

Quality Assurance of Training in Diagnostic and Therapeutic Gastrointestinal Endoscopy

Dr Adam Vaughan Haycock
BSc MBBS MRCP

Thesis Submitted for the degree of
Doctor of Medicine (M.D.)

Imperial College London
Division of Surgery, Obstetrics, Reproductive
Biology & Anaesthetics

2010

Abstract

Previous evidence has shown that standards of performance in gastrointestinal endoscopy are variable and that there are disparities in training outcomes. Many changes have been made recently to both training and assessment of endoscopy in the UK. However, no prospective methods of evaluating their outcome have been put in place. The aims of this research were to evaluate current and new training processes and assessments in order to quality assure the outcomes and improve the training process. Two audits were undertaken demonstrating improvements in colonoscopy training over 5 years within a single region and in trainee perceptions of their training nationally. Two studies were done investigating a novel computer colonoscopy simulator for assessment of colonoscopic skills, demonstrating excellent construct validity. A multi-centre randomised controlled trial evaluated the use of this simulator in novice training, which was shown to be equivalent to standard bed-side training with a high degree of skills transfer to real-life colonoscopy. Assessment tools for therapeutic endoscopic procedures were developed, validated and used to quality assure a course in therapeutic endoscopy. This course resulted in significant improvements in practical skills for three of the four therapeutic procedures following training. Web-based training and assessment modules for lesion recognition at capsule endoscopy were developed, validated and piloted. This demonstrated the effectiveness of using new training methodologies for skills improvement in this area. A training course for radiographers in virtual colonoscopy was developed and the training evaluated. This demonstrated competence in practical performance and improvements in knowledge and interpretative skill. Finally, two qualitative studies on non-technical skills in endoscopy were undertaken in order to widen the assessment domains from purely knowledge and skill. An interview study provided the basis for development of a non-technical skills taxonomy and a video-analysis study resulted in production of a marker system for professional behaviour within gastrointestinal endoscopy.

Dedication

*To all those who have given me encouragement and support
throughout, most especially my parents, my wonderful wife Sarah,
and my lovely daughter Olivia.*

Acknowledgements

I am greatly indebted to a large number of people for their help and support in this work. I have benefited from their enthusiasm, experience and imagination. I am extremely fortunate that during my time at the Wolfson Unit I have also made many good friends and colleagues.

In particular I wish to acknowledge my supervisors Mr Paris Tekkis, Dr Siwan Thomas-Gibson, and Dr David Burling. I am deeply appreciative of all their support and assistance. I would especially like to thank Siwan, whose encouragement, enthusiasm and commitment has been exceptional and entirely unrecompensed. I am eternally grateful for allowing me to 'twist her arm' into supervising me despite initial reservations. I would also like to thank Dr Brian Saunders and Dr Christopher Williams for placing their faith in me and supporting this research when so many other priorities could have taken precedence.

I would not have been able to complete this work without the unending support from all the individuals in the Wolfson unit. Christopher Williams for his encouragement and enthusiasm in the ongoing simulator development; Steve Preston, whose IT and media skills were essential to many of the projects; Eric Tripoli, who never said 'No'; Maggie Vance, whose encouragement and support never wavered; Jean Mannings, whose organisational skills made light of many tasks; Philippa Youd and Ana Ignjatovic, my delightful comrades in training; James East, for his inspirational academic achievements and Aymer Postgate for keeping me grounded.

I wish to thank all those who contributed to a number of the projects: Chris Fraser, Ripple Man, Nicky Palmer, Noriko Suzuki, Ed Despott, Aine Fitzpatrick, Suresh Sivanesan, Mariann Baulf. I am also greatly appreciative of the continuous support given by the nursing and clerical staff within the unit.

I would like to thank many people outside the unit who contributed to many of the projects, particularly the ESCAPE trialists, Jelle Haringsma, Arjun Koch, Evelien Dekker, Foke van Delft, Lucio Petruziello, and Mark Jackson. I am deeply grateful to

those consultants who were willing to recount difficult incidents for the interview study; Sue Clark, Peter McDonald, Jeremy Nightingale, Simon Gabe, Meron Jacyna, Jonno Hoare, James Lindsay, Mark Appleyard.

I am grateful to Paul Basset for his statistical advice and analysis in a number of the projects. They greatly contributed to the interpretation and evaluation of results.

I would like to thank Olympus Keymed and Schrodgers for their support in funding this research.

I am extremely grateful to all the trainees who allowed themselves to be assessed, videoed and generally scrutinised in the name of research. There are too many to list here, but their willingness to participate was vital. I would finally like to thank all the patients who consented to having a trainee perform their procedure, which in this day and age shows remarkable trust and understanding.

Table of Contents

ABSTRACT	2
DEDICATION	3
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS	6
FIGURES.....	14
TABLES	16
APPENDICES	18
CHAPTER 1: INTRODUCTION.....	21
1.1 Overview.....	21
1.2 Endoscopy training: recent developments and drivers for change	21
1.2.1 Poor performance outcomes	21
1.2.2 Disparities in training outcomes	22
1.2.3 Changes to doctors' working lives	23
1.2.4 NHS Bowel Cancer Screening (BCS) Programme	24
1.2.5 National training project.....	24
1.2.5.1 <i>Aims and delivery</i>	25
1.2.5.2 <i>Outcome measures</i>	25
1.2.6 Assessment of competency	26
1.3 Quality Assurance in education	27
1.3.1 Background.....	27
1.3.2 Adult learning theory	27
1.3.2.1 <i>Bloom's Taxonomy</i>	28
1.3.2.2 <i>Experiential learning</i>	29
1.3.2.3 <i>Learning style</i>	30
1.3.2.4 <i>Competency and expertise</i>	31
1.3.2.5 <i>Learning and training</i>	32

