PETRUSA, E. R., GRIFFIN, K. L., HOBBS, G. W. & TAEKMAN, J. M. 2009. Assessing teamwork in medical education and practice: relating behavioural teamwork ratings and clinical performance. *Medical Teacher,* 31**,** 30-8.

WYND, C. A., SCHMIDT, B., SCHAEFER, M.A. 2003. Two Quantitative Approaches for Estimating Content Validity. *Western Journal of Nursing Research,* 25**,** 508-518.

XIAO, Y., HUNTER, W.A., MACKENZIE, C., JEFFIRES, N.J. AND HORST, R.L. 1996. Task complexity in emergency medical care and its implications for team coordination. *Human Factors,* 38**,** 636-645.

YOUNGBLOOD, P. 2008. Design, development, and evaluation of an online virtual emergency department for training trauma teams. *Simulation in Healthcare,* 3**,** 146.

YULE, S., FLIN, R., MARAN, N., ROWLEY, D., YOUNGSON, G., PATERSON-BROWN, S., YULE, S., FLIN, R., MARAN, N., ROWLEY, D., YOUNGSON, G. & PATERSON-BROWN, S. 2008. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World Journal of Surgery,* 32**,** 548-56.

YULE, S., FLIN, R., PATERSON-BROWN, S., MARAN, N. & ROWLEY, D. 2006. Development of a rating system for surgeons' non-technical skills. *Medical Education,* 40**,** 1098-104.

ZAUSIG, Y. A. 2009. Inefficacy of simulator-based training on anaesthesiologists' non-technical skills. *Acta Anaesthesiologica Scandinavica,* 53**,** 611.

APPENDIX A – Materials for the Interview Study

Interview study topic guide

Topic	Key points/phrases
1. Introduction	• Introduce myself as an ED specialist registrar working in London and also doing research with the Clinical Safety Research Unit. • Explain the aims of the interview • Emphasise that there is strict confidentiality and that there are no right or wrong answers • Explain that if they don't understand the question they can ask for clarification • State **Participant Identification Number** for the record
2. Staff Grade	*Establish whether the participant is a doctor or a nurse and what grade they are.*
3. Importance of research topic (I)	• Do you think there is a lot of pressure on staff working in the Emergency Department? • Why?
4. Identify stressors	• What makes you feel under pressure when at work? *Think in terms of workload, other staff, types of patients, equipment etc* *If they say they never feel under pressure, may ask whether they feel overstretched or other similar terms. Alternatively, could ask if they think there are factors that appear to put pressure on **other** staff members.* • Can you recall a situation when you felt particularly under pressure? Can you describe what happened?
5. Effects on the Individual (I)	• Have you noticed that you, as an individual, behave differently when you're under pressure? *May ask them to think in terms of communication, working more as an individual/more collaborative etc.* • Do you do different things, or do things in a different way/ different priority when you are under pressure? For example, do you take on jobs that others normally do, or ask others to things you would normally do? *e.g. to help improve patient flow when it is very busy* • *May ask* if they can identify the trigger if they state that they change into different mode of working when under pressure.

6. Effects on the Individual (II)	• Do you think that working under pressure can sometimes make people work better? • Can you recall someone who appears to thrive on stressful conditions? How does their behaviour change when they appear to be under pressure? or • Can you think of someone who is good to work with when the department is really busy and staff are stressed? What is it that makes them a good person to have around?
7. Effects on the Individual (III)	• Have you noticed when other staff members in the department appear stressed (in a negative way)? How does their behaviour change? • How does that affect you? or • Can you think of someone you donôt like to work with when the department is busy? What makes them a difficult person to have around?
8. Error	*Precede question with statement that research has found that error in medicine is usually not solely down to an individual but due to organisational factors etc.* • Do you remember a time when you have made a mistake or underperformed because youôve felt under pressure at work?
9. Size of the problem	• How often do you feel youôre working under pressure? *(most days, a few times a week, once a week, rarely etc)*
10. Effects on the Team (I)	• Have you noticed that the ED team works/interacts differently when there is a lot of pressure? What changes? *Eg coordination, communication etc* • *(If in a senior position)* Do you direct your team differently when the team is under pressure? How? Or • What do the leaders of the department (the nurse in charge or registrar) do when there is increased pressure that is different to normal? •
11. Effects on the Team (II)	• Have you seen the ED team work particularly well when there was a lot of pressure? What happened?
12. Effects on the Team (III)	• Can you recall a time when the ED team didnôt function very well under pressure? What happened?

13. Coping strategies	• Have you got a particular way of helping you cope with situations when there is a lot of pressure at work? • How did you learn that? • Have you seen staff try to help each other to cope with high pressure situations? What did they do?
14. Further questions	• Is there anything you would like to add? • Do you have any questions?
	State that all the interviews will be transcribed and then analysed. Explain that we may come back to check that the way we have analysed your answers is the intended meaning. Offer Occupational Health leaflet.

Imperial College London

Information Sheet for Participants

You will be given a copy of this Information Sheet

1. **Study title: Experiences of staff working in the Emergency Department**

2. Introduction

You are being invited to take part in a research study. Before you decide to take part in this study it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Please ask questions if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

3. What is the purpose of the study?

The purpose of the study is to understand the factors that contribute to pressure on staff working in the Emergency Department and how this affects the way they work. It is hoped that this information will be useful to develop ways of helping the ED team function and improve patient care.

4. Why have I been chosen?

In order investigate this we plan to interview 25 members of staff who are interested in the subject and willing to take part. The participants will be a mixture of medical and nursing staff from the Emergency Department at St Maryôs Hospital.

5. Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be asked to sign a consent form and will be a given a copy of this and the information sheet. If you decide to take part you are still free to withdraw at any time and without giving a reason.

6. What will it involve?

A mutual convenient time will be arranged to conduct the interview. The interview will take up to an hour but may be less depending on what you have to say. The interview will take place in the one of the meeting rooms in the Emergency Department and will be tape-recorded. There are no right or wrong answers to any of the questions that the researcher will ask you.

7. What are the possible disadvantages of taking part?
We do not anticipate any disadvantages or risks as a result of taking part in this study. You will be offered support from Occupational Health if the interviews bring to light any stress-related issues you feel you would like to discuss further. A leaflet explaining the services available and contact details will be given to you at the end of the interview. If you are already receiving help for stress-related issues it is probably not a good idea to take part in the study as we do not want to interfere with any ongoing treatment or counselling you are receiving.

8. What are the possible benefits of taking part?
The information we get from this study will form the basis of future research with the aim of developing ways to help staff cope with high pressure conditions in the Emergency Department.

9. Will taking part in this study be kept confidential?
All the information which is collected about you during the course of the research will be kept confidential. Any information about you that is recorded will be identified by number rather than name to maintain confidentiality. Your answers will be tape-recorded during the interview and this will then be typed out. All data will be kept secure in a locked cabinet and password protected computers and only dedicated research staff will have access.

During the course of the interview, you may mention a clinical incident that has occurred. If this has not been reported, we may discuss with you whether it would be a good idea to report it so that we can improve the service we provide. This will only be considered if it is a significant incident and may be done confidentially. No information will be passed onto a third party without your prior knowledge.

10. What will happen to the results of the research study?
The results of the research will be written up as a report and it may be published in a research journal. You will not be identified in any report or publication.

11. Who is organising and funding the research?
The study is organised by the Clinical Safety Research Unit, Imperial College London, which is part of the Centre for Patient Safety and Service Quality. The research project has been funded by the National Institute for Health Research.

12. Who has reviewed the study?
Barnet Enfield & Haringey Research and Ethics Committee has reviewed the study.

13. Contact for further information
If you would like further information about this study, please contact Dr Lynsey Flowerdew on 0207 594 3105 or email l.flowerdew@imperial.ac.uk.

You will be given a copy of this information sheet and a signed copy of the consent form to keep.

Imperial College London

Consent Form

Title of study: Experiences of staff working in the Emergency Department

Participant Identification Number for this trial:

- I confirm that I have read and understand the information sheet for the above study. I have had the opportunity to consider the information and ask questions and these have been answered satisfactorily.
- I understand that my participation is voluntary and that I am free to withdraw at any time, without any reason, and without my legal rights being affected.
- I consent to the processing of my personal information for the purposes of this study only and understand that my information will not be used for any other purpose by the investigators. I understand that my information will be treated as strictly confidential and handled in accordance with the provisions of the Data Protection Act 1998
- I understand that data collected during the study may be looked at by responsible individuals from Imperial College or certain regulatory authorities, where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records.
- I consent to take part in the above study.

