THE ROLE OF SIMULATIONS IN THE ASSESSMENT OF TECHNICAL PROFICIENCY AND COMPETENCE IN SURGICAL TRAINING

By

KRISHNA MOORTHY

MBBS, FRCS

Department of Surgical Oncology and Technology, Imperial College, St. Mary's campus, London, UK

A thesis submitted to the University of London for a Doctorate in Medicine (MD)

2004

Supervisors: Prof. Sir Ara W. Darzi and Prof. C Vincent

Department of Surgical Oncology and Technology, Imperial College, St.Mary's campus, London, UK

ABSTRACT

Surgical training has responded to the challenges of reduced training opportunities, shortened working hours and financial pressures by using simulations primarily for the acquisition of basic psychomotor skills. However, one of the major drawbacks has been the absence of evidence that this paradigm of training actually has any benefit. This thesis has described the use of objective measures of skills such as dexterity analysis systems, virtual reality simulators and video-based expert analysis in the assessment of task and procedural simulations. It has described the development of multi-component assessment protocols which should aid the process of formative feedback during training.

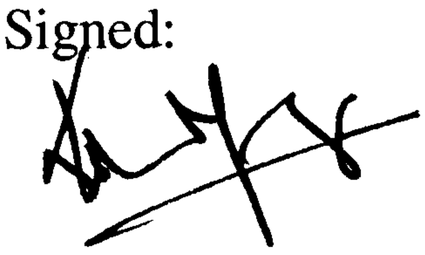
This thesis is designed along a skills-assessment escalator which progresses from the use of simple task based simulations for the assessment of technical proficiency to more complex team based simulations for the assessment of other aspects of surgical competence. The assessment of performance during both a 'quiet' procedure and the bleeding crisis, using objective measures of performance, revealed variability in performance both between and among groups of trainees of varying grades. This thesis has established the face validity of the simulated scenarios by demonstrating the realism of the surgical model and participant's experience in a Simulated Operating Theatre.

In the future, objective measures of technical proficiency and surgical competence can be used to evaluate the benefit of simulation- based training, demonstrate the transfer of skills from simulated environments to real procedures and correlate performance in simulations with performance in real life- the final affirmation of the validity of simulations.

DECLARATION

I hereby declare that I am the sole author of this thesis and that all work within is my own, except where work was carried out in collaboration with other individuals, who are appropriately credited. I authorise the library of the University of London to lend this thesis to other institutions or individuals for the purpose of research.

Signed:



Date: 26/7/24

Krishna Moorthy

MBBS, FRCS

PEER REVIEWED PRESENTATIONS AND PUBLICATIONS

Multimodality assessment of psychomotor skills in endoscopic suturing. Validity and correlation between its individual components. Moorthy K, Munz Y, Dosis A, Bello F, Chang A, Darzi A. Presented at the 11th annual conference of EAES, Glasgow, June 2003. Abstract In Press- Surgical Endoscopy

A human factors analysis of technical and team skills during procedural simulations. Moorthy K, Munz Y, Adams S, Pandey V, Martin S, Poore L, Lo B, Edwards L, Yang GZ, Darzi A. Presented at the annual meeting of the Association of Surgeons of Great Britain and Ireland. Manchester, May 7-9, 2003. Abstract in Br J Surg; 90 (Suppl 1): 88-89.

The use of virtual reality simulators in multi-modality assessment of technical skills in lower gastrointestinal endoscopy: a pilot study. Moorthy K, Munz Y, Orchard T, Gould S, Rockall T, Darzi A. Presented as an oral presentation at the annual conference of the Society of American Gastrointestinal Endoscopic Surgeons (SAGES), Los Angeles, March 12-15, 2003. Abstract in Surgical Endoscopy. 2002; 17 (Suppl): S 198.

Learning the skills of flexible sigmoidoscopy- the wider perspective. Kneebone R, Nestel D, Moorthy K, Taylor P, Bann S, Munz Y, Darzi A. Medical Education. In Press

ACKNOWLEDGEMENTS

To Prof. Sir Ara Darzi for his inspiration and guidance. To Prof. Charles Vincent for his advice and for helping me maintain focus. To Yaron Munz for his patience with my often insane ideas and for all those brainstorming sessions that led to even more ideas. To Lee Edwards, Shirley Martin, Lorraine Poore, Vikas Pandey and Shabnam Undre without whose support the simulations in the Simulated Operating Theatre would have been impossible. To Roger Kneebone and Simon Bann who were instrumental in helping me with the inception of some of the studies that are integral to this thesis. To Dr. Dave Peck for his advice and patience. To all my medical colleagues for participating as subjects in my studies. To my sister and her husband for providing me with not just food and shelter but even a few light moments at the end of many a disappointing day. And most of all, to my wife, Leena, for the sacrifices that she has had to make over the last two years and for her unfailing good humour and unstinting faith in me.

CONTENTS

Page

ABSTRACT 2

DECLARATION 3

PEER REVIEWED PRESENTATIONS AND PUBLICATIONS 4

ACKNOWLEDGEMENTS 5

CHAPTER 1: INTRODUCTION

1.1 Training in surgery 21

1.1.1 History 21

1.1.2 Training in surgery in the UK 22

1.1.3 Challenges to the current model 24

1.1.4 Innovation in surgical training: the role of simulations 28

1.1.5 Transfer of skills from simulations to real procedures 34

1.1.6 Task analysis and the development of a structured surgical curriculum 35

1.2 Assessment of skills 37

1.2.1 Assessment in surgery

1.2.1.1 The RITA process during Calman training 38

1.2.1.2 Principles of assessment 38

1.2.1.3 Current methods of assessment 41

1.2.1.4 Recently developed objective methods of assessment 43

1.2.2 Assessment in endoscopy 50

1.3 Technical proficiency and surgical competence 52

1.4 Human factors in high-reliability organisations and medicine 55

1.5 Human factors in the Operating Room (OR) and surgery 58

1.5.1 The OR team and inter-personnel dynamics 58

1.5.2 Leadership in the OR 61

1.5.3 The Input-Process-Output model 62

1.5.4 Individual attitudes in the OR 64

1.6 Enhancing safety in aviation: Crew Resource Management (CRM) training 66

1.6.1 Evolution of CRM 66

1.6.2 The fifth and current generation of CRM 67

1.6.3 Evaluation of non-technical skills 67

1.6.4 Studying the positive effects of CRM 69

1.7 CRM in medicine 70

1.7.1 Anaesthesia CRM (ACRM) 70

1.7.2 Assessment of performance during ACRM and measurement tools 72

1.7.3 Effectiveness of ACRM 74

1.7.4 Validity and reliability of ACRM 75

1.7.5 Value of simulation based training 76

1.8 The Imperial College experience (background to the research project) 77

1.9 Rationale behind the research 78

1.10 Project design and structure- Skills and Assessment Escalator 80

1.11 Aims of the research project 82

CHAPTER 2: SIMPLE TASK BASED SIMULATIONS FOR THE TRAINING AND ASSESSMENT OF TECHNICAL PROFICIENCY

2.1 Introduction 84

2.2 Aims 87

2.3 Materials and Methods 88

2.3.1 Subjects 88

2.3.2 The simulated task 88

2.3.3. Assessment of performance 89

2.3.4 Data analysis 90

2.4 Results 93

2.5 Discussion 100

CHAPTER 3: HIGHER FIDELITY PROCEDURAL SIMULATIONS FOR THE ASSESSMENT OF TECHNICAL PROFICIENCY- ENDOSCOPY

3.1 Introduction 105

3.2 Aims 108

3.3 Materials and Methods 109

3.3.1 The subjects 109

3.3.2 The Virtual Reality simulator 109

3.3.3 The Virtual Endoscopy Suite 110

3.3.4 Assessment of performance 111

3.3.5 Data analysis 112

3.4 Results 116

3.4.1 Construct validity of the simulator's metrics 116

3.4.2 ICASE score 116

3.4.3 Correlation between the ICASE score and the simulator 117

3.5 Discussion 121

CHAPTER 4: TEAM BASED PROCEDURAL SIMULATION FOR THE ASSESSMENT OF SURGICAL COMPETENCE

4.1 Introduction 125

4.2 Aims 129

4.3 Materials and Methods 130

4.3.1 The Simulated Operating Theatre 130

4.3.2 The subjects 134

4.3.3 Study design 134

4.3.4 Performance measures and validity of the simulation 136

4.3.5 Data analysis 141

4.4 Results 142

4.4.1 Technical skills assessment 142

4.4.2 Essential items checklist 147

4.4.3 Non-technical skills assessment 148

4.4.4 Communication- Utterance Frequency 151

4.4.5 Correlation between UF and communication score 155

4.4.6 Correlation between technical and non-technical skills scores 155

4.4.7 Face validity of the simulation 157

4.4.8 Debriefing data 157

4.4.9 Summary of results 158

4.5 Discussion 159

CHAPTER 5: ASSESSMENT OF CRISIS MANAGEMENT AMONG SURGICAL TRAINEES

5.1 Introduction 166

5.2 Aims 171

5.3 Materials and Methods 172

5.3.1 The subjects 172

5.3.2 Study design 172

5.3.3 Simulated operating theatre 172

5.3.4 Assessment of Performance 176

5.3.5 Data analysis 180

5.4 Results 181

5.4.1 Validity and reliability 181

5.4.1.1 Face validity of the simulation 181

5.4.1.2 Construct validity of the assessment measures 182

5.4.1.3 Reliability of the assessment measures 183

5.4.2 Effect of the crisis on performance 190

5.4.2.1 Technical skills 190

5.4.2.2 Non-technical skills 190

5.4.3 Content validity of the non-technical skills assessment 192

5.4.4 Additional observations 193

5.4.4.1 Observations of unsafe practice 193

5.4.4.2 Examples of safe practice 194

5.4.4.3 Monitoring 194

5.5 Discussion 195

CHAPTER 6: OVERALL DISCUSSION

6.1 Assessment of technical proficiency 200

6.1.1 Importance of multi-component assessment protocols 201

6.1.2 Validity and reliability of the assessment measures 203

6.2 Assessment of surgical competence 203

6.2.1 Face validity of the simulated scenarios 204

6.2.2 Validity of the assessment measures 205

6.3 Crisis management among surgical trainees 206

6.4 Transfer of skills from simulations to real procedures 207

6.5 Implications of objective assessment of competence 209

6.6 Future work 211

REFERENCES 213

LIST OF FIGURES

Page

CHAPTER 1

Fig 1A: Yerkes- Dodson curve 25

Fig 1B: Laparoscopic VR simulator 32

Fig 1C: MIST-VR tasks 32

Fig 1D: Imperial College Surgical Assessment Device (ICSAD) 47

Fig 1E: Team profile in the Operating Room 59

Fig 1F: Input- Process- Output Model 62

Fig 1G: Skill-assessment escalator 81

CHAPTER 2

Fig 2A: Box trainer 91

Fig 2B: ICSAD 91

Fig 2C: Checklist for laparoscopic suturing 92

Fig 2D: Box-plots- time taken per suture 95

Fig 2E: Box-plots- path length per suture 96

Fig 2F: Box-plots- checklist score 97

Fig 2G: Cluster box- plots for needle loading and needle driving 98

Fig 2H: Cluster box- plots for knot technique and knot quality 99

CHAPTER 3

Fig 3A: Lower gastrointestinal VR simulator 113

Fig 3B: Virtual endoscopy suite 114

Fig 3C: ICASE 115

Fig 3D: Cluster box-plots- Depth of insertion and percentage of mucosa 118

Fig 3E: Box-plots- Percentage of red-out 119

Fig 3F: Box-plots- ICASE score 120

CHAPTER 4

Fig 4A: Simulated Operating Theatre (SOT) 132

Fig 4B: Clinical data recorder (CDR) 133

Fig 4C: Sapheno-femoral junction model 136

Fig 4D: Box-plots- technical skills 144

Fig 4E: Box-plots- ICEPS score 145

Fig 4F: Scatter plot- global score vs ICEPS score 146

Fig 4G: Box-plots- Non-technical score 150

Fig 4H: Utterance frequency (UF) 153

Fig 4I: Box-plots- Surgeon- anaesthetist UF 154

Fig 4J: Scatter plot- technical and non-technical scores 156

CHAPTER 5

Fig 5A: Bleeding scenario 175

Fig 5B: Box-plots- technical ability to control bleeding 186

Fig 5C: Box-plots- non-technical skills score 186

Fig 5D: Box-plots- Decision making 187

Fig 5E: Box-plots- Utterance frequency 187

Fig 5F: Box-plots- Crisis management time measures 188

Fig 5G: Box-plots- Surrogate outcome measures 189

CHAPTER 6

Fig 6A: Stepwise learning in surgery 210

LIST OF TABLES

Page

CHAPTER 1

Table 1A: Assessment of technical skills: Watts and Feldman 40

Table 1B: Op Comp assessment 40

CHAPTER 2

Table 2A: Assessment of laparoscopic suturing 94

Table 2B: Individual components of the checklist 94

CHAPTER 3

Table 3A: Simulator's metrics 117

CHAPTER 4

Table 4A: Essential items checklist 138

Table 4B: Utterance frequency 140

Table 4C: Technical skills assessment 143

Table 4D: Non-technical skills scores 149

Table 4E: Scores for components of non-technical skills 149

Table 4F: Utterance Frequency 152

Table 4G: Surgeon- anaesthetist UF 153

CHAPTER 5

Table 5A: Questionnaire results 184

Table 5B: Non-technical skills, DM, UF 185

Table 5C: Crisis Management- time measures 185

Table 5D: Effect of the crisis 191

LIST OF ABBREVIATIONS

A: Assistant

A and E: Accident and Emergency

ACRM: Anaesthesia Crisis Resource Management

ADEPT: Advanced Dundee Endoscopic Psychomotor Trainer

ANTS: Anaesthesia Non-Technical Skills

BST: Basic surgical training

CCST: Certificate of Completion of Specialist Training

CDR: Clinical Data Recorder

CMAQ: Cockpit Management Attitudes Questionnaire

CRM: Crew/ Cockpit Resource Management

DM: Decision Making

DVD: Digital Versatile Disc

ER: Emergency Room

FMAQ: Flight Management Attitudes Questionnaire

GI: Gastrointestinal

HST: Higher surgical training

ICSAD: Imperial College Surgical Assessment Device

ICASE: Imperial College Assessment of Skills in Endoscopy

ICEPS: Imperial College Evaluation of Procedure- specific Skills

JCHST: Joint Committee for Higher Specialist Training

LAR: Leak rate ratio

LOFT: Line Oriented Flight Training

LOSA: Line Operations Safety Audit

MIST-VR: Minimally Invasive Surgical Trainer- Virtual Reality

MRCS: Membership of the Royal College of Surgeons

NASA: National Aeronautics and Space Administration

ODA: Operating Department Assistant

OR: Operating room

ORMAQ: Operating Room Management Attitudes Questionnaire

OSATS: Objective Structured Assessment of Technical Skills

OSCE: Objective Structured Clinical Examination

RCS: Royal College of Surgeons

RITA: Record of In- Training Assessment

S: Surgeon

SAC: Surgical Advisory Committee

SFJ: Sapheno-femoral Junction

SN: Scrub Nurse

SOT: Simulated Operating Theatre

SP: Simulated Patient

TC: Time to use effective control measures

TD: Time to diagnose

TH: Time to call for help

TL: Time to close the laceration

TT: Time to inform the team

UF: Utterance Frequency

VES: Virtual Endoscopy Suite

VR: Virtual Reality

LIST OF APPENDICES

Appendix 1: Checklist 238

Appendix 2: Global rating scale 239

Appendix 3: NOTECHS rating scale 240

Appendix 4: ANTS 241

Appendix 5: Simulation brief to the surgeon 242

Appendix 6: Modified global rating scale 243

Appendix 7: ICEPS 244

Appendix 8: Technical ability to control the bleeding 245

Appendix 9: Virtual Operating Theatre Non-technical Skills Assessment 246

Appendix 10: Questionnaire to assess the perception of participants 247

CHAPTER 1: INTRODUCTION

1. 1 TRAINING IN SURGERY:

1.1.1. A brief history

Training in surgery has traditionally depended on an apprenticeship model and learning 'on the job'. Trainees acquire their skills by learning under the supervision of a senior surgeon. Training in the apprenticeship model consists of three phases (1). In the first phase the trainee observes a senior surgeon. The trainee then assumes the role of the main operating surgeon and learns surgical skills under the supervision of the training surgeon. Eventually trainees are allowed to perform procedures on their own.

William Halstead was a pioneer of the apprenticeship model and acquired his expertise by travelling to numerous centres in Germany. In turn, his residents carried his methods to numerous centres in the United States (2). In addition, Halstead spent a brief period in the US where he was instrumental in establishing the residency system (3). In the United Kingdom the acquisition of surgical skills was traditionally dependant on period of apprenticeship under a 'doyen' of surgery (4), till a structured residency programme was established in Guy's Hospital, London in 1972 with the aim of making surgical training more systematic and in the process giving registrars in surgery more security and guidance (5). A further change to the structure of training came about in 1993 with the introduction of the Calman system. The stimulus for this came from the finding that, even though the structured residency system was successful in ensuring uniformity of training,

there were some senior registrars who remained in their positions for periods varying from 3-10 years without any guarantee of career progress to consultant status. One of the reasons for this was that there were not enough consultant positions for the large number of registrars. The Calman system was set out to address this discrepancy by ensuring that Calman numbers given out one year matched the predicted requirement of consultant positions in the future. The intention of the Calman system was also to make training more structured and supervised instead of relying primarily on experience (6). The system defined a curriculum for each specialty and a minimum training period at the end of which the trainee would qualify to enter the specialist register and qualify to become consultant.

1.1.2 Training in surgery in the UK

On completion of a year of house-jobs, those intending a surgical career enter a 2-3 year Basic Surgical Training (BST) programme. During this period they rotate between different specialities usually consisting of Accident and Emergency, orthopaedics and general surgery depending on the programme and the hospitals they train in. On completion of the BST programme, they appear for the Membership of the Royal College of Surgeons (MRCS) examination which consists of a set of multiple choice questions, viva voce and clinical examinations. Once they have completed their BST training and have successfully completed the MRCS examination, they are eligible to apply for a Calman training number and enrol into a Higher Surgical Training (HST) programme. The HST training programme lasts for a period of 6 years of which one year can be taken out to spend in a centre of excellence in the specialty of interest. The first 3 years are

spent in training in general surgery with exposure to all its sub-specialities such as colorectal, vascular, breast and upper gastrointestinal surgery. The last two years in the programme are usually spent in training in an area of sub-specialty interest. On completion of the HST programme, they appear for an exit examination in general surgery and are usually expected to declare an area of sub-specialty interest. On passing the exit examination and completion of the 6 years, they become eligible for a Certificate of Completion of Specialist Training (CCST) and are eligible to apply for consultant posts.

Thus, training in the UK is similar to the residency system in the US, though slightly longer, and has been made more structured and goal-based with the introduction of the Calman system. However, training is still based on an apprenticeship model, where trainees acquire their skills under the supervision of many consultants as they usually move from hospital to hospital within a region on an annual basis. Training is primarily dependant on learning skills in the operating theatre. The apprenticeship model is thus dependant on the surgical and teaching skills of the trainer as well as his/her commitment to training and the availability of time and resources. This model of training referred to as 'learning-by-doing' (7) or the 'see one, do one, teach one' model (8) is time-tested and over the decades has ensured the free exchange of skills and knowledge from one generation to the next. Over the years, however, there has been an acknowledgement of the limitations of training solely by the apprenticeship model and there are a number of reasons to augment this model of training with other training methods.

1.1.3 Challenges to the current model of training

1) Resources:

In the United States, training 1000 residents in the operating theatre costs nearly 50 million dollars a year (9). In the present environment of managed healthcare with increasing pressure on hospitals to save costs, one simple solution would be to reduce trainee's exposure to cases and for most procedures to be undertaken by senior surgeons resulting in a larger number of cases being performed in one operating session (8). In order to increase efficiency and at the same time ensure that surgical training is not compromised, it would be essential to develop structured training curricula and offer trainees other training opportunities in addition to training in the operating theatre (9).

2) Risk to patients

There are a number of studies that have challenged the commonly held belief that procedures performed by surgical trainees result in poorer outcomes for patients (10;11). However, there is abundant evidence to indicate that there is a high risk of complications during the early phase of a surgeon's learning curve (12;13). It has been clearly demonstrated that the advent of laparoscopic surgery in the late eighties was accompanied by an increase in the incidence of procedure-related complications (14;15). This was essentially due to a number of surgeons delivering this form of care without suitable training. One author commented on the fact that laparoscopic surgery was introduced without the legitimacy of a residency based training programme(16). This also prompted Cuschieri to comment about the 'uncontrolled expansion of surgical

endoscopic practise which amounted to the biggest un-audited free-for-all in the history of surgery' (17).

3) Unsuitable training environment

The main priority in the operating room (OR) is patient care rather than training. In addition some trainees may view the OR as a stressful and hostile environment (8;18). Time pressures from OR scheduling, the presence of other OR team members and the fear that the case will be 'taken over' by the trainer at any stage places the trainee in a fairly stressful situation. The Yerkes- Dodson curve of stress vs. learning shows (Fig 1A) that learning occurs best in an environments of moderate stress with lower or higher levels of stress being counter-productive to the learning process (19).

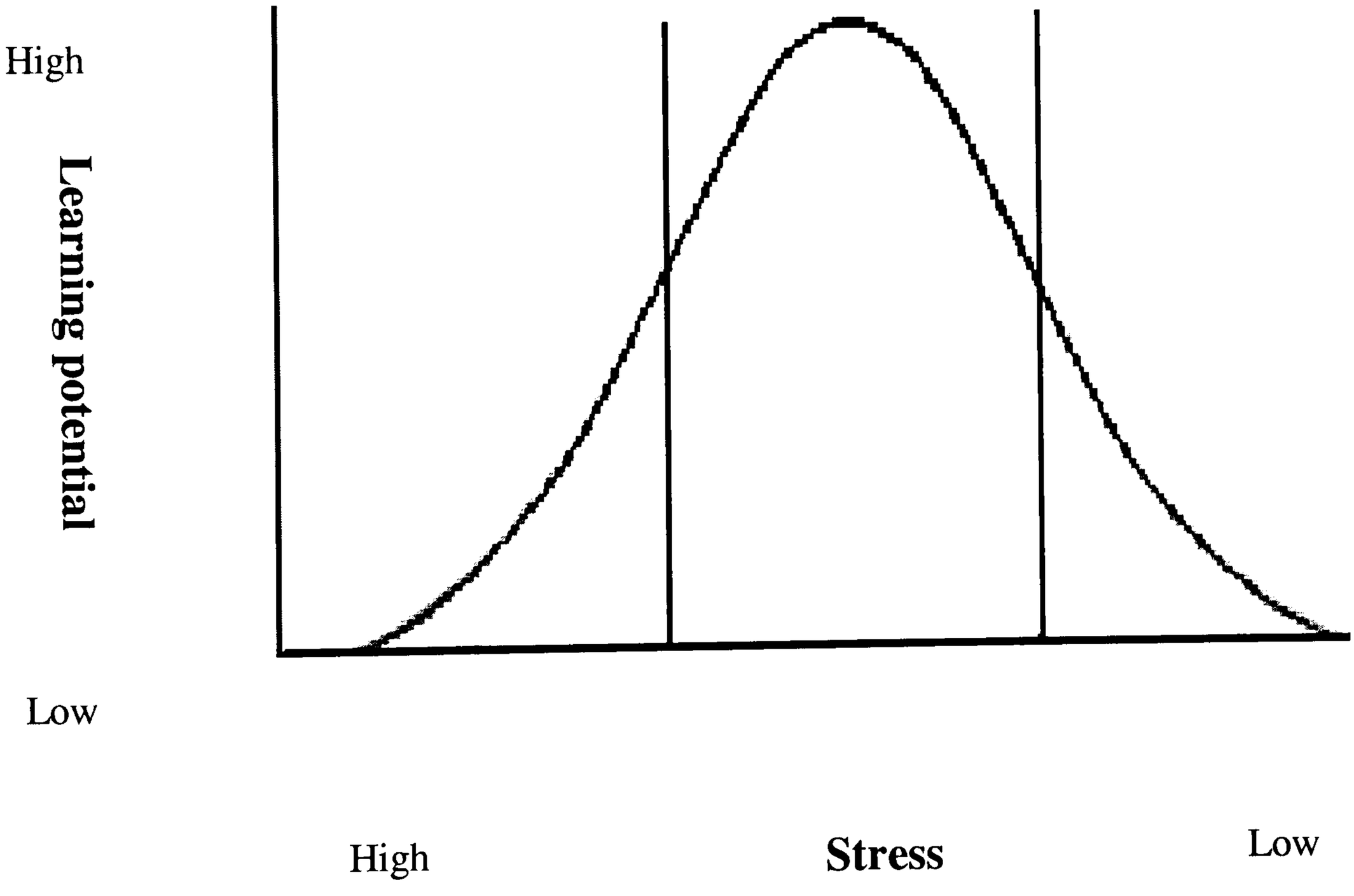


Fig 1A: Yerkes- Dodson law: Inverted 'U' shape curve- areas of high and low stress result in a low learning potential

4) The changing structure of health care delivery.

Pressure on hospitals especially in the UK has led to more cases being performed on a day surgery basis (20), with nearly 65-70% of elective procedures in the UK and the US being performed as day procedures (21). There is a commonly held belief that the success of this change in the delivery of health care depends on cases traditionally considered suitable for training, such as inguinal hernias, varicose veins and laparoscopic cholecystectomy (22) being performed by senior surgeons. This can have great implications for training (23) with reduced training opportunities. Managed care in the US has also placed significant financial constraints on hospitals and surgeons (24).

5) The threat of malpractice claims and the publication of League tables

There has been a increase in the number of medico-legal claims against surgeons over the past few years (18). In addition, there has been a recent move to publish league tables for surgeons and hospitals with regards to morbidity and mortality (25;26). Both these issues have the potential to result in a scenario where senior surgeons would be compelled to perform more procedures. This would be at the expense of training. It is also likely that the publication of league tables would results in doctors becoming reluctant to treat the very sick who are at a higher risk of complications and death (25). This would reduce a trainee's exposure to an adequate and varied case-mix which is essential for training in surgery.

6) Changes to training- Working hours

Over the past decade there has been a move to reduce the working hours of junior doctors (trainees) in Europe with the implementation of the European Working Time Directive (27) and in the US, where a similar ruling was first implemented in New York state (28;29). Though studies have shown that this results in an improvement in the life style of trainees(29) , there is some concern amongst surgical trainees that such a move would result in a reduction of their training opportunities and experience (30).

7) Changes to training- Calman

To a large extent, trainees appear to be satisfied with the Calman system (6). However, there have also been anxieties, especially amongst the surgical specialities, that the shorted training period would result in reduced experience and skills amongst trainees (31) (32).

8) The impact of new technologies

An increase in the incidence of complications with the introduction of laparoscopic surgery emphasizes the need to undergo structured training prior to the adoption of a new technique. In the case of laparoscopic surgery, a number of surgeons undertook procedures without acquiring the basic skills essential for the safe and effective practise of this technique. They had to acquire expert hand-eye coordination, adapt to a two dimensional image and rely more on their visual rather than haptic cues (33;34). The lack of supervised training and the random adoption of this technique resulted in a number of adverse events especially damage to the bile duct due to the careless use of the diathermy

in Calot's triangle (35). The evidence that there was a learning curve for this procedure and the fact that complications were lower at institutions that performed a large number of these procedures (15), led to the recommendation to provide adequate training in order to reduce the incidence of adverse events. The integration of new technologies such as robotic surgery provides further challenges to resident training as some of the more traditional resident procedures get performed by senior surgeons during their learning process (36).

1.1.4 Innovation in surgical training: the role of simulations

The above-mentioned challenges to training in the past decade have compelled the surgical community to consider changes to the ways in which surgery is taught. The need to augment training in the operating room (OR) with other methods, has led to the development of courses and workshops where trainees could acquire their skills in a 'non-threatening and unhurried environment'(18). Learning in the operating theatre provides little opportunity for practise and reflection and this is one of the strongest stimuli for taking training in surgical skills into non-OR environments (1;4). *Simulation* can be defined as a device or exercise that enables the participant to reproduce or represent, under test conditions, phenomena that are likely to occur in actual performance (37). Surgical simulations intended to help trainees acquire basic psychomotor skills can be synthetic, animal tissues or virtual reality. Using these various models, a number of courses have been established across the globe, including those conducted by the Royal Colleges of UK (38).

Training courses in surgical technical skills

The Royal Colleges in the UK run courses for trainees according their level of training and experience. The intercollegiate basic surgical skills course is now mandatory for all basic surgical trainees in the UK. Over three days trainees are instructed on the methods of performing basic skills such a knot tying and suturing to tasks such anastomosis and excision of lesions. The transition from one skill to the other is performed in a structured manner with strict adherence to a format developed by the Royal Colleges. The training is thus uniform across a number of centres offering such courses and trainees practise these skills on synthetic bench and animal cadaveric models. Over a period of time the department of education at the Royal College of Surgeons of England, has developed an Intermediate Skills course, applying similar principles as the basic skills course, to teach more complex tasks and procedures on bench and animal cadaveric models to higher surgical trainees.

The use of live animals is banned in the UK (1) but is still commonly used by some skills centres (39). The advantages with live animal models are the amount of realism they provide during operative surgery (40). The disadvantages are the lack of anatomical simulation, cost, moral issues and concerns about infectious diseases (4).

Training courses in laparoscopic surgery

The concerns regarding the random introduction of laparoscopic surgery emphasized the need for structured training courses (41). The intent was to teach surgeons the skills essential for laparoscopic surgery such as hand-eye co-ordination under the supervision of skilled surgeons. A number of these courses use animal cadaveric and synthetic

models but in some skills centres, procedures on live models are considered essential (41). Kirwan et al were the first to describe the use of pigs for learning and practising laparoscopic cholecystectomy (41;42). However with the development of virtual reality simulators it is likely that the need for training on animal models will reduce (43;44) and their use will be difficult to justify (45). It has already been demonstrated especially in open surgery that synthetic models are equal to animal models in teaching trainees surgical skills (4). There are now a number of established laparoscopic skills courses across the globe and a number of expert surgeons have established curricula for teaching basic skills in box-trainers (46;47).

Virtual Reality in surgical training

Virtual reality (VR) is defined 'as a collection of technologies that allow people to interact efficiently with 3D computerised databases in real time using their natural senses and skills (48). This definition highlights that fact that the interaction no longer involves the use of mounted headsets and gloves which were the features of VR in the eighties and early nineties. The main feature of VR is that it allows real time interaction for the user through an interface. This interface could be a laparoscopic frame (Fig 1B) with modified laparascopic instruments.

There are presently a number of VR simulators available for laparoscopic surgery. Most of the VR simulators developed so far are predominantly task trainers where trainees acquire the basic skills associated with laparoscopy. The success of this lay in the ability of some surgeons, working in collaboration with psychologists, to carry out task analyses of surgical procedures. In addition, it has been technologically easier to develop VR

simulators for laparoscopic surgery as compared to open surgery because of the ergonomics associated with the former. Laparoscopic surgery is monitor based, and the user requires hand-eye coordination and depth perception and has to adapt a two-dimensional view from the monitor and extrapolate that to the real three dimensional surgical environment. There is blunting of haptic feedback and the user has to rely more on visual cues.

The MIST-VR (Minimally Invasive Surgical Trainer- Virtual Reality) system was one of the first VR laparoscopic simulators developed as a task trainer. The system's development was the result of collaboration between surgeons and psychologists who performed a task analysis of laparoscopic cholecystectomy. This revealed a 'tool-kit' of skills needed for the successful performance of the procedure (48). These were then replicated in the virtual domain by producing 3-D images of geometric shapes, which users could manipulate (Fig 1C).

More recently, VR simulators have been made more realistic with the incorporation of tissue features such as tissue deformability, bleeding and haptic (touch) feedback (49). The next generation of laparoscopic VR simulators are more complex (include force feedback mechanisms) with the potential that in the future complete procedural simulations will be possible. This group includes several simulators, some of which have been validated, such as the Xitact LS500 (50;51) and the Reachin RLT (Reachin Technologies, Sweden) (unpublished data).

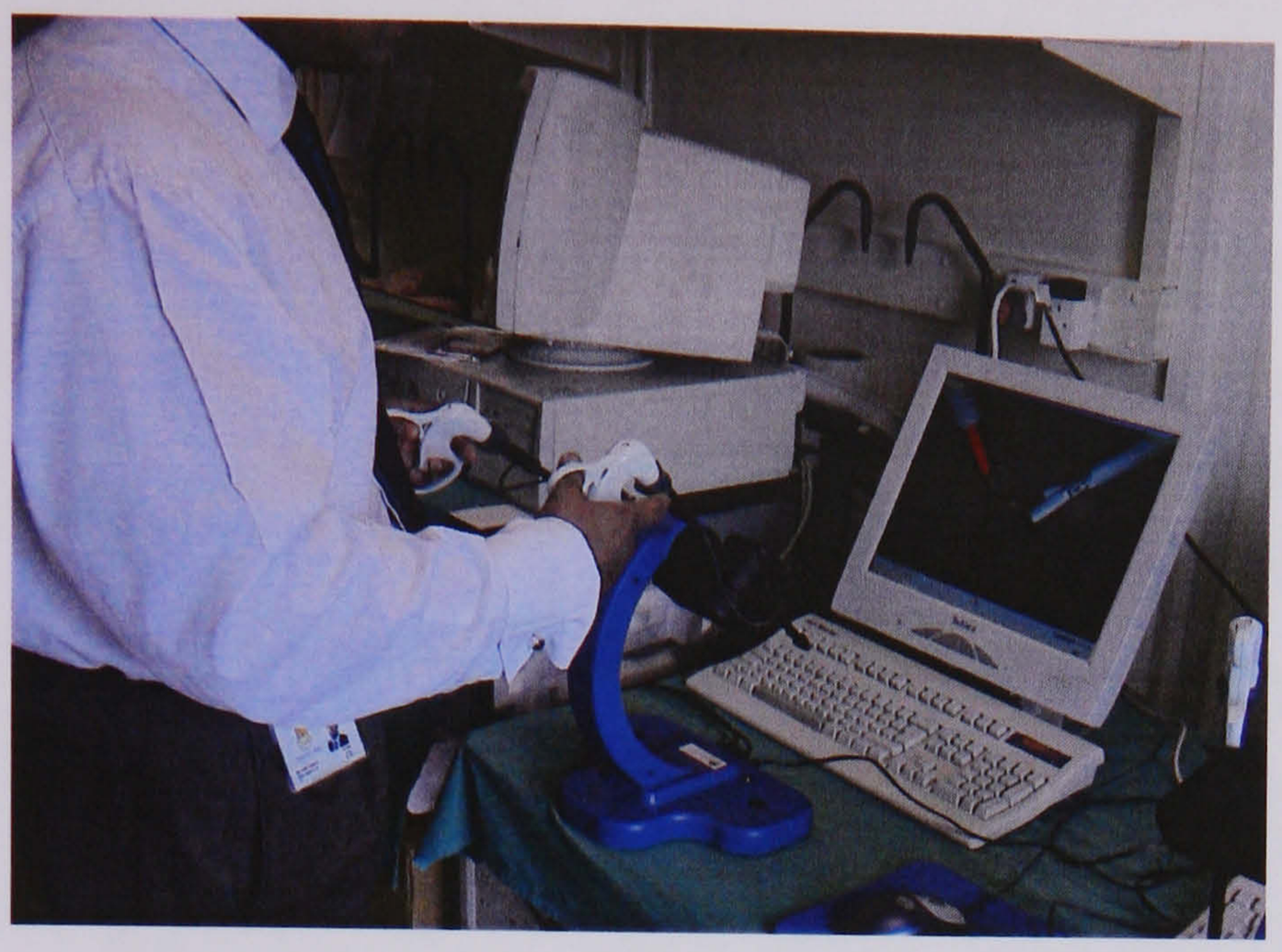


Fig 1B. Laparoscopic VR simulator- User interacts with the software through a laparoscopic frame

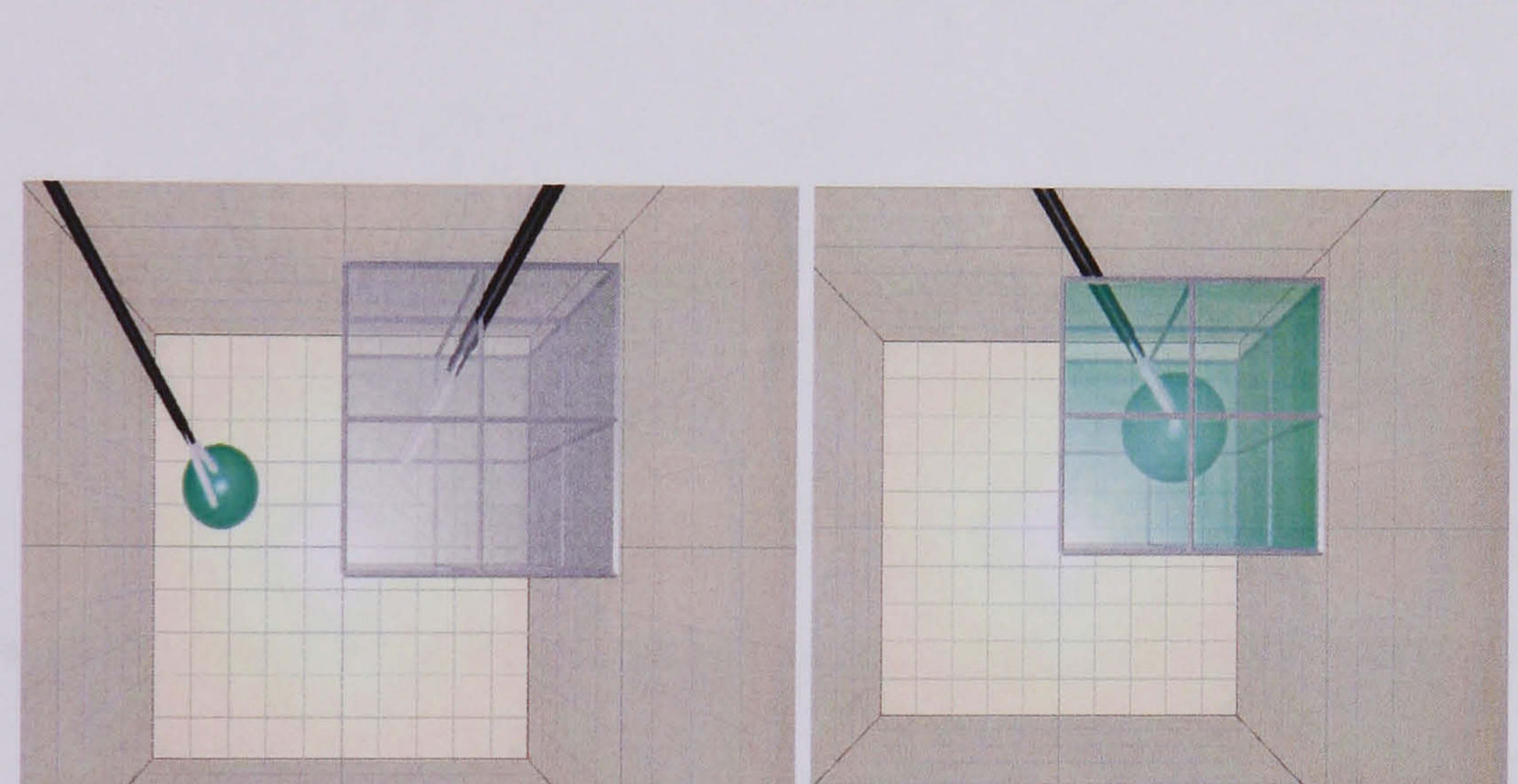


Fig 1C: MIST-VR tasks: The two above figures are from the acquire-place task and lower two figures, from the traversal task

Training in endoscopy

Endoscopic procedures are in many ways similar to laparoscopic surgery and require the mastery of unique skills. Endoscopists have to work without stereoscopic vision, adapt an image from a 2-D monitor to a 3-D perspective, and develop hand-eye coordination and depth perception. More importantly endoscopists have to perform procedures as quickly as possible as patients are most often only lightly sedated. A large part of the learning curve in the initial stages is spent in acquiring these basic psychomotor and visuo-spatial skills.

Training has so far consisted solely on training on real patients in endoscopy suites following the Halstadian "see one, do one, teach one" model. In addition to all the problems associated with this model, some problems unique to endoscopy are the need for the trainee to have access to an endoscopy suite and the fact that most first time procedures are carried out on real patients, with the associated risk of such an approach. It has also been found by a study called the Clinical Outcomes Research Initiative in the USA that training by this method is also expensive (52). The study concluded that training in the endoscopy suite lengthened a procedure by 10-37% and that in a model of 4000 endoscopic procedures per year, this would reduce department income by $500,000 to $1,000,000 causing time and financial strain on teaching institutions.

Training with simulations provides an ideal alternative to the current model of training. There are a few synthetic models available on which novices can acquire their basic skills (53). However these are quite low fidelity simulations and are not realistic enough in their anatomical representation and tissue feel. Although animal models are available for both lower gastrointestinal (GI) endoscopy and gastroscopy, there are a number of

problems associated with them. They require special facilities, are all for single use and there are ethical and hygiene issues associated with them (54). It is for these reasons that most endoscopy suites across the globe do not possess such models.

Virtual reality simulators have a potential to address this training need. There are now a few virtual reality simulators for upper (55) and lower gastrointestinal endoscopy (56) and bronchoscopy (57), which have been demonstrated to possess satisfactory face validity (a measure of how close a simulation is to real life) (56). They are quite real in their anatomical simulation, possess force feedback and some advanced simulators contain facilities for therapeutic interventions. Using these simulators, novices can acquire their basic skills in a stress free environment and practise extensively.

1.1.5 Transfer of skills from simulations to real procedures

The past decade has seen the development of a number of simulations for psychomotor skills training. The value of these training alternatives is presently assumed but this would only be demonstrated by studies that show a transfer of skills from simulations to real procedures. It is believed that the extent of the transfer from simulated models to real procedures is determined to a large extent by the realism of the simulation (4). The higher the fidelity of the simulation, the higher will be the likelihood of transfer of skills learnt away from the operating room (18).

However, in their study evaluating the extent of transfer from box-trainers to laparoscopic cholecystectomy, Scott et al found that (24) there was a significant improvement in the performance of laparoscopic cholecystectomy in a group of residents even after training on simple tasks in a box trainer. Anastakis et al similarly found that

there was an effective transfer of skills in open surgery from bench models to a human cadaveric model (58). Thus box-trainers and bench models are effective adjuncts to experience gained in the operating room especially with regards to the acquisition and practice of psychomotor skills. Virtual reality is still an emerging technology, as a result of which there have been few studies examining transfer to real procedures. Tuggy et al (59) have demonstrated the transfer of skills from the VR simulators to real sigmoidoscopy. They showed that those trained on VR simulators made fewer directional errors (inability to keep the lumen of the bowel in the centre) and spent lesser time without clear vision as compared to the controls who did not receive any VR training. Similarly Seymour et al have shown an effective transfer of skills to laparoscopic cholecystectomy using MIST-VR. They showed that trainees who had undergone VR simulator training were six times less likely to make skill based errors and were 29% faster than the non-VR trained group (60).

1.1.6 Task analysis and the development of a structured surgical curriculum

The role of task analyses is crucial to the development of a structured surgical curriculum. It has been said that ' task analysis provides a blue print of the things a student should master if he is to reach the objectives that have been set' (61). Task analysis is considered to be one of the most important steps in the development of simulation based training curricula by other high reliability organisations such as the military and aviation (62). In surgery it was previously believed that the skills of an expert surgeon were not easy to identify and that complex actions were difficult to break

down into components (63). However, task analysis in surgery, has been made possible by the ability of some surgeons to deconstruct surgical procedures to specific skills and tasks which underpin surgical proficiency. An international workshop of expert surgeons has attempted to define the key components of surgical performance (64). Surgical ability is the natural state or condition of being capable. Skill is a developed state of proficiency or dexterity. A task is a piece of work to be undertaken and a procedure is a sequence of steps. Following this a taxonomy of these components has been developed. For example, tracking, translation and aiming are considered to be abilities in laparoscopic surgery, suturing and knot tying are examples of skills, excision and anastomosis are tasks and laparoscopic cholecystectomy is a procedure (64).

Simulations have provided the surgical community with the option of making surgical training structured and curriculum based. Such a curriculum should define a set of goals and a level of proficiency that has to be achieved in a certain period of time. It has been suggested that training should begin on simple models in non-threatening environments before skills are transferred to the operating theatre (65). The ability to deconstruct surgery into components mentioned earlier such as abilities, tasks and procedures permits the development of a curriculum that can be tailored to the needs of a trainee and the level of training. For instance training courses for novice trainees would focus on the teaching of skills such as knot tying, suturing, and transfer. Those for intermediate level trainees would focus on tasks such as anastomosis, excision and basic procedures. While those for advanced trainees would focus on more complex procedures (64). Thus, surgical training becomes more structured. It begins in a non-OR environment with possibilities for feedback, repeated practice and the ability to make mistakes and learn

from them without any drastic consequences to patients. When basic psychomotor skills are adequately mastered in the skills laboratory, the trainees can then be allowed to begin their learning in the operating theatre (65).

In addition to the ability to acquire basic psychomotor skills in non-threatening skills laboratories, simulations have the added benefit to introducing feedback during training in a formal, more organised and focussed manner. Training solely in real operating theatres possesses the limitation that assessment of performance is subjective and global rather than being focussed and task specific.

1.2 ASSESSMENT:

1.2.1 Assessment in surgery:

It has been emphasized that surgical education should be based on a model of adult learning (18). Just as in adult learning, surgical education should be less prescriptive and trainees should be able to set their goals and objectives and chart their own progress (18). One important aspect of self- directed learning is the need for continuous evaluation and feedback. Feedback is essential because deficiencies in training and performance are difficult to correct without feedback (66). In addition, it is also necessary that this feedback be as objective as possible as surgical trainees already have a certain perception of their abilities which may be difficult to influence. In the apprenticeship model, trainers are also responsible for the assessment of the skills of a trainee. In this capacity they have the larger responsibility of ensuring the highest quality of surgical care in the future. However, the assessment by trainers is subjective. In a study of training in colonoscopy

Cass et al, demonstrated a wide discrepancy in the objective assessment of skills and the subjective assessment of trainers (67).

1.2.1.1 The RITA process in Calman training

At a congress of expert surgeons there was a consensus that all surgical trainees should be assessed at least on an annual basis (68). While ensuring that progress during training is based on objective criteria, objective assessment methods also serve as quality assurance in the future (68). The annual RITA (record of in-training assessment) consists of the subjective rating of trainers with regards to technical skills and other characteristics such as communication skills, enthusiasm for research and team skills. While the RITA process is a step in the right direction in order to make assessment regular, structured and criteria based, there is a clear need to make the assessment of technical skills more objective. The need for objective assessment of technical skills is made all the more crucial by the recently mooted concept of revalidation of senior surgeons and for resolution of conflicts between trainers and trainees during the RITA process.