1.3.3 Quality assurance process	32
1.3.3.1 Assessment	33
1.3.3.2 Evaluation	34
1.3.4 Application of educational theory	35
1.4 Quality Assurance in endoscopy training.....	36
1.4.1 Background.....	36
1.4.2 Flexible sigmoidoscopy.....	36
1.4.3 Colonoscopy	37
1.4.4 Upper GI endoscopy	37
1.4.5 ERCP and other therapeutics.....	38
1.4.6 Trainee's learning experiences	39
1.4.7 The trainer	39
1.4.7.1 Qualities of an endoscopy trainer.....	40
1.5 Use of simulation in endoscopy training and assessment.....	42
1.5.1 Background.....	42
1.5.1.1 Use of simulation for training	42
1.5.1.2 Use of simulation for assessment	42
1.5.2 Current types of simulators.....	43
1.5.3 Evidence for use of simulation	44
1.5.3.1 Types of study	44
1.5.3.2 Lower GI endoscopy	45
1.5.3.3 Upper GI endoscopy	47
1.5.3.4 ERCP.....	47
1.5.3.5 Therapeutic endoscopy.....	48
1.5.3.6 Other.....	48
1.5.4 Quality assurance needs	48
1.6 New technologies for endoscopy training	49
1.6.1 3-D imager	49
1.6.2 E-learning.....	50
1.6.2.1 Capsule endoscopy training	50
1.6.3 Virtual colonoscopy	51
1.7 Attitudes, behaviours and professionalism.....	53
1.7.1 Background.....	53
1.7.2 Human factors research	54
1.7.3 Importance in medicine.....	55
1.7.4 Quality Assurance needs.....	55
1.8 Summary	56

1.9 Hypothesis	57
CHAPTER 2: EVALUATION OF CURRENT ENDOSCOPY TRAINING.....	59
2.1 Background	59
2.2 Comparative survey on colonoscopy training	59
2.2.1 Background and Aims	59
2.2.2 Methods	60
2.2.2.1 Statistical analysis	60
2.2.3 Results	60
2.2.4 Discussion	64
2.3 National training survey	65
2.3.1 Background and Aims	65
2.3.2 Methods	66
2.3.3 Results	66
2.3.4 Discussion	70
2.4 Conclusions	71
CHAPTER 3: CONSTRUCT VALIDITY OF A COLONOSCOPY SIMULATOR.....	73
3.1 Background and Aims	73
3.2 Methods	75
3.2.1 Statistical analysis	76
3.3 Results	76
3.3.1 Construct validity	76
3.3.2 Face validity	78
3.3.3 3-D Imager	78
3.4 Discussion	82
3.5 Conclusions	85
CHAPTER 4: USE OF A COLONOSCOPY SIMULATOR IN TRAINING	87
4.1 Background and Aims	87
4.2 Methods	88
4.2.1 Participants	88

4.2.2 Pre-training	89
4.2.2.1 Questionnaire.....	89
4.2.2.2 Simulation pre-training assessment.....	89
4.2.3 Training	89
4.2.3.1 Simulator training.....	89
4.2.3.2 Patient-based training	89
4.2.4 Post training.....	90
4.2.4.1 Simulator assessment.....	90
4.2.4.2 Patient-based assessment	90
4.2.4.3 Feedback	91
4.2.5 Outcome measures.....	91
4.2.6 Statistical analysis.....	91
4.3 Results	92
4.3.1 Post-training simulator assessment.....	96
4.3.2 Patient-based assessment	96
4.3.3 Feedback	100
4.4 Discussion	100
4.5 Conclusions	102
 CHAPTER 5: CRITERION VALIDITY OF A COLONOSCOPY SIMULATOR	 103
5.1 Background and Aims	103
5.2 Methods	103
5.3 Results	104
5.3.1 Discussion	105
5.4 Conclusions.....	108
 CHAPTER 6: THERAPEUTIC ENDOSCOPY TRAINING.....	 109
6.1 Background and Aims	109
6.2 Methods.....	110
6.2.1 Participants	110
6.2.2 Interventions	110
6.2.3 Outcome measures.....	112
6.2.4 Statistical analysis.....	113

6.3 Results	114
6.3.1 Study participants	114
6.3.2 Expectations of trainees	117
6.3.3 Effect of lectures	117
6.3.4 Effect of hands-on skills training	117
6.4 Discussion	120
6.5 Conclusions	123
 CHAPTER 7: E-LEARNING IN CAPSULE ENDOSCOPY	 125
7.1 Background and Aims	125
7.2 Methods	126
7.2.1 Module development	127
7.2.2 Module validation	129
7.2.3 Web-based training pilot	129
7.2.4 Statistical analysis	130
7.3 Results	131
7.3.1 Validation study	131
7.3.1.1 <i>Participants</i>	131
7.3.1.2 <i>Construct validity</i>	132
7.3.1.3 <i>Content validity</i>	133
7.3.1.4 <i>Feedback</i>	134
7.3.2 Web-based training pilot	134
7.3.2.1 <i>Participants</i>	134
7.3.2.2 <i>Content validity</i>	135
7.3.2.3 <i>Feedback</i>	135
7.4 Discussion	136
7.5 Conclusions	137
 CHAPTER 8: VIRTUAL COLONOSCOPY TRAINING FOR RADIOGRAPHERS	 139
8.1 Background and Aims	139
8.2 Methods	140
8.2.1 Participant details	140
8.2.2 Course design	140
8.2.3 Course evaluation	141

8.2.4 Statistical analysis	145
8.3 Results	145
8.3.1 Reaction (feedback)	145
8.3.2 Learning (knowledge)	145
8.3.3 Learning (skills)	146
8.3.4 Behaviour (image interpretation).....	147
8.4 Discussion	147
8.5 Conclusions	149

CHAPTER 9: PROFESSIONALISM IN ENDOSCOPY – NON-TECHNICAL SKILLS TAXONOMY DEVELOPMENT

9.1 Background and Aims	151
9.2 Literature review	151
9.2.1 Current non-technical skills taxonomies and behavioural marker systems	151
9.2.2 Non-technical skills identified	153
9.2.3 Development of behavioural marker systems.....	153
9.3 Methods	155
9.3.1 Triangulation	155
9.3.1.1 Knowledge audit	156

THE FOLLOW-UP PROBE QUESTIONS THEN AIMED TO ELICIT DETAILS REGARDING SKILLS DEVELOPMENT AND USE OF NON-TECHNICAL SKILLS IN CRISIS SITUATIONS (SEE

9.3.1.2 Sorting task	157
9.3.1.3 Rating task	157
9.3.1.4 Critical incident analysis.....	157
9.3.2 Phases of data collection.....	158
9.3.2.1 Focus group	158
9.3.2.2 Semi-structured interviews	158
9.4 Analysis	159
9.4.1 Familiarisation	160
9.4.2 Identifying a thematic framework	160
9.4.3 Indexing	161
9.4.4 Charting	161