_____	_____	_____
Name of Participant	Signature	Date

_____	_____	_____
Name of Interviewer	Signature	Date

Contact Details

Westminster
Primary Care Trust

Occupational Health Service Information

For Staff of Imperial College Healthcare NHS Trust

Accessing the Service

For further information on any of the above in your area, please contact your local Occupational Health Unit.

St. Mary's Hospital
Praed Street, London W2 1NY

Telephone: 020 7886 1099
Fax: 020 7886 6677

Business Hours:
Monday - Friday
8:00am - 4:30pm

Occupational Health Personnel @ St. Mary's Unit

Christine E. Hunter
Head of Services

Dr. Roberto Ledda
Consultant

Neal Gething
Clinical Psychologist

Michelle Wallis
PA to Head of Services

Rafique Ahmad
Data Co-ordinator

Fax: 020 8962 7691



Sharon Blagrove-Campbell

A Healthier Future Is Within Your Reach

Occupational Health Westminster

Occupational Health Westminster is here to support you. Your health is our concern, to enable us to do this we provide the following services:

Occupational Health Services

- Pre-employment Screening
- Workstation Assessment
- Work Protection Immunisation Services
- Rehabilitation At Work Programme Programme
- Health& Safety
- Sickness Absence Referr al
- Health Awareness
- Mental Health
- Lifestyle Screening

Access To Services

The St. Maryôs Occupational Health Unit offers an appointment only service for your convenience during the following times:

Mondays: 8:00am - 4:00pm
Wednesdays: 8:00am - 4:00pm
Thursdays: 8:00am - 12:30pm
Fridays: 8:00am - 4:00pm

To access any of the Occupational Health services, please phone **020 7886 1099/6799** to book an appointment. Appointments are for access to services associated with Occupational Health and not for general medical care. Please attend A&E in an emergency or contact your G.P. for general poor health issues.

Mental Health

Confidential counselling is available from trained counsellors for work-related or personal issues. This is a free service and self-referrals are accepted.

Fitness Facilities

Well equipped fitness centres are available at each of the Occupational Health Units for the promotion of health, fitness and rehabilitation. After a comprehensive induction, staff have access t o:

- Multi-gym
- Circuit Training
- Rehabilitation Programmes
- Cardiovascular Equipment
- Free Weights
- Back Care Program
- Nutrition Advice
- Personalised Training

Come and enjoy the professional and friendly atmosphere at reduced rates for staff. Follow the path to fitness and have a personal exercise programme created just for you with the help of our team of qualified instructors. For further information, please phone **020 7886 1288** (St. Maryôs Unit).

Travel Advice & Vaccination

Information and guidelines are updated regularly and provided by the *Medical Advisory Service for Travellers Abroad*. Vaccines are offered at reduced rates for staff and their partners. Appointments are required for this service. This service includes vaccination for:

- Hepatitis A & B
- Japanese Encephalitis
- Rabies
- Yellow Fever
- Diptheria, Polio & Tetanus
- Meningitis
- Typhoid

Coding Framework

Tree Nodes				Definition and Examples
a) Is the ED a high pressure environ?				Yes, no, sometimes etc
b) Stressors	Individual factors			eg tired, hungry, personal problems etc
	Organisational & Management	Clinical governance		Pressure to get it right for the patient or external pressure to perform
		Human resources		Planned staffing levels perceived to be insufficient (not including unavailability of staff)
		Targets	4 hr targets	Pressure caused by government 4 hr target
			targets for junior doctors	Supposed to see 850 patients in 6 month job
	Patient factors	Accompanying individuals		eg factors due to family or friends accompanying patient, eg relative shouting at staff for long wait
		Condition		Complexity, seriousness, co-morbidities etc
		Personal		Personality, language (inappropriate, shouting etc)
	Team components	Leadership		Ineffective leadership
		Supervision, seeking help & teamwork		Availability of senior/junior staff, responsiveness of snr/jnr staff, support from other departments
		Verbal communication		Communication between jnr and snr staff, between specialities, between specialities, handover etc.
		Written communication		Incomplete, absent info, unavailable/illegible notes, inadequate documented management plan etc
	Work Environment	Bed shortages		Pressure caused by unavailability of beds in the hospital
		Environment		Temperature, noise, light etc
		Equipment & facilities		Malfunction/ unavailability or equipment, problems with lab, pod system not working etc
		Staffing		Unavailability, use of temporary staff (agency, locum), skill
		Workload		High no. of patients, unable to take rest breaks

c) How does individual change under pressure	Back-up behaviour			ie doing someone elseôs job for them, eg push trolleys, doctor emptying commode
	Communication			How staff change the way they communicate (eg junior doc asks fewer questions)
	Delegation			Eg doctor asks nurse to do bloods, registrar does not see patients but hands out cards to junior doctor
	Departmental Coordination			How senior staff change the way they coordinate the running of the department, eg use of computer/white board, ward rounds etc.
	Leadership			How they change they way the lead more junior staff (usually NIC, reg or consultant) eg allocating tasks
	Task related			How staff change the tasks they do or the order etc. E.g. writing in notes, referral and use of investigations
e) Others positive bvr or attitudes	Attitude			eg calm, hard working, willing to help others
	Behaviour			the behaviour of the staff member (excluding leadership behaviours) eg helping others, being proactive etc
	Communication			Eg remaining calm when communicating with others, keeps voice down
	Leadership behaviours			Eg specific allocation of tasks, ask for plan off junior doctor in helpful way ,
f) Others negative bvr or attitudes	Attitude			Apparent attitudes of other staff when working under pressure, eg unhelpful, donôt appear to care
	Behaviour			The behaviour of other staff members (excluding leadership behaviours) eg nurses not doing obs, staff trying to avoid work
	Communication			Communication practices of other staff members eg, shouting, rude, giving inadequate info. etc
	Leadership behaviours			Eg how board round is run, allocating staff etc
g) Example of Error under pressure				Example of a time when they felt they had underperformed or made an error due to working under high pressure
h) How often high pressure				How often staff feel they are working under pressure, eg everyday, most days, rarely
i) Effects on the team				How working under pressure can affect the team as a whole (may be positive or negative)
k) Coping with high pressure	How staff cope			Strategies staff have learnt to cope with working under pressure, eg take a 5 min breather, vent to someone etc
	Who teaches coping			Who taught staff coping strategies, ie learnt it themselves, senior instructed them etc
l) Ideas for interventions				Ideas staff have to help the team work under pressure, including practical tips, team training etc

Appendix B – Materials for development of a provisional tool

List of skills taken from other observational assessment tools

Behaviour/ Skill Group	Main skill
1. Providing/ setting/ maintaining standards Cross-monitor team member actions Engage in error-correction actions	Maintaining standards
2. Assessing capabilities/ understanding team needs	Assessing capabilities
3. Workload management/ identifying and utilising resources/manage team resources and workload	Workload management
4. Co-ordinating activities with team members	Coordinating team members
5. Exchanging information/ communication	Communication
6. Leadership/ establish leadership	Leadership
7. Prioritising	Workload management
8. Primary and secondary triage	Decision making & Workload management
9. Authority and assertiveness/ appropriate assertiveness/ coping with pressure	Authority and assertiveness
10. Conflict Solving	Conflict solving
11. Team Building/ cultivate team climate	Team building
12. Support of others	Support of others
13. Decision-making a. definition and diagnosis b. option generation/ identifying options c. balancing risks/ selecting/ communicating options d. implementing/ reviewing/ re-evaluating	Decision making
14. Conduct situational planning	Decision making
15. Situational Awareness a. gathering information/ notice b. recognising/ understanding/ communicating c. projecting/ anticipating/ think ahead	Situational awareness
16. Planning and preparing/organise the team	Sit. Awareness (anticipation)
17. Establishing a shared understanding	Sit. Awareness(communicating)
18. Engage in team improvement strategies/ teaching	Teaching

Comparison of College of Emergency Medicine generic skills curriculum and corresponding non-technical skills