1.2.1.2 The principles of assessment.

Any method of assessment should be **valid and reliable**. There are various subtypes of **validity** (18;69). *Construct validity*: the extent to which a test is measuring what it is supposed to measure. Or more simply, the ability of the test to discriminate between people of varying experience/ expertise. *Content validity* is the extent to which the

domain that is being measured, is measured by the assessment tool. Eg. while trying to assess technical skills we may actually be testing knowledge. *Concurrent validity* is the extent to which the results of the assessment tool correlate with the gold standard for that domain. *Face validity* is the extent to which the examination resembles real life situations. *Predictive validity* is the ability of the examination to predict future performance. **Reliability** is a measure of a test to generate similar results when applied at two different points. This is referred to as test-retest reliability (34). When assessments are performed by more than a single observer, there is another type of reliability test referred to as inter-rater reliability which measures the extent of agreement between two or more observers (18).

In the assessment of surgical technical skills, Watts and Feldman (70) have suggested that there are currently five methods available with varying degrees of reliability and validity (Table 1A). According to this table, it is clear that the first two methods of assessment of technical skills, which are methods presently used, have poor reliability and unsatisfactory validity. In addition, the difficulties associated with the objective assessment of technical skills have resulted in a situation where the assessment of surgical trainees consists solely of the assessment of their clinical skills and their knowledge.

The Surgical Advisory Committee in the UK has appreciated the need to increase the extent of emphasis placed on the assessment of technical skills. With that intention it has recommended the use of an assessment tool referred to as the Operative Competence assessment with consultant surgeons (faculty) assessing their trainees on procedures performed during a fixed training period (Table 1B). The assessment consists of the

trainee's ability to carry out procedures on a 5-point rating scale. This is a modification of an assessment method introduced by Savage and Darke (71).

Method of assessment	**Reliability**	**Validity**
Procedure lists with logs	Not applicable	Poor
Direct observation	Poor	Modest
Direct observation with criteria (eg. checklists)	High	High
Animal models with criteria	High	Proportional to realism
Videotapes	High	Proportional to realism

Table 1A: Methods of assessment of technical skills with their validity and reliability- Watts and Feldman

U	Unknown
A	Competent to perform the procedure unsupervised (can deal with complications)
B	Does not usually require supervision but may need help occasionally
C	Able to perform the procedure under supervision
D	Unable to perform the procedure under supervision.

Table 1B. 'Op Comp' assessment: For the assessment of technical skills in surgical trainees- Joint Committee for Higher Surgical Training (JSHST)- General Surgery

The 'Op Comp' assessment is a step undertaken by the Joint Committee for Higher Surgical Training (JCHST) with the aim of making technical skills assessment an important aspect of surgical training in the UK and to also make it more objective than the methods currently employed.

1.2.1.3 Current methods of assessment in surgery:

Assessment of knowledge:

The focus of examinations such as the Fellowship/ Membership of the Royal College of Surgeons (FRCS/ MRCS) is mainly on the knowledge and clinical skills of a trainee. However such examinations fail to assess the technical ability of a trainee. Studies have shown that there was no relationship between the American Board of Surgery in Training Exam (ABSITE) and clinical skills in an Objective Structured Clinical Examination (OSCE) examination (72) and technical skills (73). Thus examinations have poor content validity for the assessment of technical skills.

Time taken for surgery.

Prior to the advent of anaesthesia time was of essence (74). However, as Lister noted 'anaesthetics have abolished the need for operative speed and allow time for attention to detail' (75). However, as time is the easiest parameter to measure it is still used as a surrogate marker of expertise. It has been used for the assessment of technical skills in laparoscopic surgery on bench models (33). It is also used for determining the learning curves for surgical procedures (76). While operative speed is a desirable surgical trait in

order to lower the time spent under anaesthesia, it fails to assess the quality of surgical performance. While it could be used as a measure of skill on standardised bench models, it is an unreliable measure when used during real procedures, due to the influence of various other factors such as patient characteristics.

Operative Log Books:

All surgical trainees are expected to maintain a log of their operative experience, which has to be submitted at the time of examinations, interviews for placements and during the annual assessment. However, it has been found that they are merely indicative of procedural performance and not a reflection of operative ability and thus lack content validity (68;70).

Direct observation by trainers/ other surgeons.

Assessment by observation, as currently occurs in the operating room is subjective and not based on specific criteria. It is not only influenced by the subjectivity of the observer and would have thus possess poor test-retest reliability, but also suffers from poor inter-rater reliability as even experienced senior surgeons do not possess the same level of agreement while rating the skills of a trainee (18).

Outcome after surgery

Outcomes after surgery are sometimes used as surrogate markers of operative ability. The league tables published in the media are a crude form of outcome analysis. The incidence of post-operative morbidity can also be indicative of operative skill (25). In a study of

recurrence rates after rectal cancer surgery, multi-factorial analysis revealed that surgical expertise was one of the key determinants of outcome (77). It has also been shown that surgeons who perform high volumes of procedures in colo-rectal cancer (78;79) and upper gastrointestinal surgery (80) produce better outcomes than low volume surgeons . However, it is yet to be demonstrated whether the superior outcomes are purely due to operative expertise or due to a combination of other factors that influence outcomes such as the availability of specialised teams. Some surgeons have also expressed caution in the acceptance of league table results on morbidity and mortality without including other factors, especially patient factors in the analysis (26).

1.2.1.4 Recently developed objective assessment methods of technical skills

Use of checklists and global scores

In 1970 Kopta suggested the use of check-lists for the assessment of technical skills amongst orthopaedic trainees (66). Since then check lists have been used across a number of specialities including laparoscopic surgery (81). The availability of set criteria against which technical skills can be assessed makes the assessment process more objective, valid and reliable (70). It has been said that 'check lists turn examiners into observers of behaviour rather than interpreters of behaviour, thereby removing the subjectivity of the evaluation process' (82). Other groups have used similar objective assessment methods. Winkle et al reported the reliability and validity of global rating forms and operation specific check lists in the evaluation of operative skills amongst residents in the operating theatre (83).

OSATS (Objective Structured Assessment of Technical skills)

The wide acceptance of the Objective Structured Clinical Examination (OSCE) led to the development of a similar concept for the assessment of technical skills by a group in Toronto called OSATS (84). The OSATS assessment consists of six stations where residents/ trainees perform procedures on live animal or bench models in fixed time periods (85). The assessment of performance during the performance of tasks is done using an operation or task specific check lists (Appendix 1) and a global rating scale (Appendix 2) as initially developed by the group (83). The checklists consist on 20- 40 items considered essential for the competent performance of the task at the relevant station. The global scale consists of seven components of operative skills which are marked on a 5-point Likert scale. The middle and the extremes are anchored by clear descriptors to help in the criterion- referenced assessment of performance as suggested by Watts and Feldman. A trainee's score is the average of the score on the seven dimensions interpreted as a percentage. These seven components are generic to any operation/ procedure.

Both live animal operating and bench models have been used for the OSATS assessment. There were no differences in the performance of trainees in both formats of the examination (85). Datta et al demonstrated a strong correlation between performance on a bench model and a real procedure with no difference in the global rating scores for both platforms of assessment (86). Their study consisted of a group of trainees performing a simulated sapheno-femoral junction high-tie on a synthetic model and then performing the same part of a real varicose veins procedure. Both these procedures were videotaped and played back to two observers.

The only drawback to the performance of the OSATS assessment are the resources and time involved in getting a number of staff surgeons to observe the performance of trainees. It has been suggested that retrospective video watching of performance may be a way forward. It does not involve the presence of multiple faculty raters at the same time and also makes the assessment blinded. Using this method of assessment on bench models, Datta et al demonstrated the construct validity of the global rating scales (87). It is apparent that the use of both checklists and global rating scales involve the presence of multiple faculty raters or hours of video watching. With increasing time constraints especially in the UK, it will be difficult for faculty members to find time to carry out such assessments. Thus, there is a need to measure technical skills by systems which will generate real time data in terms of skills and dexterity. The acknowledgement of the limitations of video-based assessments gave rise to the development of dexterity analysis systems.

Dexterity analysis systems

A technically skilled surgeon is one who executes a task quickly, is economical in movements as well as being precise. While methods such as global rating scales and checklists evaluate the technical ability of the trainee in executing certain aspects of a task/ procedure, they fail to quantify the skill of the surgeon in terms of dexterity parameters such as accuracy and economy. Motion analysis systems have been used for evaluating the efficacy of therapeutic interventions in patients with Parkinson's disease. As these systems can measure hand movements they could also be used for the assessment of surgical dexterity.

ICSAD (Imperial College Surgical Assessment Device)

This system consists of a commercially available electromagnetic tracking system (Isotrak II, Polhemus Inc, USA), which is connected to a portable computer. The Isotrak II system (Fig 1D) consists of an electromagnetic field generator and two sensors which are attached to the dorsum of the surgeon's hands at a standardised position. Special software is used for data retrieval and analysis. This consists of a software called 'Track 3D' to collect positional data and a software called 'Track' for converting this data to dexterity measures.

In addition to the number of movements made by each hand' the system can also calculate the time taken for the task, the distance travelled by each hand and the average speed. Studies have demonstrated the construct validity of ICSAD with respect of a range of laparoscopic tasks. There is a strong correlation between dexterity and previous laparoscopic experience on a simple task in a box trainer (88), and on more complex tasks such as laparoscopic cholecystectomy on a porcine model (89). Experienced and skilled laparoscopic surgeons are more economical in terms of the number of movements and more accurate in terms of target localisation and thus use significantly lower path lengths, which is a measure of accuracy. Datta et al have demonstrated the construct validity of ICSAD for open bench model tasks (90) and have also shown a strong correlation between ICSAD and global rating scales (87).





Fig 1D: Imperial College Surgical Assessment Device (ICSAD). Electromagnetic motion tracking system for the assessment of surgical dexterity. **A**-Field generator that generates the electromagnetic field; **B**-sensors that are worn on the dorsum of the surgeon's hands

Other motion analysis systems

Movement tracking can be based on electromagnetic, mechanical or optical systems. The Advanced Dundee Endoscopic Psychomotor Trainer (ADEPT), another motion tracking system was originally designed as a tool for the selection of trainees for endoscopic surgery. Studies have demonstrated the validity and reliability of ADEPT (91;92). Optical motion tracking systems depend on infrared sensors which are attached to the surgeon's arms (93;94). These systems consist of infrared cameras surrounded by infrared light emitting diodes. Markers, which are placed on the limb of a surgeon, reflect the infrared light. Software is used to extrapolate the positional data of the markers to data on movement analysis. The disadvantages of such optical systems are that they suffer from line of vision disturbances with consequent loss of data. To overcome this shortcoming several cameras and markers are required, making these systems bulky and expensive. Signal overlap also prevents the use of markers on both limbs making these systems more restrictive.

Virtual Reality simulators

VR simulators are computer-based systems which generate output data in the form of performance measures such as the number of movements made in the execution of a task. At an international workshop held two years ago a group of experts reviewed all methods of assessment and suggested, based on the data available, what should be considered as output-measures for technical skills assessment (64).

Included were parameters such as economy of movement, path length, time to completion and execution errors. Currently there are VR systems for laparoscopic surgery and

endoscopy that can generate these measures, thereby providing an objective assessment of technical skills in a wide range of subspecialties. The MIST-VR system, which has been extensively validated in several centres (34;95;96), is perhaps the only commercially available validated system that can presently be used for basic laparoscopic skills assessments. This system generates results in terms of time to completion, economy of movement, path length and errors. Upon task completion the data is readily accessed, thus providing users with instantaneous and objective feedback of performance. By storing the data in databases, further improvements can be continuously monitored and progress can be assessed. This data can be then used for educational, training and research purposes.

Error scoring methods

One way of measuring technical ability may be to examine performance for the number of errors executed during the procedure. One study assessed performance while trainees performed a small bowel anastomosis on a porcine model and generated a cumulative error score (97). Eubanks et al developed a minor and major error score of the technical errors executed during a laparoscopic cholecystectomy in addition to a structured checklist. Such errors were weighted according to their severity (81). Using a categorisation of errors based on the External Error Modes framework, Joice et al identified errors executed during the performance of laparoscopic cholecystectomy. This categorisation of errors includes procedural errors which are executed either because of the omission or the re-arrangement of steps during the procedure or execution errors which are due to the failure of the surgeon to execute a step correctly (98). As such error

based assessment methods are a modification of checklists, they add objectivity to the assessment process.

Outcome analysis on bench model tasks.

As outcomes after surgery are difficult to ascribe solely to surgical technique and as adverse outcomes from poor technique may not be apparent for many years such as recurrence after cancer resections, some researchers have suggested the idea of using outcome measures after performance on bench models. Szalay et al assessed the 'quality of the final product' after performance on six different bench models (99). This outcome analysis consisted of a four item global rating scale on a 5-point Likert scale which assessed completeness, aesthetics, function and overall quality of performance. This was performed by senior surgeons who were blinded to the skill of the trainee. The investigators found that the method possessed construct validity. Datta et al assessed the leak rates after performance of a vascular anastomosis on a bench model and found that there was significant correlation between the leak rate and surgical dexterity (100). Hanna et al studied that quality of knots performed laparoscopically using a tensiometer and derived a knot quality score as an index of knot reliability (101).

1.2.2 Assessment in endoscopy:

In addition to understanding the procedure and the equipment, the ability to take consent, recognise pathologies accurately and understand the principles of sedation, procedural proficiency underpins competence in lower GI endoscopy (67). The problems associated

with training and assessment in endoscopy are greater than even those associated with surgery. While surgical tasks can be trained and assessed on suitable simulations, the lack of availability of realistic synthetic or animal based simulations in endoscopy makes assessment even more difficult. In addition technical proficiency in lower GI endoscopy has been difficult to define. As time taken for the procedure is unreliable due to patient variability, one measure of procedural proficiency used commonly has been the caecal intubation rate. However, even using this as a criteria there has been very little consensus on the number of procedures required to achieve a caecal intubation rate of at least 90% (67).

In addition, one study by Cass et al showed that greater than previously recommended numbers are required to achieve caecal intubation levels of 90% (102). Thus the number of procedures performed seems to be an unreliable indicator of skill.

Outcome measures such as complications rates may be better determinants of expertise (103) but as complication rates are extremely low (104), a large number of procedures need to be performed, in order to acquire reliable data. Therefore, such data would only be useful for the re-credentialing of experienced endoscopists and not for formative assessment during training.

For the latter, objective and unbiased assessment of procedural activity is essential (102;105), but currently there are no accepted methods. In one of the most comprehensive studies on endoscopy training to date, Cass and colleagues demonstrated a wide discrepancy between the subjective assessment of proctors and the objective assessment of technical skills using four objective measures: traversing the splenic flexure, intubating caecum, recognising abnormalities and correctly identifying them

(106). This emphasizes the need to make the assessment of procedural skills more objective.

In addition to their training potential, VR simulators also provide an objective record of performance. The computer program automatically provides a comprehensive log of each procedure, including the time taken, the amount of bowel wall visualised and the volume of air insufflated, together with the incidence of serious complications such as bowel perforation. This information can be presented immediately after the procedure, and can also be used to track each learner's trajectory. Such metrics of performance go a long way in ensuring that competence assessment in endoscopy is objective. This is particularly important, as so far there are no objective, validated measures of competence in endoscopy.

1.3 TECHNICAL PROFICIENCY AND SURGICAL COMPETENCE:

The importance of non-technical skills

While technical proficiency is an important aspect of surgery, technical ability is only one of the skills required to perform an operation successfully (107). In addition to which teamwork, communication, judgement and leadership underpin surgical competence. Assessment should therefore not be based solely on technical skills as this is only one aspect of surgical competence (4). While technical skills training ensures task proficiency, it cannot address deficiencies in performance resulting from lapses in communication and decision-making. In surgery just as in other specialities of medicine, very little attention is paid to training in these skills.

Satava et al have attempted to develop a taxonomy of surgical competence. According to them, surgical competence has six components. There are knowledge, patient care, interpersonal and communication skills, professionalism, practise-based learning and improvement, and system-based practise. These six components have been developed by the Accreditation Council for Graduate Medical Education (ACGME) and its Residency Review Committees (RRC) in surgery and the American Board of Medical Specialities (ABMS) (108). Surgical proficiency on the other hand is a measure of performance in each individual component. So, technical proficiency is only one component of surgical competence.

In addition to the need to assess other facets of surgical competence, research in the assessment of technical skills has demonstrated the existence of a ceiling effect (109) when tasks are performed in a laboratory environment. This could mean two things. Either that the tools that are currently available are limited in their sensitivity to detect differences in skill or it could also mean that most surgeons acquire the appropriate extent of technical skills by a certain stage of their training. The progression from that stage to expert performance is then dependant on decision-making, knowledge, experience and other cognitive traits.

Training in all the specialities of medicine in addition to surgery, places a large emphasis on technical and clinical skills without much emphasis on the training of other skills that are crucial. This is synonymous to the position the aviation industry was in about 20 years ago. The focus was on technological developments to the aircraft and on technical training of pilots till research sponsored by NASA (National Aeronautics and Space Administration, USA) revealed that 70% of errors were due human causes. NASA's

research revealed that a large number of errors were a result of failures in inter-personal communication, decision making and leadership (110). The realisation of the importance of these skills led to the development of new initiatives in training in aviation with an equal emphasis on technical skills and other non-technical skills with the aim of reducing errors caused by communication and team coordination.

The similarities that exist between aviation and the medical specialities has already prompted certain specialities such as anaesthesia and intensive care to adopt some of aviation's strategies towards training of their personnel and towards the understanding of errors. The successful performance of an operation in the operating theatre and the flying of an aircraft involve complex inter-personnel/ inter-professional interactions. Teamwork is essential (111) and communication is an essential aspect of effective teamwork. Reason sees the similarity between aviation and medical specialities such as intensive care and anaesthesia existing at two levels (112).

<u>At the ' sharp end'.</u>

This is the similarity that exists between the medical professional and the pilot at the level of the human- machine/ human-patient interface. The similarities include dynamic environments with multiple sources of information which have to be rapidly processed, actions often have immediate consequences, there is considerable human- technology interaction and there are often multiple players with different goals.

At the 'latent' level.

These are the organisational and cultural characteristics that indirectly influence individual and team performance. These consist of the attitudes between different individuals and behavioural norms established in these work settings.

Confidential surveys of flight personnel (113) revealed that there were clearly established hierarchal structures where juniors were not encouraged to make suggestions and that pilots seldom solicited help from them. These surveys also revealed that pilots considered that their 'decision-making was as good in emergencies as under normal conditions, that they can leave behind their personal problems, and that they perform effectively when fatigued' (113). A survey of personnel in the operating theatre revealed similar findings (114). This survey found that 70% of senior surgeons denied the effect of stress and fatigue on their performance and decision-making and that senior surgeons were more likely to support steep hierarchies where juniors do not question the actions and decisions of seniors.

1.4 HUMAN FACTORS IN HIGH- RELIABILITY ORGANISATIONS AND MEDICINE.

The understanding of the way people work and developing methods in which people's performance can be improved in the essence of a discipline called Human Factors. Human factors is a discipline that spans engineering, cognitive psychology and ergonomics and emerged specifically in response to the safety concerns of other high risk industries (115). The discipline found its origins during the Second World War when

psychologists realized there was a gap between technology and the resources, including humans, required for using them. For some time human factors research focussed only on understanding and improving the man-machine interface. It has however been suggested that it is important to focus on the original spirit of human factors as defined by Chapanis in 1949 (115). Chapanis stated that 'a considerable amount of improvement can be made by altering many characteristics of the environment in which men work'.

There has been an increasing interest in the application of human factors in the understanding and prevention of medical disasters. However, the medical community has lagged behind other high-reliability organisations by many years. The reason for this is that disasters in aviation, shipping and nuclear plants cost many lives, create widespread environmental problems (116), cost these industries a considerable sum of money in lost equipment and litigation and get widely reported in the press. In comparison, medical disasters usually affect a single individual and seldom receive any publicity. However, it is only over the past 20 years that research into medical adverse events has revealed the scale of the problem. A study from Harvard (116) found that patients were unintentionally harmed by treatment in almost 4% of admissions in New York state. For 70% of patients the resulting disability was slight or temporary, but in 7% it was permanent and 14% of patients died partly as a result of their treatment. Similar findings were reported from Colorado and Utah (117). In the United Kingdom, one study carried out a review of 1014 records and found that of the 10.8% adverse events rate, half were preventable (117).

The anaesthetists have been in the forefront in the application of a human factors approach to errors. A study of peri-operative events leading to adverse incidents in

patients undergoing surgery found that 11% of patients had an inadequate pre-operative assessment or preparation. Of these, nearly 27% of the patients had an adverse incident related to the inadequacy of the pre-operative assessment (118). Another study found that nearly a third of deaths attributable to anaesthesia were related to inadequate pre-operative work-up (119). In a study evaluating the occurrence of anaesthesia related errors using a voluntary reporting system, Chopra et al found that nearly 80% involved occurrences which were considered preventable. They also found that human error was responsible for 75% of the reported errors. Lack of vigilance and failure to check were the most frequently reported factors associated with error and 45% of all reported observations were made during maintenance of anaesthesia (120). In addition, nearly 25% of anaesthetists report committing an error with drastic consequences (121).

Other human factors groups have made a significant contribution to the understanding of errors and near- misses in the operating theatre. The seminal work of de Leval and Carthey (122), which consisted of performing a multicentre observational study of events during a complex paediatric cardiac procedure, identified the incidence of major and minor events during the procedure and their subsequent impact on patient outcomes. Their research revealed that the number of such events had a strong relationship with outcome and that the inadequate compensation of these events was further predictive of death after surgery. In addition to examining the procedures for errors and near- misses, they also used a scale called the STAR (Surgical Team Assessment Record) questionnaire to assess organizational, environmental and team performance during these procedures and the effect of these factors on procedural outcome. They, however, found no correlation between these variables and outcome and believed that this was probably

because of a limitation in their study which consisted of getting the questionnaires answered by the surgical team after the procedure, introducing an element of hindsight bias. While acknowledging this they also emphasized the need to continue research in this field to explore the role of these factors in the incidence of errors in the OR.

1.5 HUMAN FACTORS IN THE OPERATING ROOM AND SURGERY

It is becoming increasingly clear that there is a need to apply a systems approach to the occurrence of errors in patients undergoing surgery just as in other high reliability organisations. Such an approach includes the need to understand the physical, cultural and social structure of the OR.

1.5.1 The OR team and inter-personnel dynamics

The operating room is a highly complex working environment which consists of different groups of individuals involved in a coordinated effort to perform complex procedures which involve a high level of skill, attention and knowledge. There are essentially two teams, each of which could be divided further into two teams (Fig 1E).

The surgical team consists of the operating surgeon and either one or two assistants depending on the complexity of the procedure. There is a surgical nursing team which consists of a scrub nurse who is scrubbed up to assist with the procedure and a 'runner' who is a nurse who acts as an 'unscrubbed' assistant to the main scrub nurse. Her role is to fetch instruments and suture material that are needed and missing and to adjust equipment like the temperature in theatre and the operating light. The anaesthetic team

consists of an anaesthetist and an operating department assistant (ODA). The ODA's role is to assist the anaesthetist throughout the procedure in the maintenance of anaesthesia. In the US, the anaesthetist is an individual who administers anaesthesia and may or may not be a qualified physician. A physician who administers anaesthesia is referred to as an anaesthesiologist. Throughout this thesis, using the UK's model, the anaesthesia physician would be referred to as the anaesthetist while his/ her assistant will be referred to as the ODA.

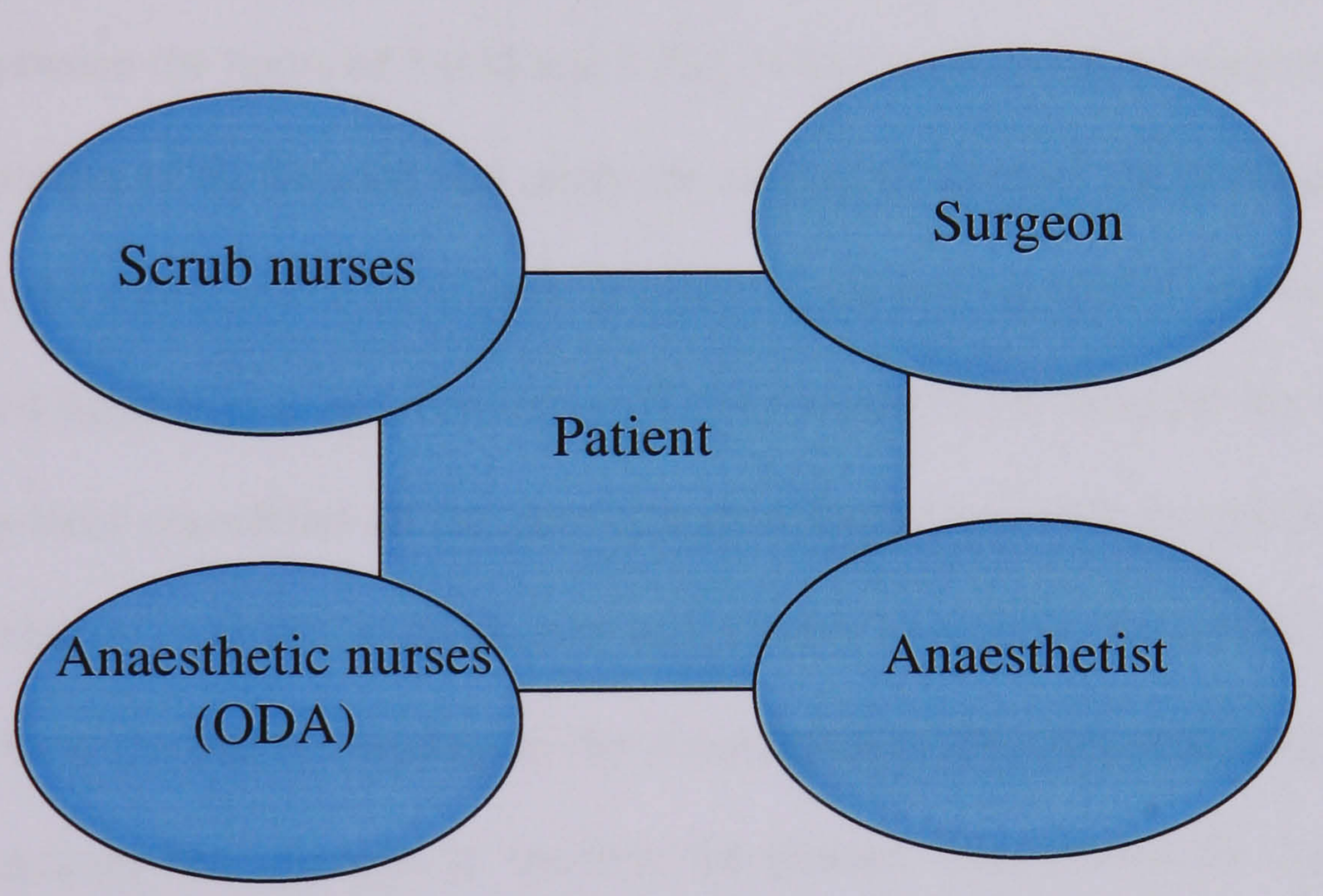


Fig 1E: Team profile in the OR

Helmreich actually sees the two teams as four components of a single team working with the common goal of maintaining the patient's well-being throughout the procedure (111). The performance of individuals in this complex inter-professional work setting depends on an individual's skills and knowledge as well as their inter-personnel dynamics.

The dynamics between the different members of the operating theatre has the potential to be strained as these groups have different agendas within their common aim of ensuring the safety of the patient. These conflicts can find their origins from the period that involves pre-operative planning and scheduling (111) to the period spent in the operating theatre. For example, surgeons would like the patients to be under anaesthesia and ready for surgery when they enter, while the anaesthetists and the nursing team would prefer the presence of the surgeon before the patient is induced. Another potential period of conflict is the time towards the end of a procedure. In the UK routine operating theatres are run only between the hours of 8 AM and 5 PM. With that in mind the surgeons would like the next patient to be induced and ready for surgery as soon as the previous case is completed. As the anaesthetists have to be present for the post-operative recovery of the patient and for a hand-over to the recovery staff, they would like to send for the next case only after they have completed all the post-operative formalities with the previous case. Often under pressure from the surgeons they might agree to send for the patient from the ward only to have the patient waiting in the anaesthesia room often unaccompanied or will have to arrange for someone to monitor the patient. This results in straining an already poorly staffed operating theatre. Often the potential delay between cases, from the surgeon's perspective, leads to the postponement of cases to the surgeon's annoyance and patient's inconvenience. Training of a junior member of staff leads to strained dynamics with the other team members as this may lead to the surgery being prolonged and the subsequent postponement of cases, thus reducing the productivity and efficiency of the OR for that day.

1.5.2 Leadership in the OR

Unlike aviation and the military where there is a formal command structure, the OR is a unique example of dual and alternating leadership. The anaesthetists and the surgeons are both equally responsible for the patients. They both perform widely different roles with some degree of overlap (123). One model of OR's leadership structure looking at the anaesthetist as the provider and the surgeon as the customer leads (124) to the perception that the surgeon's needs are paramount as a result of which the surgeon is the leader (125) . The fact that the patient is in the OR because of the surgeon's decision to operate that patient (111) gives rise to the natural impression that the surgeon is the leader. However, the importance of the anaesthetist for the maintenance of the patient's vital functions during the operation can result in anaesthetists assuming that they are the leaders. This highlights some of the strain that exists between surgeons and anaesthetists in the operating theatre. It is probably more productive to see their roles as one of dual leadership or alternating leadership where individuals assume leadership depending on the main task at hand. For example when the patient is safely maintained under anaesthesia and the surgeon is carrying out a complex procedure the surgeon is the leader. On the other hand, when the patient is experiencing certain anaesthesia related problems, the anaesthetist assumes leadership in the OR. In fact the leadership could even fluctuate to the scrub nurse, especially towards the end of the procedure when the swab, needle and instrument count needs to be performed. Such a model of dual or alternating leadership with smooth transition from one individual to the other could only result from team coordination.

1.5.3 Input- Process- Output model: An Integrated model of OR performance

The complexity of the roles of the different individuals in the OR and the resulting dynamics between them led Helmreich to develop a model for operating room performance based on an input, process, output concept (111) (Fig 1F).

Input factors such as 'organizational, environmental and individual factors set the stage for group processes that in turn lead to multiple outcomes' (111). From this model is can be seen that positive input factors would underpin effective group processes which would lead to positive individual and team outcomes in addition to good patient outcomes. These positive outcomes would then feedback and enhance both the input factors and the group processes.

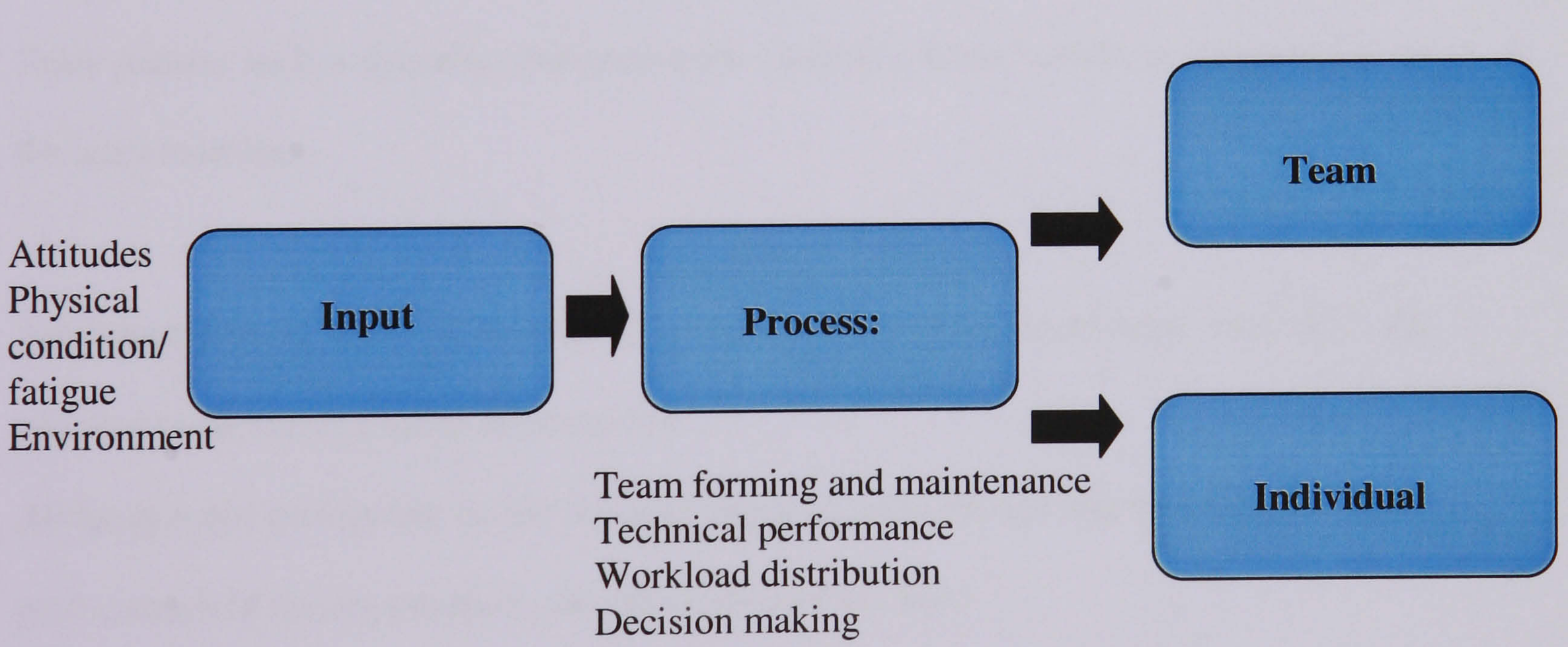


Fig 1F: Input-Process-Output model of team working in the operating theatre

Input Factors:

These can be classified as organizational, environmental, team and individual factors.

Organizational factors:

- OR efficiency, timing of procedures, work schedules,
- Financial incentives to improve, perform efficiently
- Organisational norms regarding quality control. In medicine quality control can be achieved and improved by carrying out regular audits.
- Conflict between different roles and departments as mentioned earlier

*Environmental factors*_such as noise, music, equipment in the OR and temperature.

Team factors such a dynamics between team members, team morale, and experience of the team members

Individual factors such as individual attitudes, motivation, knowledge and skill and physical condition (eg. sleep deprivation).

All these input factors can be the focus of organizational change and training to enhance performance of the organisation, the individual and the team.

1.5.4 Individual attitudes in the OR

Research in aviation has shown that the attitudes of team members regarding leadership, hierarchy, error, communication and personal abilities are highly predictive of team performance (126). People with differing attitudes regarding team structure and dynamics may find it difficult to work together as a cohesive unit. Preliminary research in attitudes in the OR has revealed some interesting findings. Modelling a questionnaire along the lines of the Flight Management Attitudes Questionnaire (FMAQ) and the Cockpit Management Attitudes Questionnaire (CMAQ), Sexton et al developed an Operating Room Management Attitudes Questionnaire (ORMAQ) to study the attitudes of OR teams towards communication, hierarchy, personal capabilities and team coordination (127). Both FMAQ and CMAQ have been found to be predictive of crew performance in aviation (128) (127).

Some of the conclusions of ORMAQ are summarised below:

- Senior surgeons were more accepting than pilots and other OR personnel of the statement that 'My decision making is as good in medical emergencies as in routine situations'.
- The anaesthesia teams were more accepting of the idea that a pre-operative team briefing would improve team performance as compared to the surgical teams.
- Surgeons were more accepting of the fact that juniors should not question the decisions made by seniors.

- 62% of surgical staff (surgeons and residents) reported high levels of teamwork with anaesthetic staff while only 41% of anaesthetists reported high levels of teamwork with surgeons.
- Surgeons perceive teamwork and communication to be of a higher quality than other team members.

Some of the attitudes of the operating room staff reflected the attitudes of the cockpit team members in the past (113). This was before the aviation industry instituted certain training initiatives to change attitudes and thus enhance the safety of flights. Just as the ORMAQ was used to survey the operating team, the CMAQ and the FMAQ were used to study the attitudes of the cockpit team at the same time. While 60% of OR team denied the effects of fatigue on performance, only 26% of flight crews did with no difference between captains, first officers and second officers (127). 6% of flight crews rejected steep hierarchies while 55% of senior surgeons supported steep hierarchies. Similarly, observational studies of teamwork in aviation and the operating room have found that teamwork in the cockpit was minimal or poor in only 10% of cockpit observations as compared to 50% of OR observations (129).

Research over the past 20 years has revealed that here have been considerably changes to attitudes among cockpit crews. This has largely been due to the results of the NASA research, which stimulated the aviation industry to initiate certain training initiatives. CMAQs have been used to evaluate changes in safety related attitudes before and after training (127;130). For example pilots have begun to acknowledge their limitations under stress and fatigue and communication and teamwork in the cockpit has improved. Crucial

to the development of a safety culture in aviation is the training program referred to as Cockpit or Crew Resource Management (CRM) training.

1.6 ENHANCING SAFETY IN AVIATION: CRM TRAINING.

CRM has been defined as ' using all available sources- information, equipment and people- to achieve safe and efficient flight operations (131). The essence of CRM training is to enhance the performance of aviation crews and help them understand the elements of human effectiveness (126). CRM training aims to help crews understand the limitations of human performance and to develop safety culture attitudes.

1.6.1 The evolution of CRM training

The first CRM training program was initiated in 1981 by United Airlines (110). Developed with the aid of management consultants who had a background in management courses amongst corporations, it essentially consisted of seminars and classroom based lectures. Gradually with time, CRM training began to include simulator training where crewmembers were given an opportunity to practise their CRM skills. The early generations of CRM training were considered by many crews to be 'charm schools' or exercises in psycho-babble (110).

In 1990, the Federal Aviation Administration introduced the Advanced Qualification Program (AQP) (110). The AQP allowed airlines flexibility in designing their training programmes with the proviso that technical training in flight simulators, referred to as

Line Oriented Flight Training (LOFT), be integrated with human factors or CRM training. Such comprehensive training in flight simulators is now mandatory for the certification of every pilot. The LOFT sessions are usually video-recorded for the purposes of post-simulation feedback.

1.6.2 The fifth and current generation of CRM training

While the obvious drive for CRM training was to reduce the occurrence of errors during flights, this somehow became secondary with the development of CRM training programs. With the fifth and current generation of CRM (132), the aviation industry has tried to re-focus on error management strategies for flight crews. Helmreich et al developed the concept of the 'error troika' (110). This consists of three countermeasures against errors. The first is error avoidance; the second is error trapping where errors are trapped before they manifest and finally error mitigation in which the drastic consequences of errors are prevented by effective error recovery. Error management focuses on getting crews to recognise the limits of their abilities and acknowledge the effect of stress and fatigue on the performance.

1.6.3 Measurement of non-technical skills

Equal importance is given to the assessment of technical abilities as well as interpersonal (or behavioural/ non-technical skills) of a pilot during LOFT. CRM skills assessment, initially developed by the University of Texas Human Factors project, was developed as a

research tool to evaluate the effectiveness of CRM training. However, with time, the use of CRM skills markers grew, with the increasing interest of airlines to demonstrate their sincerity in enhancing performance. A CRM manual served as an aid to CRM evaluators (usually training captains). CRM skills were evaluated by using exemplars (pointers) for various markers, which then served as a guide for the training captain who was assessing a pilot/ crew's performance. CRM skills assessment was modified by different airlines under the original proviso of CRM training which stated the importance of modifying CRM training to the organizational culture of different airlines.

Over time the University of Texas project collected data on all aspects of an airline's procedure in a method known as the Line Operations Safety Audit (LOSA) during different phases of real flights (114). The LOSA assessment consisted of a checklist (LOSA checklist) into which were incorporated all the behavioural markers of the CRM skills assessment. LOSA gradually evolved to become a threat and error management assessment rather than just an assessment of behavioural skills.

In 1996, the JAA (Joint Aviation Authority) Research Committee on Human Factors initiated a project sponsored by Germany, France, Netherlands and UK, to develop a behavioural skills (non-technical skills) assessment method. The European Joint Aviation Requirements (JAR) Ops NPA 16 states that '--- the purpose of such an assessment (CRM skills) is to: provide a feedback to the individual and serve to identify retraining; and be used to improve the CRM training system' (133). The NOTECHS rating method, which was developed by the project, consisted of four categories, two of which assess social skills (Cooperation; Leadership and Management) and two of which assess cognitive skills (Situation awareness and Decision making). For each of these skills,

certain elements were identified and for each element, positive and negative exemplars were included in the assessment rating scale to aid the assessors (Appendix 3).

1.6.4 Studying the positive effects of CRM training

The most convincing demonstration of the effect of CRM training would be to demonstrate a reduction in the incidence of airline crashes and 'near misses'. However, this has been difficult. The reasons for this are many- airline disasters are rare and the reporting of 'near misses' depends to a large extent on the crew's initiative (132). In addition CRM training is not effective with all airline crew members (114) and there are wide national and professional cultural differences between different airlines (132). However, using CMAQs before and after CRM training, studies have demonstrated the benefit of CRM training on the attitudes and performance of cockpit crews (130). In addition the development of Line Oriented Safety Audits (LOSA), where trained observers observe certain segments of a flight (132) for attitudes, errors and near misses, produced results which demonstrated the effectiveness of CRM training (126). As a result of CRM training, the percentage of crews rated as 'above average' increased while those rated as 'below average' decreased. In addition informal survey of flight personnel have shown that crews consider CRM training to useful and effective (126).

1.7 CRM IN MEDICINE

The similarities between aviation and medicine have encouraged teams working in the operating room (134), accident and emergency departments (ER in the US) (135), maternity suites (136) and intensive care units to incorporate CRM-like training initiatives. Anaesthesia Crisis Resource Management (ACRM) is an established mode of anaesthesia training with a number of centres in the US and the UK offering ACRM courses.

1.7.1 ACRM:

The fact that anaesthetists work in a complex, dynamic environment with a high level of interaction with technology and an environment which can rapidly change from an uneventful procedure to a crisis situation- makes the practise of anaesthesia ideal for integrating the principles of LOFT and CRM training with training in anaesthesia. The pioneering work of Gaba and colleagues from Stanford University led to the development of ACRM courses which have now been transferred to other centres across the world (137). ACRM courses are intense and comprehensive and include a didactic session of lecture and video presentations and a simulator session which is followed by a debriefing session (137). Each ACRM course consists of 4-6 participants. While one participant is in the 'hot seat', others serve as observers and one may even be called by the participant in the hot seat for help.

Didactic session

This consists of lecture presentations on the principles of error occurrence in aviation and other high reliability organisations and on crisis management. The focus during these sessions is on human performance determinants during crisis situations such as decision-making, team working and communication. The participants are shown a video of a recreated cockpit incident which results in a crash. This is followed by a video- recording of an actual anaesthesia mishap. This is followed by an open discussion on the importance of the different principles of crisis management.

Simulator session

This session takes place in a simulated operating theatre. This simulated environment is equipped as a real working operating theatre as far as possible, with monitoring equipment, surgical trays and instruments, diathermy and suction machines etc. The central feature is the anaesthetic simulator.

There is usually one participant in the 'hot seat' while the other participants are permitted to watch the simulation proceeding via a live video-link. Each participant is exposed to at least 4 different scenarios. There are at least 80 different types of scenarios that trainees can undertake depending on the type of simulator (137). These vary from airway problems to other scenarios such as malignant hyperthermia, tension pneumothorax and cardiac arrest. Occasionally, some scenarios contain two simultaneous problems where one more serious problem is masked by one seemingly minor problem.

The fidelity of the anaesthesia simulators vary with some higher fidelity simulators which are capable of responding to pharmacological actions with appropriate physiological responses.

Debriefing session

Following the simulation session, all the participants of a group undergo a debriefing session which is video- based. Discussions focus on missed cues, fixation errors, wrong interpretation of data and other elements of crisis management including leadership, team working and communication.

1.7.2 Assessment of performance during ACRM and measurement tools:

Assessment of technical skills during ACRM

In addition to training, simulations can also be used for the assessment of performance of anaesthetists (138). However, there has been considerable variation in the use of measurement tools during simulations. These measurements have varied from quantitative measures such as 'time to detect' and ' time to solve' to assessment of protocol based actions (69). Deviations from fixed protocols being classified as 'major' or 'minor'. Some simulator studies have used scoring systems developed from accepted guidelines for conditions such as malignant hyperthermia. Numerical weightings are given to each action during the simulation depending on whether those actions are of major or minor importance. Using such scoring systems, a total score is then calculated

(139) (138). Some simulator studies have used mortality of the 'patient' as an assessment tool (140).

Devitt et al developed a scoring system for the assessment of technical performance during simulations. Each problem presented to the participants was scored from 0-2. If there was no response to the problem, they scored 0. Compensating interventions scored 1 and corrective interventions scored 2 (141).

It has also been perceived that in a majority of simulation studies the focus has largely been on the assessment of the technical ability of the anaesthetists to deal with a scenario. However, Gaba et al, following the principles of CRM training, developed a method for the assessment of non-technical skills (138).

Assessment of non-technical skills during ACRM

Gaba et al emphasized the need to assess non-technical skill during simulations. According to them 'technical performance concerns the adequacy of the actions taken from a medical and technical perspective', while behavioural performance 'concerns the decision making and team interaction processes used during the team's management of a situation' (138). The behavioural rating tool was adapted from the LOSA checklist and consisted of 10 management behaviours: orientation to case, inquiry/ assertion, communication, feedback, leadership, group climate, anticipation/ planning, workload distribution, vigilance and re-evaluation. All behavioural ratings were on a 5-point scale: 1-poor performance, 2-substandard (minimally acceptable) performance; 3- standard performance; 4-good performance and 5- outstanding performance.

Fletcher et al reviewed the literature on non-technical skills or human factors skills assessment in anaesthesia and emphasized the need for skills such as communication, team working, planning, resource utilisation and decision making as essential for effective performance (142). Acknowledging the differences between aviation and anaesthesia and the difference in the organizational cultures between the US and the UK, they emphasized the need to develop a non-technical skills assessment for anaesthetists and used a cognitive task analysis approach to identify those factors anaesthetists considered to be crucial for effective management. This led to the development of a non-technical skills assessment tool for anaesthetists- Anaesthetists Non-technical Skills (ANTS) (143). This consisted of 4 categories: task management, team working, situation awareness and decision making, which were then anchored by specific elements (Appendix 4).

1.7.3 Effectiveness of ACRM

While there is no clear evidence of the reduction in anaesthesia related errors after ACRM training, using post-simulation questionnaires, Holzman et al demonstrated that while senior anaesthetist expressed some reservations about the future of simulations, nearly 95% of the trainees felt that they had learnt a lot. However, nearly 80% of senior anaesthetists and 95% of trainees felt that practise on the ACRM course could help them work more safely. Nearly half the group felt that the course should be taken every 12 months. While the debriefing session was considered to be productive by over 80% of the participants, only 30% felt the pre-simulation didactic session to be of any use (137).