9.5 Results	161
9.5.1 Participants	161
9.5.2 Triangulation tasks	162
9.5.2.1 Knowledge audit	162
9.5.2.2 Sorting and rating tasks	163
9.5.3 Critical incident analysis	168
9.5.3.1 Familiarisation	168
9.5.3.2 Identifying a thematic framework	170
9.5.3.3 Indexing	191
9.5.3.4 Charting	191
9.6 Discussion	193
9.6.1 Strengths of the study	193
9.6.2 Methodological issues	194
9.6.3 Validation of results	195
9.7 Conclusions	195
 CHAPTER 10: PROFESSIONALISM IN ENDOSCOPY - BEHAVIOURAL MARKER SYSTEM DEVELOPMENT	 197
10.1 Background and Aims	197
10.2 Methods	198
10.2.1 Selection of cases	198
10.2.2 Video recordings	199
10.2.3 Analysis	200
10.2.3.1 Coding the video	200
10.2.3.2 Interpreting the behaviours	200
10.2.4 Descriptor development	200
10.3 Results	200
10.3.1 Participants	200
10.3.2 Initial coding	201
10.3.3 Re-coding based on 'purpose'	201
10.3.3.1 Map to taxonomy	204
10.3.4 Behavioural marker system	206
10.3.5 Descriptors and example behaviours	208
10.3.5.1 Communication and Teamwork	208
10.3.5.2 Situation Awareness	211
10.3.5.3 Leadership	213

10.3.5.4 Judgement & Decision Making.....	215
10.3.6 Rating scale.....	217
10.4 Discussion	217
10.4.1 Strengths of the study.....	217
10.4.2 Methodological issues	218
10.4.3 Validation	219
10.5 Conclusions.....	219
 CHAPTER 11: DISCUSSION	 221
11.1 Summary	221
11.2 Quality assurance process	225
11.3 Simulation in endoscopy.....	225
11.4 E-learning for diagnostic procedures	227
11.5 Attitudes, non-technical skills and professionalism.....	228
11.6 Future work.....	229
 BIBLIOGRAPHY	 233
 PUBLICATIONS ARISING FROM THIS RESEARCH	 256
 APPENDIX 1 LITERATURE SEARCH ON SIMULATION IN GI ENDOSCOPY ..	 258
 APPENDIX 2 THERAPEUTIC CHECKLISTS AND GLOBAL SCORE SHEET ...	 269
 APPENDIX 3 VICTOR COURSE DETAILS AND LEARNING OUTCOMES	 274
 APPENDIX 4 VC CHECKLIST AND GLOBAL SCORE SHEET.....	 277
 APPENDIX 5 LITERATURE SEARCH FOR NON-TECHNICAL SKILLS.....	 279
 APPENDIX 6 INTERVIEW SCHEDULE FOR NON-TECHNICAL SKILLS	 283

Figures

Figure 1 Overview of thesis	20
Figure 2 Bloom's taxonomy of learning domains	29
Figure 3 Schön's cycle of experiential learning	29
Figure 4 Experiential learning cycle and learning styles.	30
Figure 5 The process of achieving competence	31
Figure 6 A simplified quality feedback loop for training	32
Figure 7 Miller's framework for clinical assessment.....	33
Figure 8 Kirkpatrick's four levels of training evaluation	35
Figure 9 Schematic model of the endoscopy learning experience (from Thuraisingam et al, 2006).....	39
Figure 10 A model of teaching endoscopy (courtesy of C. Wells).....	41
Figure 11 Theory of planned behaviour	54
Figure 12 Percentage of trainees who had ever received formal training in colonoscopy, polypectomy or extubation technique	62
Figure 13 Supervision of the trainees' first 100 colonoscopies	62
Figure 14 Level of assistance given to resolve problems during a trainee's first 100 colonoscopies.....	63
Figure 15 Survey responses by region	67
Figure 16 Survey responses by training year (start date)	67
Figure 17 Number of endoscopy sessions per week	68
Figure 18 Actual occurrence of training lists	69
Figure 19 Attendance at endoscopy training courses	69
Figure 20 External view (A) and screen view (B) of the Olympus ENDO TS-1 colonoscopy simulator	74
Figure 21 Graph of time to caecum for all cases combined.....	83
Figure 22 Flow diagram of participants through the study	93
Figure 23 Distribution of case difficulty for patient-based assessments.....	99
Figure 24 Flow diagram of the study protocol	112
Figure 25 Flow diagram of participants through each stage of the study.....	115
Figure 26 Graph of change in overall score between initial and re-assessments for each therapeutic modality.....	118

Figure 27 Screenshot from capsule endoscopy training module	128
Figure 28 Screenshot from capsule endoscopy test module	128
Figure 29 Pre and post training test performance in validation study	133
Figure 30 Summary of ViCToR course evaluation.....	143
Figure 31 Difference in mean MCQ score pre to post training for consecutive blocks of trainees.....	146
Figure 32 Focus group sorting task results - endoscopists.....	165
Figure 33 Focus group sorting task results – non-endoscopists.....	166
Figure 34 Example of video recording ‘Quad-split’ view	199
Figure 35 Map of non-technical skills to behavioural elements and categories	205
Figure 36 Behavioural marker system structure at category and element level....	207
Figure 37 Overview of Discussion	221

Tables

Table 1	Reported number of colonoscopies performed as a trainee prior to performing colonoscopy unsupervised	61
Table 2	General measures	79
Table 3	Technical measures.....	79
Table 4	Time taken to landmarks and time taken to negotiate flexures	80
Table 5	Sigmoid N-looping.....	80
Table 6	Sigmoid alpha looping.....	81
Table 7	Transverse looping	81
Table 8	Expert ratings of the simulator cases	81
Table 9	Demographics and pre-course endoscopy experience	94
Table 10	Simulator pre-training assessment.....	95
Table 11	Simulator post-training assessment	97
Table 12	Logistic regression of scores from pre-training to post-training assessments	98
Table 13	Results of patient-based assessments	99
Table 14	Results from the feedback questionnaire.....	100
Table 15	Associations between numerical simulator and patient-based results ...	106
Table 16	Associations between categorical simulator and patient-based results .	107
Table 17	Association between completion rates for simulator and patient-based results.....	107
Table 18	Demographics of participants.....	115
Table 19	Age and prior endoscopic experience of participants.....	116
Table 20	Initial skills assessment scores for subjects and controls	116
Table 21	Initial skills assessment scores for those who had the lectures before assessment (Group 1) and lectures after assessment (Group 2).....	119
Table 22	Final skills re-assessment scores for subjects and controls.....	119
Table 23	Change in score from initial assessment to re-assessment	120
Table 24	Previous endoscopy experience of validation study participants	132
Table 25	Comparison of baseline (pre-training) test performance	132
Table 26	Comparison of pre and post training test performance in validation study	133