Knowledge, Skills and Attitudes from CEM curriculum	Related Non-technical Skill
Documentation	Maintaining standards
Diagnosis	Decision making
Decision making	
Time management	Workload management
Safe prescribing	Maintaining standards
Continuity of care	Communication Decision making ï information gathering
Communication with colleagues	Communication Conflict solving
Communication skills ï referrals	Communication
Team working	Team building Communication (handover, feedback) Workload management (delegation, seeking help) Teaching
Information management	Decision making ï information gathering
Professional attributes	Team building Conflict solving
Developing others learning	Teaching
Assessment and appraisal	Support of others
Risk management	Maintaining standards
Medico-legal issues	
Confidentiality	

Academy of Royal Medical Colleges Medical Leadership Competency Framework and associated non-technical skills

Knowledge, Skills and Attitudes from the MLCF	Related Non-technical Skill
Developing self awareness	Coping with pressure
Managing yourself	Maintaining standards Workload management
Acting with integrity	Communication Maintaining standards
Developing networks	Team building Communication Decision making ï option generation
Building and maintaining relationships	Team building Authority and assertiveness
Encouraging contribution	Authority and assertiveness Conflict resolution Decision making ï option generation
Working within teams	Team building Workload management
Managing people	Teaching Support of others
Ensuring patient safety	Maintaining standards

Roles and responsibilities (from St Mary's ED handbook) and corresponding non-technical skills

Roles & Responsibilities	Related Non-technical Skill
Liaise closely with the nurse-in-charge regarding the running of the department.	Communication Situational awareness
Participate in (and lead out of hours) the board round to achieve agreement on a management plan for all patients.	Communication Decision making Supervision
Oversee the workload in majors and ensure a decision is made within 30 minutes of the patient being seen.	Workload management Decision making
Provide clinical leadership to ensure patients are not moved inappropriately to other wards or areas in the Trust if they are unwell.	Supervision Maintaining standards
Discuss and briefly review all patients before moving to CDU.	Decision making Situational awareness
Oversee the workload in minors and paediatrics and move staff as required.	Situational awareness Coordinating team members
Oversee and act as team leader for all resuscitations out of hours, or nominate another middle grade as team leader in hours if possible.	Leadership Workload management
See patients themselves to manage the workload as appropriate	Workload management
Ensure handover of all patients in Resus, majors and CDU to the Middle grade doctor who takes the bleep at handover.	Communication
Communicate with the consultant on call over any adverse conditions	Communication Workload management
Handover any outstanding issues that might need action, for example referrals to the coroner.	Communication
Inform the consultant of any doctors who phone in sick and participate in the ordering of additional locum staff as required	Communication Decision making
Be familiar with all CDU patients admitted overnight to give handover in the morning	Situational awareness
Refer to Escalation Plan where necessary	Workload management Communication

Comparison with existing methods of assessing non-technical skills

CEM Multi-source Feedback	Related non-technical Skill
Problem-solving skills	Decision making
Note-keeping; clarity, legibility, complete	Maintaining standards Communication
Is a team player	Team building
Asks for otherssô point of view	Decision making ï gathering information
Encourages discussion, empathy & sensitivity	Decision making ï communicating options Team building
Is clear and precise with instructions	Decision making ï communicating options Communication
Treat colleagues with respect	Team building
Communicates well	Communication
Can lead a team well	Coordinating team Authority and assertiveness
Takes responsibility	Authority and assertiveness
ñI like working with this doctorò	Team building
Teaching and training (5 elements)	Teaching

ACAT-EM	Related non-technical Skill
Clinical Assessment	N/A
Medical record keeping	Maintenance of standards
Investigation and treatment of the critically ill patient	Decision-making
Time management	Workload management
Management of the team	Team building
Clinical leadership	Workload management/ Supervision
Patient safety	Maintenance of standards
Handover	Communication
Overall Clinical Judgement	N/A

CEM Multisource feedback form

UNKNOWN	1	2	3	4	5
Not Observed	*Performance* ***Does Not Meet*** *Expectations*	*Performance* ***Partially Meets*** *Expectations*	*Performance* ***Meets*** *Expectations*	*Performance* ***Exceeds*** *Expectations*	*Performance* ***Consistently Exceeds*** *Expectations*

Good Clinical Care		**Rating (UNKNOWN / 1 - 5)**
1	*Medical knowledge and clinical skills*	
2	*Problem-solving skills*	
3	*Note-keeping – clarity; legibility and completeness*	
4	*Emergency Care skills*	
Relationships with Patients		**Rating (UNKNOWN / 1 - 5)**
1	*Empathy and sensitivity*	
2	*Communicates well with all patient groups*	
3	*Treats patients and relatives with respect*	
4	*Appreciates the pyscho-social aspects of patient care*	
5	*Offers explanations*	
Relationships with Colleagues		**Rating (UNKNOWN / 1 - 5)**
1	*Is a team-player*	
2	*Asks for others' point of view and advice*	
3	*Encourages discussion Empathy and sensitivity*	
4	*Is clear and precise with instructions*	
5	*Treats colleagues with respect*	
6	*Communicates well (incl. non-vernal communication)*	
7	*Is reliable*	
8	*Can lead a team well*	
9	*Takes responsibility*	
10	*"I like working with this doctor"*	
Teaching and Training		**Rating (UNKNOWN / 1 - 5)**
1	*Teaching is structured*	
2	*Is enthusiastic about teaching*	
3	*This doctor's teaching sessions are beneficial*	
4	*Teaching is presented well*	
5	*Uses varied teaching skills*	
Global ratings and concerns		**Rating (UNKNOWN / 1 - 5)**
1	*Overall how do you rate this Dr compared to other ST4-6 Drs*	
2	*How would you rate this trainees performance at* ***this stage*** *of training*	
3	*Do you have any concerns over this Drs probity or health?*	

The Acute Care Assessment Tool (ACAT-EM) form

Trainee name:			
Assessor:		**Assessor GMC no.**	
Grade of assessor:		**Date**	/ /
Setting,		Acute presentations covered (5 max for EM)	
Timing, duration and level of responsibility			

Please TICK to indicate the standard of the trainee's performance in each area	Not observed	Further core learning needed	Demonstrates good practice		Demonstrates excellent practice
			Must address learning points highlighted below	Should address learning points highlighted below	
Clinical Assessment					
Medical record keeping					
Investigation and treatment of the critically ill patient					
Time management					
Management of the team					
Clinical leadership					
Patient safety					
Handover					
Overall Clinical Judgement					

ACAT –EM	
Assessment Domains	Description
Clinical assessment and clinical topics covered	Quality of history and examination to arrive at appropriate diagnosis- made by direct observation in different areas especially in the resuscitation room. No more than 5 AP should be covered in each ACAT and this should involve a review of the notes and management plan of the patient.
Medical record keeping	Quality of recording of patient encounters including drug and fluid prescriptions
Investigations and referrals	Quality of trainees choice of investigations and referrals
Time management	Prioritisation of cases , doesn't spend too much time with any one patient
Management of take/team working	Appropriate relationship with and involvement of other health professionals
Clinical leadership	Appropriate delegation and supervision of junior staff
Handover	Quality of handover of care of patients between EM and in patient teams and in house handover including obs/CDU ward
Patient safety	Able to recognise effects of systems, process, environment and staffing on patient safety issues
Overall clinical judgement	Quality of trainees integrated thinking based on clinical assessment, investigations and referrals. safe and appropriate management, use of resources sensibly

Appendix C - Materials used to refine the assessment tool

Example of field notes

15/1/10 pm (NO BEDS)