1.7.4 Validity and reliability of anaesthesia simulations

There has been considerable controversy regarding the use of simulations for the assessment of an anaesthetist's skills. This is largely because the validity of simulations still remains to be demonstrated. The face validity of the simulation was found to be fairly high by Holzman et al who assessed realism of the simulation by questionnaires given to the participants on an ACRM course (137). 63% of senior attendings (consultants) and 75% of trainees found the simulation environment to be realistic. The construct validity of the simulation has been shown by only one study conducted by Devitt et al who demonstrated that experienced anaesthetists scored higher than trainees. The reliability of the simulations has been established by Devitt et al who showed a high inter-rater reliability between two anaesthetists rating the simulations for technical abilities at managing a crisis (141) and by demonstrating the internal consistency of the rating items (141). Gaba et al showed that the inter-rater reliability for technical performance was quite high while that for non-technical skills assessment was poorer (138). In their review article on the use of simulations for assessment Byrne and Graves concluded that ' any moves towards using performance in administering simulated anaesthesia, as assessment in training or in re-certification procedures are premature' (69).

1.7.5 The value of simulation-based training

In spite of the lack of evidence demonstrating a reduction in airline disasters with CRM training, the aviation industry has adopted CRM training to be a mandatory part of pilot training based on the face validity of the simulator experience. However, with groups such as the University of Texas Human Factors project gradually accruing data from a number of airlines, there is a likelihood that aviation may be able to demonstrate increased safety on flights as a result of CRM training. This is easy in aviation as assessment by independent observers during real flights and in simulators is a part of an ongoing annual assessment of pilots for certification. However, the demonstration of a similar benefit of ACRM training will be much more difficult in anaesthesia and for that matter in surgery as capturing data during surgery will be fraught with medico legal issues and will be expensive. In supporting the need for ACRM training programs Gaba has asserted that 'no industry in which human lives depend on skilled performance has waited for unequivocal proof of the benefits of simulation before embracing it' (144). In making a suggestion that CRM-like training should become mandatory in medicine as well, some authors suggest that 'for the analogous reasons that aviation has adopted CRM training based on face validity, healthcare decision makers may wish to consider face validity in lieu of massive research investment' (144). The benefit of ACRM training is to be found in that fact that from just 2-3 centres in the early nineties, there are over 25 centres all over world offering simulation training to their trainees.

1.8 THE IMPERIAL COLLEGE EXPERIENCE (BACKGROUND TO THE PROJECT)

The lack of objective methods of assessment of surgical competence, led the department on a quest, in the late nineties, to develop and validate novel techniques in the assessment on psychomotor skills. As the department's primary area of expertise was in minimally invasive surgery and because the advent of laparoscopic surgery had been accompanied by a high incidence of complications, the focus of the department's research was on developing techniques which were initially applicable to laparoscopic surgery. The Imperial College Surgical Assessment Device (ICSAD) was developed at this time and initial validation studies conducted by the department demonstrated the validity of this system as a tool for the assessment of skills in laparoscopic surgery. The virtual reality simulator, MIST- VR, also developed around the same time, was validated by the department as a system for the assessment of psychomotor skills in laparoscopic surgery (95).

Over the past three years the department's research interests have included the assessment of open surgical skills as well. It was around this time that the OSATS form of assessment came to the attention of the department. Research conducted by the department demonstrated the construct validity of this form of assessment on bench models such as small bowel and vascular anastomosis (90). Further studies also revealed a strong correlation between OSATS global rating scale and ICSAD's measurement of dexterity (87). The validity of these assessment methods coupled with the development of novel bench models such as small bowel and the sapheno-femoral junction and the

parallel development of virtual reality simulators led to the development of panels of tasks for the assessment of technical proficiency of junior surgical trainees. One paper suggested the creation of a competence day for junior trainees with these assessment methods integrated with didactic examinations and annual assessments such as the RITAs (145).

While considerable advances have been made in the assessment of technical skills, with increasing experience and knowledge in this field, the department has gradually begin to realise some of the limitations of these forms of assessment. In addition, it was only natural that the department's interest in the assessment of surgical competence would progress to develop methods and technologies for the assessment of the other aspects of competence in addition to technical proficiency.

1.9 RATIONALE BEHIND THE RESEARCH

Although there has been a growing interest in competence based assessment during surgical training, so far methods have largely focussed on the assessment of didactic knowledge, clinical skills and lately technical proficiency. The past decade has seen considerable development in the objective assessment of technical skills. Devices such as ICSAD and VR simulators in laparoscopy and endoscopy, however, assess technical skills from a quantitative perspective without the benefit of an assessment of performance completion or quality. Such methods are probably useful in credentialing and summative assessment but seem to lack the information necessary for formative feedback during training. Information such as the number of movements made or the distance travelled by the hands in the execution of a task may not have much relevance to a trainee who is

learning novel psychomotor skills. However, information such as the use of instruments, the technique of knot tying and tissue respect may be of greater importance to the learning process. Rating scales such as OSATS assess these generic skills but then suffer from the limitation that they are not procedure specific, thus making it difficult to use for formative feedback in individuals who have developed generic and not procedure specific skills.

In addition, the assessment of technical proficiency, currently, also suffers from the limitation that assessments occur in skills laboratories. Such an approach can be seen to be an unrealistic representation of skill as this is dissociated from the other characteristics that underpin surgical competence. These are a surgeon's decision-making abilities, inter-personal communication, team skills and leadership. However, these characteristics are developed during training in a random manner usually by observing senior surgeons.

Other high reliability organisations have shown that structured training and objective feedback using simulated scenarios can develop even these complex skills. Adapting a similar approach to medicine, there have been a few specialities that have shown that technical performance can be integrated with higher cognitive functions and team interactions to develop a more holistic approach to the learning and assessment of skills. However, one of the limitations of applying such a concept to surgery has been the inability to simulate surgical tasks and procedures. Over the years simulations have been developed for technical skills training and possess the potential to be integrated with simulated environments to teach and assess technical skills embedded in a matrix of more complex skills.

1.10 PROJECT DESIGN AND STRUCTURE

This thesis is designed to address some of the current limitations in the assessment of technical skills as previously mentioned and also to develop objective methods of assessment of skill and integrate those with techniques already available to make surgical proficiency and competence assessment more comprehensive and multi-dimensional. This project draws on the experience of the department in the use of different types of simulations with varying fidelity and also rationalises their use depending on the skill being assessed. This thesis is structured along a skills ladder starting with the use of simple task- based simulations and gradually escalating the complexity of the simulations to team based procedures. It attempts to take technical skills assessment beyond the skill laboratory to more complex environments in order to add a holistic dimension to the assessment of surgical competence. Fig 1G explains the structure of this thesis as a skills and assessment escalator.

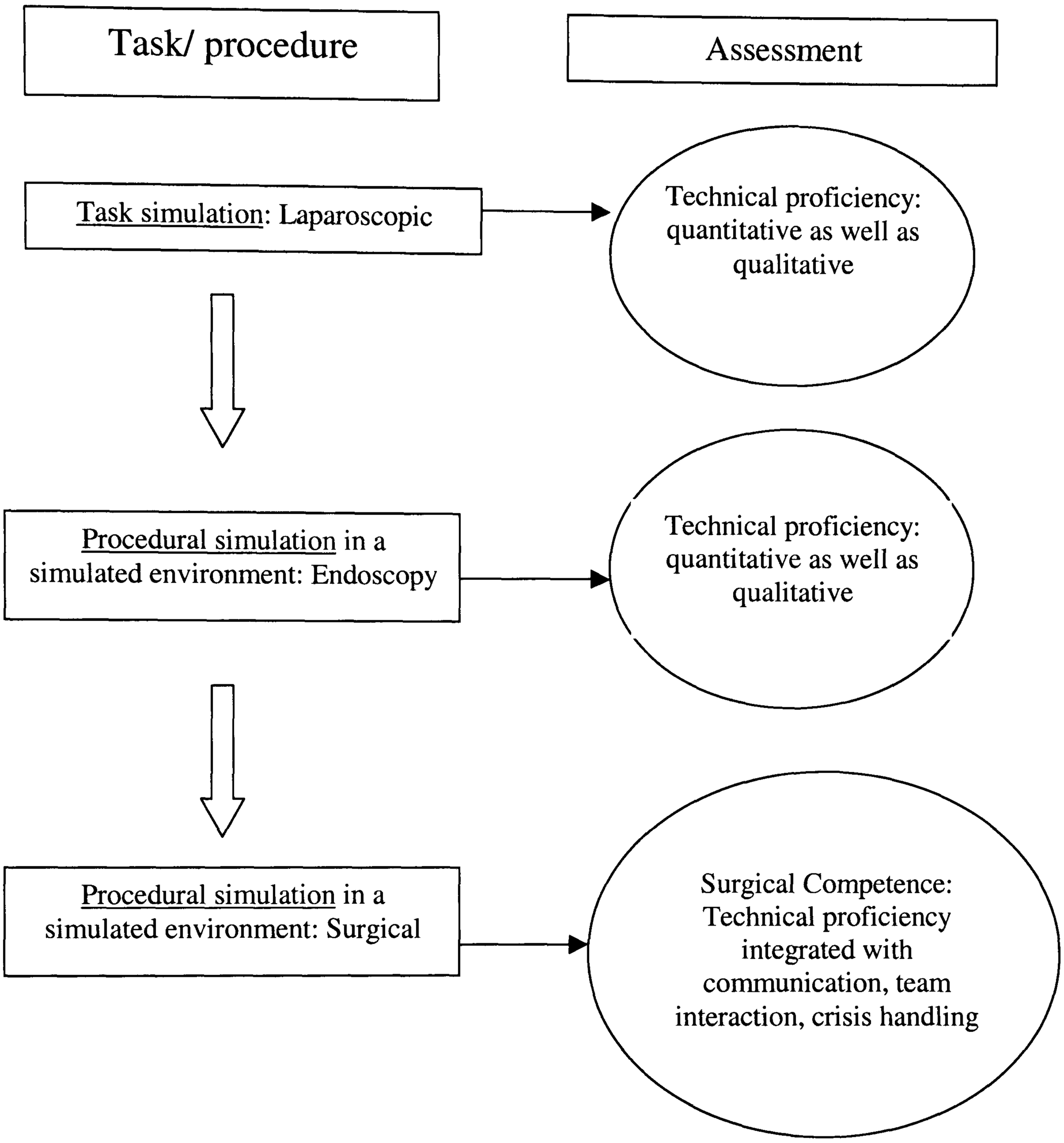


Fig 1G: **Skills and Assessment Escalator**- The thesis is structured along this escalator and demonstrates the role of simple task based simulations in the assessment of technical proficiency and progresses to complex team based simulations involving crisis simulations

1.11 AIMS OF THE PROJECT

Study 1

- Use a simple task based simulation (laparoscopic suturing)
- Develop and validate a scale for the assessment of task completion and accuracy
- Correlate that with other available measures of technical proficiency to develop a multi-component assessment of technical proficiency

Study 2

- Validate a higher fidelity and procedure based simulation (VR endoscopy)
- Adapt this to a simulated environment
- Develop and validate a scale for the assessment of technical skills with the aim of comprehensively assessing proficiency in the procedure

Study 3.

- Adapt a procedure-based simulation (open surgical task, sapheno-femoral junction high-tie) to a simulated environment,
- Conduct an observational study to assess performance during the procedural simulation combining technical skills assessment with that of communication with other team members.
- Examine the face validity and feasibility of the simulation and explore the correlation between technical and team skills

Study 4

- Following on from Study 3 develop a crisis scenario and assess technical performance, communication and team performance under crisis situations
- Examine the feasibility, validity and reliability of such scenarios.

CHAPTER 2: USING SIMPLE TASK SIMULATIONS FOR THE TRAINING AND ASSESSMENT OF TECHNICAL PROFICIENCY

2.1 INTRODUCTION:

The high incidence of complications associated with laparoscopic cholecystectomy during the early nineties (15) reinforced the fact that there is a learning curve associated with laparoscopic surgery and further that learning curves vary for different procedures. For example, the learning curve for laparoscopic cholecystectomy is around 25-50 procedures (12;146) and for laparoscopic colorectal surgery is around 50 procedures (147). However even the number of procedures required to achieve competence may vary (147;148) reflecting the lack of consistency and reliability of this being as a measure used to determine learning curves. A majority of reports on learning curves use time taken for the procedure as the criteria that plateaus with repeated performance (76). Time is a crude measure as it depends on not only the skill of the surgeon but even on patient variability. Therefore, if surgeons have to receive accreditation for procedures, as has been suggested (149), the process of should be made more objective and reliable.

The haphazard introduction of laparoscopic surgery with its adverse consequences and the suggestion that skills achieved early in the learning curve should be acquired outside the OR, emphasized the need for structured training at established skills centres (41;149;150). The intent of skills courses is to teach surgeons the skills essential for laparoscopic surgery under the supervision of skilled surgeons (33;151). There are now a number of established laparoscopic courses with curricula for teaching basic skills in

box-trainers using simple task simulations (33;152;153). Courses in laparoscopic skills vary from basic courses, that use simple drills and aim to teach novices the skills of laparoscopic surgery such as transfer, bimanual coordination and depth perception, to intermediate skills courses that teach procedures such as cholecystectomy and advanced courses that teach complex tasks and procedures. The attendance of laparoscopic skills courses, especially the laparoscopic cholecystectomy course, is increasingly being considered to be mandatory before a trainee can be allowed to perform a supervised real procedure.

The appreciation of the importance of training using simple simulations hasn't, however been accompanied by a parallel effort to enhance the quality of feedback that is provided during these training courses. Technical skills assessment in laparoscopic surgery in addition to being essential for learning skills and credentialing surgeons, is also important in order to evaluate the impact of training interventions such as skills courses (18). Over the past few years there have been major developments made in the objective assessment of technical skills. Systems such as ICSAD have permitted the objective evaluation of surgical dexterity and this system has been validated for the assessment of both simple and complex bench laparoscopic tasks (88;89). Checklists have been developed and used for the assessment of task performance mainly on open bench and animal cadaveric models (85;87). The availability of set criteria against which technical skills can be assessed makes the assessment process more objective, valid and reliable. It has been said that 'check lists turn examiners into observers of behaviour rather than interpreters of behaviour, thereby removing the subjectivity of the evaluation process' (82).

Preliminary research has, however, led to the realisation that ICSAD and video-based assessments such as checklists, when used in isolation, can have certain limitations. ICSAD is solely a measure of dexterity and while it can help in understanding skill and expertise in greater detail, it has certain limitations as a feedback tool. Video-based assessments involve experts watching videos or rating the performance live. Such an assessment involves considerable concentration and to a certain extent is not as objective as ICSAD. However, checklists possess the potential to be useful feedback methods as the learner can be given focussed feedback on steps performed correctly and those that need attention and practise. Such feedback can even be video-based to enable the learners to appreciate their own strengths and limitations- an important aspect of adult learning. Thus while these assessment measures have limitations when used alone, it is possible that when used together they would make the feedback process more comprehensive.

Over the years the role of laparoscopic surgery has expanded and is being used to perform a number of procedures. In order to achieve this, increasing numbers of surgeons are going to need to acquire not just basic skills abut even complex skills such as laparoscopic suturing (153). Laparoscopic suturing is considered to be a skill which every laparoscopic surgeon must possess (153). It is a highly complex task which demands knowledge and a high level of bimanual dexterity, hand-eye coordination and depth perception (33;154). Despite its complexity, it is a task that can be taught by using even a low technology simulation such as a box/ pelvi trainer on synthetic models.

2.2 AIMS

- Develop and validate a checklist for the assessment of laparoscopic suturing skills using a simple task simulation.
- To explore the relationship between the check list and ICSAD with the aim of developing a multi-component assessment protocol in laparoscopic suturing.

2.3 MATERIALS AND METHODS

2.3.1 The Subjects

Surgeons of varying levels of experience in laparoscopic surgery were enrolled into the study. Informed consent was obtained from all the participants. The surgeons were divided into three groups.

Group 1 consisted of experts who frequently performed laparoscopic suturing during operative procedures or frequently practiced in box-trainers

Group 2 consisted of surgeons who possessed laparoscopic skills from having performed laparoscopic procedures mainly cholecystectomy and had some experience in laparoscopic suturing

Group 3 consisted of trainees who had performed fewer than 10 laparoscopic cholecystectomies or none but did not possess any laparoscopic suturing skills.

2.3 2 The simulated task

All the surgeons were asked to place between 1-3 sutures on a synthetic suture pad (Annexe Art, UK) placed within a laparoscopic box trainer (Fig 2A), within a period of 15 minutes. There were pre-determined marks on the suture pad on either side of an incision through which the participants were instructed that the needle should enter and exit. They were also instructed to exit through the incision before going through the opposite side. The angle of the laparoscope (0^0 Karl Storz UK), the distance of the

laparoscope from the suture pad, the port positions, the suture material (3-0 polyglactin, USSC, USA) and the needle holders were all standardized.

The trainees from group 3 were given video based instruction on laparoscopic suturing by an expert surgeon, prior to performing the task.

2.3.3 Assessment of performance

(A) ICSAD (Imperial College Surgical Assessment Device)

This is an electromagnetic motion tracking system where the positional data from sensors worn on the dorsum of a surgeon's hands (Fig 2B) are converted into meaningful dexterity data using special software. Previous studies have shown that path length is a discriminator of skill (construct validity) for simple (88) and complex (89) laparoscopic tasks. Thus time and path length per suture were used as the dexterity parameters for this study.

(B) Checklist (Fig 2C)

All the tasks were video taped and the first suture of each participant was played back to two independent blinded observers who rated the task performance using a checklist.

The checklist was developed after discussions with 3 expert surgeons and consulting a book on laparoscopic suturing techniques (155). The categories of the checklist are similar to the categories of assessment of suturing skills developed by Mori et al (156). The checklist consists of 6 categories expanded into 29 items, which are scored as a 1 or 0. These categories are Needle Loading (NL), Needle driving (ND), pulling the suture

through (PS), Technique of Tying knots (KT), Knot slippage (KS) and Knot quality (KQ). The first two categories are present as two phases as the subjects were requested to exit though the incision prior to pushing the needle through the opposite side. The total score is expressed as a percentage. For purposes of standardization, the participants were requested to load the needle on the needle holder and place the instruments on the suture pad before ICSAD was started and they were then asked to commence the task. As result, two of the items on the check list relating to needle positioning before the task commenced were eliminated from the analysis. Comparison between the groups consisted of analyzing the total checklist score as well as the six individual categories.

2.3.4 Data analysis

Non-parametric tests were used for statistical analyses. Construct validity was established by analysing differences across the all the groups using Kruskal-Wallis test. We also performed an intergroup analysis using Mann-Whitney 'U' test to explore for any differences between the groups and also between the subjects in group 3 who possessed some laparoscopic skills and those that did not.

Correlation between the level of experience and the dexterity data (path length) and the checklist score were done using the Spearman's Rank test. Spearman's Rank test was also used to demonstrate correlation between the path length and the checklist score. We determined the inter-rater reliability between the observers for the checklist score using Cronbach's alpha coefficient.

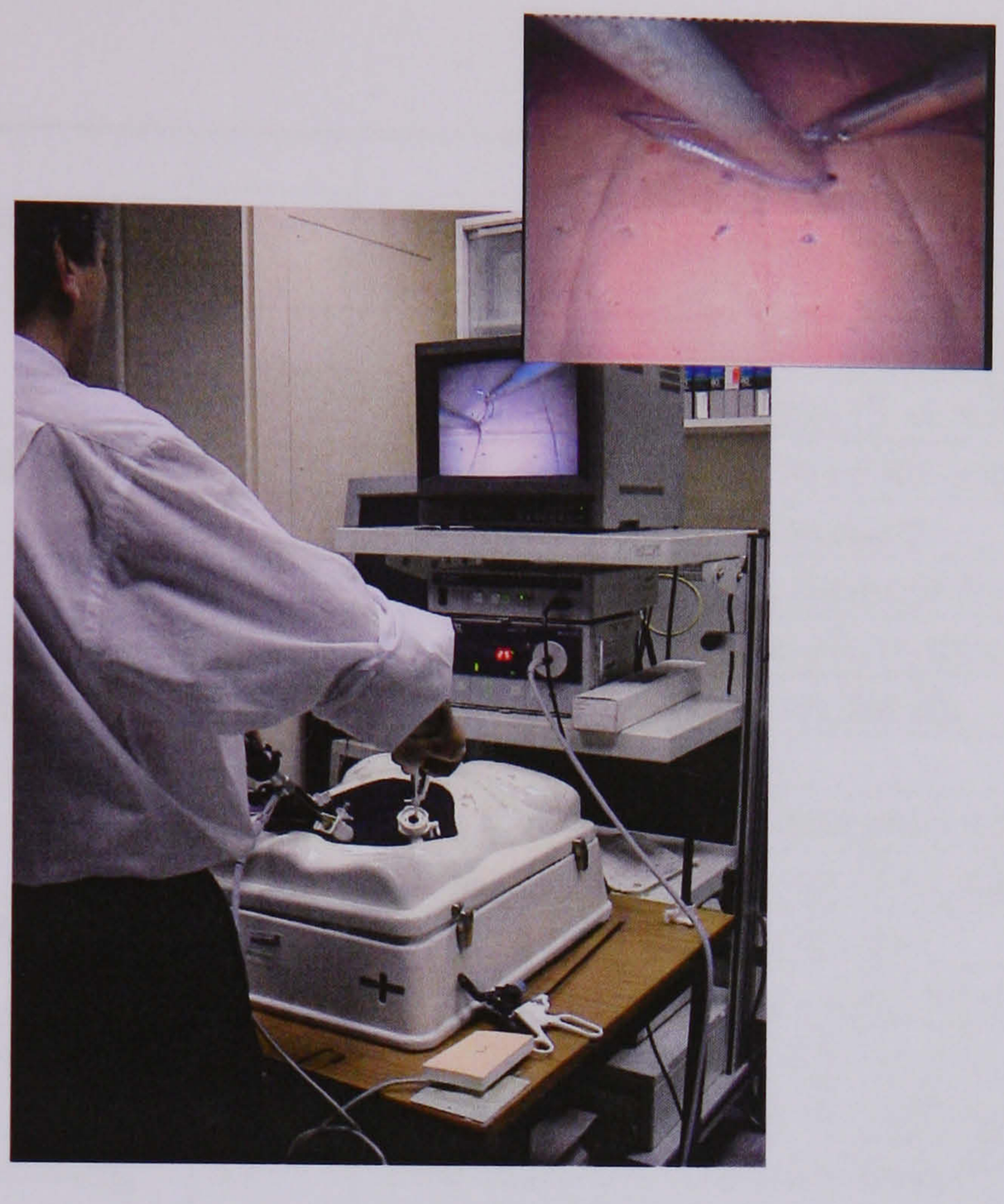


Fig 2A: Laparoscopic task simulation: laparoscopic suturing on a synthetic pad with pre-determined marks placed within a box-trainer

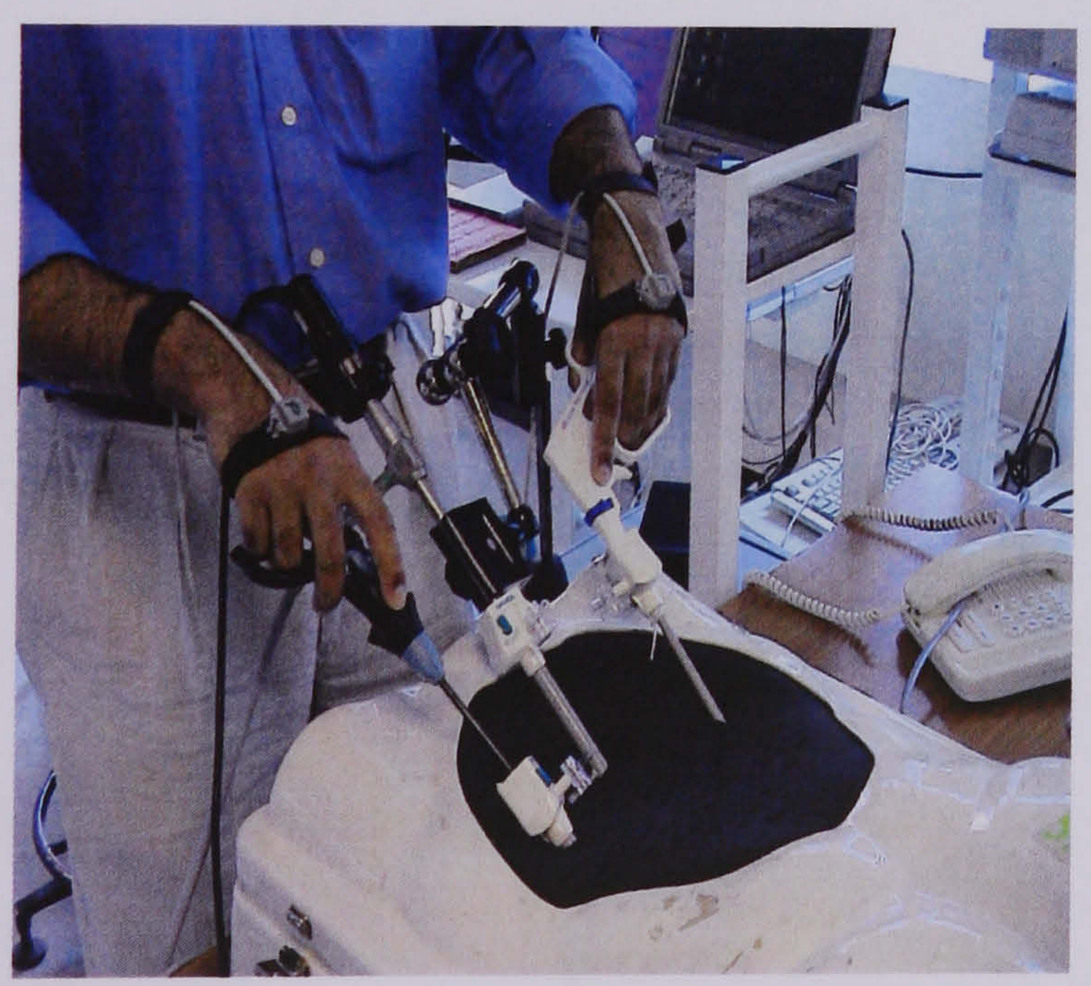


Fig 2B: ICSAD sensors worn on the dorsum of the surgeon's hands performing the laparoscopic task

			Y	N
(A-1) Needle Loading (NL) Entry to incision	1	Held at 1/2 to 2/3 from the tip		
	2	Angle- 90^{o} +/- 20^{o}		
	3	Uses tissue or other instrument for stability		
	4	Attempts at positioning (3 or <3)		
(B-1) Needle driving (ND) through tissue-1 (entry to incision)	5	Entry at 60^{0}-90^{o} to the tissue plane		
	6	Driving with one movement		
	7	Single point of entry through tissue		
	8	Removing the needle along its curve		
(A-2) Needle loading (NL) Incision to exit	9	Held at 1/2 to 2/3 from the tip		
	10	Angle- 90^{o} +/- 20^{o}		
	11	Uses tissue or other instrument for stability		
	12	Attempts at positioning (3 or <3)		
(B-2) Needle driving (ND) (incision to exit)	13	Driving with one movement		
	14	Removing the needle along its curve		
(C) Pulling the suture through (PS)	15	Needle held on needle holder in view at all times		
	16	Using pulley concept or walking along the suture		
(D) Technique of knotting (KT)	17	Two-handed-Overwrap/ underwrap followed by same OR if one handed- one followed by the other		
	18	Correct C loop (no S or O loops)		
	19	Smoothly executed throw, no fumbles		
	20	Correct inverse C loop (no S or O loops)		
	21	Smoothly executed throw, no fumbles		
	22	Knot squared (capsized reef/surgical)		
	23	Correct third C loop (no S or O loops)		
	24	Smoothly executed throw, no fumbles		
(E) Knot slippage after two throws (KS)	25	Knot left loose to slip		
	26	Knot slippage attempts 3 or < 3		
(F) Knot quality (KQ)	27	All throws squared		
	28	Secure knot		
	29	All knots laid on the side (not over the incision)		

Fig 2C: Checklist for laparoscopic suturing.

2.4 RESULTS

There were 6 surgeons in group 1, 7 in group 2 and 13 in group 3. There was a significant difference across the three groups for the time taken per suture (p=0.000) (Fig 2D), for the path length per suture (p=0.000) (Fig 2E) and the checklist score (p=0.000) (Fig 2F). Table 2A gives details of the median and inter-quartile range data for all three groups for the three measures. Analysis of the individual categories of the checklist revealed that there was a significant difference across the groups for 4 of the six components except PS and KS (Table 2B; Figures 2G and 2H)

Analysis between the groups also revealed that there were significant differences in between groups 1 and 2 for path length (p=0.01) and the checklist score (p=0.01) but not for time taken per suture (p=0.06). There were significant differences in between groups 2 and 3 for all three measures: time (p=0.001), path length (p=0.000) and checklist score (p=0.001).

Comparison between the groups for the individual checklist categories revealed that though group 1 scored higher than group 2 (Figs 2G and 2H) for nearly all the checklist categories the difference was only significant for KT (Table 2B). There was a significant difference between groups 2 and 3 for all the categories except for PS and KS (Table 2B, Figs 2G and 2H).

Subgroup analysis of group 3 revealed that there were no differences between the trainees who had previously performed laparoscopic cholecystectomy and those that had not-

Time (p=0.38); path length (p=0.66); check list (p=0.92).

The concurrent validity of the checklist is demonstrated by the existence of a strong and significant correlation between the path length and the checklist score (rho= -0.78, p=0.000). The inter-rater reliability or level of agreement between the expert raters for the checklist score was 0.90.

	Time (sec)	Path length (Cm)	Check list score (%)
Experts (Group 1)	128 (37.2)	1299.23 (437.7)	84.46 (9.8)
Intermediate (Group 2)	248 (131.8)	2110.04 (708.9)	63.6 (20.05)
Novices (Group 3)	667 (432.2)	10050 (7554.5)	28.8 (19.2)

Table 2A: Assessment of laparoscopic suturing. Performance of the three groups per suture. The table describes the median and the inter-quartile range of time (sec-seconds), path length (cm- centimetre) and the checklist score (%- percentage score)

Component	Kruskal Wallis 'p' value	Mann Whitney Groups 1 vs 2	Mann Whitney Groups 2 vs 3
Needle loading (NL)	0.000	0.19	0.001
Needle driving (ND)	0.000	0.11	0.001
Pulling suture (PS)	0.63	0.62	0.66
Knot technique (KT)	0.001	0.02	0.006
Knot slippage	0.10	0.33	0.26
Knot quality	0.000	0.16	0.001

Table 2B: Individual categories of the checklist- Difference across the three groups and between the groups.

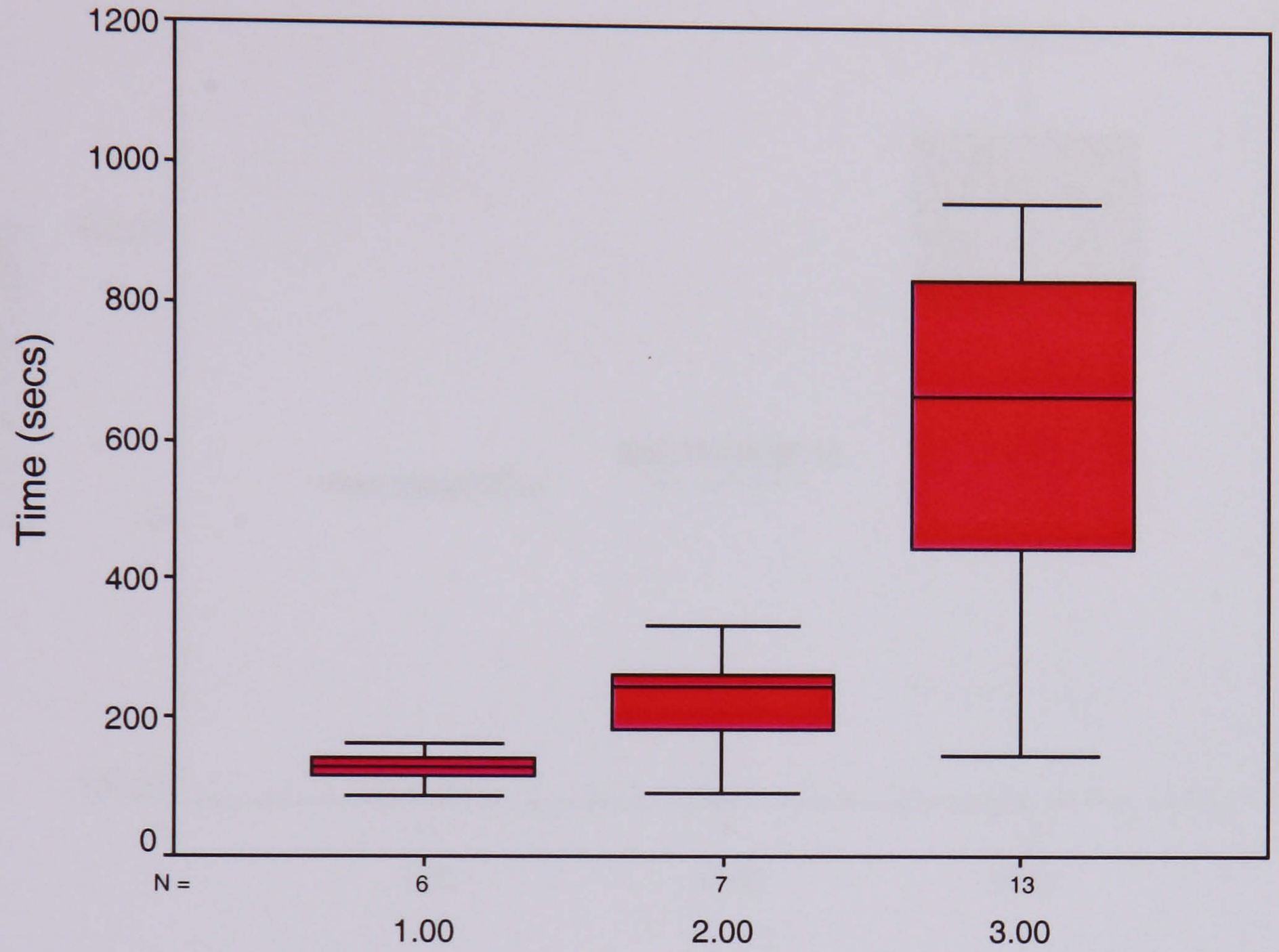


Fig 2D: Time taken per suture. Box plots. The central black line is the median, the data within the box denotes the inter-quartile range and the whiskers denote the whole range. Group 1-experts; group 2- intermediate skills, group 3- novices

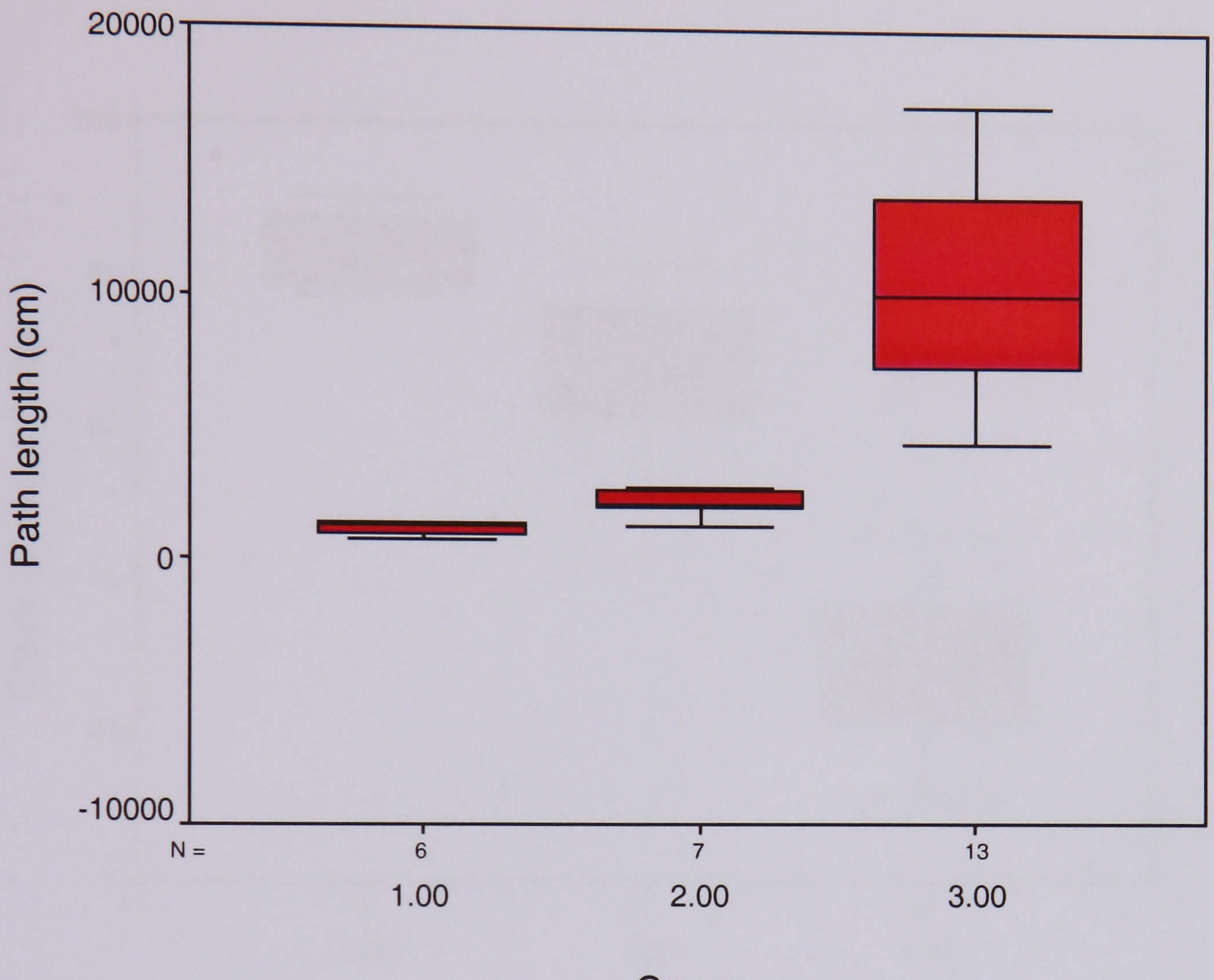


Fig 2E: Path length per suture.

It can be seen that there was a significant difference across all three groups ($p<0.001$). There were significant differences between groups 1 and 2 ($p=0.01$) and between 2 and 3 ($p<0.001$)

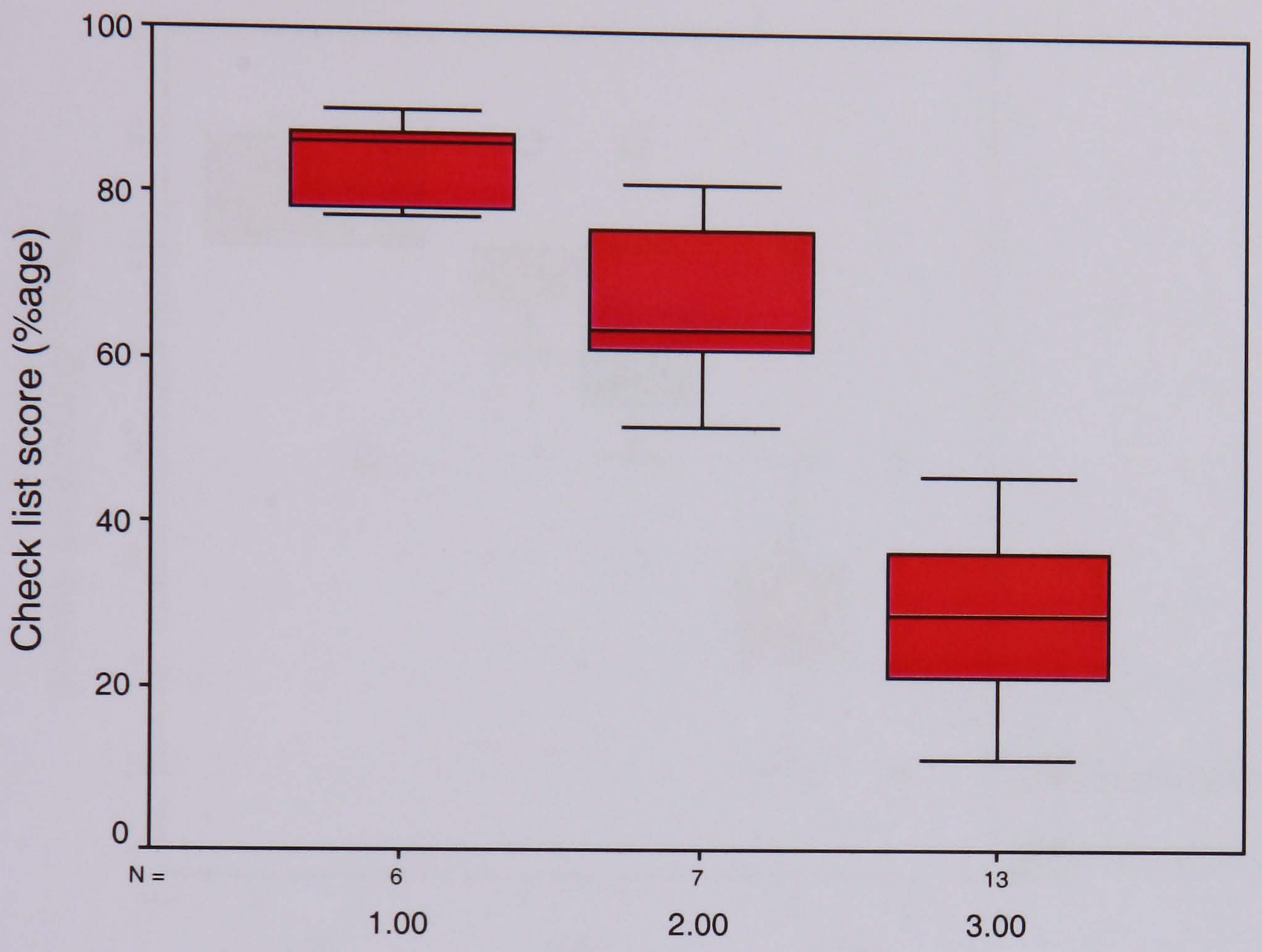


Fig 2F: Check list as a percentage score

There was a significant difference across all three groups ($p<0.001$). There were significant differences between groups 1 and 2 ($p=0.01$) and between groups 2 and 3 ($p=0.001$)

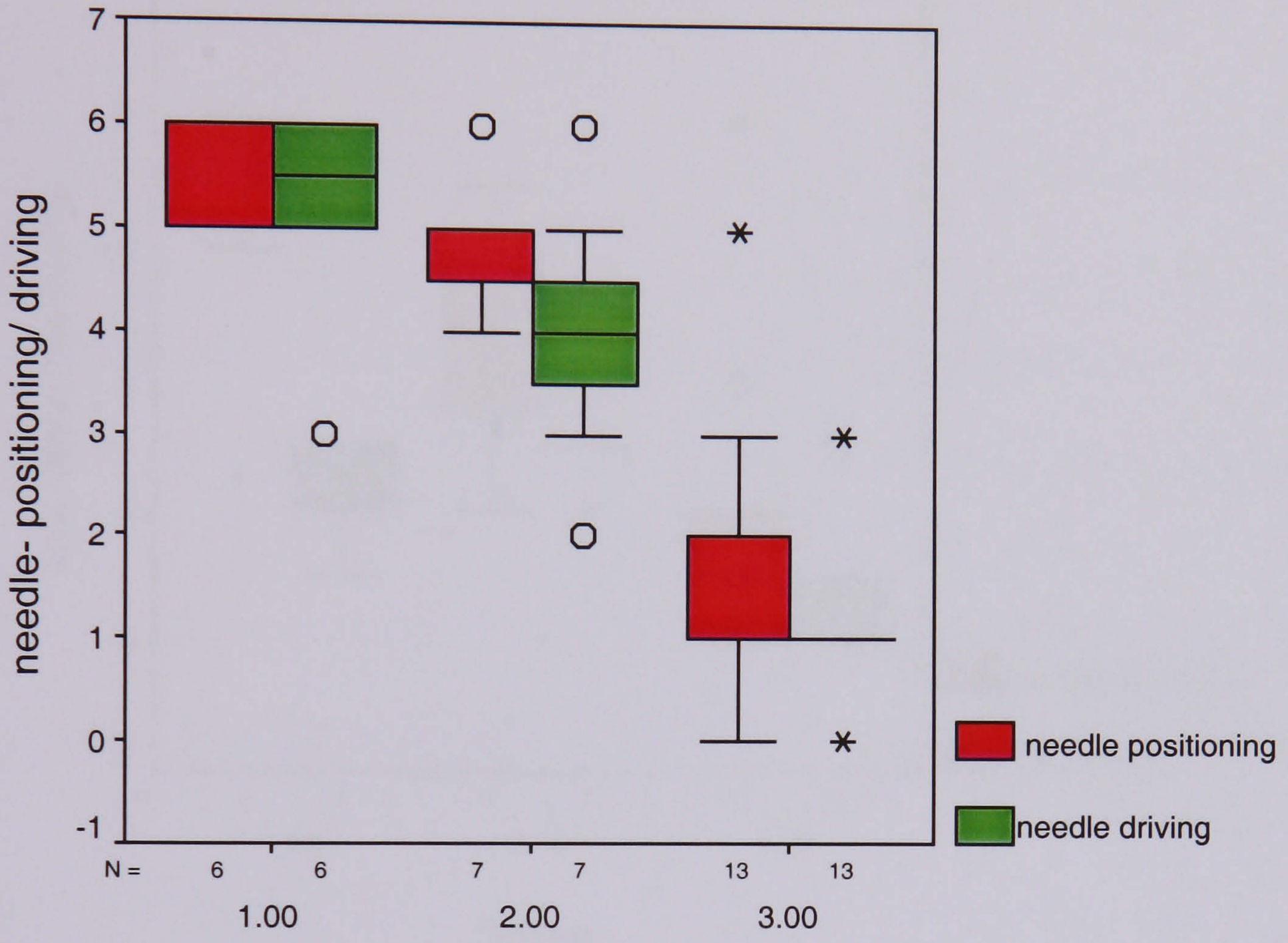


Fig 2G: Cluster box plots for needle loading (NL) and needle driving (ND). Central black line being the median, the data within the box being the inter-quartile range and the ends of the vertical line denoting the whole range. 'o'-denotes outliers; 'x' denotes the extreme observations.