Table 27	Demographics of participants in web pilot.....	134
Table 28	Comparison of pre and post training test performance in web-based pilot study.....	135
Table 29	ViCToR Course Aims.....	139
Table 30	VC interpretation test case details	144
Table 31	Summary of results for learning and behaviour evaluation	147
Table 32	Non-technical skills identified from literature search	153
Table 33	Demographics of consultants for semi-structured interviews	161
Table 34	Non-technical skills elicited from the focus group knowledge audit.....	162
Table 35	Category names given during the sorting task by individual consultant endoscopists	167
Table 36	Arrangement of sorting task skills by most common allocation.....	168
Table 37	Range of incidents recounted for the critical incident analysis.....	169
Table 38	Codes developed during initial analysis of critical incident transcripts by two independent readers.....	170
Table 39	Initial thematic framework of codes.....	172
Table 40	Non-technical skills taxonomy.....	190
Table 41	Classification of behavioural elements and categories	204

Appendices

Appendix 1 Literature search on simulation in GI endoscopy	258
Appendix 2 Therapeutic checklists and Global Score sheet	269
Appendix 3 ViCToR course details and learning outcomes	274
Appendix 4 VC checklist and Global Score sheet.....	277
Appendix 5 Literature search for non-technical skills.....	279
Appendix 6 Interview schedule for non-technical skills	283

Figure 1 Overview of thesis



Chapter 1: Introduction

1.1 Overview

A culture of seeking excellence in all aspects of medical care is firmly embedded in modern medical practice. This attitude is also now being applied to teaching and training within medicine. The evaluation of training and the assessment of trainees aims to ensure the competency of independent practitioners and therefore produce the best outcomes for patients. This process of evaluation is termed Quality Assurance (QA), and is particularly important for practical procedures where outcomes are directly related to operator skill. This thesis describes the application of this process to training in diagnostic and therapeutic gastrointestinal endoscopy.

1.2 Endoscopy training: recent developments and drivers for change

There have been many changes in endoscopic training in the United Kingdom (UK) over the last 10 years. Prior to the formation of the Joint Advisory Group (JAG) for gastrointestinal endoscopy in 1994, there was little standardisation in flexible endoscopy training. Practitioners were often self-taught or, at best, taught by those who were self-taught. There was no formalised programme for endoscopic training and no standardised method of competency assessment before unsupervised practice was permitted. The JAG was initially set up to define the standards for the training of all endoscopists, no matter their professional background. Their aim, to deliver standardised, high quality training and assessment was driven by several different factors.

1.2.1 Poor performance outcomes

In 2000, the National Health Service cancer plan published a strategy to reform cancer services nationally, with particular emphasis on diagnostic services to improve earlier detection and treatment. This was driven by data showing that survival rates for many cancers, including gastrointestinal (GI) cancers, were amongst the lowest in Europe[1], with wide variation in results across the UK. Several audits demonstrated poor performance outcomes from flexible endoscopy, which clearly had an impact on diagnosis and treatment. The largest prospective

study of colonoscopy practice in the UK by Bowles et al in 2004[2] revealed poor outcomes in terms of completion and perforation rates as well as deficiencies in almost all aspects of training. The commonest reasons for failure of completion was patient discomfort (35.3%) and looping (30.3%), both of which may be addressed by improved technique achieved through better training. A multicentre study in the United States[3] concluded that “endoscopy by an inexperienced or untrained endoscopist may result in significant morbidity, mortality, and healthcare cost...” highlighting the fact that quality in endoscopic training is recognised as an important factor in healthcare provision not only in the UK.

In 2004, the National Confidential Enquiry into Patient Outcome and Death (NCEPOD) report “Scoping our Practice”[4] demonstrated a 3% overall mortality following therapeutic endoscopic procedures, rising to 6% for specific procedures such as percutaneous endoscopic gastrostomy (PEG) placement. Their recommendations were that there should be national guidelines for assuring continuing competency in endoscopy. This echoed calls by individuals within the profession who recognised the need to set standards for endoscopic teaching and training[5-7].

1.2.2 Disparities in training outcomes

Several publications have shown that endoscopy training programmes can produce wide variation in the numbers of procedures performed and the resulting competency of practitioners[2,8-11]. As expert opinion has given way to objective data, the requirements for the minimum number of procedures required for credentialing in endoscopy has risen and recognition that meeting these requirements may not be enough. These findings are mirrored by publications showing that outcomes are also variable for practicing endoscopists, reflecting their experience and procedural volumes[12].

An audit performed in 2002 looked at the experience of trainees in the North West Thames region of London with regard to the standard of training in colonoscopy[13]. North West Thames is the second largest post-graduate training region in the UK, containing six teaching hospitals and eight general hospitals, with over 30 specialist gastroenterology and 50 general surgical trainee posts. Results showed that

colonoscopy training was generally poor, with only 17% of the trainees receiving supervision for all of their first 100 colonoscopies and only 39.3% had attended a course in colonoscopy. Quality of the supervision and assessment of competence was also variable. Trainees were routinely supervising more junior trainees in endoscopy without their own procedural or training competence having been established, and trainees were not routinely keeping procedural records. These results were consistent with findings from other studies in the literature [10,14,15].

There has also been a recognition that completion of a specified number of procedures does not necessarily reflect the competency of the practitioner. The UK study by Bowles et al[2] showed that completion rates were poor and complication rates high despite over 85% of the procedures being performed by a consultant or staff grade endoscopist. Other large clinical audits have reported similar findings[16-18].

1.2.3 Changes to doctors' working lives

In 1993 the Calman report proposed reforms of the registrar training grade to bring medical training in the UK into line with the European Union[19], resulting in a change to the way in which specialist training is provided. More recent changes introduced by Modernising Medical Careers (MMC)[20,21] have continued this approach by shortening and 'streamlining' the period of training for more junior grades. However, with the introduction of the European Working Time Directive (EWTD) in 1998, the time available for training has become significantly shortened, with trainees limited to 48 hours work per week from 2009. The combination of these two changes has been estimated to shorten the time available for training to 1/5 that of previous training programmes[22]. With such reduced training hours and opportunities it is no longer possible to rely purely on experiential learning that was an integral part of the apprenticeship model. This has been recognised by the Postgraduate Medical Education and Training Board (PMETB), which has focused on maintaining standards by mandating more formalised and structured learning opportunities than were previously available[23], but the short period of time available has prompted serious concerns regarding the outcomes from training, particularly in the surgical specialities[24].