- Phones patientôs relative to give update. Goes to resus to review patient (repeat attender with seizures)
- Overhears nurses talking in next resus bay ï asks ñWhatôs that?ò - #dislocation ï asks if a fall
- Asks SHO ok as leaves resus patient
- Washes hands
- Asks NIC to go through board together. Says donôt like triage ñunwellò as donôt know whatôs wrong with them
- NIC tells plan. Corrects NIC and says patient due to go to chair in EAU (previously discussed with consultant). Checks notes and not seen by ortho yet so goes to eyeball patient and ortho with her. Checks with doctor/patient that ok to sit in chair in EAU
- Overhears other nurse saying something and confirms plan
- Tells NIC she will do EAU form
- Other reg updates that patient with confusion (alcoholic) has positive trop so canôt go to EAU and will breach ï ñthatôs fineò
- Gives EAU form to receptionist for admission
- Tells NIC she is going to resus and will come back to finish going through patients
- Tells SHO she had written up paracetamol and midazolam on drug chart as no notes to write in. Ensure other drugs are to be written up. Tells SHO to contact neuro as seizures not stopping
- Calls nurse and asks if he is in resus ï ñno, triageò but asks what she wants. Says ñfluidsò ï ñWhat do you want?ò ï ñNormal salineò ï says itôs a portocath and tell him how to use it ñif youôre not comfortable Iôll do itò. Other nurse arrives and asks ñwhat has she had?ò ï reg says that the SHO is about to sign for it. SHO says that neuro are coming
- Asks NIC to go through PTA ï listens. Finds patient who has gone home on PTS ï checks which doctor saw it ï other reg so knows she will take it off
- NIC identifies patient he knows nothing about
- Tells NIC plan for seizure patient
- Other nurse tries to interrupt but she says ñcan I just finish this?ò

Coding framework for recoding interviews and observations

Skill	Definitions and examples
Maintenance of Standards	• Examples highlighting the importance of standards such as good clinical notes, proper administration of drugs etc. • Comments related to delivering good quality care, especially with the competing pressure of government targets
Decision making - Option Generation - - Selecting & Comm. Options - Outcome Review	• Using various resources (written/verbal) to gather information and generate appropriate options for a problem. Involving the team in decision making. • Clearly stating decisions and plans (e.g. patient management plan) and explaining reasons, if necessary. Or failure to communicate a plan. • Once a decision has been made, reviewing suitability in light of new information. Confirming tasks have been done.
Coordinating the team	• Reallocating staff members to different areas of the department as the need arises (e.g. due to increased waiting time). • Organising staff breaks.
Team Building	• Characteristics and attitudes that have a positive or negative impact on the team e.g. appearing friendly and approachable • Describing the importance of team relationships (within ED team or ward staff, ambulance crews etc)
Communication	• Comments related to the importance of good communication (e.g. being concise) • Examples of how staff changed the way the communicated (e.g more direct instructions) • Examples of miscommunication or ineffective communication
Authority & Assertiveness	• Describing the use of authority to get something done (either effective use of authority or being over-assertive)

	• Speaking up when necessary or failure to speak up • Remaining calm or losing control
Situational awareness - Gathering Information - Anticipating - Informing the Team	• Surveying the environment to pick up cues that may need action as well as requesting reports from others (e.g. monitoring Symphony or board rounds) • Planning ahead and anticipating potential issues such as staffing or bed/ cubicle availability • Communicating the situation to keep team, e.g. about long waiting time
Assessing Capabilities	• Examples where staff consider a team members experience or condition and adapt their behaviour accordingly (e.g. changing level of supervision, allocating staff)
Support of Others	• Staff providing support to other team members, e.g. senior doctor supporting a junior doctor who is having difficulty making a referral • Comments related to the importance of feedback
Conflict Solving	• Descriptions of conflict within the ED team or with specialties/ radiology (e.g. refusing referral or refusing investigation)
Teaching	• Responses related to the value of teaching on the ósshop floorô

Appendix D - Validation survey materials

Questionnaire

Please indicate job title

- ○ ED consultant, senior staff grade, associate specialist
- ○ EM trainee ST 3-4, SpR
- ○ ED middle grade, senior clinical fellow
- ○ ED nurse

Read the following statements carefully. Each skill is given a definition (shaded grey) and is followed by 3 examples of observable behaviours that represent evidence of competence in the ED.

Please rate them according to how important you feel they are for an ED registrar

1	2	3	4	5
Not at all	Slightly important	Moderately important	Very important	Essential

There is space at the end of the questionnaire to add your own suggestions or feel free to make comments and suggestions in the margin as you go along.

	Important skill/ behaviour? (Please circle answer)
Maintenance of standards - subscribes to clinical and safety standards as well as considering performance targets	
Reprimands doctor for illegible notes and explains value of good note keeping	1 2 3 4 5
Explains importance of ensuring sick patient is stable rather than moving patient for organisational reasons	1 2 3 4 5
Washes hands between patients	1 2 3 4 5
Workload management - manages own and others' workload to avoid both under and over-activity	
Identifies that SHO has spent a long time on a patient and ascertains if there is a problem or just going slow	1 2 3 4 5
Practices back-up behaviour (eg does set of observations) when nurse is busy	1 2 3 4 5
Escalates to consultant when overloaded	1 2 3 4 5
Supervision and feedback - assesses capabilities and identifies knowledge gaps. Provides opportunities for teaching and constructive feedback.	
Gives constructive criticism to team member	1 2 3 4 5
Takes opportunity to teach while reviewing xray with junior	1 2 3 4 5

Gives positive feedback to junior who has made a difficult diagnosis	1 2 3 4 5
Team leadership - provides motivation and support for the team	
Reacts in a friendly way to junior doctor who asks for help even though they are busy	1 2 3 4 5
Says thank you at the end of a difficult shift	1 2 3 4 5
Motivates team to work faster during periods of high workload	1 2 3 4 5
Exchanging information - gives verbal and written information concisely and effectively	
Gives succinct clinical history when requesting radiological investigation	1 2 3 4 5
Acknowledges and repeats important message to ensure understanding	1 2 3 4 5
Gives clear referral to speciality doctor with reason for referral	1 2 3 4 5
Authority and assertiveness - behaves in an appropriately forceful manner and speaks up when necessary	
Uses appropriate degree of assertiveness when speciality refuses referral	1 2 3 4 5
States differing opinion to senior colleague regarding patient management	1 2 3 4 5
Remains calm under pressure	1 2 3 4 5
Decision making (option generation) - uses all resources (written and verbal) to gather information and generate appropriate options for a given problem.	
Checks that the patientôs carer has been involved when discussing discharge plan	1 2 3 4 5
Goes to see patient to get more information when junior is unclear about history	1 2 3 4 5
Listens to team memberôs input	1 2 3 4 5
Decision making (selecting and communicating options) - considers risks of various options and discusses this with the team	
Verbalises consideration of risk when suggesting sending home elderly patient	1 2 3 4 5
Decisive when giving advice to juniors	1 2 3 4 5
Discusses plan for patient with SHO and explains reason for change in management plan	1 2 3 4 5
Decision making (outcome review) - once a decision has been made, reviews suitability in light of new information or change in circumstances and considers new options. Confirms task is done.	
Seeks out doctor after seeing patient to see if initial plan needs revising	1 2 3 4 5

Assesses outcome of moving doctor from majors to minors to relieve waiting time	1 2 3 4 5
Reviews impact of treatment on acutely sick patient	1 2 3 4 5
Situation awareness (gathering information) - surveys the environment to pick up cues that may need action as well as requesting reports from others.	
Checks team members all turn up for shift at correct times	1 2 3 4 5
Eyeballs patients in cubicles when long wait to check for anyone who looks unwell	1 2 3 4 5
Calls mini-board round to get update on patients when is department busy	1 2 3 4 5
Situation awareness (anticipating) - anticipates potential issues, such as staffing or cubicle availability in the department, and discusses contingencies	
Identifies that there are lots of patients with injuries waiting to be triaged and allocates extra doctor to Minors	1 2 3 4 5
Recognises that Majors is almost full and discusses plan with nurse in charge to clear cubicles	1 2 3 4 5
Asks for side room early for patient who may be admitted with infective diarrhoea	1 2 3 4 5
Situation awareness (informing the team) - communicates situation to keep team ‘in the picture’ rather than just expecting action	
Updates team about waiting time	1 2 3 4 5
Keeps NIC up-to-date with plan for patients	1 2 3 4 5
Communicates to the team any new issue, such as a doctor calling in sick	1 2 3 4 5

1. Non-technical skills are considered ‘the behavioural aspects of performance which are not related directly to medical expertise’. This survey relates to team-related skills (ED team or interdisciplinary) and clinical leadership. Please add any skills or behaviour, not already listed above, a competent registrar should demonstrate.