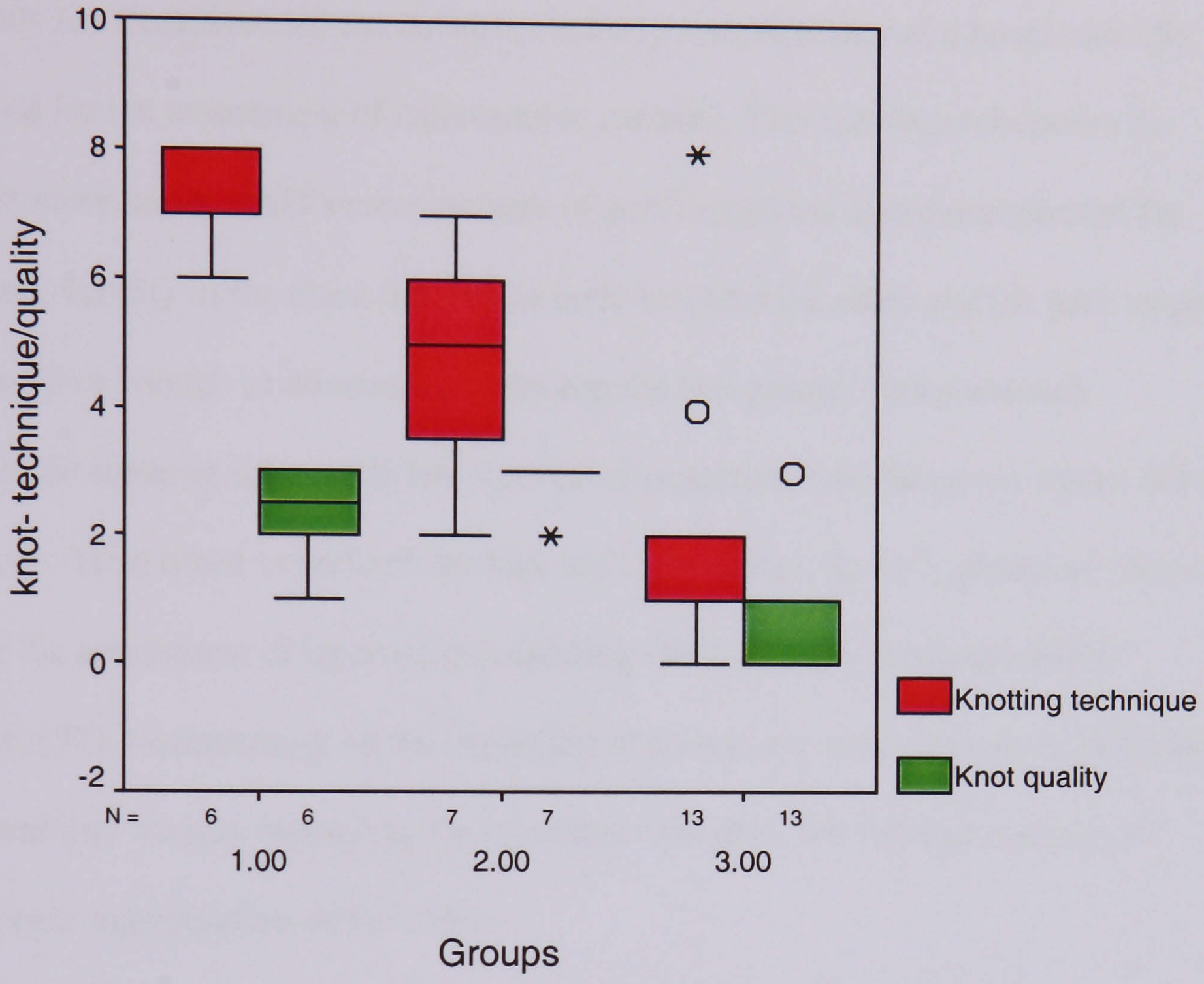


Fig 2H: Cluster box plots for Knot technique (KT) and Knot quality (KQ). Central black line being the median, the data within the box being the inter-quartile range and the ends of the vertical line denoting the whole range. 'o'-denotes outliers; 'x' denotes the extreme observations.

2.5 DISCUSSION

This study has demonstrated the construct validity and reliability of a novel checklist developed for the assessment of laparoscopic suturing. The correlation between the checklist score and ICSAD's measurement of path length has also demonstrated the concurrent validity of the checklist. While both the checklist score and the path length were sensitive enough to discriminate between the two groups who possessed laparoscopic suturing skills to different levels of expertise, time taken per suture did not reflect this. Time taken to perform the task has usually been the only objective measure used for the assessment of laparoscopic suturing skills in many previous studies (153;156;157). Commenting on the limitation of time as the sole measure of skill, Smith et al stated that 'simply measuring the operative time does not indicate mastery of laparoscopic manipulative skills' (151).

ICSAD's parameters help in quantifying dexterity and in understanding expert performance. In a previous study it has been found that expert surgeons perform a task with a lower path length which reflects their economy of motion and their accuracy (88;89). Using a novel surgical assessment device (SAD), with sensors attached to the laparoscopic forceps, Smith et al (151) demonstrated that the learning curve for manipulative accuracy, as measured by calculating the path length, was longer than the learning curve for time taken for task completion. They also suggested that accuracy is a more sensitive indicator of skill than time.

It has been stated that checklists make the assessment of psychomotor skills criteria based and thus increase the validity and reliability of the assessment process and in this study

we have shown that the inter-rater reliability is an impressive 0.90. It is believed that an assessment method with an inter-rater reliability of greater than 0.80 is even suitable for high-stakes assessment (85) such as credentialing for procedures and revalidation .

In a previous study comparing a check list and a global rating scale embedded within the Objective Structured Assessment of Technical skills (OSATS), check lists were found to possess poor reliability in comparison to the global rating scale (82). One of the reasons for the high inter-rater reliability of our checklist is probably a result of the fact that most of the items were anchored with explicit criteria for that particular item to be scored as a 1 or a 0. For example, if there was more than one entry through the suture pad, or even one fumble during the throws while tying the knot, the item was scored as a 0.

The checklist evaluates task execution for mainly psychomotor skills and bimanual coordination. However, in addition there are some items that assess safety such as the needle on the needle holder being in view at all times and the knots being squared and secured. The categories of the checklist except for pulling suture through (PS) and knot slippage (KS) were able to detect differences across the groups. One of the reasons for the apparent lack of difference for KS is that surgeons from even group 1 (experts) did not keep the needle held on the needle loader in view while pulling the suture through. This may be considered to be an unsafe act during the performance of intracorporeal sutures with the likelihood of inadvertent damage to nearby structures. The experts did score higher than the intermediate group for nearly all the categories but the difference was only significant for knot technique (KT). This is one of the most complex tasks in laparoscopic surgery and requires bimanual dexterity and probably is achieved only with considerable practise.

The assessment protocol developed in this study addresses some of the limitations of ICSAD. As mentioned earlier, ICSAD's measures are only an indicator of dexterity and the assessment gives no information regarding the accuracy and precision of the steps. For example, a surgeon could execute a laparoscopic knot without squaring the throws, take fewer throws or complete a loose knot and yet execute the knot within a short time with fewer movements and a lower path length. Thus a comprehensive multi-component assessment of performance as developed in this study would integrate the quantitative measures of dexterity using ICSAD with an expert's assessment of performance using checklists.

In the future it is also likely that synchronizing videos with ICSAD, will result in a situation where video assessment could be dependant on dexterity data. In other words, an analysis of the dexterity data could direct an expert's attention to only those frames of the video that need to be reviewed for a comprehensive assessment of skill. Our group in Imperial College is presently working on such software that will integrate and synchronize video frames with ICSAD's positional data.

The development of a multi-component task assessment could have number of implications. Most importantly, such objective measures can be used for credentialing surgeons for complex procedures. They can also be used to evaluate learning on skills courses and be use to determine the efficacy of courses in imparting skills to participants. Such assessment can also be used to evaluate the impact of certain training interventions such as virtual reality simulators.

This study used a simple task simulation (synthetic suture pad placed within a box trainer) for the assessment of laparoscopic suturing skill. Such simple task simulators are

available in most skills centres and are used to train participants in laparoscopic tasks. However, box trainers are associated with certain drawbacks especially when compared to VR simulators. The lament often is that VR simulators are expensive, but such a view does not take into consideration the cost associated with box-trainers which consist of a camera, a monitor, a light source and the need for instruments. It is cumbersome to change tasks within box trainers and during that process- standardisation is often lost. This is particularly important when such tasks are used for assessment. In comparison, VR simulators are admittedly associated with a high initial cost, but address some of the issues associated with laparoscopic box trainers. The tasks can often be changed by a 'mouse click' and the level of difficulty can be adjusted to allow a structured escalation in training. The software is run of standard desktop personal computers (PCs).

Of crucial difference is the fact that the user is presented with immediate performance feedback as the software's metrics and this feedback could have a positive impact on the learning process. This feedback consists of path length and number of movements along with a tally of skill based errors (48). However, the use of dexterity analysis systems such as ICSAD when integrated with task simulations in box-trainers could serve a similar purpose with the exception that 'error' data is not available as yet.

Even if the argument in favour of VR simulators is convincing, at the present time there are certain limitations with the currently available technology. Most simulators are task trainers as compared to the ability to perform procedures such as laparoscopic cholecystectomy on cadaveric animal models placed within box-trainers and research comparing them with box-trainers hasn't demonstrated an advantage of training on them (158).

In conclusion, in the immediate future it is likely that the use of simple simulations in box-trainers has a definite place in the training of basic and intermediate laparoscopic skills. In addition, the use of such a comprehensive assessment protocol as developed in this study could be employed for the assessment of skills during performance of tasks/ procedures on animals and during real procedures.

CHAPTER 3: USING HIGHER FIDELITY PROCEDURAL SIMULATIONS FOR THE ASSESSMENT OF TECHNICAL PROFICIENCY- ENDOSCOPY

3.1 INTRODUCTION

In the previous chapter, a simple low technology task simulation was used for the assessment of technical proficiency. The advantage of task simulations are that tasks learnt in any simulation, be it a box-trainer or a VR simulator, can be transferred to real life as tasks are components of a procedure (64). For example, a task such as laparoscopic suturing with its inherent complexities once learnt in a simulation can be used to perform any procedure. However, the transfer of skills from procedural simulations to real procedures depends to a large extent on the realism (fidelity) of the simulation.

Simulation based training in endoscopy has so far been difficult to achieve due to lack of realism associated with synthetic models and the problems associated with animal based models (54). There are now a few virtual reality simulators, which have been demonstrated to possess satisfactory face validity (a measure of close a simulation is to real life) (56). They are quite real in their anatomical simulation, possess force feedback and some advanced simulators contain facilities for therapeutic interventions. Using these simulators, novices can acquire their basic skills in a stress free environment and practise extensively.

In addition to their training potential, VR simulators also provide an objective record of performance. The computer program automatically provides a comprehensive log of each procedure in terms of measures such as the time taken, the amount of bowel wall

visualised and others. It is essential, however, that these measures are validated before these simulators are used for assessment. Studies that have validated the metrics of various VR endoscopy simulators have correlated performance on simulators with the experience of the endoscopists, thus demonstrating the construct validity of the simulators. However, this is a crude way to correlate performance in real life with performance on the simulator. An objective assessment technique that could be used both for the assessment of real procedures and performance on a simulator could demonstrate this correlation. In addition, the metrics of a simulator only quantitatively measure performance in terms of time taken, mucosa visualized etc. These measures still do not inform training and credentialing bodies on the quality of performance such as the reaction to patient discomfort, reaction to red-out, strategies for coping with difficult cases and for coping with an inadequately prepared bowel.

While there have been a few of studies validating endoscopy simulators (55;56), there have been very few studies that have demonstrated the true value of these simulators- their ability to shorten the learning curve. This is probably because any study that aims to demonstrate the transfer the skills learnt in virtual reality to real procedures will have to objectively assess skill during real procedures both before and after VR training and also have to be sensitive enough to detect even a small benefit.

Unlike surgery, which is usually practised on patients under general anaesthesia, patients undergoing endoscopy are often awake. In such a situation, technical expertise is an essential but small part of competence in endoscopy. Communication skills before, during and after a procedure are crucial.

While simulators may be effective means of training technical skills, the learning process is devoid of a more realistic simulated experience. Simulated environments have been used in medical school curriculum for a few years now. The role of simulated patients in the acquisition of clinical skills has been widely recognised (159). They have also been used for the assessment of communication and clinical skills during the OSCE format of medical school examinations. Kneebone et al have described an innovative approach to the learning of skills in surgery, that integrates technical skills learning with the training of communication skills (160). Such scenarios involve simulated patients (SPs- actors trained to play the role of patients) who are the focus of the learning experience where the learner's task is to carry out a procedure on a model while interacting with the patient. This is followed by a debriefing sessions where the learner gets feedback from the trainers and more importantly gets the 'patient's perspective' from the SP.

Such a strategy has been found to be feasible and effective for learning skills associated with simple procedures such as urinary catheterisation and wound closure (160).

3.2 AIMS OF THE STUDY

Establish the construct validity of a Virtual Reality Simulator

- Extend the concept of scenario based training and assessment as mentioned above to a more complicated simulated environment that integrates VR simulators with SPs
- Develop and validate a scale for the assessment of procedural proficiency that includes qualitative assessment.
- Demonstrate a correlation between assessment of performance using the scale and simulator's metrics.

3.3. MATERIAL AND METHODS

3.3.1 The subjects

Endoscopists with varying levels of experience were asked to perform Case 4 of the sigmoidoscopy module of a VR lower gastrointestinal simulator (Immersion Medical, USA). Informed consent was taken from all the participants. Familiarity with the procedure and the simulator were essential for enrolment into the study. Informed signed consent was obtained from all the participants.

The participants were divided into three groups.

Group 1- experts (>200 sigmoidoscopy procedures)

Group 2- intermediate (10-50 procedures)

Group 3 – novices (<10 procedures performed or attempted).

3.3.2 The VR simulator

The simulator (Immersion Medical, USA) (Fig 3A) consists of a computer with a monitor, simulation software and a proxy endoscope. The endoscope is designed to mimic a real endoscope and replicates all its functions such as the controls, the suction and insufflation buttons. The interface between the endoscope and the software programme tracks the movements of the scope tip and reproduces the forces experienced during a procedure (haptic feedback). The computer screen presents a dynamic, computer-generated image of the inside of the colon as the user navigates through the virtual colon. This image responds to movements of the endoscope, providing a realistic

representation of the operator's view during all stages of a procedure. As well as providing standard functions such as insufflation of air and cleaning of the camera lens, the endoscope responds to physical manoeuvres such as torque. It can be used to visualise, biopsy and snare lesions, red-out is simulated when appropriate, and rough handling and the inadequate or inappropriate application of torque leads to a recorded groan from the simulator simulating patient discomfort. There are a number of case simulations where each case is unique in terms of the complexity of the procedure and in the presence of pathologies.

3.3.3 The 'Virtual Endoscopy Suite'

In order to increase the realism, a 'Virtual Endoscopy Suite' (Fig 3B) was created. This consisted of a table on which a simulated patient (SP) lay on his/her side. The simulator was deconstructed and placed below the table to give an impression that the procedure was being performed on the SP. The computer's (simulator) audio signal was given to the SP through earphones so that the expressions of discomfort (groans) came directly from the SP instead of coming from the simulator. The video signal was fed to a monitor, which was placed in an ergonomic and more realistic position. After a brief, standardised period of communication controlled by the SP, the participants were asked to directly proceed to performing the procedure.

The 'virtual endoscopy suite' was created in a specially developed simulation room with its advanced video and audio recording facility. The novel technology called the Clinical

Data Recorder (CDR) allows multiple streams of video and audio data to be collected and recorded onto a Digital Versatile Disc (DVD) recording device.

The researchers view the proceedings in the 'virtual endoscopy suite' on a monitor placed in the control room. The data from the cameras is initially fed into a video-mixer (Panasonic Inc, Japan) before being viewed on the monitor. The images on the monitor thus depend on the number and the order of the cameras selected. The audio from the virtual endoscopy suite is streamed into the monitor and the recording medium. Thus the researcher/ trainers are able to view the simulation in real time as well as record the simulation onto a DVD for future review and evaluation. For the purposes of this study we used one input channel for the monitor (the endoscopic view), one close-up view of the endoscope to observe manipulation of the instrument controls and scope handling skills. The other two cameras were intended to observe the entire environment with respect to the interaction between the SP and the endoscopist.

3.3.4 Assessment of performance

Video based expert assessment

The endoscopic view and the image of the hands manipulating the endoscope were simultaneously played back to three blinded independent experienced endoscopists. A global scoring method, Imperial College Assessment of Skills in Endoscopy (ICASE- Fig 3C) was developed for performance assessment. There are 10 components on the global scale. These are movements on hands on control, force of insertion, use of torque, scope handling, time spent in red-out (without luminal vision), patient discomfort, reaction to

patient discomfort, reaction to red-out, knowledge and flow of the procedure and over-all performance. Each component was rated on a 5 point Likert scale with the extremes and the middle being anchored by clear descriptors. The scale was designed such that the middle score (3) was considered to be competent. Thus of a maximum total score of 50, a score of 30 was judged to be competent.

Simulator's metrics

The simulator's software consists of an assessment module for the assessment of the user's performance. It generates metrics for every case. These consist of time taken, percentage of mucosa visualized, percentage of time spent in red-out (without vision) and depth of insertion of the scope. It also records the extent of patient discomfort. For this study we included time taken for the procedure, the depth of scope insertion, the percentage of mucosa visualized and the percentage of time spent in red-out.

3.3.5 Data analysis

Differences between the groups (construct validity) were established using Kruskal-Wallis test. Mann- Whitney 'U' test was used to analyse inter-group differences. Cronbach's alpha was used for inter-observer reliability of the ICASE score. Spearman's correlation co-efficient was used to analyse correlation between the observer's assessment and the simulator's metrics. SPSS statistical package was used for analysis. $p < 0.05$ was considered significant.



Fig 3A: VR Lower GI simulator.

The user interacts with the simulator's software through an interface (the proxy endoscope). Once inserted into the simulator's hardware the movements of the controls of the endoscope result in simultaneous movements of the tip on the monitor



Fig 3B: 'Virtual Endoscopy suite'

The simulator is placed below the bed on which a simulated patient (SP) is lying. The top two images (monitor and scope handling) are shown to the expert observers for evaluation of the ICASE score.

Movements of hands on controls

Many unnecessary movements		Usually purposeful, some unnecessary movements		All purposeful movements
1	2	3	4	5

Gentleness of procedure (amount of force used)

Rough, dangerous, uses excessive force		Gentle usually, excessive force sometimes		Gentle at all times, appropriate but never excessive force
1	2	3	4	5

Torque used

Never		Usually when necessary		Always when necessary
1	2	3	4	5

Scope handling

Looks awkward		confident, competent scope handling appeared awkward occasionally		Skilled (expert)
1	2	3	4	5

Time spent in red-out (absence of luminal vision)

50-75% of the time	Nearly 50%	25-50%	10-25%	<10%
1	2	3	4	5

Patient discomfort

Severe discomfort for >25% of the procedure		Minimal discomfort for 25-50%		Minimal discomfort for <10%
1	2	3	4	5

Reaction to patient discomfort

Ignores discomfort, Advanced regardless of patient's discomfort		Stopped, waited, withdrew (appropriate safety measures) most times		Appropriate safety measures at all times
1	2		4	5

Reaction to red-out (absence of luminal vision)

Advanced without suitable caution		Waited/ withdrew/ insufflate most times		Waited/ withdrew/ insufflate all times
1	2	3	4	5

Knowledge and flow of procedure

Hesitated often, unsure of what to do		Usually confident, hesitated occasionally		Confident, sure of what to do at all times
1	2	3	4	5

Overall performance

poor		competent		excellent
1	2	3	4	5

Fig 3C: ICASE- Imperial College Assessment of Skills in Endoscopy. 10 components rated on a 5 point Likert scale with the extremes and the middle anchored by clear descriptors

3.4 RESULTS

3.4.1 Construct validity of the simulator's metrics

There were 20 participants (6 experts, 7 of intermediate level and 7 novices). Table 3A gives details of the simulator's metrics for the three groups. There was no significant difference across the groups in terms of the time taken for the procedure (p=0.75) and for depth of insertion (p=0.10) (Fig 3D). While there was a difference between groups 2 and 3 for depth of insertion, this was not significant (p=0.06). There was also no significant difference for percentage of mucosa visualized (p=0.21) (Fig 3D). While there was a difference between groups 2 and 3, this was not significant (p=0.18). There was a significant difference across the range in terms of the percentage of time spent in red-out (p=0.015) (Fig 3E). The difference between groups 1 and 2 was nearly significant (p=0.05) while the difference between 2 and 3 was not (p=0.11). From Table 3A it can be seen that groups 1 and 2 were comparable in terms of the time taken, percentage of mucosa visualized and depth of insertion.

3.4.2 ICASE score

There was a significant difference in the ICASE scores across the range ($p< 0.001$) (Fig 3F). The median scores of three groups were 40, 32.3 and 24 respectively (out of a maximum score of 50). There were significant differences in between Groups 1 and 2 (p=0.003) and between groups 2 and 3 (p=0.004). The inter-rater reliability for the ICASE score was 0.94.

3.4.3 Correlation between the ICASE score and the simulator's metrics

There were no significant correlations between the ICASE score and the depth of insertion (rho = 0.17, p=0.46) and the percentage of mucosa visualized (rho=0.26, p=0.26). However, there was a strong correlation between the ICASE score and the percentage of time spent in red-out (rho = -.61, p=0.004).

Parameter	Group 1	Group 2	Group 3
Time (sec)	606 (334.7)	476 (460)	525 (460)
Depth of insertion (cm)	53.5 (1.75)	54 (7.0)	48 (23)
Percentage mucosa	71.6 (26.6)	76.5 (27.8)	55 (52.9)
Percentage red-out	2.24 (3.7)	7.7 (5.1)	10.2 (7.3)

Table 3A: Simulator's metrics as median values; values in parentheses= Inter-quartile range

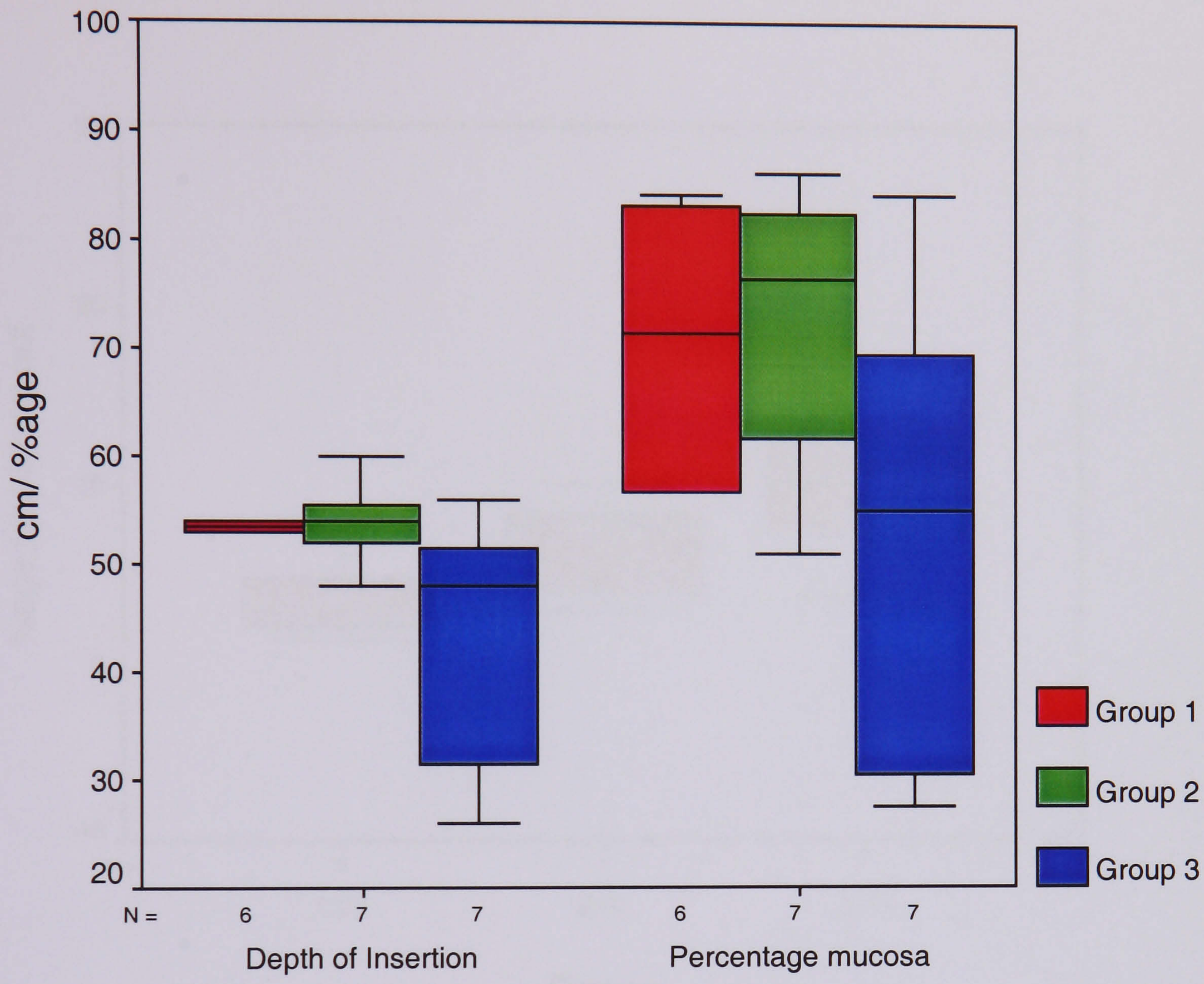


Fig 3D: Box plots. The central black line is the median, the data within the box denotes the inter-quartile range and the whiskers denote the whole range.

Group 1-experts; group 2- intermediate skill, group 3- novices

Depth of Insertion (cm): No significant differences across the groups (p=0.10)

Percentage mucosa (% age): No significant differences across the groups (p=0.21)

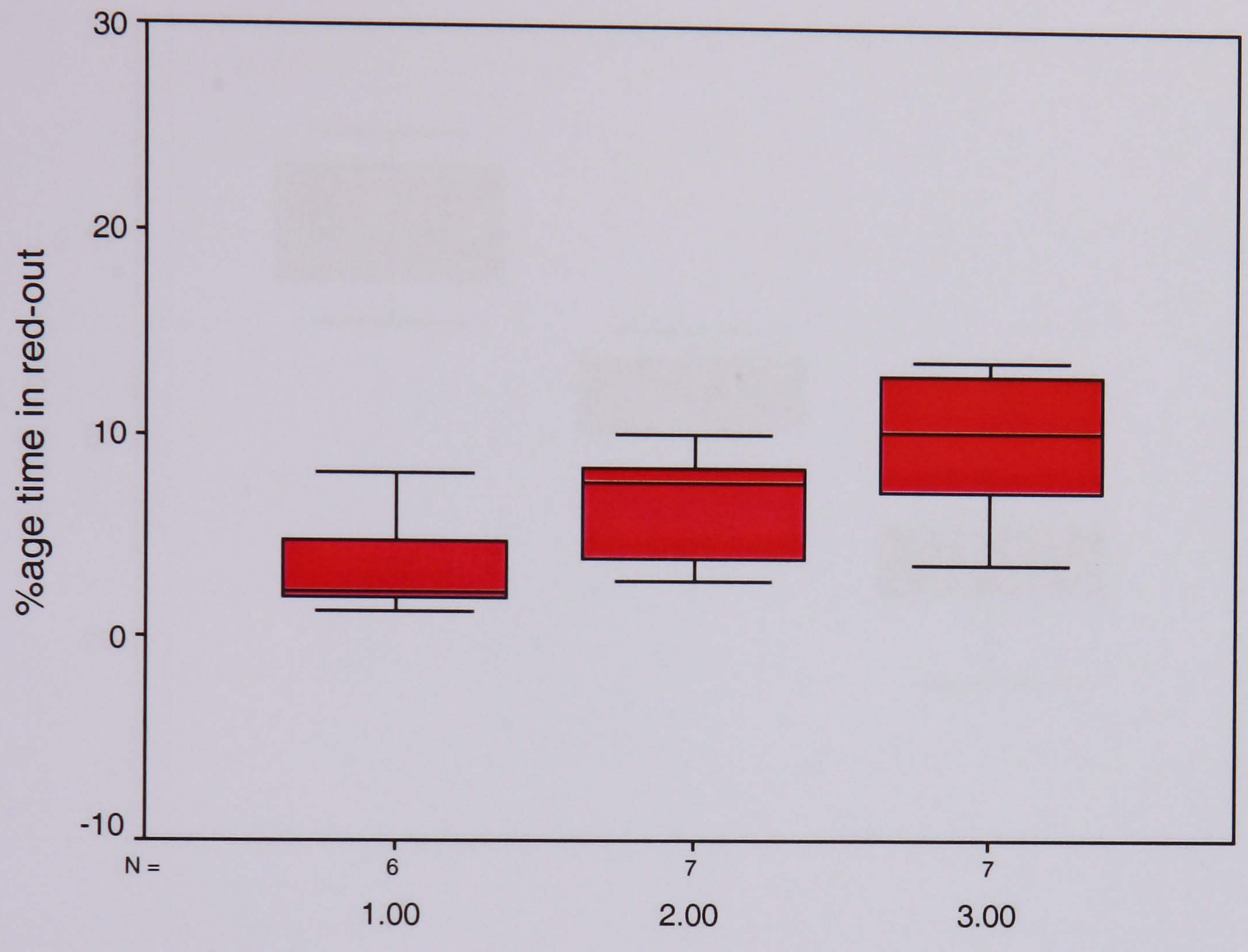


Fig 3E: Percentage time spent in red-out

Significant difference across the groups (p=0.01). Significant differences between groups 1 and 2 (p=0.05) but not between 2 and 3 (p=0.11)

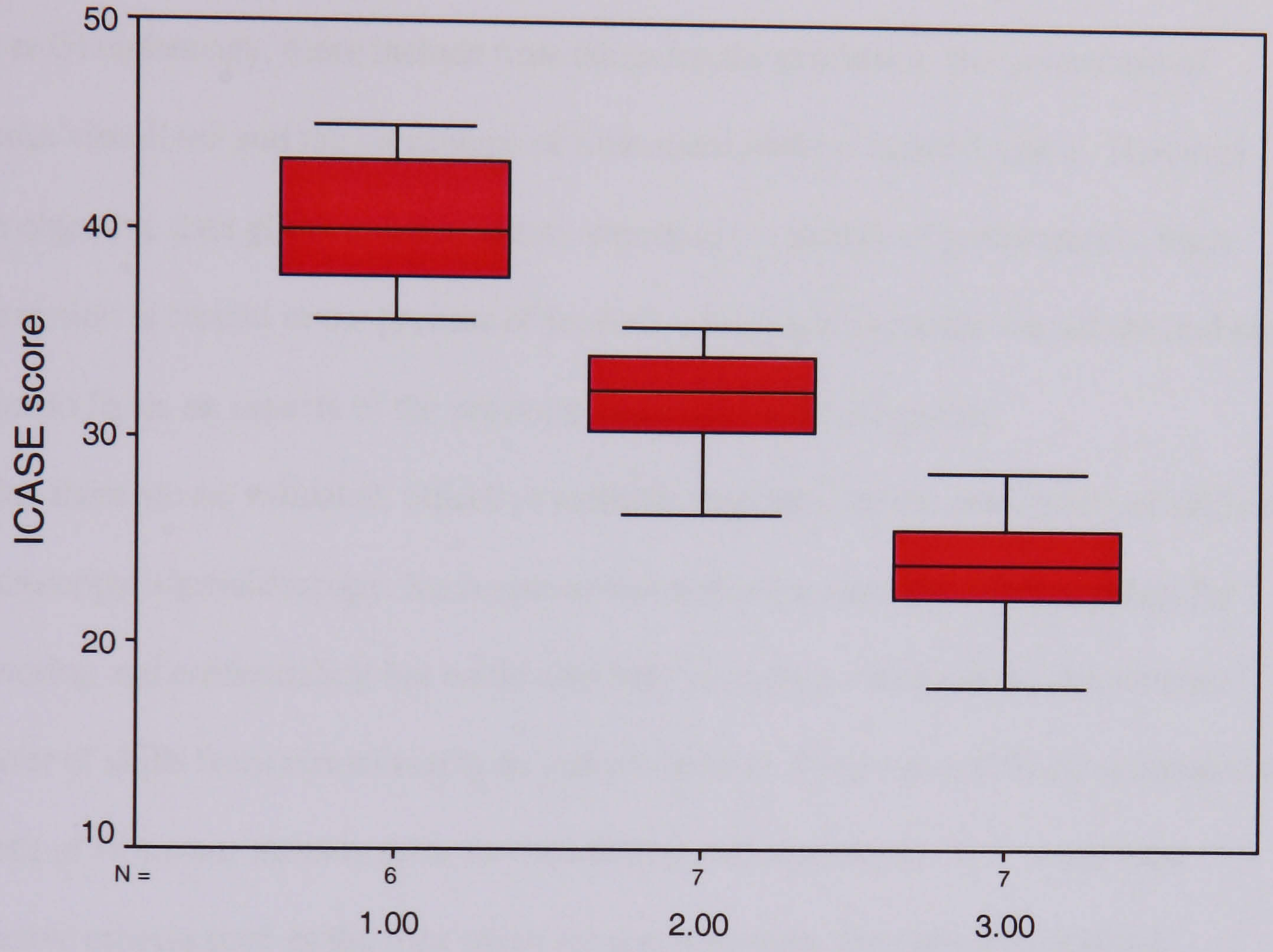


Fig 3F: ICASE score

Significant difference across the groups (p=0.001). Significant differences between groups 1 and 2 (p=0.003) and between groups 2 and 3 (p=0.004)

3.5 DISCUSSION

Most VR simulators present the endoscopist with end-of- performance metrics. In case of Lower GI endoscopy, these include time taken for the procedure, the percentage of mucosa visualized and the percentage of time spent without luminal vision. However such objective data gives no information regarding the quality of performance. Such information is crucial in the process of formative feedback to enable the trainee and the trainer to focus on aspects of the procedure that need to be improved.

So far, there are no validated, objective methods available for the assessment of skills in colonoscopy/ sigmoidoscopy. Such assessment techniques are not only important for proctoring and credentialing but could also help in studies which aim to demonstrate transfer of skills from virtual reality to real procedures. Tuggy et al (59) demonstrated the benefit of simulator training prior to undertaking real procedures. This study used objective criteria such as the time taken for the procedure, time spent in 'red-out' (without vision), number of directional errors (inability of the trainee to direct the scope towards the lumen) and percentage of mucosa visualized but failed to give any details on the observers who viewed the video-tapes, whether the assessment was blinded and the inter-observer reliability.

This study has demonstrated the construct validity and inter-rater reliability of an objective blinded method for the assessment of technical proficiency in lower GI endoscopy. A similar global scoring method, Objective Structured Assessment of Technical Skills (OSATS) has been validated for the assessment of technical skills in surgery (85) and has been found to be valid and reliable. Shah et al found that their point scoring system to assess skills in colonoscopy on patients possessed poor inter-rater

reliability (161). The study concluded that a global scoring system, such as ICASE, would be probably more reliable. Such global scores are effective objective feedback tools during formative assessment as the procedure is divided into individual components and feedback can thus focus on specific components of the procedure that need attention.

In comparison to a previous validation study (56), this study failed to demonstrate a significant difference in the performance of the three groups using the simulator's metrics. The reason for this could be that there were a smaller number of subjects in each group. With greater numbers the differences between the novices and the more experienced groups for most parameters would probably have achieved significance. It is also likely that some of the simulator's metrics are not sensitive enough to detect differences between the skilled endoscopists and those with intermediate skills. One study found that while the simulator was effective in detecting a difference between trained endoscopists and lay people it was ineffective in differentiating the skill between those who were experts and those with intermediate skills (162). Another study found that time taken for the procedure and percentage of time spent in red-out were the only metrics which were sensitive enough to detect differences between endoscopists of varying experience (163). It is important to note that ICASE was able to demonstrate significant differences across the groups and even in between the groups.

The use of real patients introduces an element of patient variability and may well account for the poor inter-rater reliability in the study conducted by Shah et al (161). While that should be the direction to pursue to make assessment in endoscopy more objective, the use of a VR simulator makes the assessment process standardised (161) and face validity

studies have shown that the VR simulator used in this study is accurate in its simulation of real procedures (56).

A use of a global scoring system in the setting of real procedures would be a significant advance in the formative assessment of training effectiveness, particularly if combined with accurate data on extent of endoscope insertion. Previous studies have demonstrated poor inter- observer agreement in terms of percentage of mucosa visualized and the extent of colonic intubation (164). Using an electromagnetic endoscope imager, one study found that the endoscopist's estimation of extent of insertion is accurate less than 50% of the time when the caecum is not intubated (165). When using the ICASE technique for the assessment of real procedures, the image of the hands and the monitor can be combined with the image from a magnetic endoscope imager. This has been referred to as the tri-split assessment (161). Thus a combination of the ICASE score and magnetic imaging technique may result in a comprehensive assessment during real procedures while the ICASE score and the simulator's metrics would result in a comprehensive assessment of skill during procedural simulations as described in this study. However, before that is possible, further research should be directed at validating the simulator's parameters and making the necessary software changes.

This study found that there was a strong and significant correlation between the ICASE score and the simulator's estimation of the percentage of time spent without vision (red-out). This maybe a reflection of the fact that both criteria assess scope handling skills and dexterity. During the procedure, inexperienced or inaccurate hand-eye co-ordination and the inappropriate amount of gas insufflation will result in the scope tip resting or being pushed against the bowel wall, causing an absence of luminal vision (59).

By integrating a simulator with a simulated patient to develop a simulated physical environment, this study has demonstrated the development and validation of an objective blinded assessment method that combines objective measures of performance with information on the quality of performance such as patient discomfort, reaction to the patient's discomfort and reaction to red-out. However, the usefulness of the 'virtual endoscopy suite' can be extended beyond assessment of procedural proficiency. One study by the department, has demonstrated the feasibility of using such a simulated clinical scenario for a holistic training approach where procedural skills are taught in the context of communication skills. Such an approach was used during a training course in flexible sigmoidoscopy. The feedback from the participants was encouraging. A number of them found the scenarios to be highly convincing and they also found that the interaction with a real person provided an important aspect to the learning experience (166).

CHAPTER 4: A TEAM BASED PROCEDURAL SIMULATION FOR THE ASSESSMENT OF SURGICAL COMPETENCE

4.1 INTRODUCTION

The previous chapter described the development and the validity of a simulated environment- the virtual endoscopy suite, which could be used for a holistic training and assessment approach in endoscopy. This concept consisted of adapting a VR simulator to a simulated environment by integrating endoscopic technical performance with interaction with a SP to enhance realism. With the use of this environment, it became apparent that the concept could be extended to procedures in the operating theatre by integrating performance on simple simulations with the presence of OR team members to assess other aspects of surgical competence, in addition to technical proficiency.

Task simulations possess the potential to structure the learning of technical skills in a manner aimed at ensuring a smooth escalation in task complexity leading to the performance of procedures in the operating theatre (65). However, even such a structured transition from the lab to the operating theatre fails to ensure that technical skills learning is linked to the training of other skills such as gloving and gowning, adherence to asepsis and inter-personnel skills such as communication and leadership skills. While simulation based learning of technical skills may allow for learning and practise in a controlled environment, an approach focussing solely on technical skills can ignore these other skills that underpin surgical competence.

In addition, the assessment of technical proficiency in surgery has been restricted to the observation of task/ procedural performance on bench and animal models in a laboratory setting. Though such methods go a long way in making the assessment of technical skills more objective, they still suffer from the drawback that these assessments are conducted in artificial non-OR environments.

CRM training in aviation (110) and ACRM in anaesthesia (137), attempt to create near-real environments to train and assess technical performance with reference to other skills. In high- risk environments such as the ones mentioned above individuals work in teams where the performance of one team member can influence the outcome of the entire mission/ procedure. In such circumstances, it is difficult to dissociate technical proficiency from team skills. The primary aim of these simulation based training programmes is to integrate technical with team skills training to make the most effective use of all resources available during management of various scenarios. Simulation- based training in other dynamic organisations has shown that it is possible to train personnel in leadership skills and decision- making (167).

Prof. Bouchier-Hayes and colleagues (personal communication) have been one of first surgical groups to extend the concept of simulation training to surgery. They developed a procedural simulation during which surgeons performed a laparoscopic cholecystectomy on an anaesthetized pig and faced situations such as a sudden drop in blood pressure on insertion of the laparoscopic ports, the presence of blood in the peritoneal cavity on inspection and a fall in blood pressure while closing the abdomen. The surgeons were assessed for their vigilance, response times to the crisis and their decision-making. The

simulations were then followed by a debriefing session. One of the problems associated with this approach in that operating on live animals is banned in the UK.

Using a live animal based 'wet- lab' Lossing et al described an approach which consisted of teaching trainees patient preparation, draping, aseptic techniques, principles of haemostasis, surgical assisting and over-all conduct in the OR (168). However, the use of live animals was seen as a major drawback of this approach. Over the past few years there have been a number of alternatives to operating on live animal such as synthetic bench models and VR simulators. Synthetic bench models have been shown to be effective in the training of technical skills (58) and using the global rating scale embedded in OSATS one study has shown a strong correlation between procedures on bench models and live operations using a bench model of a sapheno-femoral junction (86).

The global rating scale embedded within OSATS is a generic measure of operative expertise and even though one of its features assesses knowledge of the procedure, it fails to give detailed information on the steps missed by the surgeon. Procedure-specific global rating scales (169) assess knowledge of the procedure on a 5 point Likert scale and combine the advantages of global rating scales with the procedure specific information obtained from check-lists.

While surgical trainees receive feedback on their technical performance from their trainers, however subjective and non-criteria based it may be, surgical trainees rarely ever receive any feedback on their non-technical or team skills. There is very little emphasis placed on the training and assessment of these skills. Communication skills training consist solely of training in doctor-patient skills and not inter-professional skills. The

importance of non-technical skills is highlighted by the argument that errors in the operating theatre are rarely due to deficiencies in technical performance (120). They result more from impaired decision- making, absence of situation awareness, inter-personal conflicts and often from the inability to handle a crisis or stress. Any training strategy which aims to enhance team working in the OR, improve communication skills and vigilance and impart stress-management and error recovery strategies should improve performance in the OR and thus reduce the incidence of errors.

4.2 AIMS OF THE STUDY

- To establish the construct validity and reliability of the methods used for the assessment of technical skills during performance of a commonly practised procedure on a bench model.
- Develop and explore the reliability of a non-technical skills assessment method
- Pilot such a method to observe non-technical skills in surgeons and also explore any differences between grades of trainees
- Establish the feasibility and face validity of the simulation using post-simulation video-based feedback of the participants.

4.3 MATERIALS AND METHODS

4.3.1 The simulated operating theatre

The physical environment

A simulated operating theatre (SOT) (Fig 4A) was developed with the aim of replicating a real operating theatre as closely as possible with an adjacent room serving as a control room (Fig 4A). Both these rooms are separated by a one way viewing glass through which the happenings in the SOT are monitored by the researchers and other trainers involved in the simulations. However, it soon became clear that such a view was restricted to only the over-all view of the SOT and that in order to evaluate the technical skills of the surgeon as well as the fine nuances of communication and interaction between the personnel, a higher fidelity viewing and recording medium would be essential. A novel technology called the Clinical Data Recorder (CDR) (Fig 4B) was developed. The CDR allows multiple streams of video and audio data to be collected and recorded onto a Digital Versatile Disc (DVD) recording device.

The researchers and trainers view the proceedings in the SOT on a monitor placed in the control room. The data from the cameras is initially fed into a video-mixer before being viewed on the monitor. The images on the monitor thus depend on the number and the order of the cameras selected. For the purposes of this project we used only 4 cameras, the input from which was fed into the CDR. The audio from the SOT is streamed into the monitor and the recording medium. Thus the researcher/ trainers are able to view the simulation in real time as well as record the simulation onto a DVD for future review and

evaluation. This simulated environment with its high fidelity audio and video recording technology can be used as a platform for the development of simulated scenarios for training and assessment in both technical and non-technical skills. All the simulations were held in the Simulated Operating Theatre.

The equipment

In addition to all the standard operating theatre equipment, one of the primary features of the SOT is an anaesthetic simulator (SimMan, Laerdl, UK) (Fig 4A). This is a simulator of moderate fidelity, which allows manipulation of the mannequin's haemodynamic parameters through software installed on a notebook computer located in the control room. It is possible to create a number of scenarios using the software. These include hypoxia, laryngospasm, pneumothorax etc. The mannequin can be effectively intubated and attached to the anaesthetic machine. The simulator's haemodynamic data is displayed on a monitor fixed to the anaesthetic machine. This data consists of pulse oximetry, blood pressure recording, expired carbon dioxide and others. The mannequin has facilities for insertion of an intravenous cannula into the antecubital fossa.

Other facilities in the SOT include trolleys with all the required anaesthetic instruments, tubes and drugs, a surgical tray with all the relevant instruments for the procedure and a trolley with an assortment of suture materials.

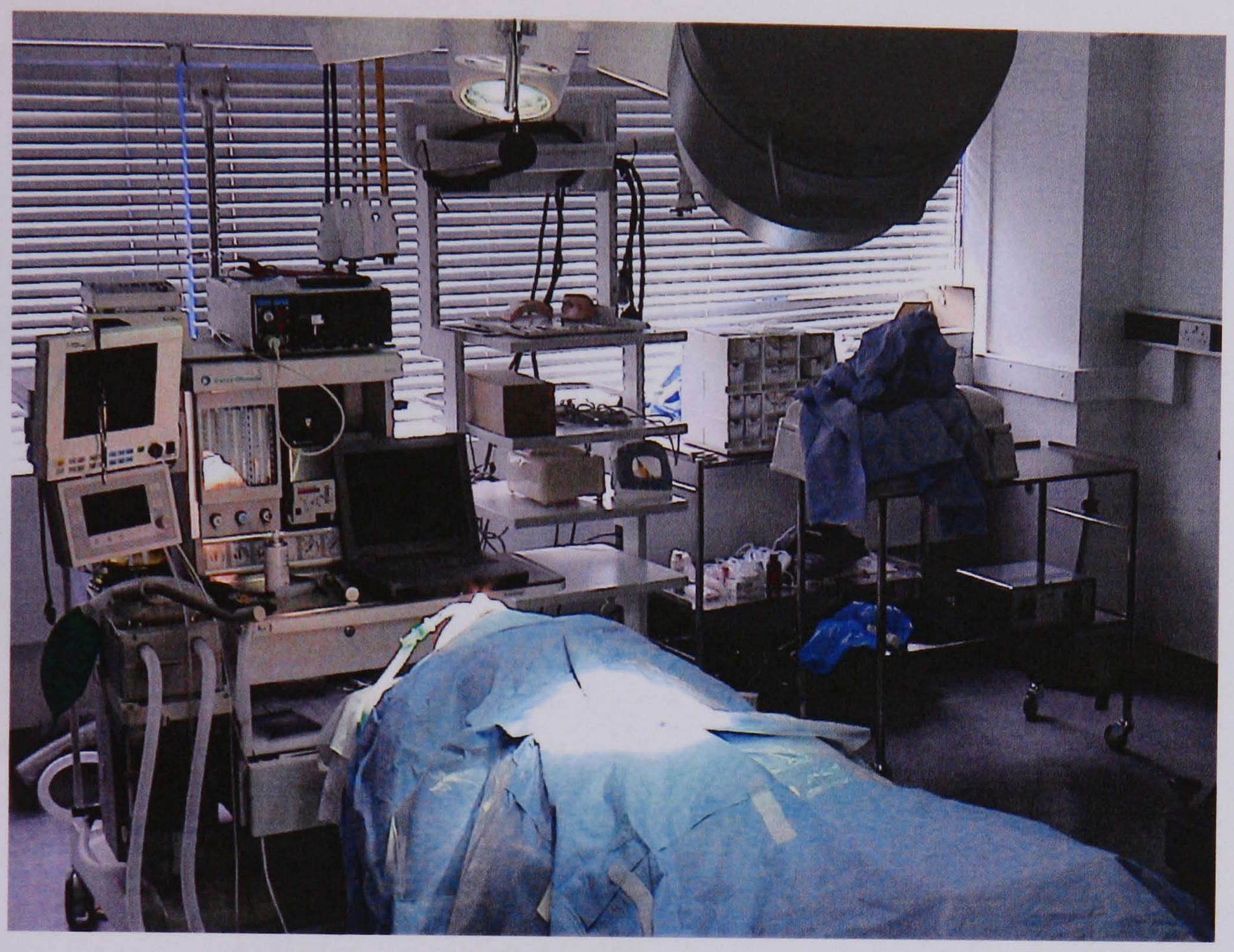




Fig 4A: The figure above shows the Simulated Operating Theatre (SOT) with the anaesthetic simulator (SimMan). The SOT is separated from the control room, shown in the lower figure, by a one-way glass. The trainers/ raters watch the simulation on monitors placed in the control room. SimMan's physiological parameters can be adjusted on a notebook computer loaded with SimMan's software to create 'scenarios'.