1.2.4 NHS Bowel Cancer Screening (BCS) Programme

Colorectal cancer is the second most common cause of cancer death in the UK with approximately 27000 new cases diagnosed each year and around 14,000 deaths in England alone[1]. As a result of data showing that survival rates in the UK are one of the lowest in Europe[25,26], in 2002 the Secretary of State for Health publicly supported the development and implementation of a national screening programme for colorectal cancer by 2007[27]. This programme is based on Faecal Occult Blood Testing (FOBT) followed by colonoscopy, and projections at the time estimated an uptake of 60%, resulting in a dramatically increased demand for colonoscopy for those 2% (300 per million) who were predicted to test positive (in the age range 60-69). The pilot data[28] also predicted that there would be an increase in the need for advanced endoscopic therapy (polypectomy) as a result of earlier detection of adenomatous polyps and early cancer, which has subsequently been shown to be the case by audit data[29].

Given the poor outcomes of national colonoscopy practice reported by audits such as the one by Bowles et al, there was a perceived need for accreditation in order to ensure the competency of individual endoscopists and the quality of participating endoscopy units. This was deemed to be necessary to ensure that the risks of undergoing a colonoscopy were acceptable in what is an asymptomatic and potentially completely healthy population.

1.2.5 National training project

Coincident with these changes, there was also a recognition that endoscopy services had been under-resourced without a national strategy for service development and training. To assist the implementation of the NHS Cancer plan and to prepare for colorectal cancer screening, the National Endoscopy Project was set up by the NHS Modernisation Agency to redesign endoscopy services, increase the number of endoscopists and tackle variation in existing approaches to training[27].

1.2.5.1 Aims and delivery

Between 2004 and 2007 the National Training Plan for Endoscopy was funded by central government to address training needs at a 'grass roots' level in England. The three aims were:

1. To establish national and regional training centres which would develop and deliver educational programmes. These include skills training for endoscopists and multidisciplinary training for the endoscopy support staff
2. To achieve a step change in improving quality and quantity of endoscopists and achieve 200 additional/newly trained endoscopists
3. To move to outcomes based commissioning and sustainable education and training strategies

Training units in England were invited to competitively bid for training centre status. Ten regional and three national training centres were established with a government budget of £8.2 million, enabling them to offer free and structured training courses in endoscopy[30]. These courses have now become essential components of gastroenterology training in the UK. The foundation skills course and the basic colonoscopy course have both achieved greater than 100% delivery rates compared to plan, with over 1400 endoscopists trained nationally[31]. The JAG published formal guidelines for training[32], including recommendations on facilities, supervision, course attendance and minimum yearly experience. A national competency framework was also published[33] which details the necessary knowledge, skills and attitudes for all procedures in gastrointestinal endoscopy.

1.2.5.2 Outcome measures

Unfortunately, no procedures to prospectively evaluate the effectiveness of the central training programme were put into place and only one publication outlines the effectiveness of a training course which subsequently became a JAG-recognised course[34]. Chapter 2 (p.59) details two studies done to determine whether the unprecedented level of investment that was directed towards the improvement of endoscopic training has had any effect.

1.2.6 Assessment of competency

There has also been a revision of the assessment process within the whole of medicine in the UK. To ensure that trainees have attained competency prior to independent practice as a consultant, assessment systems have been mapped onto each curriculum[23]. These include knowledge based assessments (known as Speciality Certificate Examinations), and workplace based assessments such as the mini clinical evaluation exercise (mini-CEX), Directly Observed Procedural Skills (DOPS), Case based discussion (CbD) and Multi Source Feedback (MSF).

Within gastrointestinal endoscopy, the JAG has developed DOPS forms for assessment of technical skill in upper GI endoscopy, colonoscopy, flexible sigmoidoscopy, endoscopic retrograde cholangiopancreatography (ERCP) and therapeutic endoscopy[35]. These are now used both formatively and summatively for trainees, and during the timescale of this research, guidelines for formal accreditation in endoscopy by JAG were published and implemented (<http://www.thejag.org.uk/Forms/tabid/72/Default.aspx>). DOPS are also used alongside a multiple choice questionnaire on colonoscopy during the Bowel Cancer Screening colonoscopy accreditation process. A multiple-choice based exam including endoscopic topics is also now part of the competency assessment of gastroenterology trainees. The Speciality Certificate Examination for gastroenterology was run for the first time in 2008 (<http://www.mrcpuk.org/SCE/Specialties/Pages/Gastroenterology.aspx>), and is mandatory for Specialist Trainees (ST's) to pass prior to their Penultimate Year Assessment (PYA). Successful candidates can apply for the postnominal MRCP(UK)(Gastroenterology).

There are also now guidelines for the training standards expected at all endoscopy units which train endoscopists. These are part of the Endoscopy Global Rating Scale (GRS)[36], a biannual audit of every endoscopy unit in England. In order to meet these standards, trainers are encouraged to attend a 'Training the Trainers' course, which aims to develop the teaching skills of practicing endoscopists. Research on what constitutes a good trainer of endoscopy has been done by Wells *et al.* (see p.39) with the aim of developing a trainer assessment tool. The National Endoscopy

Team have also now developed preliminary forms for MSF-type assessments of both the endoscopy team and individuals (personal communication), although the only one so far put into practice is for individuals performing JAG unit accreditation visits.

1.3 Quality Assurance in education

1.3.1 Background

In any industry, it is important to have processes in place to ensure that goods and/or services satisfy customer requirements; that they are 'fit for purpose'. This process is termed Quality Assurance (QA). For service provision, it is necessary to have measurable outcomes in order to identify good practice and make recommendations for improvement. In the NHS, this is termed clinical governance, and is defined as "the system through which NHS organisations are accountable for continuously improving the quality of their services and safeguarding high standards of care, by creating an environment in which clinical excellence will flourish" (www.dh.gov.uk). These processes clearly also apply to higher education, where there is a responsibility to maintain standards and quality of training. The JAG was originally set up to maintain standards in gastrointestinal endoscopy, and is now overseen by regulatory bodies such as the Postgraduate Medical Education and Training Board (PMETB) and the Joint Royal Colleges of Physicians Training Board (JRCPTB).

The following sections give a broad overview of the theory behind how adults learn and acquire new skills and the application of QA processes to training and assessment.

1.3.2 Adult learning theory

There are several key features of adult learning which should be actively addressed by a teacher or training programme or they can obstruct the learning experience[37]. The main characteristics are:

- Adult learning is **purposeful**
- Adults are **voluntary** participants.