   _____
   _____
   _____

2. Do you think it is important that these skills are explicitly and specifically assessed during training?

   _____

3. Registrars - Would you like more feedback from consultants about how you lead the department and supervise the team?

   _____

4. Do you think it would be valuable for consultants to actively observe registrars running the department

Original and revised behavioural marker statements

Skill	Original Behaviour Statement *Low response rate (>4 missing values) <0.75 CVI >=0.75 CVI	CVI	Revised Behaviour Statement
Maintenance of Standards	• Reprimands doctor for illegible notes and explains value of good note keeping • Explains importance of ensuring sick patient is stable rather than moving patient for sake for organisational reasons • Washes hands between patients	0.61 0.98 0.86	• Notices doctorõs illegible notes and explains the value of good note-keeping • Washes hands (or uses alcohol gel) before and after reviewing patients
Workload Management	• Identifies that SHO has spent a long time on a patient and ascertains the reason • Practices back-up behaviour (e.g. does a set of observations) when nurse is busy • Escalates to consultant when overloaded	0.79 0.55 0.88	• Ensures both themselves and other team members take appropriate breaks • Deals with interruptions efficiently
Supervision & Feedback	• Gives constructive criticism to team member • Takes opportunity to teach whilst reviewing patient with junior • Gives positive feedback to junior who has made a difficult diagnosis	0.82 0.76 0.88	
Team Building	• Reacts in a friendly way to SHO asking for help even though they are busy • Says thank you at end of shift • Motivates team to work faster during periods of high workload	0.93 0.78 0.70	•Motivates team, especially during stressful periods

Quality of Communication	• Gives succinct clinical history when requesting radiological investigation • Acknowledges and repeats message to ensure understanding • Gives clear referral to speciality doctor with reason for admission*	0.74 0.75 0.86	• Gives and accurate and succinct handover of the department • Ensures important message is heard correctly • Gives clear referral to speciality doctor with reason for admission (e.g. SBAR)
Authority & Assertiveness	• Uses appropriate degree of assertiveness when speciality refuses referral • States differing opinion to senior colleague regarding patient management * • Remains calm under pressure	0.88 0.79 0.98	•Willing to speak up to senior staff when concerned
Option Generation	• Checks that the patientôs carer has been involved when discussing discharge plan • Goes to see patient to get more information when doctor is unclear about history • Listens to team memberôs input	0.69 0.93 0.93	•Seeks help when unsure •Encourages team membersô input
Selecting & Communicating Options	• Verbalises consideration of risk when suggesting sending home elderly patient • Decisive when giving advice to juniors • Discusses plan for patient with SHO and explains reasons for change in management plan	0.81 0.84 0.91	• Verbalises consideration of risk when sending home patient
Outcome Review	• Seeks out doctor after seeing patient to see if provisional plan needs revising • Assesses outcome of moving doctor from majors to minors to relieve waiting time • Reviews impact of treatment on acutely sick patient	0.75 0.68 0.95	• Ensures priority treatment is given to patient

Gathering Information	• Checks team members all turn up for shift at correct times • Eyeballs patients in cubicles when long wait to check for anyone who looks unwell • Calls mini-board round to get update on patients when department busy	0.79 0.86 0.73	• Uses Patient Tracking System to monitor the state of the department
Anticipating	• Identifies that there are lots of patients with injuries waiting to be triaged and allocates extra doctor to minors • Recognises that Majors is almost full and discusses plan with NIC to clear cubicles • Asks for side room early for patient who may be admitted with infective diarrhoea	0.75 0.90 0.72	• Identifies busy triage area and anticipates increased demand • Discusses contingencies with nurse-in-charge during periods of overcrowding • Prepares trauma team for arrival of emergency patient
Understanding & Communicating	• Updates team about waiting time • Keeps nurse-in-charge up to date with plan for patients • Communicates to team any new issue such as a doctor calling in sick	0.52 0.91 0.83	• Communicates a change in patient status to relevant speciality doctors

Appendix E - Materials for the training day study

Programme

Leadership and non-technical skills for EM physicians

At the end of the study day, participants will be able to:

1. Identify the skills and behaviours that contribute to safe and effective performance in the ED
2. Implement a tool for assessing leadership and non-technical skills in the ED
3. Consider how reflection and feedback can be undertaken more effectively
4. Understand the methods available for training EM leadership

Why is leadership important?	Ruth Brown	9.30 ï 10.00
Assessing non-technical skills in the workplace ï a new tool	Lynsey Flowerdew	10.00 ï 11.00
Coffee		11.00 ï 11.15
Identifying the skills necessary for leading on the ósh op floorô (filmed scenarios 1 &2)	Ruth Brown	11.15 ï 12.45
Lunch		12.45 ï 13.45
Application of the assessment tool (filmed scenario 3)	Lynsey Flowerdew	13.45 ï 14.15
Improving the debrief	Kevin Reynard	14.15 ï 15.15
Teaching leadership skills	Ruth Brown	15.15 ï 16.15

Non-technical skills are ñthe cognitive, social and personal resource skills that complement technical skills, and contribute to safe and efficient task performanceò.

Leadership Study Day – Feedback on the Assessment Tool

1. Job Title/grade _____

2. Years working in EM _____

3. Did you find the assessment tool easy to use on the filmed scenarios? ☐ yes ☐ no If not, why not?

4. Do you feel you have had sufficient training to begin assessing trainees in your own department? ☐ yes ☐ no If not, what else would help?

5. Do you think you would be able to explain to others in your department how to use the tool? ☐ yes ☐ no If not, what else would help?

5. How could the tool be improved (e.g. revised rating scale, new skill definitions, different structure, etc.)?

6. Some departments already observe their traineesô non technical skills during a shift (e.g. ACAT-EM or intermittent, ad hoc observation of registrars). If this is done in your department how does it differ from the assessment tool provided?

7. Can you think of any advantages of direct observation (e.g. being followed by an observer) compared to current methods of assessment?

8. Can you think of any disadvantages of this kind of direct observation of non-technical skills?

9. ***Consultants only*** - Would you use the tool for assessing trainees in your own department? ☐ yes ☐ no Please explain your choice.

10. ***Trainees only*** ï Would you assess other colleagues using the assessment tool? ☐ yes ☐ no Please explain your choice.

11. ***Trainees only*** ï Would you like someone to use the tool to observe and assess your leadership and non-technical skills? ☐ yes ☐ no Please explain your choice.

12. Has todayôs sessions on identifying and assessing leadership skills made you reflect on your own practice? If so, please give an example(s). ☐ yes ☐ no

Session Feedback

	Please circle a score- 1 being not relevant, 5 being very relevant
Session 1 09:30 - 10:00	**Why is Leadership Important?** **Facilitator:** Ruth Brown score 1 2 3 4 5
Session 2 10:00 - 11:00	**Assessing leadership and other non-technical skills in the workplace – a new tool** **Facilitator:** Lynsey Flowerdew score 1 2 3 4 5
Session 3 11:15 ï 12:45	**Identifying the skills necessary for leading on the 'shop floor'** **Facilitator:** Ruth Brown score 1 2 3 4 5
Session 4 13:45 ï 14.15	**Practice using the assessment tool using filmed scenarios** **Facilitator:** Lynsey Flowerdew score 1 2 3 4 5
Session 5 14:15 ï 15:15	**Improving the debrief** **Facilitator**: Kevin Reynard score 1 2 3 4 5
Session 6 15.30 ï 16.30	**Teaching leadership skills** **Facilitator:** Ruth Brown score 1 2 3 4 5

Additional Comments

Appendix F - Further data for the multicentre study

Description of statistical tests used

Reliability

ICC has been used by other research teams to evaluate behavioural marker systems (Graham et al., 2010, Yule et al., 2008). This was used in preference to the inter-rater agreement method used in Chapter 8 because ICC is considered a superior test (it could not be used for the training day study for lack of repeated measures design).

Of the six types of ICC (Shrout, 1979) the two-way random effects model was used for this analysis. This type is appropriate when the targets are deemed a random effect (i.e. the observed registrars are randomly chosen) and the judges (i.e. observers) are selected from a larger population of judges. Furthermore, it was chosen to test for ôAbsolute Agreementõ to include judge variance in the calculation (rather than ôConsistencyõ where systematic variance due to the raters is irrelevant) (McGraw, 1996).

As with other reliability coefficients, there is no standard acceptable level of reliability using the ICC. Chin (Chin, 1990) recommends that any measure should have an intra-class correlation coefficient of at least 0.6 to be useful and other research teams have used 0.7 as an acceptable level (Yule et al., 2008, Graham et al., 2010).