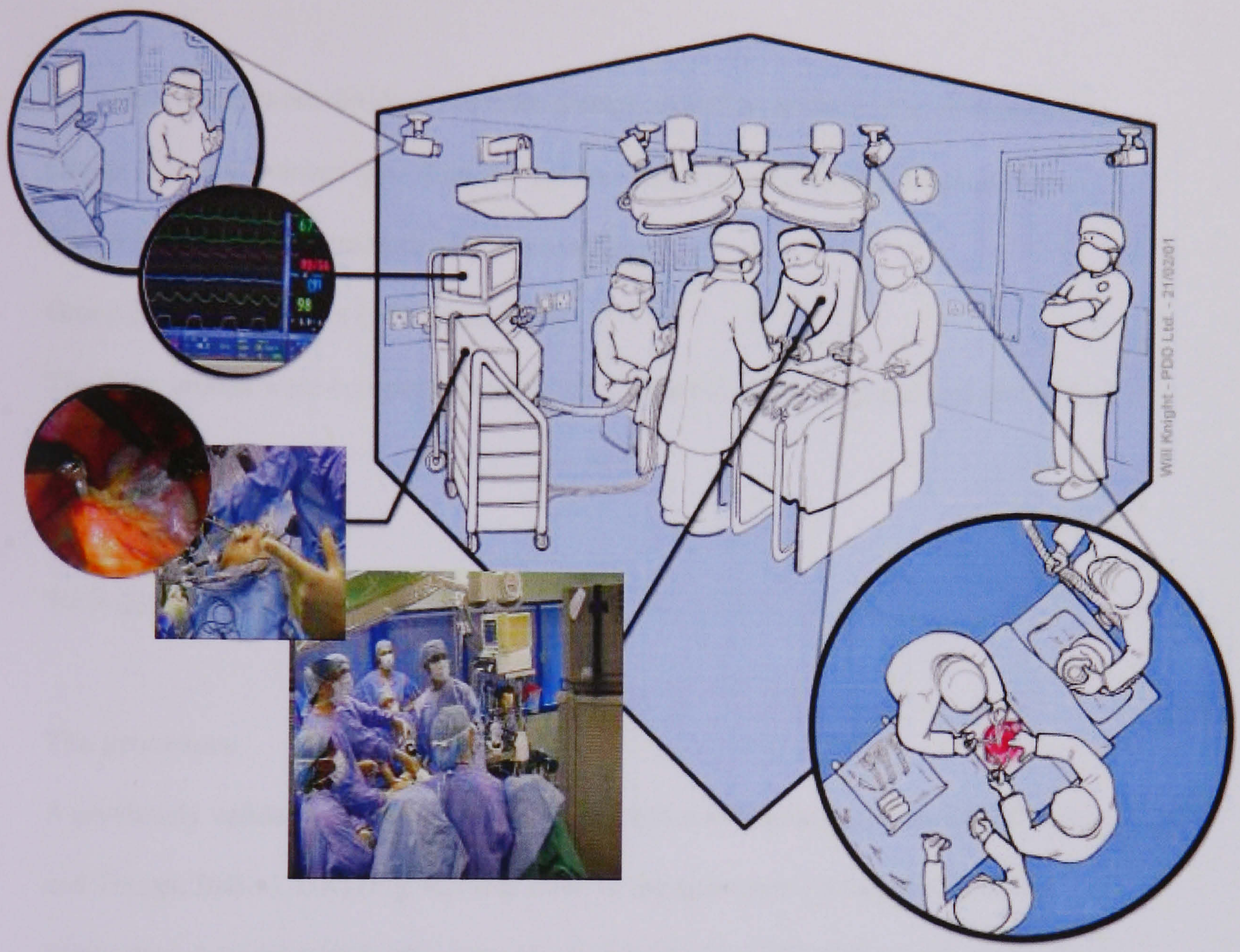


Fig 4B: **Clinical Data Recorder (CDR)** is a recording device which allows multiple streams of video and audio data and the parameters displayed on the anaesthetic monitor to be recorded onto an encrypted Digital Versatile disc (DVD) for future review. The video data can be collected from a number of cameras installed in the SOT to be able to review any aspect of the surgical environment.

4.3.2 The subjects

Surgical trainees were divided into three groups according to their level of experience.

Group 1: Junior trainees (less than 20 sapheno-femoral junction high tie procedures)

Group 2: Middle level trainees (20-50 procedures)

Group 3: Senior trainees (>50 procedures)

The three groups were compared to each other in terms of the technical and non-technical skills.

4.3.3. Study Design

The procedure

A previously validated synthetic model of the sapheno-femoral junction (SFJ) (Limbs and Things, Bristol, UK) (Fig 4C) was fixed to the anaesthetic simulator. This is a silicon-based model with fairly accurate simulation of the SFJ with a layer of simulated skin overlying a layer of superficial fascia. It consists of a 'saphenous vein' with four tributaries connected to a 'femoral' vein. These are set in a cast of silicon which resembles fat. The model was then draped with surgical drapes and held with surgical towel clips. The trainees were requested to perform the simulated procedure in the SOT.

The simulation

A standardised theatre team consisting of an anaesthetist, an ODA, a scrub nurse, a 'runner' nurse (assistant to the scrub nurse) and an assistant first entered the SOT and assumed their positions. The surgeons entered the simulation area through the control room adjacent to the operating theatre. They were briefed about the scenario (Appendix 5) and made familiar with the theatre environment through the one- way glass. They were then asked to sign a consent form for participating in the study with a clause requesting their confidentiality. They then entered the operating theatre under the assumption that they had scrubbed and proceeded to gown and glove.

While performing the simulations there was a pre-programmed hypoxia scenario was created to assess the awareness of the surgeon towards the patient's condition and the anaesthetist. This was done by the simulation controller/ research fellow by manipulating the software of the anaesthetic simulator. The anaesthetists were requested not to prompt the surgical trainee about the crisis, but were asked to do so when prompted by the simulation controller. This was usually communicated as a signal by the simulation controller when the surgical trainee had failed to acknowledge the existence of a problem. All the simulations were recorded onto a DVD disc for subsequent analysis by independent raters.

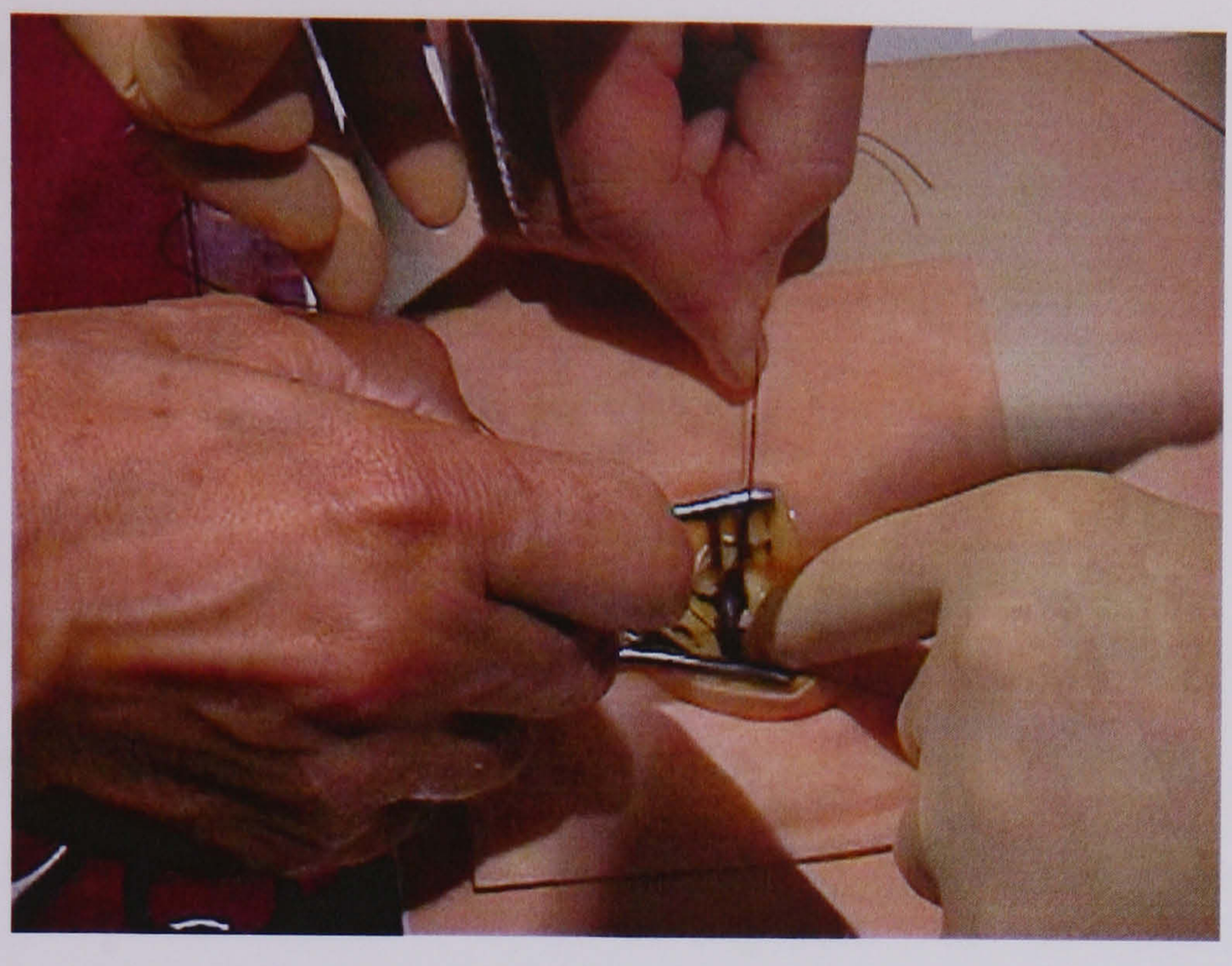


Fig 4C: Synthetic model of the Sapheno-femoral junction (SFJ)

Post simulation feedback/ Debriefing

This was given to the participants subsequent to the simulation. They were given a feedback of their technical skills by a surgeon and non-technical skills by a human factors expert with a background in simulations in the nuclear industry and the trained research fellow. These were all video-based sessions. The simulations were time-marked by the assessors during the assessment process, and these segments were played back during the feedback session to mainly highlight features of communication and vigilance.

4.3.4 Performance measures and validity of the simulation

Assessment of technical skills

One close-up image of the operative site and another of the surgeon's hands were shown to three independent blinded observers instructed in the use of the global rating scale

(OSATS) (Appendix 6). The scale used in this study is a slight modification of global rating scale embedded in the OSATS format of assessment of technical skills as developed by Reznick and group (84). This consists of 8 components of surgical skills rated on a 5 point scale with the middle and the extremes anchored by explicit descriptors (85).

In addition to the generic global rating scale they also assessed technical performance by using a procedure-specific global rating scale (Appendix 7)- Imperial College Evaluation of Procedure- specific Skills (ICEPS). This was developed by the department and consists of 7 steps of the procedure- sapheno femoral junction high-tie, which are rated in a manner similar to the global rating scale.

Essential item checklist.

This consisted of certain parameters (Table 4A) which are critical to safe and effective performance in the operating theatre. Some of these parameters are embedded in the technical skills assessment (Op Comp) guide for trainers as published by the Surgical Advisory Committee (SAC) in general surgery in the UK (170). Each item in the checklist was analysed separately. One independent observer (a surgeon) carried out the checklist assessment. This checklist consisted of 4 essential aspects of team communication which are considered to be important during a procedure (E, H-J).

	Step	Y	N
A	Gowning		
B	Gloving		
C	Checking consent and side with team		
D	Patient positioning (Trendelenburg)		
E	Asking anaesthetist to start		
F	Handling of sharps (safely)		
G	Adheres to asepsis throughout the procedure		
H	Waiting for swab/ instrument check prior to closure		
I	Acknowledgement of swab count		
J	Informs anaesthetist of closing		

Table 4A: Essential items check list.

Assessment of non-technical skills

A detailed review of non-technical skills rating scales revealed two that were developed for the assessment of non-technical skills in pilots. These were the LOSA checklist (110) and the NOTECHS rating scales (133). These scales were considered to be too complex to be used in their entirety when the aim was only to pilot the assessment of non-technical skills in surgery. However, a few elements of the LOSA checklist, which were considered to be relevant to surgery, were used with the intention of exploring the feasibility and inter-reliability of non-technical skills assessment. These consisted of the rating of

behaviours such as pre-operative preparation, communication, vigilance (situation awareness) and leadership, on a 5 point Likert scale with the two extremes being anchored by descriptors. This assessment was carried out by two independent assessors; a human- factors expert with experience in simulations in the nuclear industry and the research fellow who had been trained by the former. During the training period, the first five assessments were conducted together to ensure uniformity of assessment. Subsequently the assessments were performed independently to permit the evaluation of the inter-rater reliability.

Communication- Utterance frequency (UF)

In addition communication was also assessed by performing a communication count ('utterance frequency'-UF) of the surgeon with other team members. This method is similar to an approach used in an earlier study observing communication in the operating theatre (171). The utterance frequency was defined as the number of episodes of communication per minute.

Table 4B describes the various types of communication, which have been modified from a classification described by Coiera (personal communication). A total utterance frequency was then calculated for the communication between the surgeon (S) and the anaesthetist (An), assistant (A) and the scrub nurse (SN). This is described as S-An, S-A, and S-SN respectively.

A	Enquiry made
B	Received information
C	Received an enquiry
D	Gave information
E	Gave instructions
F	Acknowledge/ appreciate

Table 4B: Communication- Utterance Frequency (UF)

Participants assessment of the realism of the simulation: Face validity.

The participants were asked two questions to establish the face validity of the simulation.

- Was the simulation realistic- Did you suspend disbelief and feel that you were in real operating theatre?
- Was the model a realistic representation of a real procedure?

The two questions were meant to evaluate the face validity of the simulation. The participants were requested to answer the questions as disagree, not sure and agree.

Questions during the feedback session:

In addition, during the debriefing session , the participants were asked four other standardised questions:

- The SOT simulation is a good method for training technical skills
- The SOT simulation is a good method of training technical skills
- Is preoperative briefing essential?

- Is a pre-operative instrument check essential

4.3.5 Data analysis

As the data was not found to have normal distribution for most measures and as the groups were quite small, non-parametric tests have been used for analysis. Differences in between groups for the technical and non-technical skills were done using the Kruskal-Wallis test. Spearman's Rank correlation test was also used to determine the correlation between the global rating and the procedure specific assessment of technical skills and to determine the correlation between the technical skills score and the non-technical score. A correlation co-efficient of lower than 0.30 was considered to be a weak correlation, between 0.30 and 0.50 was a moderate correlation and greater than 0.50 was a strong correlation (172). The inter-rater reliability (the level of agreement between the observers) was determined by using Cronbach's alpha coefficient. Chi-square test was used for analysis of the essential items checklist data.

4.4. RESULTS:

There were 27 surgeons who participated in the study. Informed consent was taken from all the participants. There were 11 trainees in group 1(junior), 9 in group 2 (intermediate) and 7 in group 3 (senior). There were 21 males and 6 females and all the participants were right handed.

4.4.1 Technical skills assessment

Inter-rater reliability for the technical assessment

The inter-observer reliability (alpha) between the three expert observers for the global rating scale was 0.90 and for the ICEPS score was 0.94.

Difference between the groups: construct validity

There was a significant difference in the global score across the whole range of trainees (p=0.002) (Table 4C) (Fig 4D). While the difference in the global score between groups 1 and 2 was significant (p=0.01), the difference between groups 2 and 3 was not significant (p=0.31), although group 3 trainees did score higher (Table 4C).

There was a significant difference across the three groups for the ICEPS score as well (p=0.002) (Fig 4E). While there was a significant difference in the scores between groups 1 and 2 (p=0.012), this was not observed for the scores between 2 and 3 (p=0.56) although group 3 did score higher than group 2 (Table 4C).

Correlation between the global and the ICEPS score

There was a strong correlation between the two measures of technical skills. The correlation coefficient was 0.87.

	Global score (median) (Maximum score=40)	**ICEPS score (median) (Maximum score= 35)**
Group 1 (Junior)	19.66 (4.66)	17.88 (4)
Group 2 (Intermediate)	25.6 (8.8)	22.6 (6.33)
Group 3 (Senior)	28.66 (8)	24.33 (4.66)

Table 4C: Technical skills assessment: describes the median and inter-quartile range (within the parentheses) of the Global and ICEPS scores of the three groups

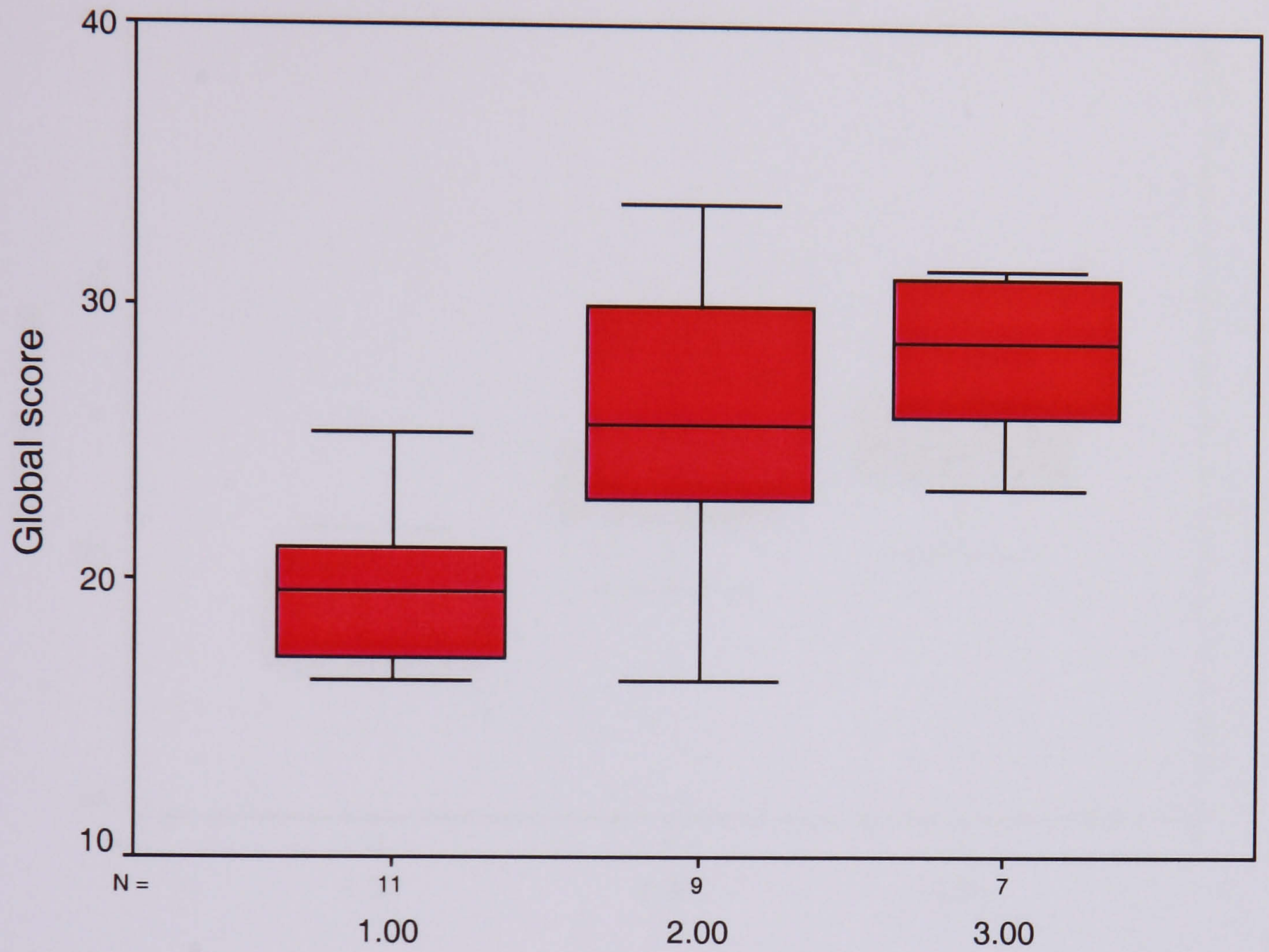


Fig 4D: Global score for the three groups. There is a significant difference across the three groups (p=0.002). Significant difference between 1 and 2 (p=0.01) but not between 2 and 3 (p=0.31)

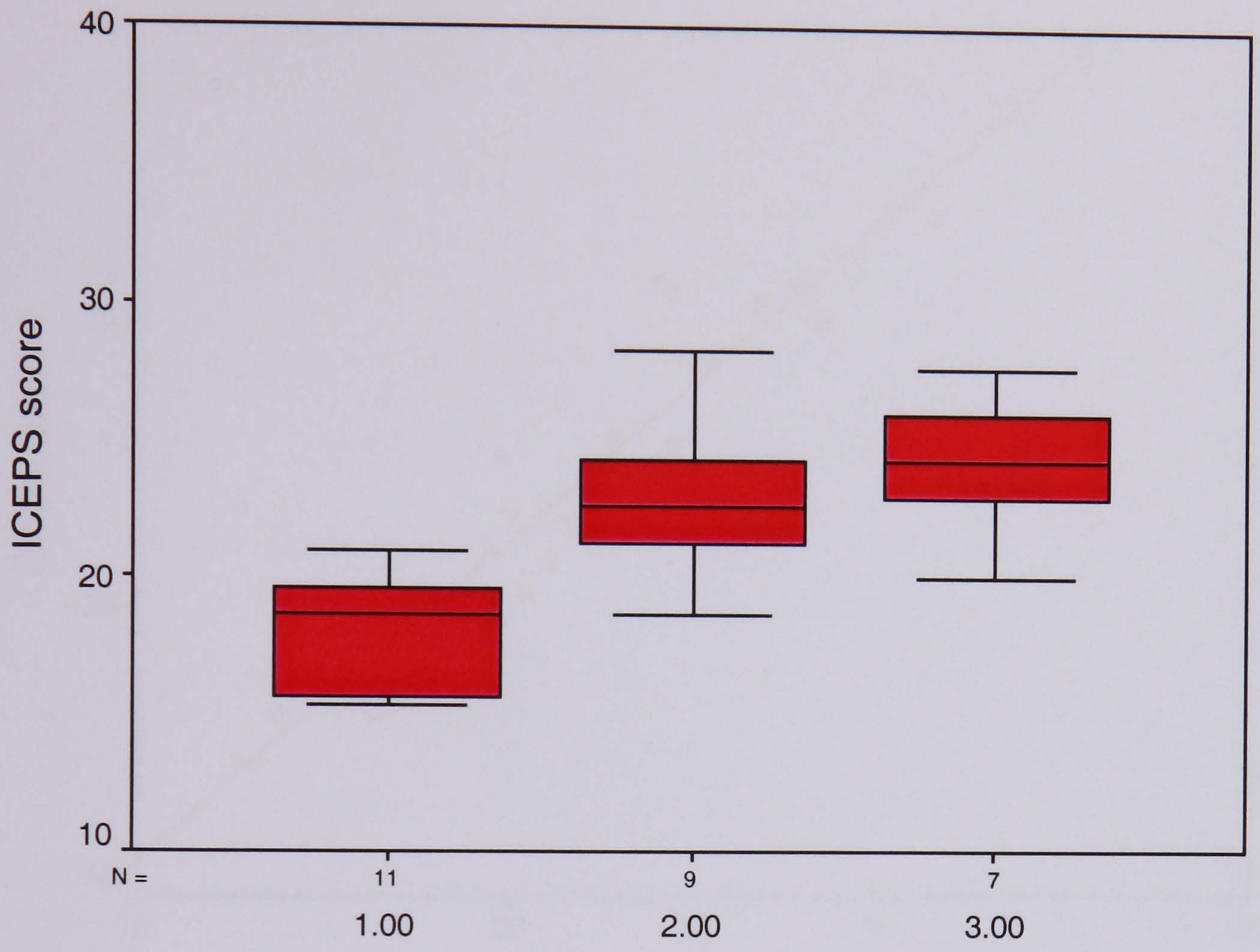


Fig 4E: ICEPS score for the three groups. There is a significant difference across the groups (p-0.002). Significant difference between 1 and 2 (p=0.01) but not between 2 and 3 (p=0.56)

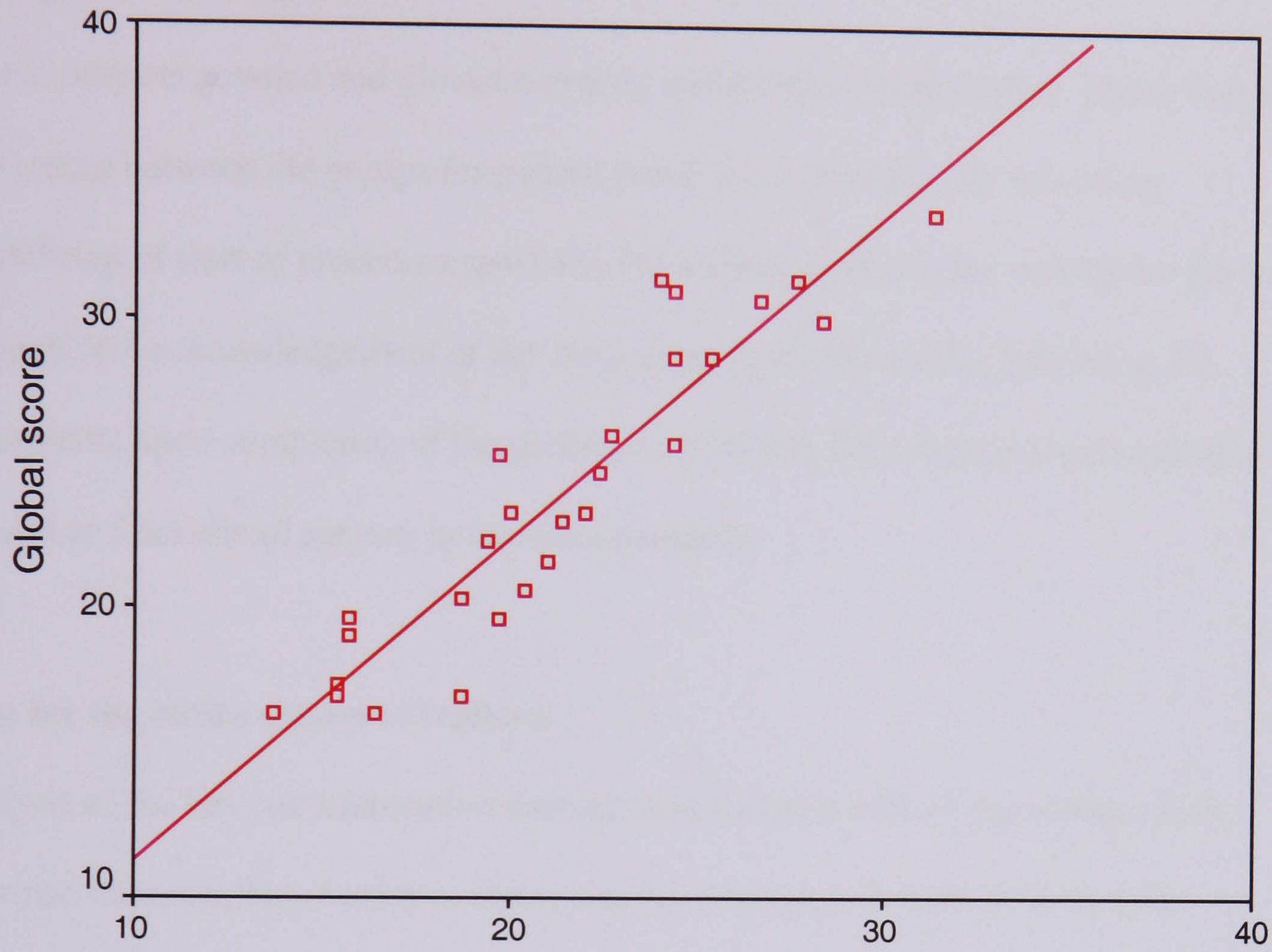


Fig 4F: Scatter plot of the global score vs ICEPS score. Correlation coefficient= 0.87. Global score on the 'Y' axis and the ICEPS score on the 'X' axis.

4.4.2 Essential items checklist

Inter-group differences

All the subjects gowned and gloved correctly and handled sharps safely. There were no differences between the groups for patient positioning (p=0.46), for informing anaesthetist of start of procedure (p=0.43), for asepsis (p=0.19), for waiting for the swab count (0.56), acknowledgement of the swab count (p=0.36) and for informing the anaesthetist upon completion of the procedure (p=0.50). Only 4 participants checked the consent and the site of surgery before the procedure.

Data for the entire cohort of trainees

Analysis of the key communication features revealed that 74% of the entire cohort informed the anaesthetist prior to taking the incision but only 21% informed the anaesthetist before proceeding to close the wound. Only 8% of them waited for a swab count from the scrub nurse before starting but 79% of them acknowledged the swab count when informed. 7.4% of them positioned the patient in the Trendelenburg position and 88% of them followed all aseptic precautions. Thus, poorest scores were observed for patient positioning, waiting for swab count and informing anaesthetist prior to closure.

4.4.3 Non-technical skills assessment

Inter-rater reliability for the non-technical assessment

The inter-rater reliability between the two independent observers was 0.84 (alpha).

Inter-group differences for the total non-technical skills score

There was no difference in the total non-technical global score across the three groups of trainees (p= 0.70). Fig 4G also shows the variability within the groups. Table 4D gives details of non-technical global scores for the three groups.

Inter-group differences for individual components

There were no significant differences across the groups for any of the components except for leadership (p=0.008). While there was a significant difference between groups 1 and 2 (p=0.003), the difference between groups 2 and 3 was not significant (p=0.54).

Data for the whole cohort

For the whole cohort of subjects the lowest scores were obtained for pre-operative preparation and vigilance. The scores, expressed as percentages, were 35.8% and 45.9% respectively. The score for leadership for the whole cohort were 56.6%. Table 4E gives details of the various components of the non-technical rating scale and the scores obtained for those components.

	Total Non-technical global score (max score=35)
Group 1	16 (7)
Group 2	16.5 (5.75)
Group 3	15 (8)

Table 4D: Non-technical skills score. Median values, value in parentheses- Inter-quartile range.

	Pre-op preparation (Max=15)	**Commu-nication (Max=5)**	**Vigilance (Max=10)**	**Leadership (Max=5)**
Group 1	5 (3)	3.5 (1)	4 (4)	2 (0.5)
Group 2	4.5 (2.75)	3 (0.5)	5 (1.75)	3 (1)
Group 3	4.5 (5)	4 (1.5)	4 (1.5)	3 (1.5)

Table 4E: Scores for the various components of the non-technical skills assessment

Max- maximum possible score. Median score, Value in parentheses- Inter-quartile range

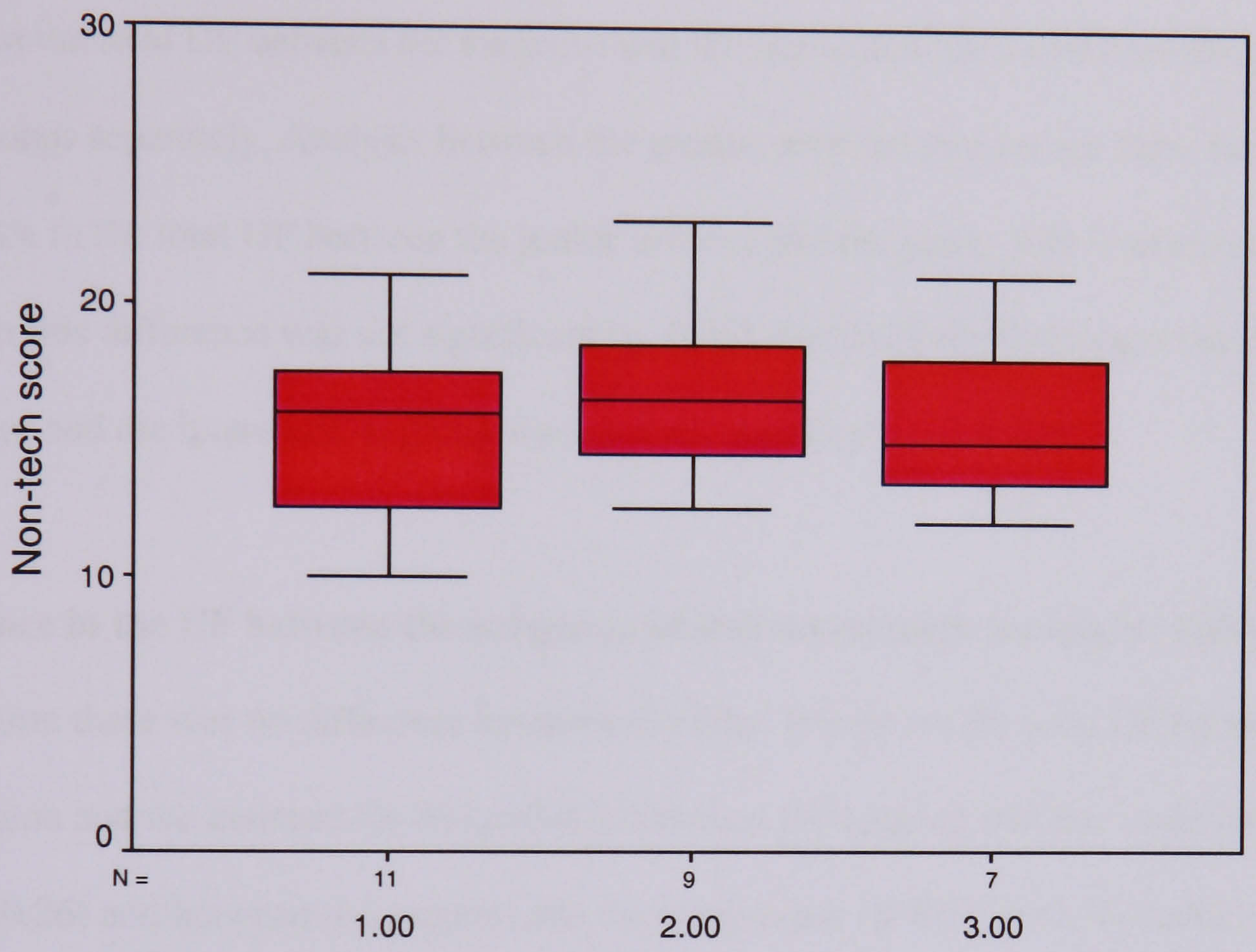


Fig 4G: Total Non-technical skills score- sum of the scores for the four skills. No difference across the groups (p=0.70)

4.4.4 Communication: 'utterance frequency' (UF)

Difference in the Total UF across the groups

There was no difference in the total UF across the three groups (p=0.20). Table 4F describes the total UF between the surgeons and the individual team members for the three groups separately. Analysis between the groups revealed that for the there was a difference in the total UF between the junior trainees and the group with intermediate skill, but this difference was not significant (p=0.06). Similarly the difference between the senior and the intermediate group was also not significant (p= 0.47).

Difference in the UF between the surgeon and individual team members (Table 4F)

In addition these was no difference between the three groups for the total UF between the surgeon and the assistant (S-A) (p=0.47), between the surgeon and the anaesthetist (S-An) (p=0.26) and between the surgeon and the scrub nurse (S-SN) (p=0.18) (table 4F). There were no differences between groups 1 and 2 for S-A (p=0.34), S-An (p=0.19) and S-SN (p=0.06) UF. There were also no differences between groups 2 and 3 for S-A (p=0.56), S-An (p=0.47) and S-SN (p=0.62) UF.

Data for the entire cohort (Fig 4H)

There was a significant difference in the UF between the surgeons and the individual team members with the UF being significantly lower for S-An as compared to S-A and S-SN (p<0.001). While there were significant differences between S-A and S-An (p<0.001) and S-SN and S-An (p<0.001), the difference between S-A and S-SN was not significant

(p=0.09). This means that the lowest communication occurs between the surgeons and the anaesthetists.

Surgeon-anaesthetist UF

The pre-scenario communication consisted of asking for the height of the table to be adjusted and for permission to start and was consistent within and across the groups. It can also be seen from the table (Table 4G) that a large part of the total communication between the surgeon and the anaesthetist was during the hypoxia scenario. In addition, there was a great variability in the UF during and after the scenario (Fig 4I).

There was moderate correlation between the UF and the vigilance score during (rho= 0.36, p=0.09) and after the hypoxia scenario (rho=0.46, p=0.03).

	Total UF	**S-A**	**S-An**	**S-SN**
Group 1	3.2 (1.8)	1.5 (0.9)	0.34 (0.36)	1.6 (0.63)
Group 2	4 (2.25)	1.27 (0.6)	0.41 (0.49)	1.9 (1.1)
Group 3	3.3 (2.6)	1.3 (0.9)	0.46 (0.61)	1.5 (1.5)

Table 4F: Utterance Frequency (UF) data for the three groups. Median values are shown. Values in parentheses are the inter-quartile range. S-A: surgeon- assistant; S-An- surgeon-anaesthetist; S-SN- surgeon- scrub nurse

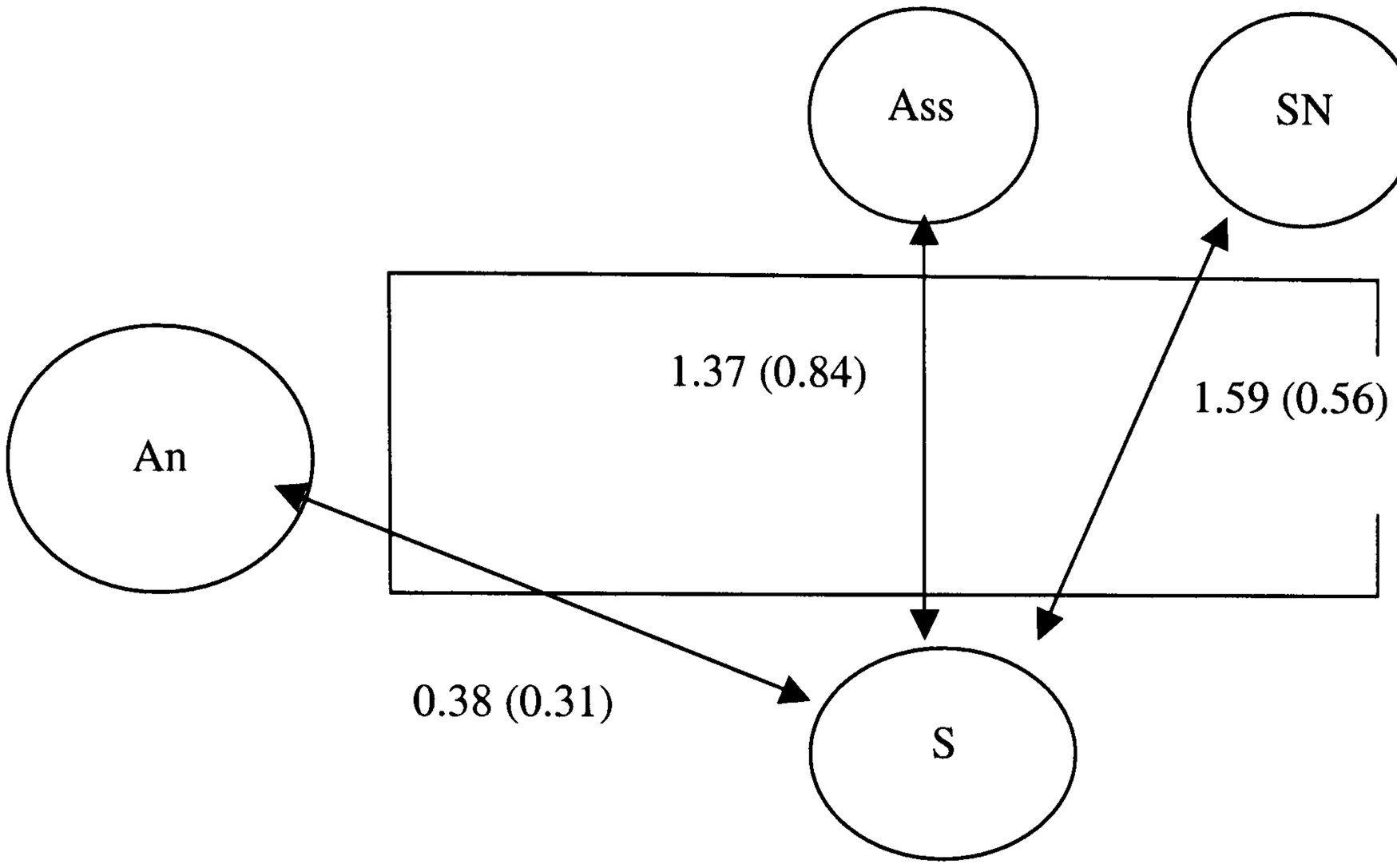


Fig 4H: Communication: 'utterance' frequency (UF) of the surgeon with other team members. S- Surgeon, An-Anaesthetist, A- Assistant, SN- Scrub Nurse

Pre-scenario	0.24 (0.37)
During the scenario	2 (2.27)
Post-scenario	0.15 (0.69)

Table 4G: Utterance frequency: surgeon-anaesthetist (S-An)

Median values are shown in the table; values in parentheses are the inter-quartile range

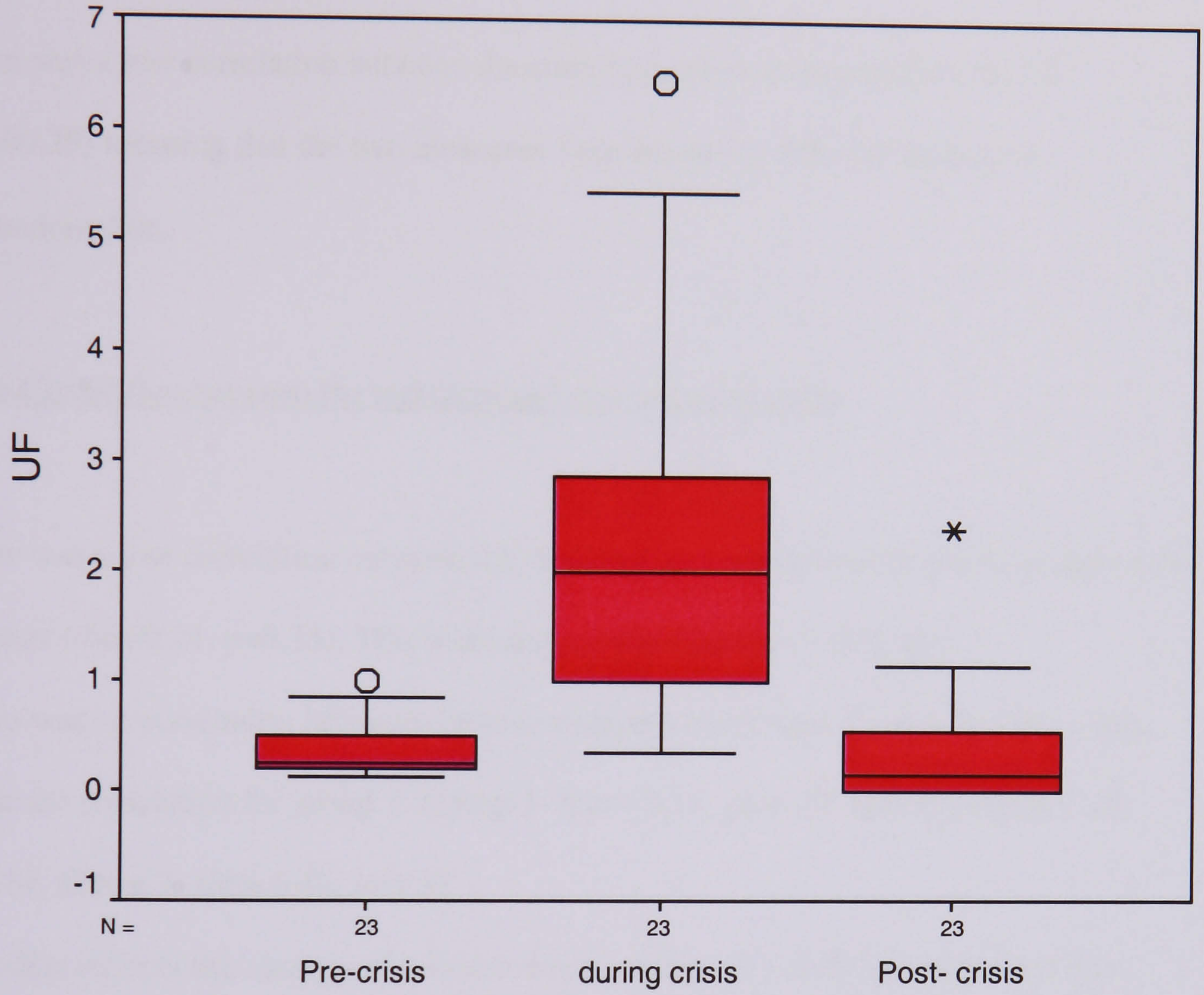


Fig 4I: UF between the surgeon and the anaesthetist- before, during and after the hypoxia scenario. Nearly half the S-An UF was during the crsis. From this figure it can be seen that there was a great variability in the UF during and after the crisis. 'o' is the outlier and 'x' is the extreme observation

4.4.5 Correlation between the UF and the communication score in the non-technical assessment

There was a low correlation between the communication score and the total UF (rho=0.29) meaning that the two measures were assessing different aspects of communication.

4.4.6 Correlation between the technical and non-technical skills

There was a low correlation between the technical and non-technical global scores of the subjects (rho=0.24, p=0.23). This is evident on the scatter plot (Fig 4J).

There was no correlation between the two measures for groups 1 and 2, but there was moderate correlation for group 3. Group 1- rho= -0.13, p= 0.70; Group 2- rho= 0.29, p=0.44; Group 3- rho= 0.44, p=0.32

This data reflects the absence of a link between a surgeon's technical skills and non-technical skills during the performance of a simulated procedure. The correlation is lowest for the junior groups and increases with increasing experience. Looking at the individual scores from Table 4C and 4D and Fig 4J, this can be explained by the fact that the junior trainees scored low for technical skills but some of them scored quite high in their non-technical skills resulting in a poor correlation between the two measures.