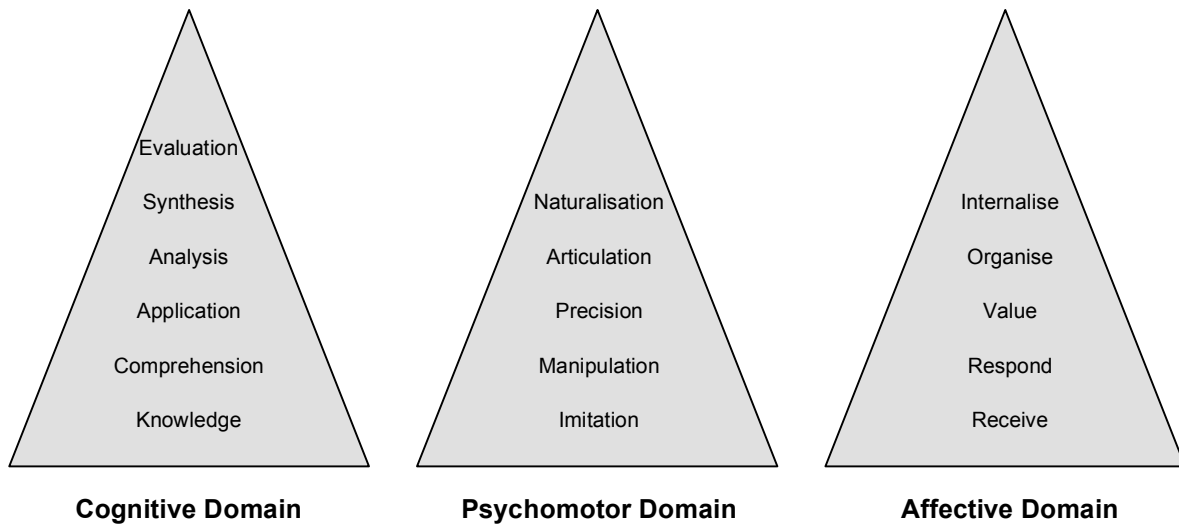
- Adults require **meaning and relevance**
- Adults require **active involvement**
- Adults require clear **goals and objectives**
- Adults need **feedback**
- Adults need to be **reflective**

These features are based on some of the theoretical models of learning outlined below. When designing and evaluating training programmes, these characteristics should be taken into consideration. For example, most courses and conferences produce learning objectives for their delegates to encourage goal-setting, and incorporation of problem-based learning (PBL) into undergraduate curricula is now almost universal to encourage active engagement in learning.

1.3.2.1 Bloom's Taxonomy

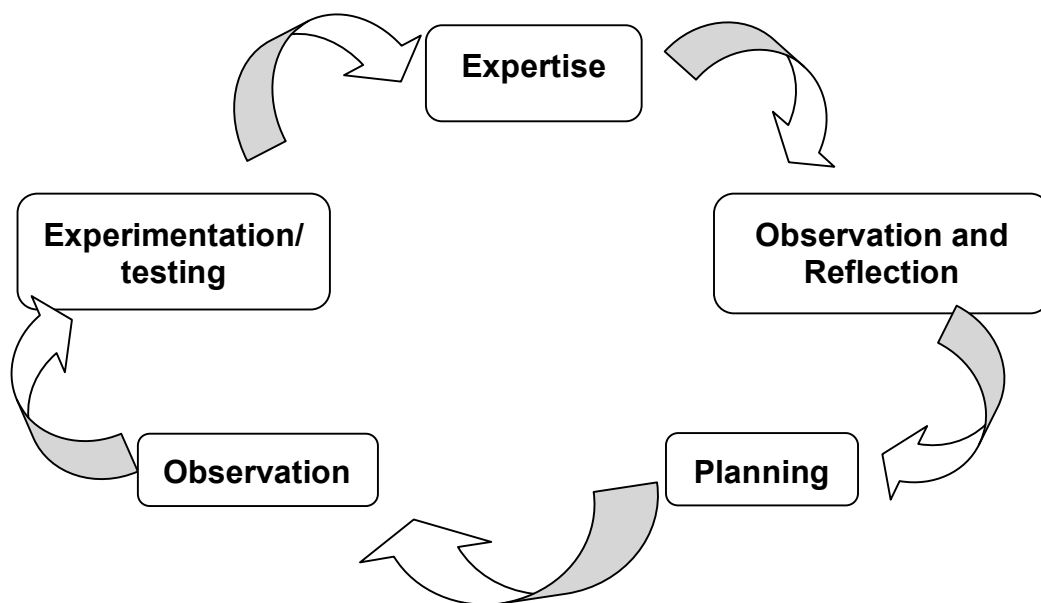
Bloom's Taxonomy of Educational Objectives, or Bloom's Taxonomy of Learning Domains (1956) is a system of categories of learning behaviour that aims to assist in the design and assessment of educational learning[38]. It identifies three main domains of educational activities: cognitive, affective and psychomotor. It provides a structure for planning, designing, assessing and evaluating training and learning effectiveness, and forms the theoretical underpinning for the 'knowledge, skills and attitudes' methodology that is in widespread use today.

Each domain has a hierarchical structure ranging from the most simple to the most complex. The cognitive domain comprises factual learning, from knowledge (remembering previously learned information) through to evaluation (the ability to judge the value of material for a given purpose). The psychomotor domain refers to skill acquisition, from imitation (copying actions) to naturalisation (mastery of activity and related skills at strategic level). The affective domain refers to feelings, emotions and behaviour, from receiving (being open to experience) through to internalisation (adopting a belief system or philosophy).