All ICC values given are single measures. Theoretically, the ICC is a value between 0 and 1, with values closer to one representing higher reliability. However, some statistical packages (including SPSS 18 which was used for this analysis) allow a negative value to be calculated. This typically happens when the between-subject variation is relatively small (i.e. the observed registrars all perform similarly) compared to the within-subject variation (i.e. large variation in scores between different raters).

Speraman rho is a non-parametric test that uses ranked data to determine if there is a correlation between scores on two different occasions.

Measures for internal reliability/consistency (for example, using Cronbachôs alpha) were not performed as the tool is not designed to measure a single construct.

Comparison between hospitals

The Kruskal-Wallis Test is a non-parametric test that uses mean rankings to compare more than two independent groups. It is used to test the null hypothesis that all populations have identical distribution functions against the alternative hypothesis that at least two of the samples differ with respect to the median. The Monte Carlo method of conducting the analysis was chosen as this is more accurate than the Asymptotic option and the data were too large to run exact tests. Bonferroni corrections were used to ensure Type 1 errors (i.e. false positives) did not exceed 0.05.

References

CHIN, S. 1990. The assessment of methods of measurement. *Statistics in Medicine,* 9**,** 351-362.

GRAHAM, J., HOCKING, G. & GILES, E. 2010. Anaesthesia Non-Technical Skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? *Br. J. Anaesth.,* 104**,** 440-445.

MCGRAW, K. O. W., S.P. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods,* 1**,** 30-46.
SHROUT, P. E., FLEISS, J.L. 1979. Intraclass correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin,* 86**,** 420-428.

YULE, S., FLIN, R., MARAN, N., ROWLEY, D., YOUNGSON, G., PATERSON-BROWN, S., YULE, S., FLIN, R., MARAN, N., ROWLEY, D., YOUNGSON, G. & PATERSON-BROWN, S. 2008. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World Journal of Surgery,* 32**,** 548-56.

Sources of rater error

1. Simple mistakes made with categorisation of behaviours

Also, some apparent incorrect categorisation was resolved by discussion with the original observer. For example, the statement ñExplained patient plan to ED consultantò was coded in Authority and Assertiveness and a second investigator questioned whether this would be better in DM ï Selecting and Communicating Options. The first observer explained that the consultant was judged to be unnecessarily interfering with a patientôs management and the trainee needed to assertively defend their decision-making.

2. Overlapping definitions of a skill

Whilst the assessment tool is designed so that behaviours fit exclusively into only one skill, inevitably some overlap exists. For example, ñLooks up drug doses in the BNFò was coded in the skill Maintenance of Standards as the observer thought the trainee probably had an idea of the dose but wanted to check and make sure it was correct. However, this could also be included in the skill DM ï Option Generation as the registrar may be using the BNF as a source of information because the dose is unknown. This difficulty arises where the observer is unable to ascertain the traineeôs reasoning for a particular behaviour. A common example included statements like ñSought out SHOôs who were taking a long time and discussed patient plansò. This may have been coded in Workload Management, as the behaviour ensures doctors are kept working efficiently, but it could also be categorised as Supervision and Feedback, as it involves overseeing the decision-making of more junior staff. Also ñAggressive when dealing with anaesthetistò affected the rating in both Team Building and Authority and Assertiveness. Sometimes a single observed behaviour could be legitimately coded into two skills as more than one skill is being demonstrated. For example, ñRefuses to send patient photo to Plastics via personal email or mobile phoneò reflects both Maintenance of Standards and Authority and Assertiveness.

3. Misunderstandings of the definition of a skill

Some apparent overlap was resolved by review of the definition and exemplar behaviours for the skill. For example, ñclear reason givenò or ñnot specific about who should do whatò should be included in the skill DM ï Selecting and Communicating Options. One observer documented ñDidnôt follow-up effect of prescribed GTNò in Maintenance of Standards but this would be better categorised in DM ï Outcome Review.

4. Concealed cognitive processes and reasoning

As previously mentioned, it was sometimes difficult to categorise a behaviour without understanding the underlying reason for the action. For example, commonly, doctors were heard saying ñHow are you getting on?ò or ñAre you OK?ò to a more junior doctor. This could be because they know the doctor had not been feeling well (Team Building), or because they are concerned the doctor was taking a long time with a patient (Workload Management) or because they know the doctorôs knowledge base is lacking (Supervision and Feedback).

5. Failure to rate objectively

This was sometimes because the observer was too generous and sometimes too harsh. . One observer wrote ñVery polite to nursesò and then only scored 5. The second investigator wondered if a higher score was warranted and the original observer agreed. Sometimes, a second investigator noted that similar behaviours were documented for two different registrarôs assessments but the ratings given were very different. Occasionally, no reason for the lack of objective scoring was given or the observer gave a vague reason like ñit was instinctò. One observer said that they started scoring more severely because they had previously been accused of being too lenient. Sometimes the observer admitted to be influenced by high or low scoring in other skills. For example, one observer said it was because ñthe registrar was so

niceò. Occasionally, the original observer was able to give additional reasoning for the rating that was not documented.

6. Observed behaviour not recorded/judged

It also became apparent in the post-assessment discussions that occasionally the two observers had seen the same behaviour but only one had thought to comment on it. For example, a registrar was treating a very sick child in the resuscitation room and one observer gave additional points for the fact the doctor remained calm and the other observer didnôt think to comment but agreed that they should have.

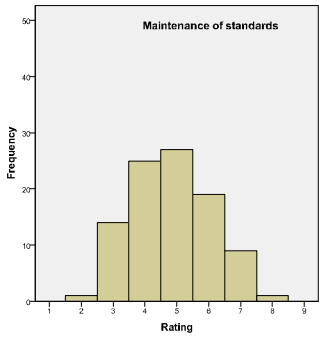
7. Disagreement between observers regarding what constitutes good and poor behaviour

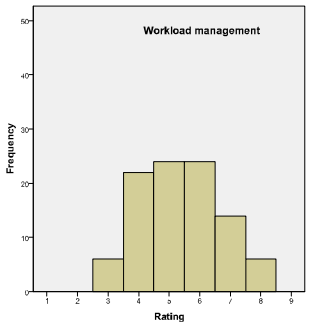
Review of assessment sheets sometimes revealed apparent disagreements about good and poor behaviour. This was sometimes because one observer had more experience of non-technical skills. For example, one observer scored a registrar down on Exchanging Information because of repeated use of vague instructions like ñCan I have a bag of fluid?ò and ñLet me know if there is a problemò. After explaining the reasons this type of communication can lead to errors, the second observer agreed that this was poor behaviour. Only once did the observers continue to disagree after discussion. This involved disparate ratings for the skill Maintenance of Standards. One observer said that although hand washing was not adequate, other standards such as documentation were good so, overall the behaviour was good. The other observer said that they felt unable to give an overall positive score with inadequate hand washing and that a score of five was more appropriate.

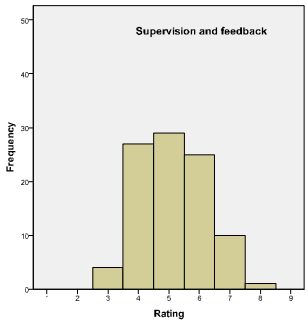
8. Observer not familiar with the department

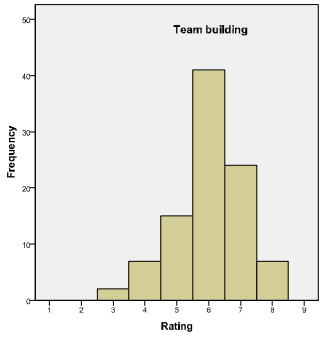
Within each pair of observers, one was fully accustomed to the department and the other was a visitor. This meant that one observer was at a disadvantage when interpreting some situations and behaviours. Examples related to unfamiliar referral patterns, availability of in-patient notes and systems for referring patients to CDU etc.