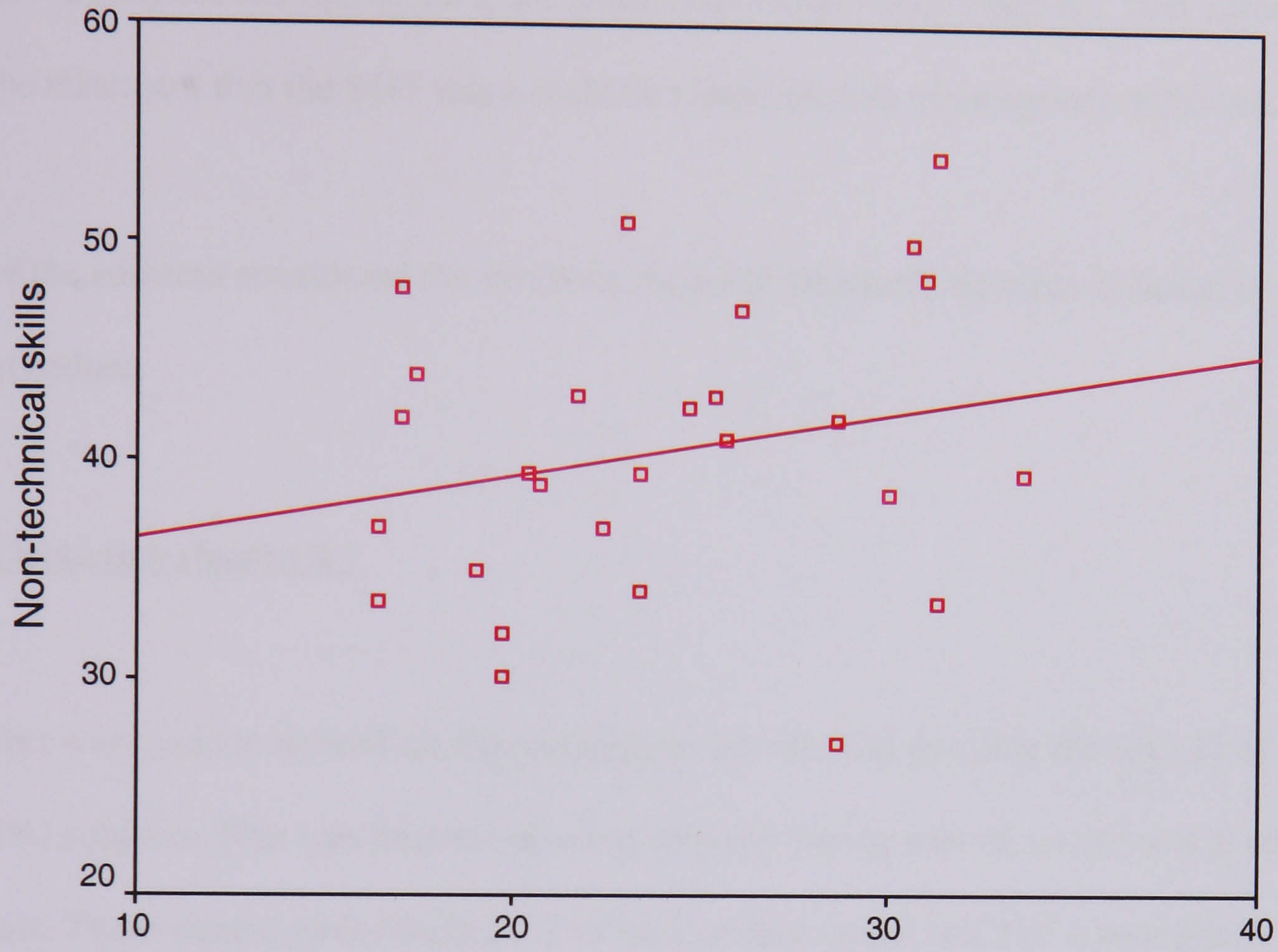


Fig 4J: Correlation between technical and non-technical scores.

Correlation coefficient- 0.24.

4.4.7 Face validity of the simulation.

90% of the subjects felt immersed in the simulation (suspended disbelief). They agreed with the statement that the SOT was a realistic representation of an operating theatre.

50% of the subjects considered the synthetic model to be a realistic representation of a real procedure.

4.4.8 Debriefing (feedback):

An effort was made to debrief all the participants but this was possible for only 17 of the 27 (63%) subjects. This was because of some subjects having moved on to posts at other hospitals. There were 6 junior trainees, 8 of intermediate grade and 3 of senior grade. All sessions were within 8 weeks of simulations.

- 80% of the trainees considered the model to be a good method of training technical skills. The junior trainees (100%) agreed more with this statement as compared to the intermediate (75%) and the senior trainees (65%).
- 88% of the trainees considered the SOT to be a good environment for training of team skills. 3 trainees were not sure.
- 60% of the subjects considered pre-operative instrument check to be unnecessary for a routine procedure such as sapheno-femoral junction high-tie. They felt that they were right to assume that there was a set of

dedicated instruments for the procedure which should have been available to them. All of them considered the pre-operative instrument check to be the responsibility of the scrub nurse.

- 75% of the subjects considered pre-operative briefing with the rest of the team to be inessential for routine/ 'minor' procedures such as sapheno-femoral junction high-tie.

4.4.9 Summary of results

Although 90% of the trainees considered the SOT to be a realistic representation of a real operating theatre, only 50% considered the model to be a realistic representation of a real procedure. In spite of the low rating for the face validity of the model, 80% still considered the model to be good method of technical skills training. 88% of the entire cohort considered the SOT environment to be suitable for training in team skills. There were two methods used for the assessment of technical skills. One was a global rating scale of generic skills and the other was a procedure specific assessment. There was a significant difference across the three trainee grades and a strong correlation between the two measures. There was also significant difference between the junior trainees and the intermediate group but not between the latter and the seniors.

In terms of the non-technical skills assessment, the essential items check list and the utterance frequency there were no differences between the three grades of trainees. A low score was achieved by the entire cohort for pre-operative preparation and vigilance. There was a low correlation between the technical skills and the non-technical skills.

4.5 DISCUSSION:

Due to the generic nature of the global rating scale and its resultant inability at assessing procedural details, we developed a procedure specific global rating scale, ICEPS. The strong correlation between the generic and the procedure specific scales may indicate the fact that the use of both scales is unnecessary. However, it is essential to understand that they both assess different aspects of procedural skill. Together, they possess the potential to make formative assessment of technical skills more complete.

Using both these methods it was demonstrated that there was significant difference across the grades but no difference between middle and senior level trainees. Other studies have commented on the existence of this ceiling effect while performing on bench models in skill laboratories (109). In this study, it was hypothesized that the two groups would be at different stages of learning of psychomotor skills and as a result even though the difference between the two groups is not apparent in the setting of a skills laboratory, the enhanced realism of the SOT would compromise the performance of the middle level trainees. The reasoning behind this hypothesis was based on the structure of psychomotor skills learning which has been described to occur in three phases (4;18). The first is the *phase of cognition*, where the trainee learns the steps involved in the execution of a task. The next is *the phase of association*, where knowledge of the task is integrated with psychomotor skills and lastly the *phase of automation,* where knowledge and technical skills are 'hotwired' and the task is performed at a subconscious level (4). It was believed that trainees from group 2 would have been in the later stages of the stage of association, while those in group 3 would have been in the stage of automation as result of which the

effect of performing in the SOT would have had a negative effect on the performance of group 2 trainees.

The reasons for the demonstration of this ceiling effect could be manifold. It could be that the assumption was wrong and that even group 2 trainees were in the stage of automation. It is also possible that the global rating scales are not sensitive enough in differentiating the differences in performance. The use of the dexterity analysis system, ICSAD, was considered but data on dexterity from the current version of ICSAD is a cumulative score for the entire task/ procedure without the ability to gather data on different stages of the procedure. Thus, because it was likely that the data would have been confounded by the time spent while waiting for instruments and for the anaesthetic (hypoxia) crisis, it was decided not to use ICSAD for this study.

One of the reasons for the apparent failure to make non-technical skills a focus of training among medical specialities has been the difficulty encountered in the objective assessment of these skills to enable structured feedback. This study piloted the use of a simple non-technical skills assessment scale, which consisted of a few measures considered important for team performance in the OR. The aim of this study was to assess the feasibility of the assessment of non-technical skills in the complex operating theatre environment, evaluate the inter-observer reliability, explore any difference between different grades of trainees and also use these measures to observe non-technical skills in trainee surgeons.

The use of the CDR with its video and audio recording facility aids the assessment of inter-personnel and communication skills. While two cameras were used to evaluate the surgeon's technical skills, two other cameras were used to observe interaction between

the team members. The non-technical skills assessment used in this study was found to have a high inter-observer reliability. This is probably because there were only a few measures on the scale, resulting in a lower likelihood of variability between the two assessors. The benefit of having only a few measures in a non-technical skills assessment has been suggested by Gaba et al who reasoned that one of the reasons for the low inter-rater reliability in their observation study of anaesthesia simulations, was the presence of multiple measures (138).

The non-technical skills assessment revealed that there were no differences between the three grades for all measures except for leadership skills. The senior trainees scored higher than their junior counterparts. Data for the entire cohort revealed that most trainees scored very low for vigilance. 14 trainees did not notice the anaesthetist's handling the hypoxia crisis and had to be prompted by the anaesthetist. It has been found that a large number of adverse events during the maintenance of anaesthesia are due to problems with the vigilance of the anaesthetist (120) and situation awareness training is considered to be an important aspect of anaesthesia simulations (173). This particular condition, however, has never before been studied amongst surgeons. The lack of awareness of the patient's condition could be either because most surgeons focus solely on completing their task (123) or may be a reflection of a cultural norm where they rely on information on the patient's condition from the anaesthetist. This needs to be explored by further studies which should examine the importance of the surgeon's awareness on patient safety in the OR. Error management in aviation places a large emphasis on the fact that team members maintain open lines of communication and cross-check each other (132). One observational study of team co-ordination emphasized the role of 'activity monitoring'

where one team member watches another complete their tasks. However even, just observing was found to be problematic and the authors recommended the need for explicit verbal communication (174).

In addition to the assessment of communication as a single measure on a 5- point Likert scale in the non-technical assessment, a communication 'utterance' frequency (UF) was also used. The utterance frequency, based on a similar method used by a previous study (171), was used because of its objectivity and simplicity and also because it helps in measuring the extent of communication with other team members.

It was found that there were no differences between the different grades for the total UF and for the UF between surgeons and individual team members. However, general observational data for the entire cohort revealed some interesting findings. It was observed that communication between the surgeons and anaesthetists was significantly lower than that with other team members. This is not a surprising finding considering the fact that a majority of the communication between the surgeon and the assistant and the scrub nurse consisted of instructions or requests for instruments. It was also observed that a large component of the communication between the surgeons and the anaesthetists was during the hypoxia scenario. Of particular interest was the finding that with a third of the study cohort, there was no communication between the surgeon and the anaesthetist in the post-scenario period suggesting the lack of vigilance of the patient's condition even after the crisis. There was also a moderate and significant correlation between the surgeon-anaesthetist (S-An) UF after the hypoxia scenario and the vigilance score. This reflects the fact that those who had a high S-An utterance frequency also scored high for vigilance, indirectly validating the vigilance score with a more objective measure.

Communication was also assessed for some key elements of communication between team members that is considered to be essential. Though there were no differences between the three groups, observations revealed that while the subjects were consistent in asking the anaesthetist about the appropriateness to start, only a quarter of them informed the anaesthetist about the completion of the procedure. This was considered to be an essential part of the communication between the surgeon and the anaesthetist, as it prepares the anaesthetist to start instituting measures to reverse the state of anaesthesia at the optimum time. It was also observed that surgeons proceeded to closing the surgical wound without the swab and instrument count being delivered to them. In both these instances it maybe be assumed that activity monitoring between the team members would be adequate but Xiao et al have advocated the need for training in communication to make 'certain verbalisations mandatory' for smooth team coordination (175).

The scores for pre-operative preparation were low for all three groups. This is a reflection of the absence of pre-operative briefing, absence of pre-operative instrument check and the inadequacy of introduction to the other team members. The absence of briefings in operating theatres has been noted by earlier studies (111). Using a 4 point scale and comparing communication in the cockpit with that in the OR, Sexton et al found that while below standard briefings occurred in only 23% of cockpit observations, they occurred in nearly 90% of OR observations (129). This was in spite of the fact that in a questionnaire to OR personnel (ORMAQ) a majority of respondents considered pre-operative discussion between team members to be important to enhance team working in the OR (127).

A majority of the study's participants believed that the scrub nurse should be responsible for the pre-operative instrument check. Most operating theatres in the UK maintain a list of instruments specific for each procedure and for each surgeon depending on their preferences. It is probably as a result of this that most of the participants considered a pre-operative check to be unnecessary. However, a majority of them mentioned that they usually do carry out a check when they start a post at a new hospital till they became acquainted with the local system and many of the subjects agreed that it should be essential before every major procedure.

A number of subjects also felt that an instrument check was the responsibility of the scrub nurse. While one senior trainee commented that he would be justified in blaming the scrub nurse for the missing instruments, one middle level trainee said he would regularly check the presence of some essential instruments such as self-retaining retractors and suture materials before every procedure. Thus, though a majority of subjects in this study (60%) considered an instrument check to be either unnecessary or the scrub nurse's responsibility there does not seem to be much consensus amongst surgeons regarding the importance or the need for this particular feature of the procedure.

In summary, the simulation revealed certain cultural norms such as the fact that briefings are absent and the pre-operative instrument check is usually considered to be the responsibility of the scrub nurse. The face validity of the simulated environment was found to be quite high as nearly 90% of the trainees felt immersed in the simulation. However the face validity of the model was satisfactory to only 50% of the cohort. In spite of this, 80% of the trainees considered the model suitable for technical skills training. Nearly 90% of the cohort considered the SOT to be suitable environment for

team skills training. The simulation also revealed the need to improve the vigilance of surgeons and the communication between the surgeons and the anaesthetists.

CHAPTER 5: ASSESSMENT OF CRISIS MANAGEMENT AMONG SURGICAL TRAINEES

5.1 INTRODUCTION

Chapter 4 described the use of the Simulated Operating Theatre (SOT) for the assessment of technical proficiency and some aspects of non-technical skills. The assessment was conducted in a 'quiet', low workload environment, where task performance and team interaction were the only demands placed on the surgeon's abilities. However, just as in other high reliability professions such as aviation and anaesthesia, these relatively 'quiet' periods are often interspersed with high 'workload' or crisis situations (112) which place an additional demand on the practitioner's competence.

The incidence of intra-operative complications such as damage to major vascular structures and organs during surgery varies, but it is generally considered that most of these crisis events are rare. The incidence of iatrogenic trauma depends to a large extent on the procedure, the proximity of major organs and the expertise of the surgeon. The incidence of damage to major vascular structures and organs increases with increasing procedural complexity. While the incidence of major vascular damage is 10% with resections of the pancreatic head (176) and the incidence of splenic trauma, 5- 20% with major GI procedures (177), the incidence of femoral vein damage during varicose veins surgery is less than 1% (178). While surgical training in these procedures emphasises the importance of avoiding such major complications, trainees are seldom taught how to manage these complications when they occur. As successful management depends on the

crisis management skills of the surgeon and the surgeon's experience, outcomes can often be varied. In addition the management of these situations is based on anecdotal experience, from watching a mentor manage a similar situation, or from one's own personal experience where the surgeon may even have learnt by making mistakes. Such learning is thus random and is based on the need to have faced similar situations previously. However, in comparison to anaesthesia, where there are often protocols for the management of even rare crisis conditions, there are very few situations in surgery with fixed management plans. 'Learning by making a mistake' is a common adage in surgical practise. Unfortunately such a learning paradigm is often at the cost of patient safety.

Simulations offer the advantage of allowing a trainee to learn the consequences of a mistake and acquire the knowledge to manage a crisis without any harm to a patient or a planeload of passengers, with the benefit of performance feedback (114). In aviation, simulator based training or Line Oriented Flight Training (LOFT) can be used to train flight crews to manage crisis scenarios making best use of the available resources. This is one of the key aspects of Crew Resource Management (CRM) training (179). Repeated practise in these simulator-based scenarios equips the pilot and the flight crew with the knowledge and the skills to manage these situations when they occur in real life. Anaesthesia CRM (ACRM) (134) is based on a similar principle and allows anaesthetists to learn the skills of dealing with both common and rare crisis scenarios. The emphasis of both CRM and ACRM training in not only on technical skills but also on team skills such as communication, vigilance and decision making (138). The aim is to train the

individual in team skills as well as train teams to work together in the most effective manner.

CRM training, in its present form, has error management as its primary focus (110). Helmreich has described error management as an error troika. This consists of *error avoidance* where safety principles are used to prevent the occurrence of errors, *error trapping* where errors are trapped before they occur often by vigilance and team monitoring and *error mitigation* where errors are rapidly diagnosed and effectively managed before they give rise to drastic consequences. CRM training in its present generation (sixth) 'stresses the fact that effective error management is the hallmark of effective crew performance and that well managed errors are the indicators of effective performance' (110).

From a surgical context, de Leval et al have suggested that error recovery strategies are just as important as error prevention measures (122). Their observations of major and minor errors during a paediatric cardiac procedure indicate that ' the surgeon's diagnostic skill, knowledge of the various surgical strategies to correct a problem, and communication with the rest of the team are important pre-requisites of error compensation'. They described error compensation as a form of error recovery whereby 'a strategy to remedy the situation is implemented before negative consequences ensue' (122). Thus, just as in aviation, error management is a crucial aspect of surgical performance. Surgeons and OR teams that are superior in error recovery are more likely to have better patient outcomes.

In addition to training in error management strategies and offering individuals the opportunity to practise rarely occurring crises, using objective assessment measures,

simulation- based training offers participants objective feedback of their performance on completion of a scenario (114). Assessment of performance during ACRM has consisted of using measures such as time taken to recognise and solve a problem (140;180), deviations from protocol and scoring systems (141). The assessment of non-technical skills during anaesthesia simulations has recently become a major focus. Gaba et al modified the LOSA checklist developed by NASA and the University of Texas Human Factors Project for the assessment of flight crews. 10 crisis management behavioural markers: orientation to case, inquiry/ assertion, communication, feedback, leadership, group climate, anticipation/planning, workload distribution, vigilance and re-evaluation. These were rated on a 5- point ordinal scale as follows. 1=poor performance; 2= substandard (minimally acceptable) performance; 3= standard performance; 4= good performance and outstanding performance (138). Fletcher et al developed and validated an Anaesthesia Non-Technical Skills (ANTS) assessment tool which consists of 4 categories and their elements. These categories were Task management, Team working, Situation Awareness and Decision Making (143).

The development of anaesthesia simulators has made ACRM courses feasible. Since the development of the first anaesthetic simulator- Sim One (181), further technological advances have resulted in the development of higher fidelity simulators. There are now simulators that provide physiological and pharmacological responses to drugs and pre-programmed scenarios (182). In comparison, it is difficult to simulate surgical procedures and more importantly simulate surgical crises. VR simulators are presently task trainers and many synthetic models lack tissue realism. Animal models may provide the opportunity to develop crisis situations such as vascular trauma, but there would be

considerable ethical problems and their use is banned in the UK. However, with increasing interest in the use of synthetic models for the training and assessment of technical skills (1), some groups across the world have taken the initiative to work closely with their industry partners to enhance the realism of such models. This chapter describes the development of a surgical crisis simulation using a synthetic model and the assessment of its management from both a technical and a non-technical perspective.

5.2 AIMS OF THE STUDY

- Establish the face validity of a crisis simulation for surgical trainees
- Establish the construct validity (difference between groups) and reliability of the measures used in the assessment of performance.
- Assessment of the participants' experience of the simulation

5.3 MATERIALS AND METHODS

5.3.1 The subjects

There were 16 surgeons who were divided into two groups:

Group 1(Senior trainees): performed > 50 saphenofemoral high tie procedures (n=8)

Group 2 (Junior trainees): performed 20-50 procedures (n=8)

Exclusion criteria- trainees who had performed less than 20 procedures were excluded from the study as it was assumed that they would not possess the skills or the knowledge to handle the bleeding scenario.

5.3.2 Study design

The two groups of trainees were assessed during the performance of a sapheno-femoral junction high-tie procedure on the synthetic bench model used in the previous study. The simulated procedure was divided into two phases. Phase 1 was the quiet phase before the bleeding was started and phase 2 was the surgical crisis (bleeding phase). The two groups of trainees were compared for their technical and non-technical skills using various measures described later.

5.3.3 The Simulated Operating Theatre (SOT), described in the previous study was used as the simulated environment for this study as well.

The model of the sapheno-femoral junction was further modified after the moderate response it had received from the participants of the previous study.

The following changes were made:

- The simulated vessels were made of silicon and contained a lumen into which simulated blood could be instilled.
- The position of the tributaries within the model in relation of the saphenous vein, the femoral vein and in relation to each other were improved to enhance anatomical simulation.
- The subcutaneous tissue was made a bit more realistic by altering the properties of the silicon used.

The Simulation

The simulation scenario descried to the participants was similar to the one described in the previous study (Appendix 5). The only changes made to the scenario were that the participants gloved and gowned in the control room prior to entering the SOT and there was a crisis built into the scenario (bleeding). The hypoxic event was maintained to enable assessment of vigilance/ situation awareness of the surgeons. In addition the bleeding scenario was the prominent feature of this study.

The Bleeding scenario (Fig 5A)

A 5 mm laceration was made in the medial aspect of the 'femoral vein' in the model, which was placed over the right 'groin' of the mannequin (anaesthetic simulator). The 'femoral vein' was then connected to a tube which ran along the walls and the floor of the

SOT which was concealed from the surgeon's view. This tubing was then connected to an intravenous line, which was in turn connected to a saline bag filled with simulated blood (Annexe Art, UK).

At a standardised point during the procedure, the simulation controller started the bleeding by opening up the IV tubing connected to saline bag. The pressure on the bag was strictly controlled and maintained at 20 mm Hg with the aid of a level one rapid transfuser (Level 1 Technologies, UK). The bleeding was maintained till the following factors gave the impression that the laceration had been effectively closed:

- The slowing down of the simulated blood appearing through the surgical wound
- The assessment of control by viewing the procedure on the monitor placed in the control room and concurrence with the assistant in the SOT. This was by means of a signal between the simulation controller and the assistant.

The simulation was stopped under the following circumstances:

- Bleeding effectively controlled
- Trainee called a senior surgeon/ vascular surgeon due to inability to control the bleeding or refusal to do so due to lack of confidence.

Feedback

The participants received an objective, video-based and criteria-referenced feedback of their performance either soon after the simulation or within 2 weeks. The technical feedback was provided by the research fellow with reference to the opinion of 5 experts (vascular surgeons) who were consulted for the development of the assessment measures. The experts were asked to highlight the various measures that they believed that a trainee

should take for the successful management of a femoral vein laceration. A human factors researcher provided the non-technical feedback.

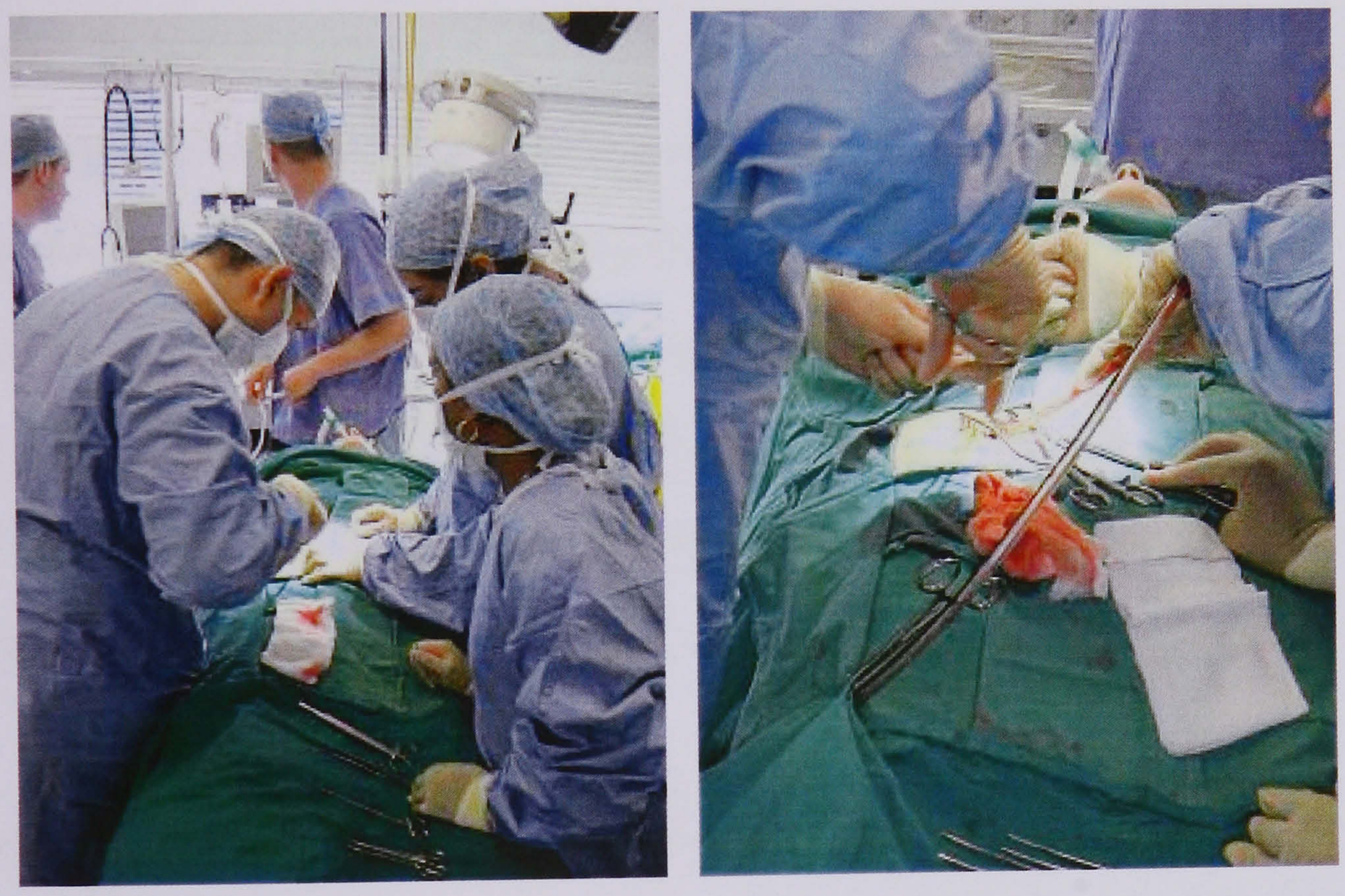


Fig 5A: Simulated operating theatre with team controlling bleeding from the femoral vein

5.3.4 Assessment of performance

In addition to assessing performance during the simulations for the purpose of the feedback, all the sessions were recorded onto DVD and played back to the assessors. The data from these assessments were then used as the study data.

1. Generic Technical skills

5 components of the global rating scale (Appendix 2) as developed by Reznick et al (84) were used for the assessment of generic technical skills during Phase 1 and 2. These were rated on a 5-point scale. This consisted of tissue respect, time and motion, instrument handling, suture handling and use of assistants/ retractors. Two surgeons instructed in the use of the global rating scale carried out this assessment. The scores were expressed as a percentage.

2. Technical ability to control the bleeding/ repair the laceration

A global rating scale, based on the global rating scale developed by Reznick et al for the assessment of generic technical skills, was developed for the assessment of the ability of the participants to control the bleeding and repair the laceration. The 7 categories on this scale were: identification of source of bleeding, exposure, control of bleeding, clear vision of laceration, technique of repair, knowledge of instruments and over-all performance. The scale was developed after consulting the 5 experts who were asked to identify all the measures required for the successful management of a femoral vein laceration. Following the identification of the key steps which formed the basis of the 7

categories, the scale was shown to the experts and suggested modifications were incorporated to develop the final version of the scale (Appendix 8). Two blinded observers performed this assessment.

3. Non-technical skills

The NOTECHS rating scale (Appendix 3) developed by the JARTEL project was modified for the assessment of the non-technical skills of the surgeons (Appendix 9). The scale consisted of 4 categories for the assessment of non-technical skills during Phase 1 and there was an additional category for the assessment of decision making/ crisis handling (DM) during Phase 2. Each of these categories consisted of 3-5 elements. Some of them were anchored by examples to facilitate the assessment process. The elements were rated on a 6- point scale. Thus the maximum possible score for the 15 elements which could be consistently rated was 90. A human factors researcher and the research fellow who was trained by the former performed the assessment. The training and calibration consisted of the first 5 simulations being observed together. The other simulations were all assessed independently to facilitate evaluation of inter-rater reliability.

4. Communication count and 'utterance' frequency (UF)

As described in the chapter 4 a communication utterance frequency was carried out. The utterance frequency was assessed separately for Phases 1 and 2. The communication during the hypoxia scenario was excluded from the analysis as the aim was only to detect any change in UF during the bleeding crisis.

5. Objective time- measures of crisis management

These were time taken to diagnose the cause for the bleeding as the femoral vein laceration (TD), time taken to informing the team (TT), time taken to calling the consultant/ help (TH), time to use appropriate measures to control bleeding (TC) successfully (use of pressure with finger on either side of the laceration or vascular clamps) and time to close the laceration (TL).

Time taken to calling for help was converted to a score of 1 or 0. From the discussions with the 5 experts it emerged that all junior trainees should call the consultant immediately on detection of the femoral vein laceration. On the other hand they believed that senior trainees could proceed to close the laceration but should call for help if the blood loss was greater than 500 ml, if the trainee had attempted closure once without success and if the there was any concern about the 'patient's' condition. Thus for the junior trainees, a score of 1 was given only if help was requested (calling a consultant) on diagnosis of the laceration. If this was not done or done late they scored 0. Senior trainees who did not call the consultant when necessary (as mentioned above) scored 0.

6. Surrogate outcome measures.

The various assessment measures mentioned previously can be seen as process measures.

In the absence of any definite outcome measures, two surrogate outcome measures namely blood loss and leak rate ratio were used.

a. Blood loss:

The total blood loss was calculated by measuring the blood in the suction apparatus and by weighing the swabs.

b. Leak rate ratio (LAR):

The femoral veins were fixed to a special apparatus and a 50:50 solution of glycerol: normal saline was run through the femoral vein at a pressure of 20mm Hg which was set with the help of a pressure transducer. The leak rate (fluid lost per minute) of each repair was divided by the ideal leak rate to calculate a leak rate ratio (LAR). The denominator (ideal leak rate) was the leak rate of a repair performed by an expert vascular surgeon under conditions as ideal as possible. This was with the vein was removed from its silicon cast and placed on a jig and held to the jig with clips. The expert performed two repairs and the average leak rate from the two repairs was assumed to be the ideal leak rate. The repairs were performed with a double ended 5-0 polypropylene in a continuous manner.

Assessment of participants perception:

A questionnaire was designed to ask the participants about their simulation experience. The participants were presented with 11 statements (Appendix 10) designed to evaluate the face validity (realism) of the model, the simulation environment and the bleeding, their perception of the value of the simulation as a training/ assessment exercise and their perceived benefit of the feedback. They were asked to agree or disagree with the statements on a 6- point Likert scale. Scores between 1-3 were considered to be

disagreements, and those between 4-6 to be agreements. The scores of the junior and senior trainees were also analysed separately.

5.3.5 Data Analysis

As the data was not found to have normal distribution for most measures and as the groups were quite small, non-parametric tests have been used for analysis. Fisher's test was used to analyse the differences between the questionnaire results of the two groups of trainees (seniors and juniors) and analysis of the difference between the two groups for calling the consultant. Mann-Whitney 'U' test was used to test the differences between the two groups for the various measures. Inter-rater reliability between the observers was analysed using Cronbach's alpha co-efficient. Cronbach's alpha was also used to assess the internal consistency of the elements in the non-technical skills rating scale.

The differences between Phase 1 and Phase 2 for the generic technical skills score, non-technical skills score and UF were compared using the Wilcoxon Sign Ranks test. Correlation between the measures was done using Spearman's Rank correlation co-efficient.

5.4 RESULTS

5.4.1: VALIDITY AND RELIABILITY

5.4.1.1 Face validity of the crisis simulation: questionnaire results

Table 5A gives details of the median scores for the questionnaire categorised as face validity (realism), usefulness for training, usefulness for assessment and the benefit of feedback. There were consistently high levels of agreement for all the questions. A score of 4 or more was taken to denote agreement with the various statements.

A majority of the participants found the model, the simulated operating theatre environment and the bleeding scenario to be realistic. 93.3% of the participants found that simulation suitable for the training of technical skills and 80% of them found the simulation suitable for team skills training. Greater than 80% of them considered the crisis simulation suitable for the assessment of their technical and team skills. All the participants found the feedback useful. When the scores for the individual questions were compared between the two groups, there were no significant differences for any of the questions except for the fact that junior trainees found the technical skills feedback more useful than the senior trainees ($p=0.04$).

5.4.1.2 Construct validity of the assessment measures (differences between the groups) and variability within the two groups

Technical ability to control the bleeding

There was a significant difference between the two groups. The senior trainees scored significantly higher than the juniors (median +/- inter-quartile range-24 +/- 5.9 vs 17.7 +/- 6; p=0.008). From Figure 5B it can be seen that there was considerable variability within both the groups.

Non-technical skills:

Table 5B compares the non-tech skills of the two groups during phase 2. The only difference of note between the two groups was for UF. There was a moderate correlation between the level of experience and UF (rho=0.45) as compared to weak correlations for all the other non-technical skills. Examining Figure 5C, reveals that there was a large variability within the groups for non-technical skills. The variability within group 1 was particularly high for decision- making and UF (Fig 5D and 5E).

Crisis Management- time measures (Table 5C)

There was a significant difference between the two groups for TD (p=0.01), TC (p=0.008), TL (p=0.01) and TH (p=0.007). The difference was not significant for TT (p=0.48). From Figures (5F i-iv) it can be seen that there was considerable variability within group 2 for all these measures. There was considerable variability within group 1 for informing the team (after diagnosis of the laceration).

Surrogate outcome measures:

Even though there was a greater blood loss associated with group 2 as compared to group 1 (median +/- interquartile range- 500 +/- 425 vs 300 +/- 100) this was not significant (p=0.08). There was also a higher Leak rate ratio (LAR) associated with group 2 as compared to group 1 (median +/- interquartile range- 4.46 +/- 13.6 vs 1.6+/- 1.6), but again this was not significant (p=0.25). There was considerable variability especially within group 2, for these two measures (Fig 5G i-ii).

5.4.1.3 Reliability of the assessment measures

Internal consistency (reliability) of the non-technical skills asessment.

The alpha coefficient between the 4 categories on the non-technical assessment scale was 0.87. The internal consistency between the 4 categories and the decision making (DM) score was 0.90. This means that the four categories were related and were measuring the same general construct- non technical skills.

Inter-rater reliability:

High levels of agreement between the different observers were achieved. The alpha coefficient (inter-rater reliability) between the two surgeons observing technical performance was 0.92 for technical ability to control the bleeding and was 0.84 for the generic technical skills assessment. The inter-rater reliability for the non-tech skills was 0.90 for the first five simulations that were assessed together for calibration purposes and 0.84 for the others. Inter- rater reliability for decision- making was 0.86.

Category	Question	Median score Max score=6	Agreements (%)
Face validity	The model is a realistic representation of the real procedure	4 (1)	93
	The SOT is a realistic representation of a real operating theatre	5 (0)	100
	The bleeding is a realistic simulation	4 (1)	86.6
	I would behave in the same way even in real life	5 (2)	100
Use for training	The crisis simulation is a good training opportunity for training technical skills	5 (1)	93.3
	The crisis simulation is a good training opportunity for training team skills	5 (1)	80
Use for assessment	The simulation is a good method to assess my technical skills	4 (1)	80
	The simulation is a good method to assess my team skills	5 (0)	93.3
Repeat the simulation	I would benefit by repeating the simulation again	5 (2)	93.3
Feedback	I found that technical skills feedback useful	5 (2)	100
	I found the team skills feedback useful	6 (1)	100

Table 5A: Questionnaire results- The table describes the median values of the answers for the questions. The values in parentheses are the inter-quartile range. The last column describes the percentage of subjects who agreed (score 4-6 on the 6-point scale) with the statement.

	Group 1	Group 2	p
Non-technical score (total)	65.2 (25.6)	58.75 (30)	0.46
Communication	16 (8)	14.5 (8.75)	0.91
Vigilance	12.75 (6.6)	10.5 (8.25)	0.34
Team skills	15 (7)	15.75 (10)	0.83
Leadership	19.75 (2.6)	18.75 (3.4)	0.19
DM score	21.25 (6.4)	19.25 (6.5)	0.37
UF	4.8 (4.0)	3.07 (1.52)	0.07

Table 5B: Non-tech skills, DM and UF- Table describes the median values of the scores for non-tech skills, DM and UF. The values in parentheses are the inter-quartile range.

The last column is the level of significance using Mann-Whitney 'U' test to compare the two groups.

	Group 1	Group 2	p
Time to diagnose (TD)(sec)	95 (113.75)	207.5 (167.5)	0.001
Time to inform (TT)(sec)	220 (272.5)	292.5 (353.7)	0.48
Time to control (TC) (sec)	215 (231.25)	905 (498.7)	0.008
Time to close (TL)(sec)	680 (120)	945 (498.7)	0.01

Table 5C: Crisis Management- time measures in median and inter-quartile range; sec- seconds

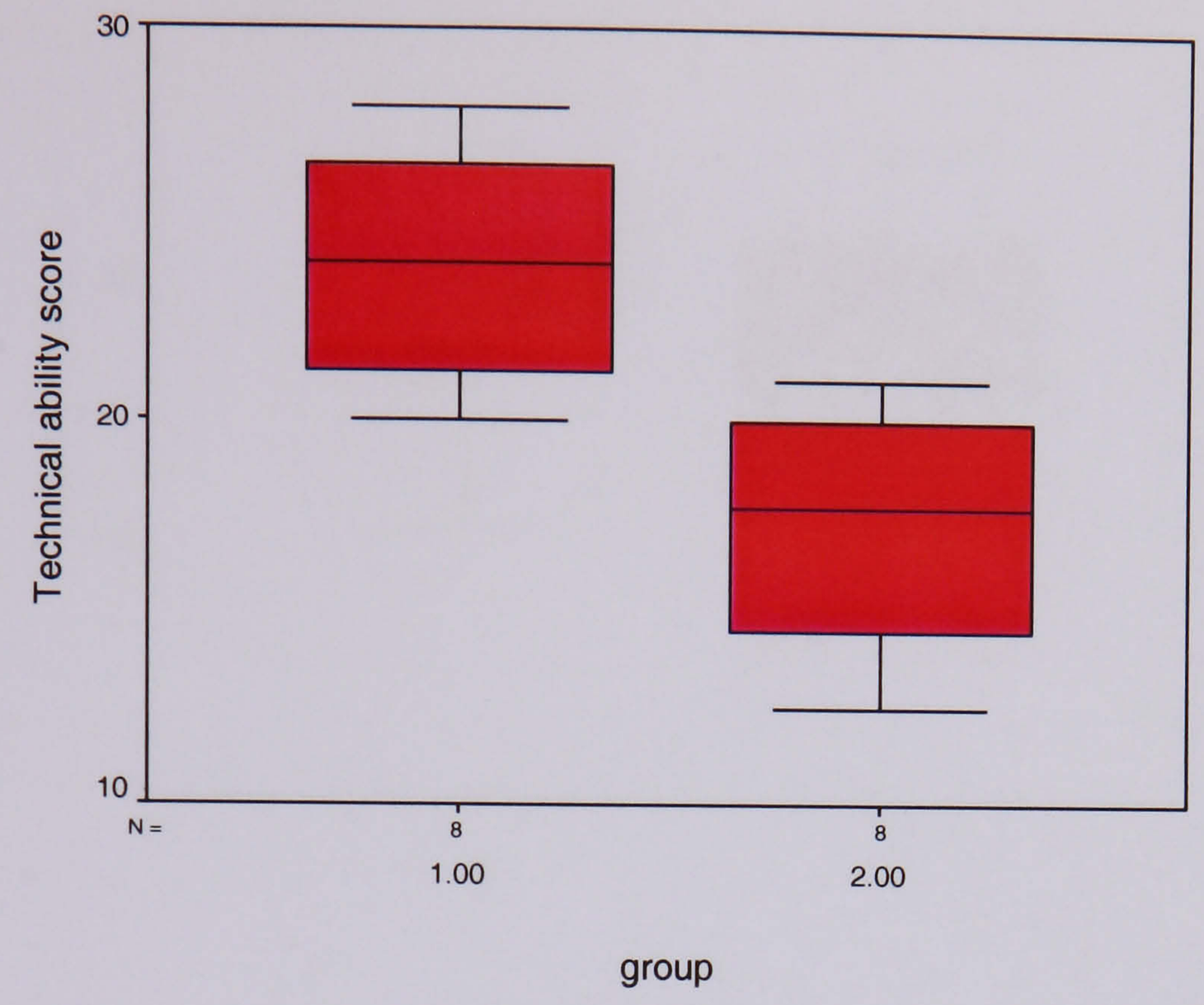


Fig 5B: Technical ability to control the bleeding (maximum score=35). Box plots showing the variability between and within the two groups. Group 1-senior trainees, group 2-juniors. Significant difference between groups (p=0.008)

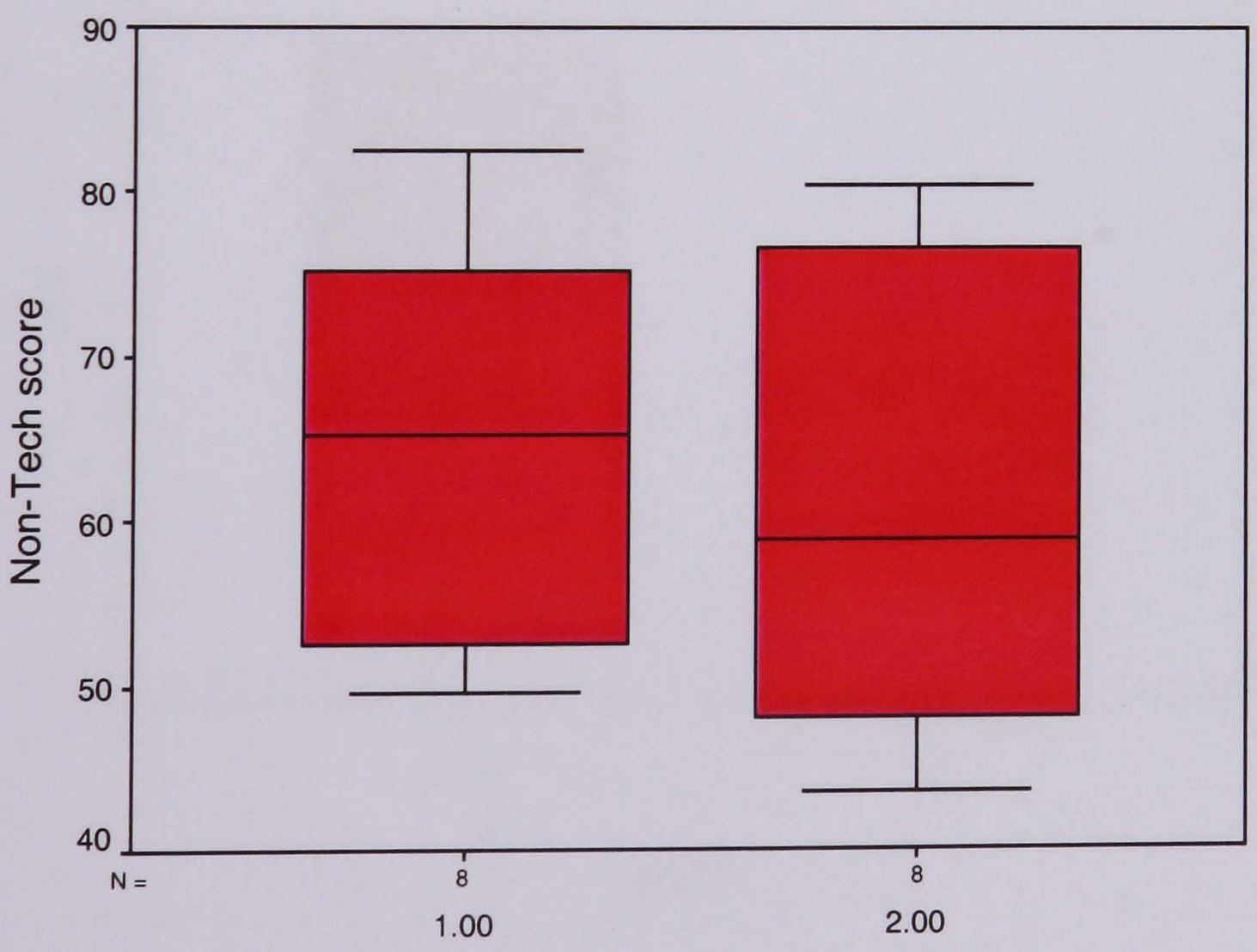


Fig 5C: Non-technical skills (maximum score=90). Box plots showing the variability between and within the two groups. No difference between the two groups (p=0.46)

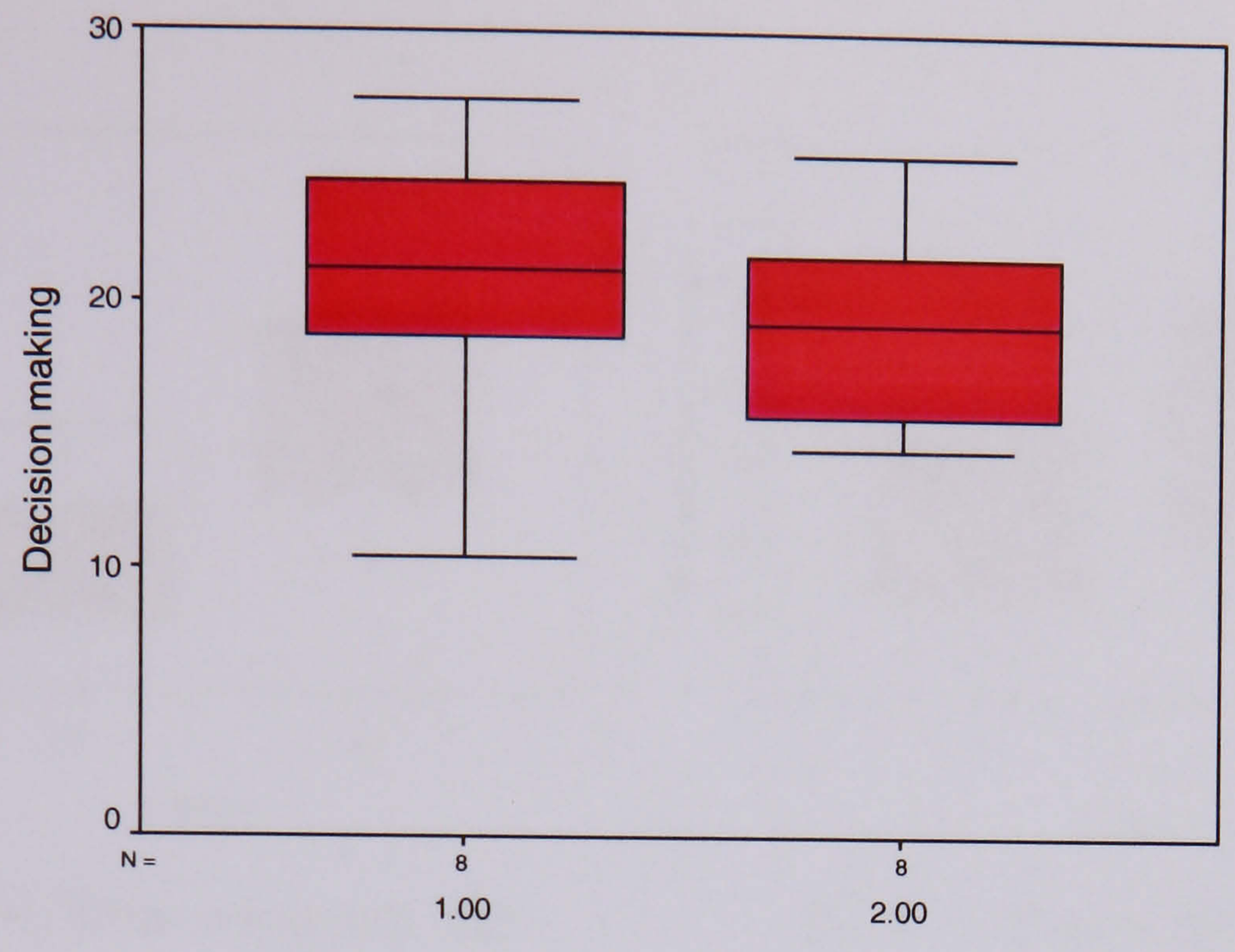


Fig 5D: Decision making (maximum score=30). Shows greater variability within group 1-senior trainees. Difference between the groups (p=0.37)