Figure 2 Bloom's taxonomy of learning domains

1.3.2.2 Experiential learning

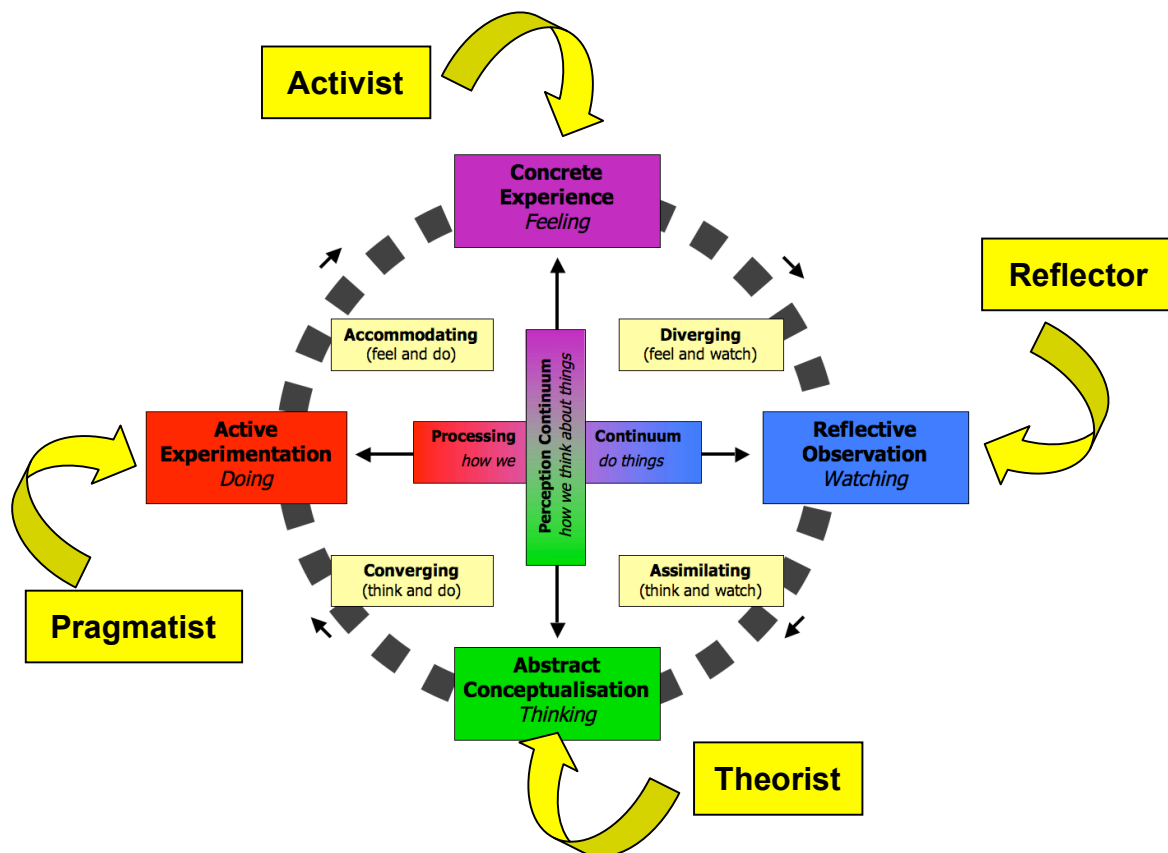
Several theoretical models, such as the theory of experiential learning described by Schön (Figure 3), recognise that learning is often most effective when based on experience. This is consistent with the observation that learning is an active process, and that reflection and planning are part of a cyclical process linking concrete experience and abstract concepts.

Figure 3 Schön's cycle of experiential learning

1.3.2.3 Learning style

It is also evident that adults do not all learn in the same way. People will have preferences in the way that they can best assimilate knowledge, skills and experiences, and this is sometimes termed a learning 'style'. Many different classification systems have been developed, for example separation into 'deep', 'strategic' and 'surface' learning (Ramsden 1992, Biggs 1987, Entwistle 1981). However, one of the most useful of the models available, described by Kolb in 1984, links the experiential learning process with learning style. This suggests there are four stages of learning which follow each other, and that individual's learning preferences lie on two ranges of learning; perceiving and processing. Individuals may have inherent preferences for a single type of learning (activist, reflector, theorist, pragmatist), or several (accommodating, diverging, assimilating, converging) depending on where they lie on the continuum (Figure 4).

Figure 4 Experiential learning cycle and learning styles.

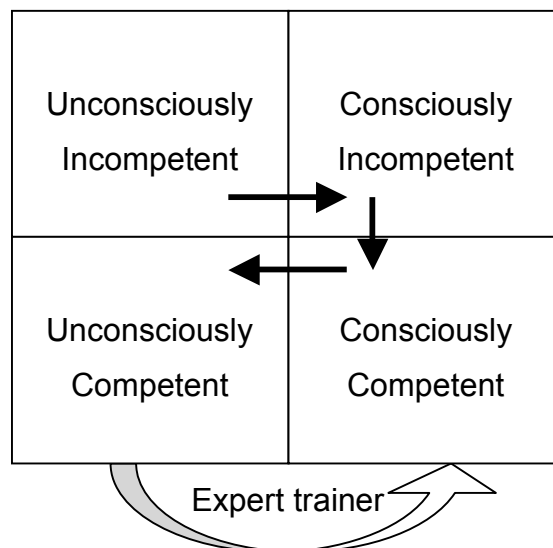


Adapted from Chapman (2006, based on Kolb's learning styles 1984), and Honey and Mumford (1982).

1.3.2.4 Competency and expertise

As they gain experience, learners pass through different stages of ability with the goal of eventually attaining expertise or mastery of the process. However, for complex or difficult tasks this may take a long period of time, requiring extensive experience. The aim of training is therefore often not expertise, but the development of competence. Competence can be defined as the ‘minimum level of skill to safely and proficiently perform a task’. A learner goes through a process in achieving competence as shown in Figure 5 (adapted from Peyton[37]).

Figure 5 The process of achieving competence



Most competent practitioners perform at the unconsciously competent level and many trainees bypass the consciously competent stage to move directly to the unconsciously competent stage where they stay and then act as trainers. An expert trainer, however, is required to pass back into the zone of conscious competence in order to understand and be able to explain what a trainee needs to learn, thus increasing the likelihood that the trainee will successfully pass through the consciously competent phase. Conscious competence may also be required to attain expertise or mastery, as performance of a task unconsciously is likely only to maintain skill levels. Expert performance requires awareness of weaknesses and active engagement in deliberate practice focused on improving particular tasks[39].

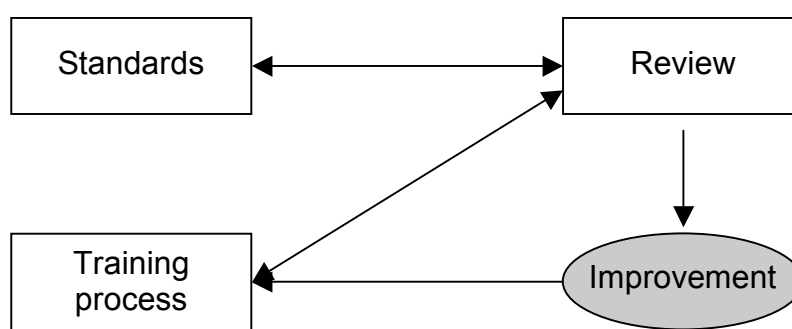
1.3.2.5 Learning and training

There is an important distinction between learning and training, which can have a significant impact on the outcomes and evaluation processes. Learning can be defined as ‘the act, process or experience of gaining knowledge or skill’ (Collins dictionary 1995, HarperCollins). Training is goal-oriented, with the aim of developing new skills, knowledge or expertise which can be used to improve performance at some specified task. Learning is therefore a necessary but not sufficient condition for training to take place[40].