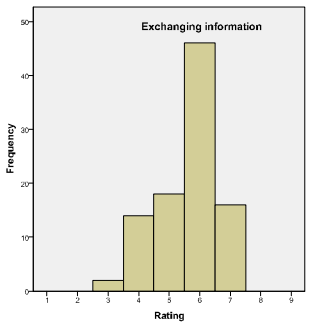
Bar charts showing frequency of rating for each skill

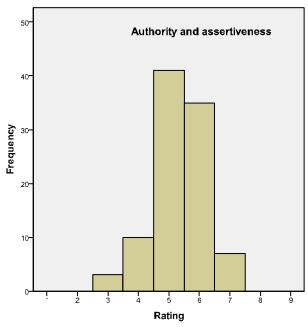


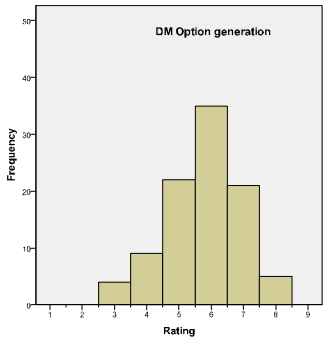


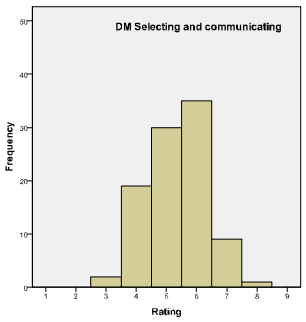


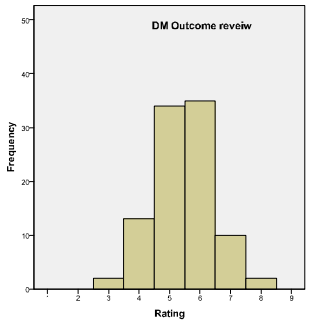


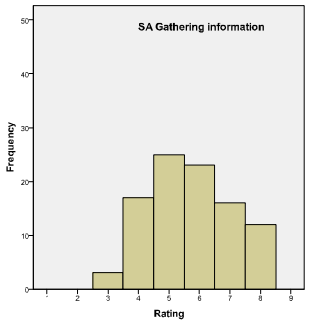


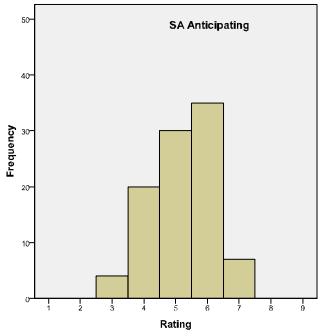


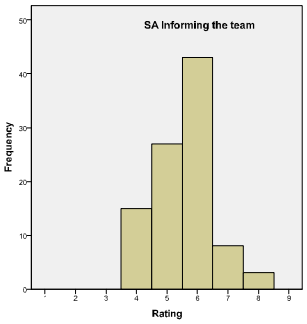












Description of good and poor behaviours observed for each skill

1. Maintenance of Standards

Positive examples often related to good documentation (for example notes timed, dated, signed; contemporaneous notes; explained importance of good documentation to SHO). There were also examples regarding the safety checks registrars used (checked details of drugs with patient; checked details on x-ray to ensure reviewing correct films; checked dose in BNF). Registrars were seen ensuring other staff complied with maintenance of standards (asked doctor to redo mistaken venous ABG). Most other comments referred to correct use of guidelines and good infection control procedures.

Usually, although not always, participants were seen to wash hands after contact with patients if no gloves were worn but if gloves were worn this was often omitted. Registrars infrequently washed or used alcohol gel before contact with patients. Other examples of poor behaviour were related to safety of sharps (stabbed Vacutainer bottles with a needle and syringe; re-sheathed a syringe; registrar passed needle and syringe directly to a colleague; inappropriate disposal of sharps) and leaving unlabelled bloods on a table. Documentation was also an issue with registrars failing to write contemporaneous notes (in the absence of an emergency), overly brief documentation and sometimes **no** documentation. Observers noted failure to follow national guidelines (including trauma, head injury and seizure guidelines) and safety issues regarding prescribing of medication (gave unnecessary verbal orders not followed up with written prescription; prescribed unfamiliar drugs without checking doses; delay in giving analgesia). A nurse was seen telling the registrar that a patient had become ñextremely tachycardicò and the doctor nodded but failed to act.

2. Workload Management

Examples of good workload management included prioritisation (prioritised a patient in pain), multi-tasking (bleeped and writes in notes while waiting) and delegation (asked other registrar to take a blue-call; locum asked for help preparing for catheter and the registrar delegated to a nurse). There were also examples where registrars were seen trying to encourage efficiency and effective prioritisation in other team members (sought and helped a doctor who was taking a long time with a patient; asked radiographer to prioritise a c-spine xray as the patient was likely to need a CT; asked a porter to prioritise transfer to CT instead of ward transfer). Observers commented on registrars who negotiated with team members (registrar offered to look at an ECG if the nurse got some medication for a patient) and some doctors who devised systems to help keep track of jobs (kept a notebook in pocket listing jobs). Registrars were frequently seen ensuring team members, both doctors and nurses, received appropriate breaks.

Registrars were observed getting unnecessarily overloaded with tasks (ôover-supervisedô competent team members; failed to delegate to nurses/technicians; failed to use shop-floor consultants) causing delays which at times appeared to compromise patient care. Conversely, some doctors were slow and inactive even during periods of overcrowding in the department. Other examples of poor behaviour highlight inefficient working practices (two senior doctors accompanied a patient to CT; told an SHO to book a bed then the registrar did it himself; unnecessarily interrupted an SHO during a patient consultation).

3. Supervision & Feedback

There were many good examples of registrars coaching junior doctors through the decision-making process and giving brief, ad hoc teaching sessions to team members. Registrars assessed capabilities of other doctors, nurses, medical students and locum doctors (asked nurse if able to do IV lines; checked locum doctorôs experience in the ED; asks student if he was confident doing a log roll). Occasionally registrars sought out doctors to get feedback about patients they had handed over or referred.

Observers commented on doctors who helped out more junior staff with referrals and requests (told SHO to report back if any problem getting a CT) and occasionally positive feedback was noted (told SHO it was a good decision to do an x-ray; ñYouôve already done a lying-standing BP, thatôs helpfulò). There were no examples documented relating to constructive criticism apart from suggesting revised management plans.

Registrars frequently missed ideal teaching opportunities when reviewing cases with junior doctors and interested nurses (failed to teach on cardiogenic versus neurogenic collapse; did not explain diagnosis of flutter on ECG to interested nurses). There were also examples when registrars did procedures on a junior doctorsô patient that could have been done by the doctor under supervision (registrar did nasal packing even when SHO offers to do it under supervision; registrar did femoral stab while doctor watched). One noticeable and relatively common observation was registrars who appeared reluctant to review patients for junior doctors or who failed to identify juniors who lacked confidence in their diagnosis or management and wanted a review (only reviewed a child with burns when SHO says ñDo you want to see it?ò; didnôt offer to review patient/ hip xray when SHO lacked confidence about diagnosis). Observers commented on a failure to assess capabilities of junior staff (told doctor to replace a suprapubic catheter without checking they were capable). On more than one occasion, medical students were observed standing around, looking uncomfortable in the department and apparently ignored by the lead registrar. Observers occasionally commented on lack of feedback to team members (did not give any feedback to new SHO who presented a case).

4. Team Building

There were few specific, positive examples of Team Building documented. Comments usually referred to the registrar being friendly and approachable and sometimes mentioned the use of team membersô names, for example when making requests. Assessment sheets included notes on registrars who introduced themselves to newly arrived resuscitation team members and on completion, thanked the team. Comments also included registrars who helped out, or were particularly friendly with

inpatient doctors in the department. Other examples included doctors who were polite when interrupted.

Registrars rarely scored less than five for Team Building (9/96). Some examples included registrars who appeared ñmiserableò or ñwithdrawnò or who were sarcastic with nursing staff.

5. Exchanging Information

Many examples of good communication were related to clear and concise referrals and listening attentively to junior doctorsô presentation of cases. Registrars also asked appropriate clarifying questions and repeated important messages (reiterated patient was not aware he had a malignancy). Clear instructions, for example, ñWeôll transfer the patient over to the trolley then weôll get the handover from the paramedicò on the arrival of a blue-light patient.