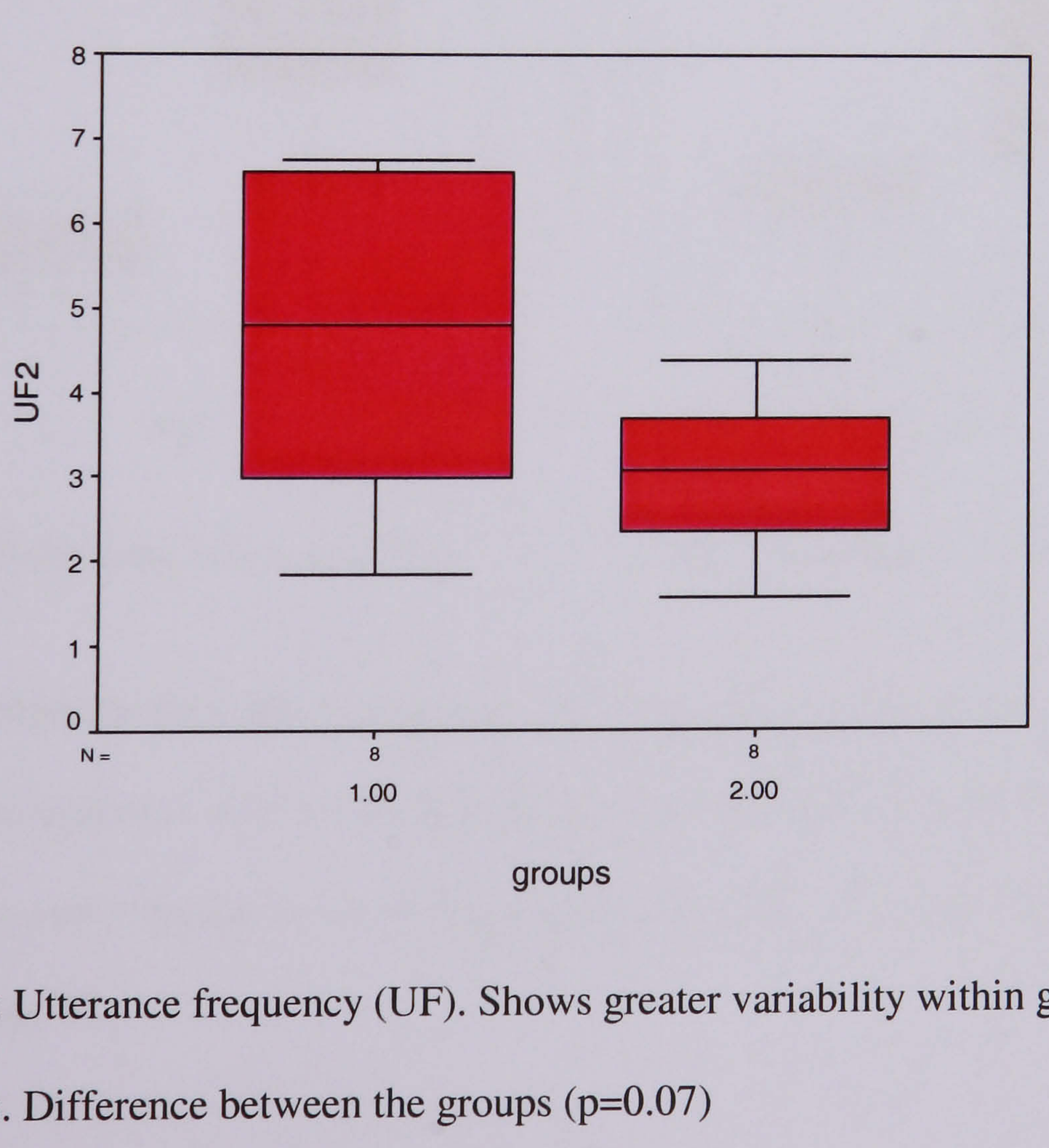


Fig 5E: Utterance frequency (UF). Shows greater variability within group 1-senior trainees. Difference between the groups (p=0.07)

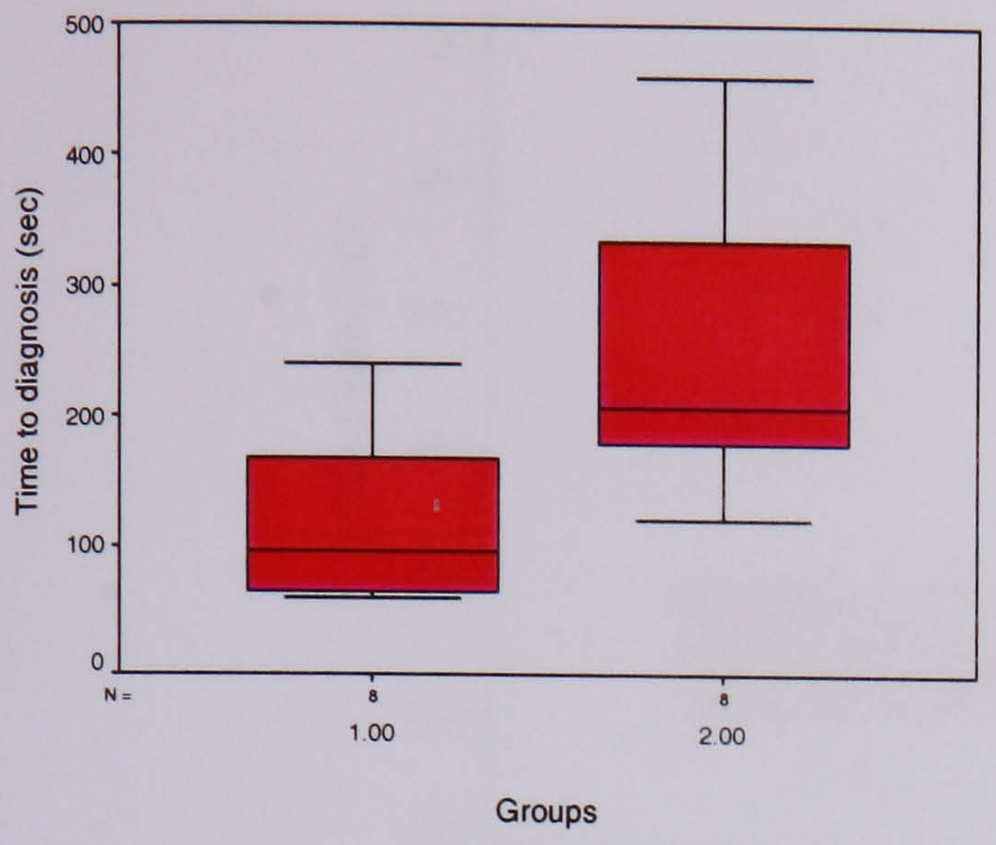


Fig 5F-i: Time to diagnose (TD)

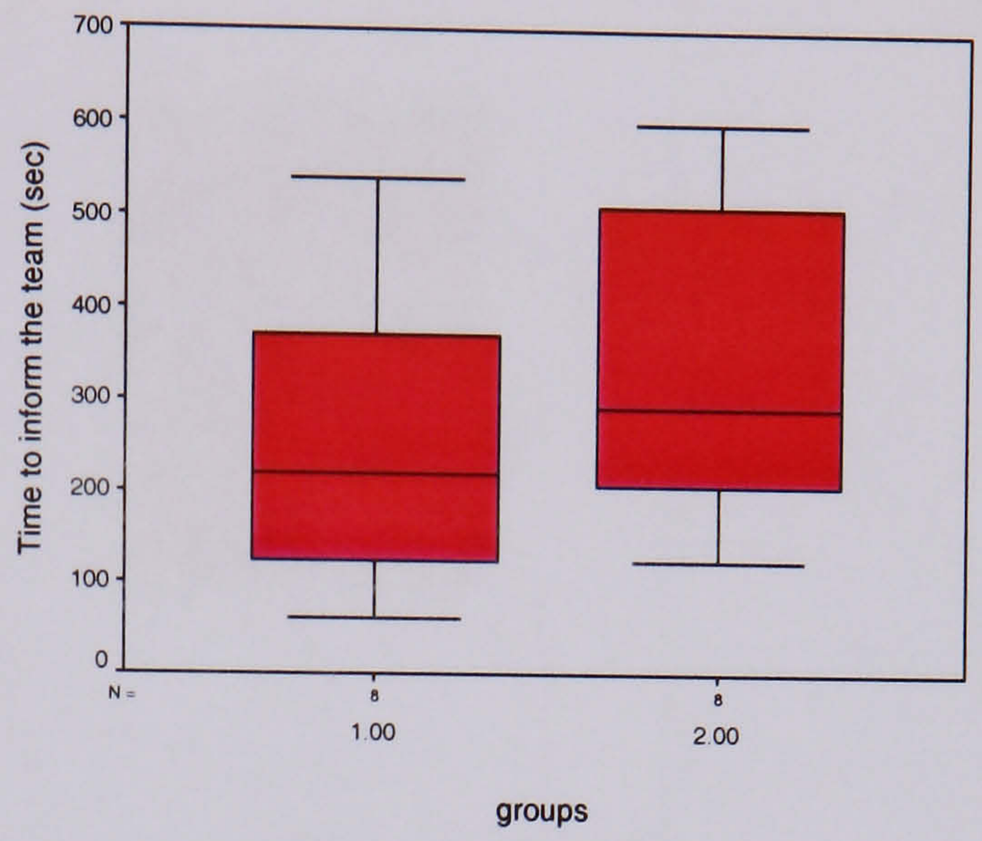


Fig 5F-ii: Time to inform (TT)

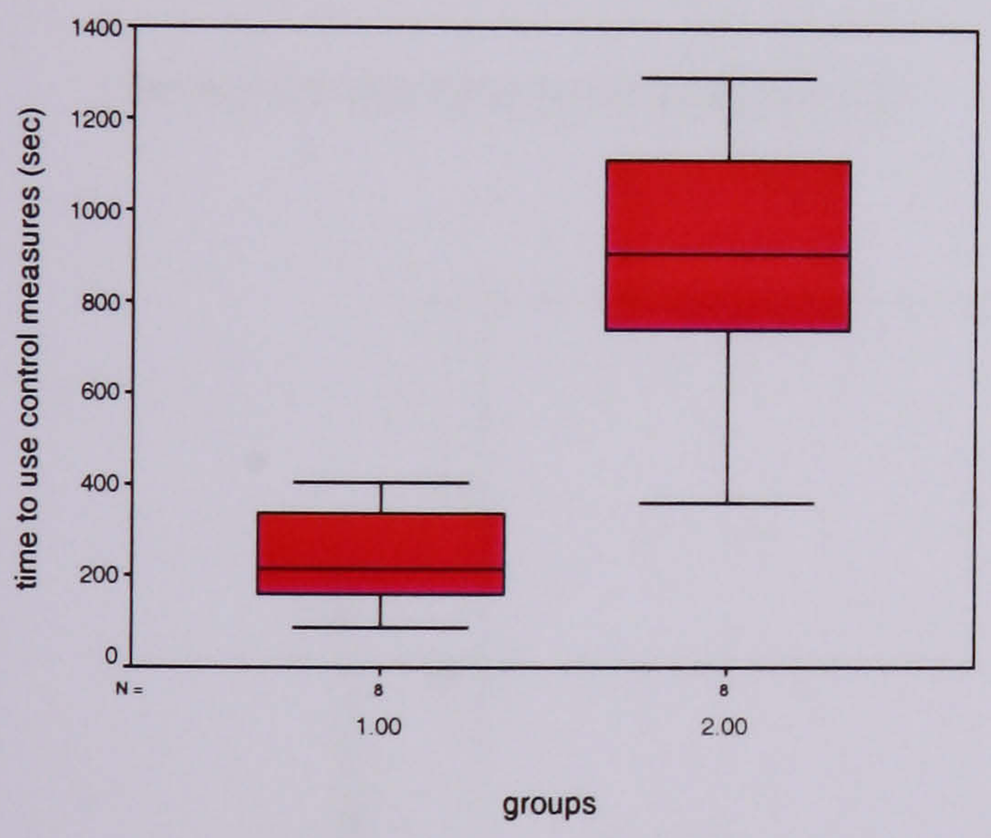


Fig 5F-iii: Time to control (TC)

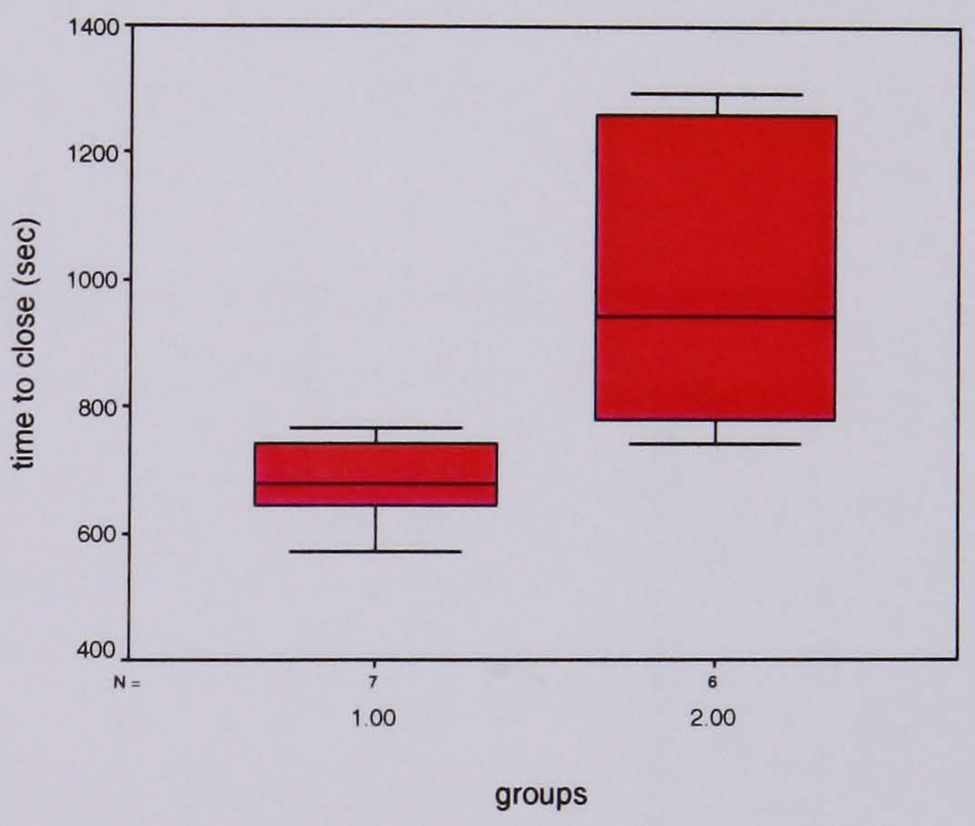


Fig 5F-iv: Time to close (TL)

Fig 5F: Box-plots for the crisis management time measures. Shows the variability in TT for group 1 (senior trainees). Also shows that there is a great variability in performance for all measures for group 2 (junior trainees). Significant differences between the groups for all measures except TT.

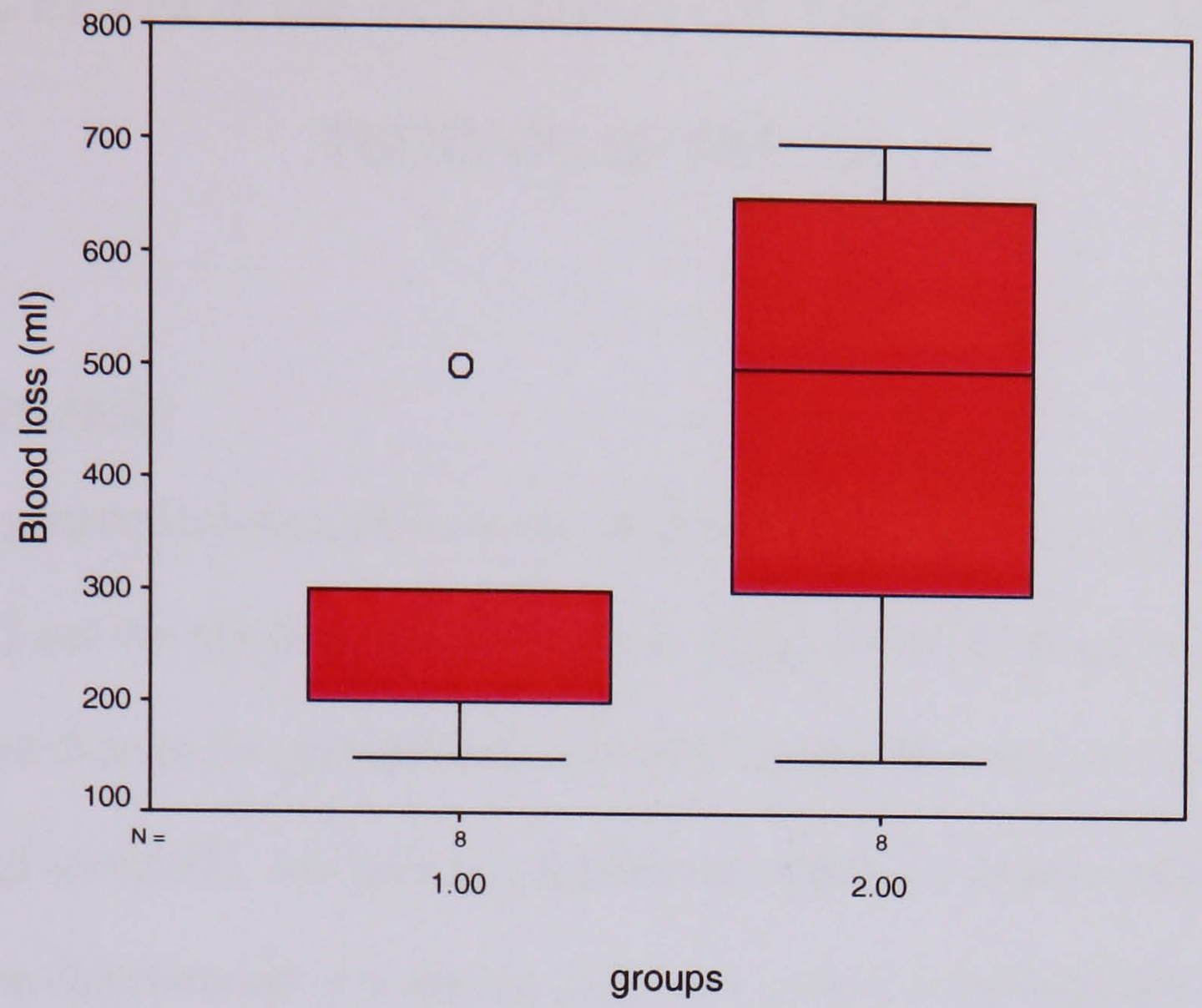


Fig 5G-i: Box plots showing the blood loss (ml). Outlier is also shown. Difference between the two groups (p=0.08)

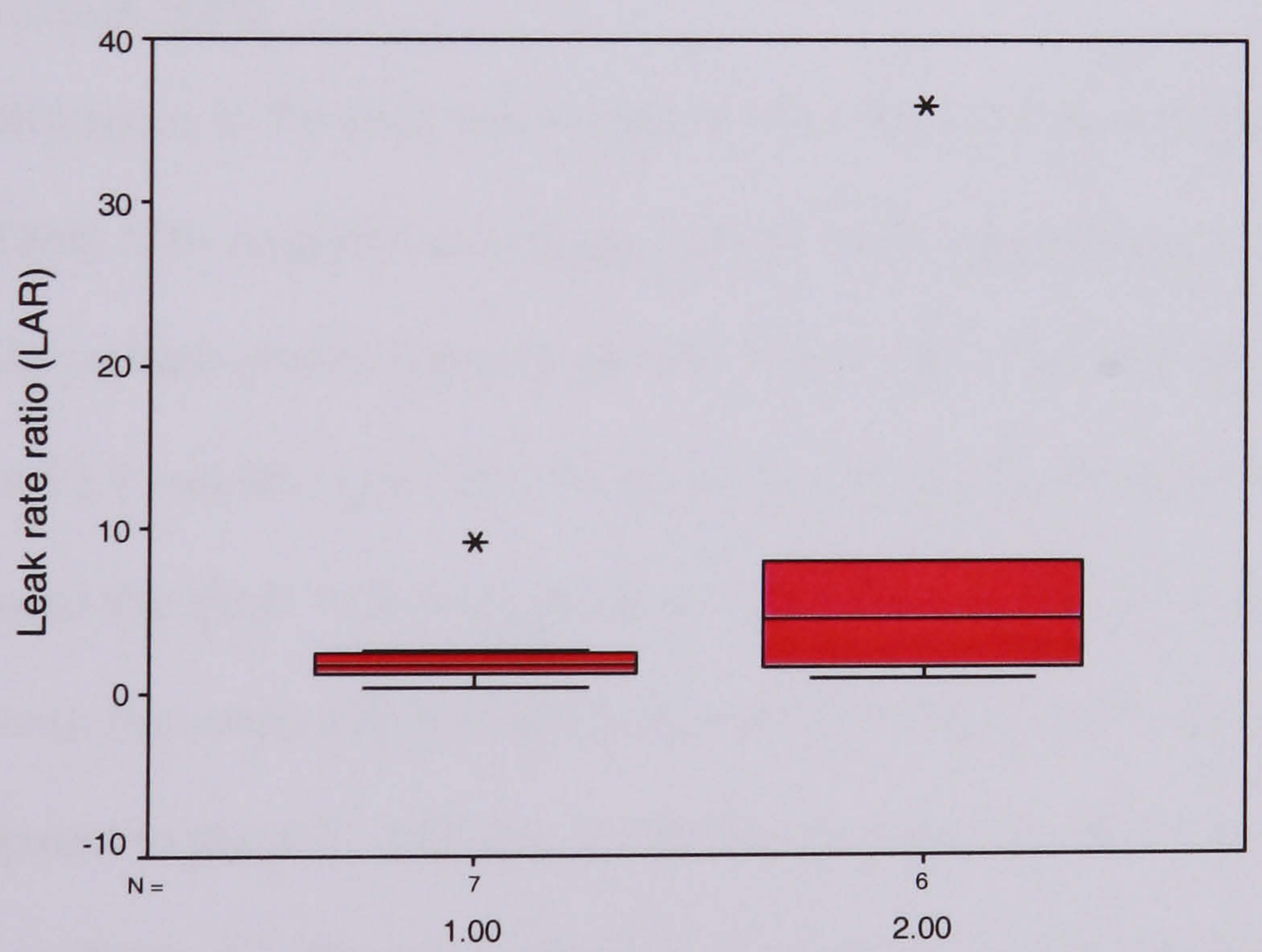


Fig 5G-ii: Box plots showing the Leak rate ratio (LAR). Extremes are also shown. Difference between the two groups (p=0.25)

5.4.2 THE EFFECT OF BLEEDING ON TECHNICAL AND NON-TECHNICAL SKILLS

5.4.2.1 Technical skills

The scores for generic technical skills were compared between the non-crisis or quiet phase (Phase 1) and the bleeding phase (Phase 2). From Table 5D it can be seen that there was a significant drop in the generic technical skills during the crisis as compared to the pre-crisis period (p=0.003). Analysis of the effect according to the level of experience revealed that the deterioration in technical skills was similar in both groups (group 1 median scores- 18.5 and 17.2 (p=0.02); group 2 median scores- 15.7 vs 14.5 (p=0.05).

5.4.2.2 Non-technical skills

There was no difference in the total non-technical score between the two phases for the entire cohort (Table 5D). Analysis according to the level of experience revealed that the effect was similar in both groups (group 1 median scores- 69.7 and 65.2 (p=0.78); group 2 median scores-73.7 and 58.7 (p=0.20). Closer examination of each of the individual categories revealed that there were no significant differences in the scores between phase 1 and 2. There was, however, a drop in the quality of the communication (p=0.03) in phase 2 as compared to phase 1. Analysis of the effect according to the level of experience revealed that the effect was similar in both groups for all the categories.

There was no difference in the UF between the two phases. However, analysis according to grade revealed that there was a significant fall in UF in group 2 (p=0.04) as compared to group 1 (p=0.48).

Skill	**Pre-crisis**	**Crisis**	**p**
Generic technical skills	17.25 (3.4)	15.25 (3.4)	0.003
Non technical skills			
• **Total score**	72.25 (25.7)	64 (27.4)	0.28
• Communication	19 (4.4)	14.75 (8.75)	0.03
• Vigilance	14.75 (6.4)	11.25 (7.9)	0.63
• Team skills	18.5 (9.6)	15 (8.1)	0.62
• Leadership	20 (3.4)	19.5 (2.25)	0.975
Utterance frequency	4.6 (2.6)	3.6 (2.6)	0.35

Table 5D: The effect of the crisis of the generic technical skills and non-technical skills.

5.4.3 CONTENT VALIDITY OF THE NON-TECHNICAL ASSESSMENT: CORRELATION BETWEEN OBJECTIVE MEASURES AND NONTECHNICAL SKILLS ASSESSMENT

There were strong correlations between UF and communication (rho=0.52), UF and decision- making (rho=0.54) and UF and vigilance (rho= 0.54). There were only moderate correlations between UF and leadership (rho= 0.45) and UF and the total non-tech score (rho=0.47). The moderately high levels of correlation between UF, an objective measure of communication, and some of the categories of the non-technical assessment can be considered to be a form of validation of the objectivity of the latter. There were also moderate correlations between the total non-tech score and time taken to diagnose the problem (TD) (rho= -0.44) and inform the team (TT) (rho= - 0.47). However, closer analysis revealed that there were strong correlations between communication and TT (rho= -.0.54) and vigilance and TT (rho= -0.61). There were also strong correlations between the DM and TT (rho= -0.68) and a moderate correlation between DM and TD (rho-.46). Thus time taken to inform the team had a close relationship with communication, vigilance and decision- making. This indicates that those trainees who scored higher for these categories on the non-technical skills assessment were also more likely to inform the team sooner. In addition the strong correlation between the time to inform the team and the non-technical assessment also validates the latter as the time measure corroborates the objectivity of the non-technical assessment.

5.4.4 ADDITIONAL OBSERVATIONS

In addition to the various measures used for the assessment of performance, the observers also made some additional observations which were considered to denote unsafe or safe practises. These observations were considered useful in helping understand the variability in performance of the subjects in between and within the groups.

5.4.4.1 Observations of unsafe practise

- One junior trainee applied a traumatic clamp (non-vascular) on the femoral vein which led to a femoral vein tear and another junior trainee applied traumatic artery forceps to the edges of the laceration to control the bleeding. Both trainees called for help only after executing these unsafe steps.
- Four junior trainees showed evidence of *tunnelled vision*. They were so focussed on closing the laceration that they did not pay any attention to the patient's condition or on communication with anaesthetist. In all four instances there was a blood loss in excess of 500 ml.
- One junior trainee did not call for help in spite of being unable to control bleeding with two sutures, which cut through the vein wall.

5.4.4.2 Examples of safe practise

- One senior and one junior trainee acknowledged their limitation at closing the laceration and called for help but only after using effective measures to control the bleeding. Incidentally both of them scored high in the few parameters that could be assessed for technical ability to control the bleeding.

5.4.4.3 Monitoring

- Only 6 subjects asked about the extent of blood loss and that too only once. There were 4 instances where the surgeons assumed that the blood loss was much lesser than what had been measured without obtaining confirmation.
- There was a great variability in the number of times the surgeons asked about the patient's condition. This varied from 0-7 times during the bleeding, Median and inter-quartile ranges for the senior trainees was 2 (3) and for juniors was 2 (2.5).

5.5 DISCUSSION

This study has demonstrated the face validity of a novel training and assessment environment for surgical trainees which integrates technical skill assessment with other skills crucial for effective performance during a surgical crisis simulation. All the participants found the simulated operating theatre environment to be realistic and 90% found the model and the bleeding scenario to be realistic as well. The high level of realism associated with the model is an improvement from the moderate response that the previous model had received in the earlier study.

This study has also established the construct validity of the simulated environment and the bleeding crisis by demonstrating a difference between senior and junior trainees. In addition to the fact that there was a difference in between the two groups for generic technical skills and technical ability to control the bleeding, the performance of the senior trainees were superior even in terms of the crisis management time measures. They were faster at diagnosing the problem, using effective control measures and at closing the laceration. This is similar to the findings from simulations in anaesthesia which found that time to diagnose a problem and institute remedial measures was faster amongst senior anaesthetists as compared to juniors (140;180). In addition, in this study the senior trainees scored higher than juniors for a number of non-technical skills, even though the difference was not significant. The difference was greatest for the utterance frequency (UF) during phase 2.

The performance of the senior trainees was also superior in terms of patient safety as compared to the juniors. Junior trainees were more likely to perform actions detrimental

to patient safety such as the blind and inappropriate use of traumatic clamps and they were also more likely to focus on closing the laceration at the cost of blood loss. They were also more likely to fail to recognise their limitations and call for help. Other than using the simulation just as an opportunity to teach trainees the skills to control major venous haemorrhage, it can even be used to teach them consequences of their unsafe or inappropriate actions. In fact, some consider the main advantage of simulations to be the ability to make mistakes and learn from them (183), without actually causing any harm to patients .

There was a wide variability in performance amongst both groups of trainees. This is similar to the findings of simulation based performance amongst anaesthetists (180). The true value of simulations probably lies in the ability to use them to identify poor performers and offer them the opportunity to practise using simulations till their performance improves to the expected standards.

This study has also demonstrated the effect that crisis scenarios have on technical and non-technical skills. There was a significant deterioration in the generic technical skills, a significant drop in the quality of communication and a significant drop in the utterance frequency among the junior trainees during the bleeding phase as compared to the 'quiet' phase. Similar to the results in chapter 4, there was no correlation between technical skills and non-technical skills during the 'quiet' (phase 1) phase. However, these two measures had a moderate correlation during the crisis phase and there was a strong correlation between the technical ability to control the bleeding and the utterance frequency. A finding of deterioration in skills on exposure to a crisis is important as it objectively demonstrates the fact that assessment of performance during ' quiet'

procedures cannot be assumed to be reflective of performance even under crisis situations. It is likely that skills and behaviour deteriorate under crises situations as trainees are not frequently exposed to them. Other high reliability organisations such as aviation and the military practise crisis handling 'beyond the attainment of a criterion level of performance' to a stage of 'over- learning' where responses become automated (184). Such a strategy may be useful in surgical training as well to ensure a consistently high level of performance amongst surgeons with respect to their technical and non-technical skills under stressful conditions.

We have used blood loss and leak rates as surrogate outcome measures of performance. A few previous studies have shown that surrogate outcome measures reflect technical expertise. Datta et al used leak rates after a bench model vascular anastomosis and showed a strong correlation between skill and outcome (100). Similarly Szalay et al showed the correlation between OSATS and outcome analysis by using a 4 component, 5 point scale to assess the final product (99). Our study has shown that there was a non-significant but appreciable difference between the two groups. Both blood loss and leak rates were higher in the junior group. However, of greater importance is the variability within the senior trainee group for these two measures.

In terms of validating the non-technical skills assessment scale it is interesting to note that there were strong correlations between the objective crisis management time measures and elements in the decision making category. This is an indirect reflection of the reliability of the two observers and a validation of the objectivity of the non-technical assessment. The high inter-rater reliability between the two observers is also probably because the first five simulations were assessed together to obtain satisfactory calibration

between the two (143). However, even after excluding the data from these five simulations the inter-rater reliability was an impressive 0.84.

This study has also indirectly demonstrated the benefit of the simulation by assessing the perceptions of the participants. A majority of the trainees considered the simulation to be useful for the training and assessment of their technical and non-technical skills and found the feedback following the simulation beneficial. Participants' assessment of simulation- based training is considered to be a powerful method of assessing the value of such training in aviation and anaesthesia (126;137). This is especially important as studies to demonstrate a transfer of skills from such complex simulations to real procedures is difficult and will take a great amount of effort and time.

While having been successful in demonstrating the face validity of the simulation and the construct validity and inter-rater reliability of the assessment measures, this study has certain limitations that will need to be addressed by further research. As the scale is a modification of one used for aviation, the content validity of the assessment will need to confirmed, probably by using a task analysis approach similar to Fletcher et al or using a Delphi type questionnaire (143). Further research will also be required to assess which elements on the scale measure the same feature and thus reduce the number of elements by using a method such as factor analysis or principal component analysis (124).

This is the first study to have extended the concept of crisis simulation training to surgery by describing the development of a simulated environment for the training and assessment of crisis handling skills among surgeons. There is a potential to use such an environment to teach safety principles such as the recognition of one's limitations and the importance of monitoring blood loss and the patient's condition during the process of

handling a surgical crisis. The study has also drawn focus on the need to make performance during crisis situations criteria and protocol based. This would enable performance between surgeons to become consistent and establish a benchmark to which trainees have to reach, aid team coordination, aid the decision- making process and potentially improve patient outcome.

.

CHAPTER 6: OVER-ALL DISCUSSION

General surgical training has evolved over the past few years to acknowledge the limitations of training solely in the operating theatre. It has responded to the challenges of reduced training opportunities, shortened working hours and financial pressures by using simulations primarily for the acquisition of basic psychomotor skills. Various training courses and drills have been developed to help trainees acquire their skills in training centres away from the pressures of the operating theatre. Such an approach is also crucial to patient safety as the early part of a surgeon's learning curve is spent not on real cases but in a skills laboratory. However, one of the major drawbacks has been the lack of evidence that this new paradigm of surgical training, using simulations, actually benefits the trainee, the patient, or the healthcare system. The benefits of simulation based training as still largely assumed. Objective measures of skill during simulations can be used to demonstrate this benefit in addition to which they can be used to design curriculum, plot learning curves and make progression during training criteria based.

6.1 Assessment of technical proficiency

These studies have described the use of objective measures of skill such as dexterity analysis systems, VR metrics and video-based expert assessment to assess technical proficiency during task and procedural simulations. The project has been designed along a skills and assessment escalator and describes how simulations varying from simple task based simulations to more complex procedure based ones can be integrated with various

assessment measures to develop multi-component protocols of technical skills assessment.

Research over the past few years has led to the realisation that most methods of technical skills assessment suffer from certain drawbacks when used alone. Methods that quantify skill such as dexterity analysis systems and VR simulators possess the potential to be used as tools for summative assessment (pass/fail) and can be used to make progression during training criteria based, but suffer from the fact that they are less meaningful than expert criteria based assessments in formative feedback. Criteria-based expert assessments such as the global scale and checklist embedded in the OSATS format, on the other hand are not as objective as the former and are labour intensive as they involve man power and long periods of concentration while viewing performance. But measures such as instrument/ endoscope handling, the use of torque/ assistants, and tissue handling/ use of force and flow of the procedure are easier for the learner to understand as formative feedback measures. A multi-component assessment approach that combines the advantages of quantitative measures with expert assessment makes formative feedback during training more comprehensive and objective than it presently is.

6.1.1 **The importance of multi-component assessment protocols**

The development of multi-component assessment protocols in the assessment of technical proficiency should have a number of implications. Objective measures of skills can be used for **formative assessment during surgical training**. A paradigm of training and

assessment would allow trainees to receive objective feedback on a regular basis and help trainers identify poorly performing trainees and plan remedial measures.

Measures such as those validated in this thesis can also be used to **evaluate learning on skills** courses. Presently end-of-course evaluation of skills is subjective on nearly all courses. The use of objective measures would allow trainees to gain insight into the skills that they have learnt, identify aspects that need attention during post-course practise and more importantly **help course organisers ensure standardisation of courses and evaluate their efficacy.** With respect to endoscopy, there are a number of centres offering short courses in basic skills in endoscopy. The popularity of these courses is gradually increasing. However, there has never been any evidence that these courses actually achieve the purpose of helping participants acquire the necessary skills. A combination of a validated simulator's metrics and an objective expert assessment such as ICASE would comprehensively feedback performance at the end of courses and also help in demonstrating the benefit of these courses objectively. More importantly it would help **course organisers alter course curriculum** according to the performance of participants.

The ICASE method makes the assessment of technical skills in endoscopy more detailed than a simulator's metrics as it assesses **procedural performance for higher cognitive abilities** such as reaction to a patient's discomfort and reaction to the absence of clear vision. In addition, the ICASE assessment can be used for the blinded assessment of technical skills even during real procedures and thus objectively **demonstrate the transfer of skills from VR simulators to real procedures**.

6.1.2 **Validity and reliability of the assessment measures**

Strong correlations between various measures such as ICSAD's measures of dexterity and the check list score for laparoscopic suturing and between the global rating scale and ICEPS during the performance of the bench model sapheno-femoral junction dissection establish the concurrent validity of the check list and ICEPS against established methods such as ICSAD and the global rating scale of OSATS. Concurrent validity is only one aspect in the validation of objective methods of assessment. This thesis has also demonstrated the construct validity and the inter-rater reliability of the checklist for laparoscopic suturing and ICASE for the assessment of technical skills in endoscopy.

6.2 <u>Assessment of surgical competence: the use of simulated environments</u>

Moving up the Skills and Assessment Escalator, this thesis has also for the first time drawn focus to the assessment of technical skills in the context of other skills that are just as important as technical proficiency. Technical proficiency is only one aspect of surgical competence. Drawing on the lessons learnt from other high reliability organisations such as aviation and from the experiences of even other medical specialities such as anaesthesia, this thesis has emphasized the need to integrate technical skills assessment with other skills such as communication, leadership and decision-making.

In addition to the use of the Virtual Endoscopy Suite (VES) for the assessment of technical proficiency, the environment can also be used for the training and assessment of communication and interaction with the patient during the performance of a procedure. With the development of the VES it became apparent that a similar environment could also be developed for team based procedural simulations. The Simulated Operating Theatre (SOT) is a novel surgical environment that moves technical skills assessment out of skills laboratories into a near- realistic OR environment and integrates technical proficiency with team skills. This thesis has described the development and feasibility a complex multi-professional simulated environment around a synthetic surgical simulation and an anaesthetic simulator of moderate fidelity.

6.2.1 **Face validity of the simulated scenarios**

The face validity of the SOT was established by the high scores obtained from the participants both after the 'quiet' and the 'crisis' simulations. Over 90% of the participants agreed that the SOT was realistic in terms of the environment and the personnel within it. The realism of the synthetic model was initially found to be accurate by only half the trainees but after suitable changes, 90% of the participants agreed with the statement that the model was an accurate representation of a real procedure. 86.6% of the participants agreed that the bleeding scenario was realistic.

6.2.2 **Validity of the assessment measures**

The two SOT studies have also demonstrated the construct validity of the various assessment measures. In addition to establishing the construct validity of technical skills during the performance of the SFJ dissection during the 'quiet' study, the thesis has also established the construct validity of the various measures used to assess skill and process during the 'crisis' simulation. Senior trainees were superior in their technical ability to control the bleeding and also in terms of the objective crisis time measures. Junior trainees were also more likely to perform actions detrimental to 'patient safety'. The results of the two studies showed that the difference in technical performance across the training grades wasn't mirrored by a difference in the non-technical/ human factors/ team skills.

The importance of team skills has been borne out by the finding that poor scores were obtained during the procedural simulations for some key elements of communication with other team members, pre-operative briefings and situation awareness. Briefings and vigilance are some of the non-technical skills considered by aviation and anaesthesia to be important in team formation and coordination. Their importance to surgery is being explored by longitudinal observational studies of team- working in the operating theatre that are currently underway in some centres across the world.

The non-technical skills assessment was modified from assessment methods used in pilot training in aviation. However, further studies are needed to establish the content validity of the elements in the non-technical assessment and evaluate their relevance to the operating theatre. This could probably be performed by using a cognitive task analysis

that may involve interviewing surgeons. However, it will also be useful to ascertain the perceptions of other team members at the same time, as elements that surgeons consider to be unimportant may be important to other team members.

6.3 Crisis management training

This thesis has also drawn focus to the need for crisis management training for surgeons. The fact that surgical crises are rare makes their management random, non-protocol based and variable. This is borne out by the finding of considerable variability in the performance within both groups. Practising crisis simulations offers surgical trainees the opportunity to acquire the knowledge and the skills to handle these situations when they do occur in real life. By establishing guidelines for their management, performance between surgeons can also be made more consistent. The importance of crisis management skills assessment is also highlighted by the fact that there was a significant deterioration in skills during the crisis as compared to the 'quiet' phase of the study reflecting the fact that assessment during 'quiet' procedures may not be representative of performance during 'crisis' conditions.

The advantage of using measures of skill, process and outcome to assess performance (185) is that feedback and thus training can be more systematic. The development of objective validated measures of performance of both technical and non-technical skills enables post-simulation feedback to be focussed and structured. In addition, the feedback sessions in this study were also video-based. Such feedback sessions provide an opportunity for participants to review and critique their own performance (137).

Objective measures of performance can also used to evaluate the effectiveness of practise in such crisis scenarios. Further research should use these measures to correlate performance in simulations with performance in the operating theatre and to correlate performance measures with patient outcome.

6.4 Transfer of skills from simulations to real procedures

The benefit of developing objective assessment measures are that the benefit of training using simulations can be objectively demonstrated and performance in simulations can be correlated with performance in real life situations. One of the greatest limitations of research in the field of simulations has been the lack of evidence that skills learnt during simulations actually transfer to performance in real life. In fact there are very few studies that have evaluated the benefit of simulation training in anaesthesia. Chopra et al studied the benefit of simulator based training by evaluating performance in a simulated scenario (139). Two groups of anaesthetists were challenged with two different scenarios and were then 're-exposed' to one of the two scenarios. They found that there was a significant improvement in the in the group with prior exposure to the same scenario. Unfortunately, this study did not address the learning effect associated with being exposed to the same scenario twice.

Even though transfer studies are crucial, the demonstration of improved patient outcomes with crisis simulation training is probably going to be difficult to demonstrate in surgery and anaesthesia due to the low incidence of adverse events and the presence of a number of confounding variables that can influence patient outcome. In addition, the assessment

of performance during real procedures will prove to be expensive and will be fraught with a number of medico-legal issues regarding confidentiality. This lack of evidence of the training benefit of simulations is not unique to medicine alone. Even in aviation, there is no direct evidence that CRM training has reduced the incidence of errors and near-misses (132). However, there is abundant evidence of the positive effects of CRM training on team working in aviation and on crew attitudes and effectiveness (126;186). Adopting a similar approach in anaesthesia there is evidence of the perceived benefit of ACRM training. 6- month follow up of anaesthetists who had undergone an ACRM course revealed that their performance during a real life critical event had been improved as a result of the course (137). Thus the perceptions of participants who undertake simulation sessions should be an important part of evaluating the training benefit of simulations. In light of this it is important to note that even in this thesis the participants rated their experience in the SOT very favourably and over 80% of them agreed with the statement that both the model and the simulated operating theatre are suitable for training technical and team skills.

The development of a simulated operating theatre has the potential to help learning outside the real OR. It builds on the stepwise framework suggested by Haluck et al (8). It can be seen as a bridge between the skills laboratories and the real operating theatre (Fig 6A). Once the basic skills of dissection, knot tying and suturing are learnt in skills labs, these skills can be integrated with others such as set-up, patient positioning, use of assistance and exposure of the operative field along with other team skills in the simulated operating theatre. The ability to create a near-real environments also helps in

introducing just the right amount of stress that comes with the need to work with other team members, to enhance the process of the learning (8).

6.5 Implications of the objective assessment of surgical competence

The objective assessment of skills inevitably raises the issue of summative assessment and revalidation. Even though the studies in this thesis have described the construct validity and reliability of assessment measures such as ICSAD and video-based expert assessments, further research is needed before these measures can be used for summative assessments. Most importantly it will be important to demonstrate the predictive validity of the simulations. In other words, it will be essential to show that performance using the simulations predicts and correlates with performance in real life.

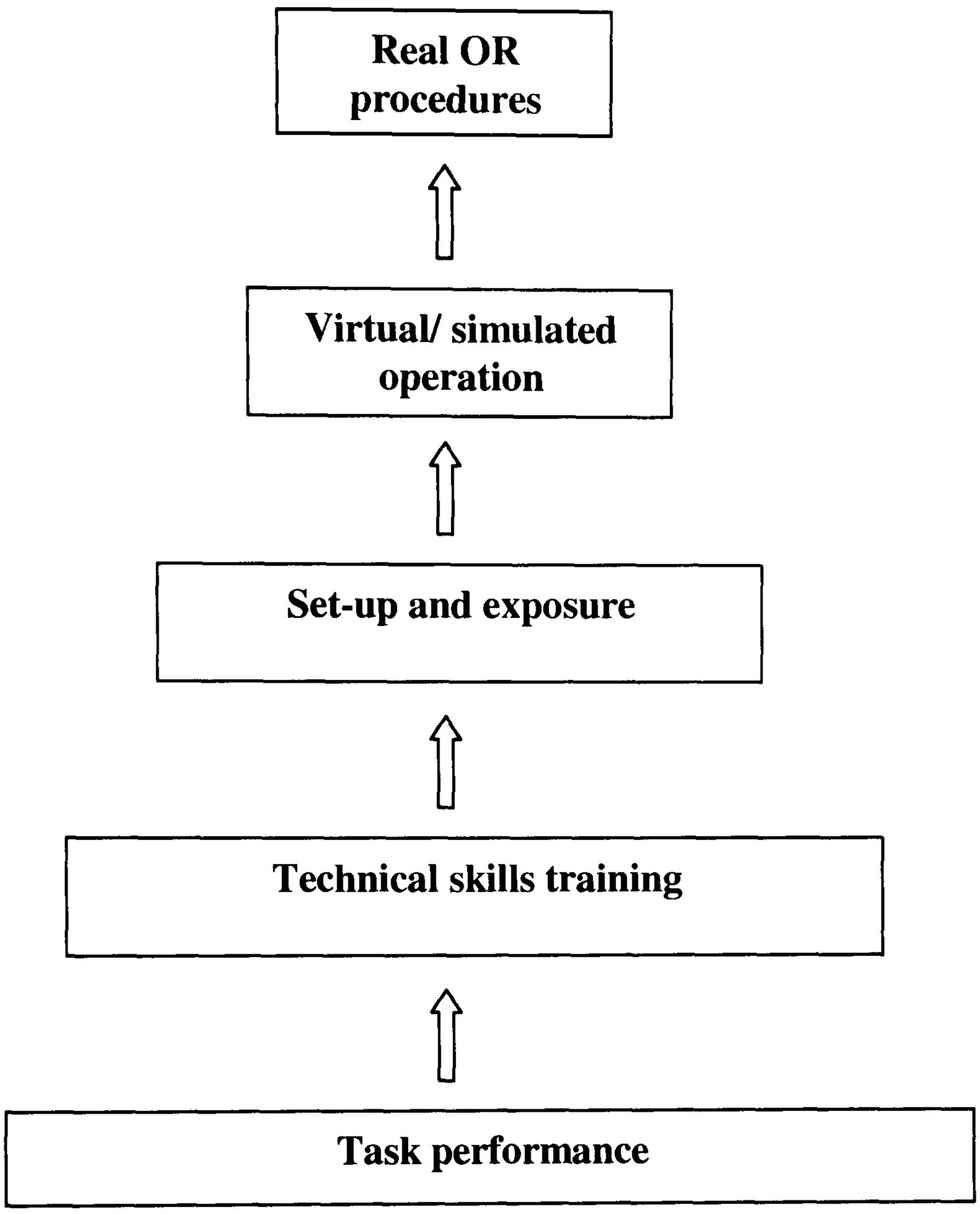


Fig 6A: Paradigm of stepwise learning in surgery. Adapted from Haluck and Krummel.

Arch Surg. 2000; 135 (7): 786-92

6.6 Future work

In the future simulation based training should be integrated with objective measures of technical proficiency and non-technical skills where necessary to demonstrate the real benefit of simulation based training. Using such measures studies can be planned that demonstrate that skills learnt in simulated environments transfer to real life situations. All the measures of skill developed and validated in this thesis can be easily applied to the assessment of performance even in real cases. This would be essential to demonstrate correlation between performance in simulations and performance in real cases- a final affirmation of the validity of simulations.

Further work will be required to address some of the current limitations in the assessment of technical proficiency. Methods to analyse outcome would be useful in any system that aims to measure surgical expertise. It will also be interesting to explore the relationship between outcome methods, expert assessments and dexterity analysis.

In the future it will be crucial to study the relationship between technical skills and non-technical or team skills and patient outcome using outcome measures such as blood loss. For example it would be interesting to explore any correlation between team coordination and situation awareness and surrogate outcome measures especially during recovery from a crisis. The simulated environment offers us the opportunity to test this, as adverse occurrences are extremely rare in surgery.

The SOT also presents us with the opportunity to train OR teams together. So far simulation training has largely involved only anaesthetists and this study for the first time extended the concept to surgeons. However, Helmreich has stated that the 'the ultimate of

aim of OR simulations should be to involve all teams to enhance overall team performance' (111). Team training in the SOT could be used to improve team coordination during both 'quiet' and 'crisis' periods. It will be interesting to evaluate the benefit of such training on subsequent performance on real cases.

The simultaneous development of both surgical and anaesthetic crises could train teams in activity monitoring and other team skills essential for the recovery from a crisis. It is possible that in the future software of surgical simulators could be linked to that of anaesthetic simulators to ensure that surgical actions are accurately mirrored in the patient's parameters.

The SOT could also provide a useful research environment to evaluate the effect of environmental conditions such as ambient temperature and noise on the performance of technical and non-technical skills of surgeons as well the performance of teams. It could also be used to train individuals and teams in the use of new technologies such as surgical robots using simulated procedures. Repeated practise and familiarisation with such systems in the SOT could shorten learning curves not only for surgeons but even for other team members. Teams could be taught to use new technologies in a coordinated manner using protocol based actions and systematic task allocation to shorten the times associated with system set-up. This would result in the efficient and safe use of OR equipment.

Reference List

(1) Macintyre IM, Munro A. Simulation in surgical training. BMJ 1990; 300(6732):1088-1089.

(2) Heuer G. Dr Halsted. Bull John Hopkins Hosp 1952; 90:1.

(3) Halsted WS. The training of the surgeon. Bull John Hopkins Hosp 15, 267-275. 1904.

(4) Hamdorf JM, Hall JC. Acquiring surgical skills. Br J Surg 2000; 87(1):28-37.

(5) McColl I. The Guy's Surgical Training Programme: a report on the first 5 years. Br J Surg 1977; 64(10):745-746.

(6) Paice E, Aitken M, Cowan G, Heard S. Trainee satisfaction before and after the Calman reforms of specialist training: questionnaire survey. BMJ 2000; 320(7238):832-836.

(7) Folse JR. Surgical education--addressing the challenges of change. Surgery 1996; 120(4):575-579.

(8) Haluck RS, Krummel TM. Computers and virtual reality for surgical education in the 21st century. Arch Surg 2000; 135(7):786-792.

(9) Bridges M, Diamond DL. The financial impact of teaching surgical residents in the operating room. Am J Surg 1999; 177(1):28-32.

(10) Naylor AR, Thompson MM, Varty K, Sayers RD, London NJ, Bell PR. Provision of training in carotid surgery does not compromise patient safety. Br J Surg 1998; 85(7):939-942.

(11) Singh KK, Aitken RJ. Outcome in patients with colorectal cancer managed by surgical trainees. Br J Surg 1999; 86(10):1332-1336.

(12) Cagir B, Rangraj M, Maffuci L, Herz BL. The learning curve for laparoscopic cholecystectomy. J Laparoendosc Surg 1994; 4(6):419-427.

(13) Feliu-Pala X, Martin-Gomez M, Morales-Conde S, Fernandez-Sallent E. The impact of the surgeon's experience on the results of laparoscopic hernia repair. Surg Endosc 2001; 15(12):1467-1470.

(14) Bernard HR, Hartman TW. Complications after laparoscopic cholecystectomy. Am J Surg 1993; 165(4):533-535.

(15) Deziel DJ, Millikan KW, Economou SG, Doolas A, Ko ST, Airan MC. Complications of laparoscopic cholecystectomy: a national survey of 4,292 hospitals and an analysis of 77,604 cases. Am J Surg 1993; 165(1):9-14.

(16) Rock JA, Warshaw JR. The history and future of operative laparoscopy. Am J Obstet Gynecol 1994; 170(1 Pt 1):7-11.

(17) Cuschieri A. Whither minimal access surgery: tribulations and expectations. Am J Surg 1995; 169(1):9-19.

(18) Reznick RK. Teaching and testing technical skills. Am J Surg 1993; 165(3):358-361.

(19) Whitman N. Student Stress: Effects and Solutions. Washington DC: Association of Study for Higher Education, 1984.

(20) Raftery J, Stevens A. Day case surgery trends in England: the influences of target setting and of general practitioner fundholding. J Health Serv Res Policy 1998; 3(3):149-152.

(21) Jarrett PE. Day care surgery. Eur J Anaesthesiol Suppl 2001; 23:32-35.

(22) Fleming WR, Michell I, Douglas M. Audit of outpatient laparoscopic cholecystectomy. Universities of Melbourne HPB Group. Aust N Z J Surg 2000; 70(6):423-427.

(23) Sharp RJ, Wellwood J. The training implications of day-case surgery. Br J Hosp Med 1996; 55(8):472-475.

(24) Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ et al. Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg 2000; 191(3):272-283.

(25) J.Howell. **Performance league tables**. BMJ 324(7336), 542.

(26) Carter D. The surgeon as a risk factor. BMJ 2003; 326(7394):832-833.

(27) Pickersgill T. The European working time directive for doctors in training. BMJ 2001; 323(7324):1266.

(28) Johnson T. Limitations on residents' working hours at New York teaching hospitals: a status report. Acad Med 2003; 78(1):3-8.

(29) Barden CB, Specht MC, McCarter MD, Daly JM, Fahey TJ, III. Effects of limited work hours on surgical training. J Am Coll Surg 2002; 195(4):531-538.

(30) Aitken M, Paice E. Trainees' attitudes to shift work depend on grade and specialty. BMJ 2003; 326(7379):48.

(31) Skidmore FD. Junior surgeons are becoming deskilled as result of Calman proposals. BMJ 1997; 314(7089):1281.

(32) Kelty C, Duffy J, Cooper G. Out-of-hours work in cardiothoracic surgery: implications of the New Deal and Calman for training. Postgrad Med J 1999; 75(884):351-352.

(33) Rosser JC, Rosser LE, Savalgi RS. Skill acquisition and assessment for laparoscopic surgery. Arch Surg 1997; 132(2):200-204.

(34) Gallagher AG, Satava RM. Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. Learning curves and reliability measures. Surg Endosc 2002; 16(12):1746-1752.

(35) Moosa AR EDVSECGDH. Laparoscopic injuries to the bile duct. A cause for concern. Ann Surg 215(3), 203-208. 2003.

(36) Mellinger JD. Meetings, millions, mentors, motives. Surg Endosc 2003; 17(3):363-364.

(37) Krummel TM. Surgical simulation and virtual reality: the coming revolution. Ann Surg 1998; 228(5):635-637.

(38) Davies RJ, Hamdorf JM. Surgical skills training and the role of skills centres. BJU Int 2003; 91(1):3-4.

(39) Saifi J, Chang BB, Paty PS, Kaufman J, Leather RP, Shah DM. An animal model for instructing and the study of in situ arterial bypass. J Vasc Surg 1990; 12(5):538-540.

(40) Christopherson WA, Buchsbaum HJ, Voet R, Lifshitz S. The canine laboratory in the training of the oncology fellow. Gynecol Oncol 1986; 23(1):26-34.

(41) Clayden GS. Teaching laparoscopic surgery. Preliminary training on animals is essential. BMJ 1994; 309(6950):342.

(42) Kirwan WO, Kaar TK, Waldron R. Starting laparoscopic cholecystectomy--the pig as a training model. Ir J Med Sci 1991; 160(8):243-246.

(43) Jakimowicz JJ. The European Association for Endoscopic Surgery recommendations for training in laparoscopic surgery. Ann Chir Gynaecol 1994; 83(2):137-141.

(44) Dunnington GL, Darosa DA. Changing surgical education strategies in an environment of changing health care delivery systems. World J Surg 1994; 18(5):734-737.

(45) Melvin DB. Alternate training methods II: Incorporating Inanimate Surgical Models. Lab Animal 28 (5), 32-36.

(46) Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg 1998; 175(6):482-487.

(47) Keyser EJ, Derossis AM, Antoniuk M, Sigman HH, Fried GM. A simplified simulator for the training and evaluation of laparoscopic skills. Surg Endosc 2000; 14(2):149-153.

(48) McCloy R, Stone R. Science, medicine, and the future. Virtual reality in surgery. BMJ 2001; 323(7318):912-915.

(49) Strom P, Kjellin A, Hedman L, Johnson E, Wredmark T, Fellander-Tsai L. Validation and learning in the Procedicus KSA virtual reality surgical simulator. Surg Endosc 2003; 17(2):227-231.

(50) Schijven M, Jakimowicz J. Construct validity. Surg Endosc 2003; 17(5):803-810.

(51) Schijven M, Jakimowicz J. Face-, expert, and referent validity of the Xitact LS500 laparoscopy simulator. Surg Endosc 2002; 16(12):1764-1770.

(52) McCashland T, Brand R, Lyden E, de Garmo P. The time and financial impact of training fellows in endoscopy. CORI Research Project. Clinical Outcomes Research Initiative. Am J Gastroenterol 2000; 95(11):3129-3132.

(53) Markman HD. A new system for teaching proctosigmoidoscopic morphology. Am J Gastroenterol 52, 65-69. 1969.

(54) Gerson LB, Van Dam J. The future of simulators in GI endoscopy: an unlikely possibility or a virtual reality? Gastrointest Endosc 2002; 55(4):608-611.

(55) Ferlitsch A, Glauninger P, Gupper A, Schillinger M, Haefner M, Gangl A et al. Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. Endoscopy 2002; 34(9):698-702.

(56) Datta V, Mandalia M, Mackay S, Darzi A. The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. Surg Endosc 2002; 16(10):1459-1463.

(57) Moorthy K, Smith S, Brown T, Bann S, Darzi A. Evaluation of virtual reality bronchoscopy as a learning and assessment tool. Respiration 2003; 70(2):195-199.

(58) Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M et al. Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 1999; 177(2):167-170.

(59) Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. J Am Board Fam Pract 1998; 11(6):426-433.

(60) Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 2002; 236(4):458-463.

(61) Anderson RC, Faust GW. Educational psychology. The science of instruction and learning. New York: Dodd, Mead and Co, 1973.

(62) Patrick J. Training: Research and practice. London: Academic Press, 1992.

(63) Kirk RM. Teaching the craft of operative surgery. Ann R Coll Surg Engl 1996; 78(1 Suppl):25-28.

(64) Satava RM, Cuschieri A, Hamdorf J. Metrics for objective Assessment. Surg Endosc 2003; 17(2):220-226.

(65) Kaufman HH, Wiegand RL, Tunick RH. Teaching surgeons to operate--principles of psychomotor skills training. Acta Neurochir (Wien) 1987; 87(1-2):1-7.

(66) Kopta JA. An approach to the evaluation of operative skills. Surgery 70, 297-303. 1971.

(67) Freeman ML. Training and competence in gastrointestinal endoscopy. Rev Gastroenterol Disord 2001; 1(2):73-86.

(68) Cuschieri A, Francis N, Crosby J, Hanna GB. What do master surgeons think of surgical competence and revalidation? Am J Surg 2001; 182(2):110-116.

(69) Byrne AJ, Greaves JD. Assessment instruments used during anaesthetic simulation: review of published studies. Br J Anaesth 2001; 86(3):445-450.

(70) Watts J, Feldman WB. Assessment of technical skills. In: Neufeld VR, Norman GR, editors. Assessing clinical competence. New York: Springer, 1985: 259-274.

(71) Savage PEA. Competence assessment of senior house officers in general surgery. Ann R Coll Surg Engl 1995; 77 ((Suppl)):246-247.

(72) Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. Am J Surg 1995; 169(4):414-417.

(73) Scott DJ, Valentine RJ, Bergen PC, Rege RV, Laycock R, Tesfay ST et al. Evaluating surgical competency with the American Board of Surgery In-Training Examination, skill testing, and intraoperative assessment. Surgery 2000; 128(4):613-622.

(74) Pugh P. Nelson and His Surgeons. 1968. Edinburgh, ES Livingstone.

(75) Ellis H. Famous operations. 1984. Pennsylvania, Harwal.

(76) Peters JH, Ellison EC, Innes JT, Liss JL, Nichols KE, Lomano JM et al. Safety and efficacy of laparoscopic cholecystectomy. A prospective analysis of 100 initial patients. Ann Surg 1991; 213(1):3-12.

(77) Porter GA, Soskolne CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. Ann Surg 1998; 227(2):157-167.

(78) Martling A, Cedermark B, Johansson H, Rutqvist LE, Holm T. The surgeon as a prognostic factor after the introduction of total mesorectal excision in the treatment of rectal cancer. Br J Surg 2002; 89(8):1008-1013.

(79) Hermanek P, Hermanek PJ. Role of the surgeon as a variable in the treatment of rectal cancer. Semin Surg Oncol 2000; 19(4):329-335.

(80) Gillison EW, Powell J, McConkey CC, Spychal RT. Surgical workload and outcome after resection for carcinoma of the oesophagus and cardia. Br J Surg 2002; 89(3):344-348.

(81) Eubanks TR, Clements RH, Pohl D, Williams N, Schaad DC, Horgan S et al. An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 1999; 189(6):566-574.

(82) Regehr G, MacRae H, Reznick RK, Szalay D. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73(9):993-997.

(83) Winckel CP, Reznick RK, Cohen R, Taylor B. Reliability and construct validity of a structured technical skills assessment form. Am J Surg 1994; 167(4):423-427.

(84) Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997; 84(2):273-278.

(85) Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. Am J Surg 1997; 173(3):226-230.

(86) Datta V, Bann S, Beard J, Mandalia M, Darzi A. Comparison of bench test evaluations of surgical skill with live animal operating performance assessments. Br J Surg 89 (Suppl 1), 92. 2002.

(87) Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. Am J Surg 2002; 184(1):70-73.

(88) Taffinder N, Smith S, Mair J, Russell R, Darzi A. Can a computer measure surgical precision? Reliability, validity and feasibility of the ICSAD. Surg Endosc 13 (Suppl 1), 81. 1999.

(89) Smith SG, Torkington J, Brown TJ, Taffinder NJ, Darzi A. Motion analysis. Surg Endosc 2002; 16(4):640-645.

(90) Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. J Am Coll Surg 2001; 193(5):479-485.

(91) Francis NK, Hanna GB, Cuschieri A. Reliability of the Advanced Dundee Endoscopic Psychomotor Tester for bimanual tasks. Arch Surg 2001; 136(1):40-43.

(92) Francis NK, Hanna GB, Cuschieri A. The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study. Arch Surg 2002; 137(7):841-844.

(93) Emam TA, Hanna G, Cuschieri A. Ergonomic principles of task alignment, visual display, and direction of execution of laparoscopic bowel suturing. Surg Endosc 2002; 16(2):267-271.

(94) Gillette JC, Quick NE, Adrales GL, Shapiro R, Park AE. Changes in postural mechanics associated with different types of minimally invasive surgical training exercises. Surg Endosc 2003; 17(2):259-263.

(95) Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. Stud Health Technol Inform 1998; 50:124-130.

(96) Gallagher AG, Richie K, McClure N, McGuigan J. Objective psychomotor skills assessment of experienced, junior, and novice laparoscopists with virtual reality. World J Surg 2001; 25(11):1478-1483.

(97) Steele RJ, Walder C, Herbert M. Psychomotor testing and the ability to perform an anastomosis in junior surgical trainees. Br J Surg 1992; 79(10):1065-1067.

(98) Joice P, Hanna GB, Cuschieri A. Ergonomic evaluation of laparoscopic bowel suturing. Am J Surg 1998; 176(4):373-8.

(99) Szalay D, MacRae H, Regehr G, Reznick R. Using operative outcome to assess technical skill. Am J Surg 2000; 180(3):234-237.

(100) Datta V, Mandalia M, Mackay S, Chang A, Cheshire N, Darzi A. Relationship between skill and outcome in the laboratory-based model. Surgery 2002; 131(3):318-323.

(101) Hanna GB, Frank TG, Cuschieri A. Objective assessment of endoscopic knot quality. Am J Surg 1997; 174(4):410-413.

(102) Cass OW. Training to competence in gastrointestinal endoscopy: a plea for continuous measuring of objective end points. Endoscopy 1999; 31(9):751-754.

(103) Traverso LW. Guidelines for granting of privileges for gastrointestinal endoscopy: focus on outcomes not number of procedures performed. Surg Endosc 2002; 16(2):229.

(104) Wexner SD, Garbus JE, Singh JJ. A prospective analysis of 13,580 colonoscopies. Reevaluation of credentialing guidelines. Surg Endosc 2001; 15(3):251-261.

(105) Cass OW, Freeman ML, Peine CJ, Zera RT, Onstad GR. Objective evaluation of endoscopy skills during training. Ann Intern Med 1993; 118(1):40-44.

(106) Cass OW FMCJ. Acquisition of competency in endoscopy skills (ACES) during training: a multi-centre study. Gastrointest Endosc 43, 308. 1996.

(107) Spencer FC. Teaching and measuring surgical techniques- the technical evaluation of competence. Bulletin of the American College of Surgeons 63, 9-12. 1978.

(108) Satava RM, Gallagher AG, Pellegrini CA. Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. J Am Coll Surg 2003; 196(6):933-937.

(109) Beard J, Rochester J, Thomas W. Assessment of basic surgical skills in the laboratory. Br J Surg 90 (Suppl 1), 90. 2003.

(110) Helmreich RL, Merritt AC, Wilhelm JA. The evolution of Crew Resource Management training in commercial aviation. Int J Aviat Psychol 1999; 9(1):19-32.

(111) Helmreich RL, Schaefer HG. Team performance in the operating theatre. In: Bogner MS, editor. Human Error in Medicine. Hillside, NJ: Lawrence Erlbaum and Associates, 1994: 225-253.

(112) Reason J. Understanding adverse events: human factors. Qual Health Care 1995; 4(2):80-89.

(113) Helmreich RL, Merritt AC. Culture at work: national, organisational and professional influences. Aldershot: Ashgate, 1998.

(114) Helmreich RL. On error management: lessons from aviation. BMJ 2000; 320(7237):781-785.

(115) Bogner MS. Introduction. In: Bogner MS, editor. Human error in Medicine. Hillside, NJ: Lawrence Erlbaum Associates Inc, 1994: 1-11.

(116) Reason JT. Human Error. Cambridge, England: Cambridge University Press, 1990.

(117) Gawande AA, Thomas EJ, Zinner MJ, Brennan TA. The incidence and nature of surgical adverse events in Colorado and Utah in 1992. Surgery 1999; 126(1):66-75.

(118) Kluger MT, Tham EJ, Coleman NA, Runciman WB, Bullock MF. Inadequate pre-operative evaluation and preparation: a review of 197 reports from the Australian incident monitoring study. Anaesthesia 2000; 55(12):1173-1178.

(119) Davis NJ. Anaesthesia related mortality in Australia 1994-1996. Melbourne: Capitol Press, 1999.

(120) Chopra V, Bovill JG, Spierdijk J, Koornneef F. Reported significant observations during anaesthesia: a prospective analysis over an 18-month period. Br J Anaesth 1992; 68(1):13-17.

(121) MacDonald JS, Peterson S. Lethal errors in anaethesiology. Anaesthesiology 63, A497. 1985.

(122) de Leval MR, Carthey J, Wright DJ, Farewell VT, Reason JT. Human factors and cardiac surgery: a multicenter study. J Thorac Cardiovasc Surg 2000; 119(4 Pt 1):661-672.

(123) Gaba DM. Human error in dynamic medical domains. In: Bogner MS, editor. Human Error in Medicine. Hillsdale, New Jersey: Erlbaum, 1994: 197-224.

(124) Le May S, Dupuis G, Harel F, Taillefer MC, Dube S, Hardy JF. Clinimetric scale to measure surgeons' satisfaction with anesthesia services. Can J Anaesth 2000; 47(5):398-405.

(125) Schisler JQ. Implementing continous quality improvement: A private practice's experience. International Anaesthesia Clinics 1992; 30:45-56.

(126) Helmreich RL, Foushee HC. Why cockpit resource management? Empirical and theoretical bases of human factors in training and aviation. In: Wiener E, Kanki BG, Helmreich RL, editors. Cockpit resource managament. San Diego, Calif: Academic Press, 1993: 3-45.

(127) Sexton JB, Thomas EJ, Helmreich RL. [Error, stress and teamwork in medicine and aviation. A cross-sectional study]. Chirurg 2000; 71(6):suppl-42.

(128) Helmreich RL, Merritt A, Sherman P, Gregorich SG, Wiener EL. The Flight Management Attitudes Questionnaire. NASA/ University of Texas/ FAA Technical Report. 93-95. 1993.

(129) Sexton JB, Helmreich R. Communication and teamwork in the surgical operating room. Panel Presentation given at the 2000 Aerospace Medical Association conference in Houston, Texas. The University of Texas at Austin Human Factors Research Project Technical Report, 00-04. 2000.

(130) Helmreich RL, Wilhelm JA, Gregorich SE, Chidster TR. Preliminary results from evaluation of cockpit resource management training: Performance ratings of flight crews. Aviation, Space and Environmental Medicine 1990; 61:576-579.

(131) Wiener EL, Kanki BG, Helmreich RL. Cockpit Resource Management. San Diego, Calif: Academic Press, Inc, 1993.

(132) Helmreich RL, Merritt AC. Safety and error management: The role of Crew Resource Management. In: Hayward BJ, Lowe AR, editors. Aviation Resource Management. Aldershot, UK: Ashgate, 2000: 107-119.

(133) Avermate van JAG, Kruijsen EAC. NOTECHS- The evaluation of non-technical skills of multi-lipot air crew in relation to the JAR-FCL requirements. Research report for the European Commission. DGVII NLR-CR-98443. 1998.

(134) Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH. Anesthesia crisis resource management training: teaching anesthesiologists to handle critical incidents. Aviat Space Environ Med 1992; 63(9):763-770.

(135) Reznek M, Smith-Coggins R, Howard S, Kiran K, Harter P, Sowb Y et al. Emergency Medicine Crisis Resource Management (EMCRM): Pilot Study of a Simulation-based Crisis Management Course for Emergency Medicine. Acad Emerg Med 2003; 10(4):386-389.

(136) Halamek LP, Kaegi DM, Gaba DM, Sowb YA, Smith BC, Smith BE et al. Time for a new paradigm in pediatric medical education: teaching neonatal resuscitation in a simulated delivery room environment. Pediatrics 2000; 106(4):E45.

(137) Holzman RS, Cooper JB, Gaba DM, Philip JH, Small SD, Feinstein D. Anesthesia crisis resource management: real-life simulation training in operating room crises. J Clin Anesth 1995; 7(8):675-687.

(138) Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crises using both technical and behavioral ratings. Anesthesiology 1998; 89(1):8-18.

(139) Chopra V, Gesink BJ, de Jong J, Bovill JG, Spierdijk J, Brand R. Does training on an anaesthesia simulator lead to improvement in performance? Br J Anaesth 1994; 73(3):293-297.

(140) Byrne AJ, Jones JG. Responses to simulated anaesthetic emergencies by anaesthetists with different durations of clinical experience. Br J Anaesth 1997; 78(5):553-556.

(141) Devitt JH, Kurrek MM, Cohen MM, Fish K, Fish P, Murphy PM et al. Testing the raters: inter-rater reliability of standardized anaesthesia simulator performance. Can J Anaesth 1997; 44(9):924-928.

(142) Fletcher GC, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. Br J Anaesth 2002; 88(3):418-429.

(143) Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. Br J Anaesth 2003; 90(5):580-588.

(144) Gaba DM. Improving anesthesiologists' performance by simulating reality. Anesthesiology 1992; 76(4):491-494.

(145) Darzi A, Datta V, Mackay S. The challenge of objective assessment of surgical skill. Am J Surg 2001; 181(6):484-486.

(146) Lekawa M, Shapiro SJ, Gordon LA, Rothbart J, Hiatt JR. The laparoscopic learning curve. Surg Laparosc Endosc 1995; 5(6):455-458.

(147) Fleshman JW, Fry RD, Birnbaum EH, Kodner IJ. Laparoscopic-assisted and minilaparotomy approaches to colorectal diseases are similar in early outcome. Dis Colon Rectum 1996; 39(1):15-22.

(148) Simons AJ, Anthone GJ, Ortega AE, Franklin M, Fleshman J, Geis WP et al. Laparoscopic-assisted colectomy learning curve. Dis Colon Rectum 1995; 38(6):600-603.

(149) Royston CM, Lansdown MR, Brough WA. Teaching laparoscopic surgery: the need for guidelines. BMJ 1994; 308(6935):1023-1025.

(150) Bailey RW, Imbembo AL, Zucker KA. Establishment of a laparoscopic cholecystectomy training program. Am Surg 1991; 57(4):231-236.

(151) Smith CD, Farrell TM, McNatt SS, Metreveli RE. Assessing laparoscopic manipulative skills. Am J Surg 2001; 181(6):547-550.

(152) Derossis AM, Bothwell J, Sigman HH, Fried GM. The effect of practice on performance in a laparoscopic simulator. Surg Endosc 1998; 12(9):1117-1120.

(153) Chung JY, Sackier JM. A method of objectively evaluating improvements in laparoscopic skills. Surg Endosc 1998; 12(9):1111-1116.

(154) Pearson AM, Gallagher AG, Rosser JC, Satava RM. Evaluation of structured and quantitative training methods for teaching intracorporeal knot tying. Surg Endosc 2002; 16(1):130-137.

(155) Cuschieri A, Szabo Z. Intracorporeal knots in endoscopic surgery. In: Cuschieri A, Szabo Z, editors. Tissue approximation in endoscopic surgery. Oxford, UK: ISIS Medical, 1995.

(156) Mori T, Hatano N, Maruyama S, Atomi Y. Significance of "hands-on training" in laparoscopic surgery. Surg Endosc 1998; 12(3):256-260.

(157) Rosser JC, Jr., Rosser LE, Savalgi RS. Objective evaluation of a laparoscopic surgical skill program for residents and senior surgeons. Arch Surg 1998; 133(6):657-661.

(158) Munz Y, Kumar B, Moorthy K, Bann S, Darzi A. Laparoscopic virtual reality and box trainers-is one superior than the other? Surgical Endoscopy. In press.

(159) McGraw RC, O'Connor HM. Standardized patients in the early acquisition of clinical skills. Med Educ 1999; 33(8):572-578.

(160) Kneebone R, Kidd J, Nestel D, Asvall S, Paraskeva P, Darzi A. An innovative model for teaching and learning clinical procedures. Med Educ 2002; 36(7):628-634.

(161) Shah SG, Thomas-Gibson S, Brooker JC, Suzuki N, Williams CB, Thapar C et al. Use of video and magnetic endoscope imaging for rating competence at colonoscopy: validation of a measurement tool. Gastrointest Endosc 2002; 56(4):568-573.

(162) MacDonald J, Ketchum J, Williams RG, Rogers LQ. A lay person versus a trained endoscopist. Surg Endosc 2003; 17(6):896-898.

(163) Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 57[2], 214-218. 2003.

(164) Marshall JB, Brown DN. Photodocumentation of total colonoscopy: how successful are endoscopists? Do reviewers agree? Gastrointest Endosc 1996; 44(3):243-248.

(165) Adam IJ, Ali Z, Shorthouse AJ. Inadequacy of colonoscopy revealed by three-dimensional electromagnetic imaging. Dis Colon Rectum 2001; 44(7):978-983.

(166) Kneebone R, Nestel D, Moorthy K, Taylor P, Bann S, Munz Y et al. Learning the skills of flexible sigmoidoscopy- the wider perspective. Medical Education. In press.

(167) RS Jensen, editor. The effectiveness of training programs for preventing aircrew 'error'. Columbus OH: Ohio State University, 1991.

(168) Lossing AG, Hatswell EM, Gilas T, Reznick RK, Smith LC. A technical-skills course for 1st-year residents in general surgery: a descriptive study. Can J Surg 1992; 35(5):536-540.

(169) Pandey V, Moorthy K, Wolfe J, Munz Y, Jackson M, Darzi A. **Procedural rating scales increase objectivity in surgical assessment**. Br J Surg 90 (Suppl 1), 14. 2003.

(170) Operative Competence ('Op Comp') forms. www.jchst.org . 2003.

(171) Calland JF, Guerlain S, Adams RB, Tribble CG, Foley E, Chekan EG. A systems approach to surgical safety. Surg Endosc 2002; 16(6):1005-1014.

(172) Cohen J. Statsitical power analysis for behavioral sciences. 2nd ed. New Jersey: Lawrence Erlbaum, 1988.

(173) Gaba DM, Howard SK, Small SD. Situation awareness in anesthesiology. Hum Factors 1995; 37(1):20-31.

(174) Xiao Y, Mackenzie CF. Collaboration in complex medical systems. Collaborative crew performance in complex operational systems. NATO Human facotrs and Medicine symposium, April 20-24, 1998. Neuilly sur Seire: NATO, 1998, 1998: 4-1-4-10.

(175) Xiao Y, Hunter WA, Mackenzie CF, Jefferies NJ, Horst RL. Task complexity in emergency medical care and its implications for team coordination. LOTAS

Group. Level One Trauma Anesthesia Simulation. Hum Factors 1996; 38(4):636-645.

(176) Kim AW, McCarthy WJ, III, Maxhimer JB, Quiros RM, Hollinger EF, Doolas A et al. Vascular complications associated with pancreaticoduodenectomy adversely affect clinical outcome. Surgery 2002; 132(4):738-744.

(177) Cassar K, Munro A. Iatrogenic splenic injury. J R Coll Surg Edinb 2002; 47(6):731-741.

(178) Miller GV, Lewis WG, Sainsbury JR, Macdonald RC. Morbidity of varicose vein surgery: auditing the benefit of changing clinical practice. Ann R Coll Surg Engl 1996; 78(4):345-349.

(179) Orlady HW, Foushee HC, editors. Cockpit Resource Management: background and overview. California: Moffet Field, 2003.

(180) Gaba DM, DeAnda A. The response of anesthesia trainees to simulated critical incidents. Anesth Analg 1989; 68(4):444-451.

(181) Abrahamson S, Denson JS, Wolf RM. Effectiveness of a simulator in training anesthesiology residents. J Med Educ 1969; 44(6):515-519.

(182) Murray DJ. Clinical simulation: technical novelty or innovation in education. Anesthesiology 1998; 89(1):1-2.

(183) Jha AK, Duncan BW, Bates WB. Simulator-based training and patient safety. http://www.ahcpr.gov/clinic.ptsafety/chap45.htm . 2003.

(184) Workload measurement. In: Farmer E, van Rooij J, Riemersma J, Jorna P, Moraal J, editors. Handbook of simulator-based training. Aldershot, UK: Ashgate, 1999: 261-268.

(185) Performance measurement. In: Farmer E, van Rooij J, Riemersma J, Jorna P, Moraal J, editors. Handbook of simulator-based training. Aldershot, UK: Ashgate, 1999: 247-260.

(186) Helmreich RL, Wilhelm JA. Outcomes of crew resource management training. Int J Aviat Psychol 1991; 1(4):287-300.

APPENDIX 1

Checklists

Small Bowel Anastomosis:
Interrupted end to end single layer anastomosis. Score one point for each correctly performed action

Tape No:_____ Item No: _____
Assessor: _____ (initials only)

	Procedural Step	Correctly performed	Incorrectly performed
1	Selects appropriate instruments (non-tooth forceps)		
2	Selects appropriate suture (absorbable)		
3	Bowel orientated- no twisting, correct stay placement		
4	Stay sutures held with artery forceps		
5	Correct needle holding technique		
6	Needle driver stabilized with good hand position		
7	Needle enters bowel at right angles< 80% of bites		
8	Single attempt at passage through bowel >90% of bites		
9	Follow through on curve of needle on entrance >90% bites		
10	Follow through on curve of needle on exit > 90% bites		
11	Adequate bites taken (>3mm from edge of bowel)		
12	Seromuscular sutures only >80% of bites		
13	Minimal damage with forceps		
14	Use of forceps to handle needle		
15	Suture spacing 3-5 mm		
16	Inverting of bowel wall with sutures		
17	Equal bites each side >80% of time		
18	Individual bites each side >90% of bites		
19	Square knots		
20	Knots placed to one side of suture line		
21	At least three throws per knot		
22	Suture cut to appropriate length		
23	Appose bowel and knots tied with correct tension		

APPENDIX 2

Global rating scale

	1	2	3	4	5
Respect For Tissue	Frequently used unnecessary force on tissue or caused damage by inappropriate use of instruments		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled tissues appropriately with minimal damage
Time & Motion	Many unnecessary moves		Efficient time/motion but some unnecessary moves		Economy of movement and maximum efficiency
Instrument Handling	Repeatedly makes tentative or awkward moves with instruments		Competent use of instruments although occasionally appeared stiff or awkward		Fluid moves with instruments and no awkwardness
Knowledge of Instruments	Frequently asked for the wrong instrument or used an inappropriate instrument		Knew the names of most instruments and used appropriate instrument or the task		Obviously familiar with the instruments required and their names
Use of Assistants	Consistently placed assistants poorly or failed to use assistants		Good use of assistants most of the time		Strategically used assistant to the best advantage at all times
Flow of Operation & Forward Planning	Frequently stopped operating or needed to discuss next move		Demonstrated ability for forward planning with steady progression of operative procedure		Obviously planned course of operation with effortless flow from one move to the next
Knowledge of Specific Procedure	Deficient knowledge. Needed specific instruction at most operative steps		Knew all important aspects of the operation		Demonstrated familiarity with all aspects of the operation

APPENDIX 3

NOTECHS Categories and elements

CATEGORY	ELEMENT
Co-operation	• Team building and maintenance • Consideration of others • Support of others • Conflict solving
Leadership and management skills	• Use of authority and assertiveness • Providing and maintaining standards • Planning and co-ordination • Workload management
Situation awareness	• Awareness of aircraft systems • Awareness of external environment • Awareness of time
Decision making	• Problem definition and diagnosis • Option generation • Risk assessment • Outcome review

APPENDIX 4

ANTS system (Anaesthesia Non-Technical Skills)

Categories	Elements
Task management	• Planning and preparing • Priortizing • Providing and maintaining standards • Identifying and utilizing resources
Team working	• Co-ordinating activities with team members • Exchanging information • Using authority and assertiveness • Assessing capabilities • Supporting others
Situation awareness	• Gathering information • Recognizing and understanding • anticipating
Decision making	• Identifying options • Balancing risks and selecting options • Re-evaluating

APPENDIX 5

Scenario and consent:

Brief to the surgeon:

The patient is a 35 year old male who needs a right sided high tie. Please do only the high tie and not the stripping.

You have seen the patient earlier and taken consent and marked the right side.

The patient is prepped and draped by your assistant.

Your assistant is a locum SHO whom you have never met before and whose clinical experience is unknown to you. The scrub nurse is an agency nurse you haven't met before.

There are sterile gloves and gowns lying on a side trolley. You have already scrubbed and now only need to proceed to wearing the gown and gloves taking all sterile precautions.

The pubic tubercle and the anterior superior iliac spine have been marked out for you and so has the SF junction.

Please take an incision and perform the high tie.

Please do not take a transfixtion suture and please just ligate the junction

There is no need to close the fascia. Please proceed with the skin closure.

There is no need to infiltrate the tissues with local anaesthetic prior to closure.

Please close the skin with interrupted 3-0 nylon which has been provided.

Your consultant is available on the phone if you need him. Your co-operation and willingness to participate is most appreciated. Please sign the consent form below.

Consent:
I hereby agree to participate in the simulated procedure in the virtual theatre. The data from the simulation will be used for research purposes and will in no way affect my clinical record or career. The data will be anonymous and accessible only to the research team. I undertake to keep the scenarios confidential and not to divulge its details to anybody else.

Signature:
Name:
Date

Appendix 6

Modified global rating scale

Respect for tissue

1	2	3	4	5
Frequently used unnecessary force on tissue or cause damage by inappropriate instrument use		Careful handling of tissue but occasionally caused inadvertent damage		Consistently handled appropriately with minimal damage to tissue

Time and motion

1	2	3	4	5
Many unnecessary moves		Efficient time & motion but some unnecessary moves		Clear economy of movement and maximum efficiency

Instrument handling

1	2	3	4	5
Repeatedly makes awkward or tentative moves with instruments through inappropriate use		Competent use of instruments but occasionally appeared stiff or awkward		Fluid movements with instruments and no stiffness awkwardness

Suture handling

1	2	3	4	5
Awkward and unsure with repeated entanglement, poor knot tying and inability to maintain tension		Careful and slow with majority of knots placed correctly with appropriate tension		Excellent suture control with correct placement of knots and correct tension

Flow of operation

1	2	3	4	5
Frequently stopped operating and seemed unsure of next move		Demonstrated some forward planning and reasonable progression of procedure		Obviously planned operation with efficiency from one move to another

Knowledge of procedure

1	2	3	4	5
Insufficient knowledge Looked unsure and hesitant		Knew all important steps of operation		Demonstrated familiarity with all steps of operation

Overall performance

1	2	3	4	5
Very poor		Competent		Clearly superior

Quality of final product

1	2	3	4	5
Very poor		Competent		Clearly superior

TOTAL SCORE_____

APPENDIX 7:

ICEPS: Imperial College of Evaluation of Procedure-specific skills

	1	2	3	4	5
Incision	1 Does not use surface landmarks. Inappropriate placement of incision. Poor handling of scalpel	2	3 Appropriate incision in terms of location and size. Looked at ease with scalpel	4	5 Uses surface landmarks to make an appropriately located and sized incision Handled scalpel expertly
Dissection	1 Appeared unsure and excessively hesitant whilst dissecting. Caused trauma to tissues. Did not dissect into the correct anatomical plane.	2	3 Controlled and safe dissection into correct anatomical plane. Caused minimal trauma of tissues. Used instrument satisfactorily whilst dissecting.	4	5 Superior and atraumatic dissection into the correct anatomical plane. Confident handling of instruments whilst dissecting.
Retraction	1 Clumsy use of retractors. Did not allow visualisation of important structures making frequent changes to retractor setting.	2	3 Good use of retraction allowing visualisation of major structures. Had to change retractor position to visualise other structures	4	5 Excellent use of retractors. Allowed good visualisation of all necessary structures. Atrauamtic
Tributaries	1 Could not or did not try to identify any tributaries	2	3 Identified all known tributaries. Did not seek other vessels	4	5 Identified all known tributaries. Sought other possible tributaries.
Haemostasis	1 Poor quality of knot tying, Knots frequently slipped or was excessively traumatic to vessels	2	3 Competent knot tying. Minimal trauma to vessels. Minimal blood loss.	4	5 Superior knot tying. Atraumatic. No knot slippage
SFJ Clearance	1 Did not identify the Saphenofemoral junction or excessively traumatic dissection around that vessel	2	3 Identified the Saphenofemoral junction. Safely dissected tissues away from vessel. Reasonable clearance of vessel. Minimal trauma	4	5 Identified the Saphenofemoral junction. Expert dissection of tissues off the vessels Atraumatic. Cleared well proximally and distally.
SFJ Ligation	1 Did not ligate the SFJ or ligated CFV or caused excessive encroachment onto CFV after SFJ Ligation	2	3 Good knot tying whilst ligating the SFJ. Minimal encroachment onto CFV following SFJ ligation.	4	5 Excellent safe and secure ligation of the SFJ. Flush ligation with no encroachment onto CFV

APPENDIX 8

Assessment of technical ability to control the bleeding

	1	2	3	4	5
Identification of aetiology	1 Does not or cannot identify source of bleeding	2	3 Recognises severity but has some difficulty in localising source	4	5 Early and accurate diagnosis of severity. Uses suction. Performs an organised check efficiently
Exposure	1 Inappropriate exposure. Inadequate use of retractors and/or assistants	2	3 Adequate incision, however repeatedly adjusting retractors	4	5 Appropriate incision (extended if necessary). Efficient use of retractors and assistants.
Control of bleeding	1 Doesn't control bleeding or uses hazardous manoeuvres to control bleeding	2	3 Adequate measures demonstrated but sometimes loses control	4	5 Proximal and distal control with finger pressure or vascular clamps
Clear vision of laceration	1 Does not adequately expose laceration prior to closure. Vision obscured by blood/instruments	2	3 Maintains clear vision with some difficulties	4	5 Maintains dry field and clearly identifies laceration prior to closure
Technique of closure	1 Poor technique for closure resulting poor technical result	2	3 Adequate venotomy closure technique	4	5 Continuous sutures starting at apex, adequate and well spaced bites
Knowledge on instruments	1 Uses the wrong instruments or inappropriate use of instruments	2	3 Familiar with all instruments and uses them appropritely	4	5 Demonstrates clear knowledge of all the correct instruments, e.g. vascular clamps, needle holder and sutures
Overall Performance	1 Unsure how to deal with this lesion, hesitant and does not mobilize assistance. Likely to result in poor outcome	2	3 Adequate identification and management of haemorrhage with some difficulties encountered .	4	5 Efficiently and confidently identifies and deals with haemorrhage without complication

Total score

APPENDIX 9: VIRTUAL OPERATING THEATRE HUMAN FACTORS RATING SCALE

NA-not applicable

1 Not done/	2 Not done well	3	4	5	6 Done very well

CATEGORY	ELEMENT							
COMMUNICATION AND INTERACTION	(a) Instructions to assistant- clear and polite	NA	1	2	3	4	5	6
	(b) Waited for acknowledgment from the assistant	NA	1	2	3	4	5	6
	(c) Instructions to scrub nurse- clear and polite	NA	1	2	3	4	5	6
	(d) Waited for acknowledgement from the scrub nurse	NA	1	2	3	4	5	6
VIGILANCE/ SITUATION AWARENESS	(a) Monitored patient's parameters throughout the procedure	NA	1	2	3	4	5	6
	(b) Awareness of anaesthetist	NA	1	2	3	4	5	6
	(c) Actively initiates communication with anaesthetist during crisis periods	NA	1	2	3	4	5	6
TEAM SKILLS	(a) Maintains a positive rapport with the whole team	NA	1	2	3	4	5	6
	(b)Open to opinions from other team members	NA	1	2	3	4	5	6
	(c) Acknowledges the contribution made by other team members	NA	1	2	3	4	5	6
	(d) Supportive of other team members	NA	1	2	3	4	5	6
	(e) Conflict handling eg. concentrates on what is right rather than who is right	NA	1	2	3	4	5	6
LEADERSHIP AND MANAGEMENT SKILLS	(a) Adherence to best practise during the procedure eg. does not permit corner cutting by self or team	NA	1	2	3	4	5	6
	(b) Time management eg. appropriate time allocation without being too slow or rushing team members	NA	1	2	3	4	5	6
	(c) Resource utilisation i.e. appropriate task-load distribution and delegation of responsibilities	NA	1	2	3	4	5	6
	(d) Debriefing the team i.e. provides details and feedback to the entire team about the procedure	NA	1	2	3	4	5	6
	(e) Authority/ assertiveness	NA	1	2	3	4	5	6
DECISION MAKING- CRISIS	(a) Prompt identification of the problem	NA	1	2	3	4	5	6
	(b) Informed team members- promptly, clearly and to all team members	NA	1	2	3	4	5	6
	(c) Outlines strategy/ institutes a plan i.e. asks scrub nurse for suction, instruments, suture material	NA	1	2	3	4	5	6
	(d) Anticipates potential problems and prepares a contingency plan eg. asks anaesthetist to order blood, calls for help	NA	1	2	3	4	5	6
	(e) Option generation- takes the help of the team (seeks team opinion)	NA	1	2	3	4	5	6

APPENDIX 10

Questionnaire to assess the perception of the participants

Level of agreement

1 Do not agree at all	2	3	4	5	6 Completely agree

The synthetic model is a realistic representation of the operation	1	2	3	4	5	6
The simulated environment is a realistic representation of an operating theatre eg communication, interaction between team	1	2	3	4	5	6
The bleeding scenario is a realistic representation of real situation	1	2	3	4	5	6
The simulated environment is a good training opportunity in technical skills eg how to deal with bleeding	1	2	3	4	5	6
The simulated environment is a good training environment in the team skills (communication, leadership etc) required to successfully manage the bleeding	1	2	3	4	5	6
The simulation is a good method for assessing my technical skills	1	2	3	4	5	6
The simulation is a good method for assessing my team skills	1	2	3	4	5	6
I would behave in the same way in real procedures	1	2	3	4	5	6
I would benefit by practising this scenario again	1	2	3	4	5	6
I found the feedback at the end of the simulation of benefit in terms of technical skills	1	2	3	4	5	6
I found the feedback at the end of the simulation of benefit in terms of team skills	1	2	3	4	5	6