There were a number of examples of poor communication such as referrals that omitted important relevant facts (referral to medics omitted important observations and past medical history) or were too brief (very little information was given to anaesthetist on arrival in the resuscitation room for an unstable patient). Registrars making referrals were also criticised for failing to make the reason for referral clear (convoluted presentation rather than stating ónew presentation of diabetesô from the outset; unclear only asking for advice re follow up rather than admission under medics). Other examples included vague instructions or questions such as ñcall me if anything goes wrongò and ñwhy is this?ò when looking at an ECGò and ñare you happy?ò to the resuscitation nurse). Another common observation was registrars who failed to listen appropriately or give adequate attention (interrupts SHO several times when presenting a case to ask questions).

6. Authority & Assertiveness

Examples of appropriate use of assertiveness often related to in-patient teams who were reluctant to accept referrals. This skill was also displayed when interacting with ED nursing staff (assertive when realised prescribed fluids had not been given for over an hour; didn't give in when charge nurse says no nurse is available to assist). Rarely, examples were documented relating to disagreements regarding patient management. There was one example of a registrar disciplining a junior doctor. Observers also commented on registrars who appeared calm in an emergency or during difficult encounters with other staff (was calm when dealing with an acute MI; remained calm when Plastics were being obstructive about a referral).

7. Decision Making - Option Generation

Overall most registrars used available resources to get further background on a patient's presentation or medical history. This included accessing ED attendance records, in-patient notes, old ECG's, previous x-rays and scans, as well as contacting family and the GP to get additional information (saw left bundle branch block on ECG so chased old ECGs and when not available accessed discharge summaries; obtained collateral history from family). Registrars also used resources such as books, the internet and other staff to help decision making when they were unsure (used internet and books to help diagnose possible patellar fracture; showed ECG to consultant to get advice). Doctors often asked pertinent questions before being asking to prescribe medication or sign documents (enquired about observations and pain when signing ECG). Registrars frequently went to see a patient directly to help make a decision, especially when assisting an SHO to construct a management plan (reviewed patient before prescribing intravenous morphine for abdominal pain).

There were also examples where this skill was not performed satisfactorily and registrars failed to get adequate background on patients (did not look at old notes to clarify indication for recent foot amputation until medical SHO receiving referral suggested it). Registrars sometimes had referrals refused because they had gathered insufficient information or evidence. Registrars sometimes made decisions without

reviewing a patient in person when the observer judged it was indicated (failed to review SHOôs patient with reduced GCS when unsure of diagnosis). Sometimes doctors failed to ask relevant questions (does not ask SHO any questions related to child protection issues when discussing a child with burns on the sole of the foot). An observer documented an incident where a registrar failed to get adequate basic investigations, such as blood electrolytes, for a patient with status epilepticus before considering transfer to CT.

8. Decision Making - Selecting and Communicating

Registrars were observed giving good explanations for decisions to the team (stateed SHO did not need to refer as long as the patientôs pain improved and verbalised reasoning; ensured nurse understood reasoning when she was instructed to organise outpatient clinic appointment). Doctors were decisive and gave clear, stepwise plans (gave clear plan and back-up plan if unable to access relevant information).

There was also an array of incidents where registrars failed to give adequate reasoning for decisions (unclear why suggested referral to in-patient team; nurse questioned diagnosis of pneumonia when patient presented with vomiting but no explanation offered; gave verbal drug order to resuscitation nurse for new patient without any explanation of indication and then left). Registrars were seen being indecisive and staff were left with no clear plan of action (diagnosed atrial flutter on ECG but left the nurse unsure what to do next, e.g. monitor, repeat observations etc.; nurse informed registrar of unstable patient and doctors simply said ñIôll make a mental noteò; no clear plan suggested when reviewing a patient with an unusual rash). One observer noted a registrar who was vaguely assessing elements of a patientôs Glascow Coma Score but no definite score was given. Registrars often failed to discuss risks with the team.

9. Decision Making - Outcome Review

Examples of good behaviour included reviewing the impact of treatment (reviewed patient after nebuliser and asked for repeat observation) and ensuring a delegated job had been completed (ensured child received prescribed analgesia in a timely manner). Registrar also followed up decisions that had been made with more junior doctors (asked SHO for an update on a patient previously discussed). Sometimes management plans were altered due to new information becoming available (changed admission plan after hearing from doctors at Chelsea and Westminster Hospital). Occasionally, Outcome Review did not take place during the period of observation but registrars instructed team members to report back (ñPlease go and find out and then come back and tell meò).

Poor examples of Outcome Review tended to relate to either failure to assess the impact of treatment (failed to reassess patient after GTN prescribed to patient with chest pain; did not review impact of treatment for child who presented with shortness of breath so nurse had to find another doctor) or delays in making a reassessment. Registrars also failed to follow-up the outcome of patients that had been discussed with junior doctors.

10. Situational Awareness – Gathering Information

Many registrars exhibited situation awareness by listening in to other staff membersô conversations (overheard charge nurse speaking to the bed manager on the phone about overcrowding and gave assistance). Examples also involved doctors noticing patients in the department who appeared to be in need of help (noticed another patient looking unwell when calling a patient from the waiting room; heard a patient who sounded short of breath behind a curtain and saw if help is needed). Doctors were seen actively checking the situation of patients (walked around the department to œeye-ballô all patients in Majors and the resuscitation room). Registrars also demonstrated good situational awareness for organisational issues (used triage notes on tracking system to get information about waiting patients; asked to go through the

board with the charge nurse and in doing so identified two patients who had missed having a bed booked; went to CDU to get an update on the bed state).

Poor situational awareness was generally demonstrated by failing to adequately monitor patients or a lack of awareness of the state of the department. There were significant patient safety issues related to the monitoring of patients (left a fitting patient alone behind a curtain; failed to adequately monitor a patient in CT). Observers commented on lead registrars who made no attempt to check the patient tracking system or get updates from staff for the whole observation period.

11. Situational Awareness - Anticipating

Many comments for the skill Anticipation related to the movements of patients, especially during periods of overcrowding (put patient in CDU while awaiting bloods to make space). Doctors also thought ahead about the potential disposition of a patient (suggested booking an ambulance transfer early because delays were common; ensured SHO was considering patientôs disposition ñEither way, the patient is likely to go homeò). Some comments related to registrars who were seen planning ahead for changes in staffing levels due to shifts finishing or trainee teaching sessions (checked when members of team were finishing and considered moving staff from Minors). Occasionally registrars could be seen anticipating the deterioration of a patient and this usually involved transferring the patient to the resuscitation room (anticipated potential clot retention in a patient with frank haematuria and advised 3-way catheter; ensured nurse monitored childôs respiratory rate and saturations to detect deterioration). Sometimes registrars were seen preparing for the arrival of an emergency ambulance (got cannulation and phlebotomy items ready).

There were few poor examples of Anticipation documented and these were mainly associated with failure to plan ahead during periods of overcrowding (the resuscitation room was full and another emergency ambulance was expected ï registrar didnôt consider moving his patient to Majors even though his patient was the most stable). Occasionally, observers commented on doctors who failed to anticipate the potential deterioration of patients.

12. Situational Awareness - Informing the Team

This skill was most commonly displayed by registrars keeping the nursing staff updated about patients (let charge nurse know about a potential problem with a referral; kept charge nurse informed of all plans for patients). Sometimes doctors were observed telling ED nurses to handover certain details to ward nurses. Registrars also kept the team up-to-date about staffing and other organisational issues (told charge nurse he was going for a break; informed staff in minors that more doctors were coming to help; brief handover was given to registrar when he returned from break).

Similarly, many examples of poor behaviour were related to a failure to tell the team about organisational issues, rather than a patientôs clinical status (fails to inform nurse in charge or consultant that he was going for a break).

Results of the Kruskal-Wallis test

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Maintenance of standards is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.315	Retain the null hypothesis.
2	The distribution of Worklaod management is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.139	Retain the null hypothesis.
3	The distribution of Supervision and feedback is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.037	Reject the null hypothesis.
4	The distribution of Team building is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.071	Retain the null hypothesis.
5	The distribution of Exchanging information is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.167	Retain the null hypothesis.
6	The distribution of Authority and assertiveness is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.086	Retain the null hypothesis.
7	The distribution of SA Gathering information is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.007	Reject the null hypothesis.
8	The distribution of SA Anticipating is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.259	Retain the null hypothesis.
9	The distribution of SA Informing the team is the same across categories of Hospital.	Independent-Samples Kruskal-Wallis Test	.298	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.