1.3.3 Quality assurance process

The quality assurance process can be problematic when applied to higher education, as it may have connotations that emphasise the notion of ‘quality as accountability’ at the expense of ‘quality as improvement’ (Bowden and Marton, 1998; Biggs, 2001). Quality as accountability does little for learning as it does not help to develop an understanding about the process of training. Quality as improvement can be helpful to understand the process, and therefore improve the outcome. A simplified model of the relationship between process and improvement is shown in Figure 6.

Figure 6 A simplified quality feedback loop for training

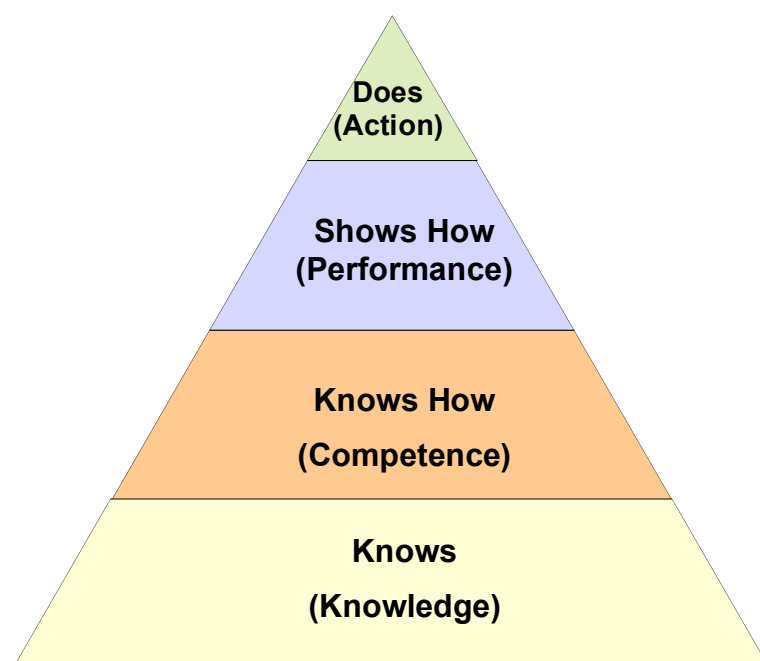


The cyclical nature of this process incorporates periods of review or reflection, where initial outcomes can be compared to subsequent outcomes. During repeated cycles, understanding of the nature of the process develops, allowing improvements to be made. At the same time, minimum standards of each part of the process can be developed because the review of the process is facilitating understanding of quality issues.

1.3.3.1 Assessment

In this thesis, the term assessment will be used to describe tests that measure individual performance. Assessments can be used formatively during the training process or summatively at the end of training to ascertain whether standards have been met. A simple framework for clinical assessment is described by Miller (Figure 7)[41].

Figure 7 Miller's framework for clinical assessment



Ideally assessment of performance and action are most pertinent, as that will give insight into how the trainee will perform in real life. However, performance may not necessarily reflect the level of learning because of other factors impinging on performance at a particular point in time.

Any tool used for assessment must be both reliable and accurate in order to perform well.

Reliability of a test is a measure of how reproducible the results are. Measures to establish reliability include *within-observer* or *test-retest agreement* (do repeated

tests produce the same scores) and *inter-observer agreement* (do independent assessors produce the same score when assessing the same examination).

Accuracy of a test is a measure of how valid the test is, or how well the test measures what it is supposed to measure. Measures to determine accuracy include:

Face validity of a test is a measure of how well the test *appears* to measure a criterion. It can be judged by both experts and laymen.

Content validity is a measure of the degree to which a test contains a representative sample of desired competencies, usually determined by expert opinion.

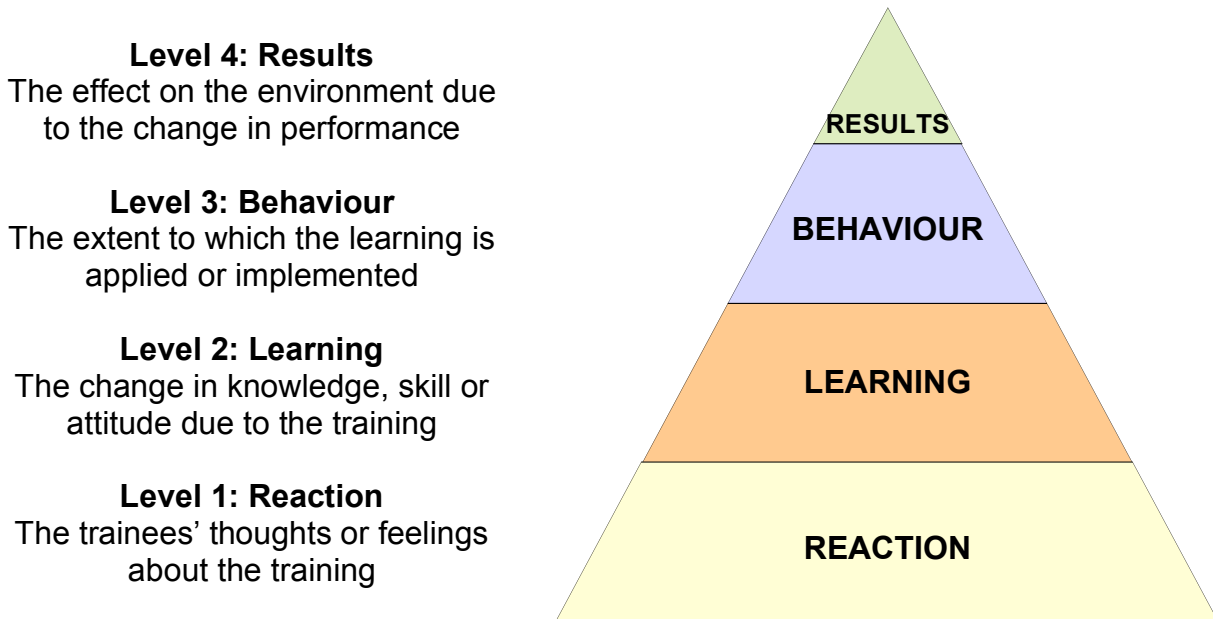
Construct validity determines how well the test measures the attributes that the examination is testing. It involves the empirical and theoretical support for the interpretation of the test, including how well it correlates with other tests it is predicted to correlate with (*convergent validity*) or how well it does not correlate with other tests it is predicted it should not correlate with (*divergent validity*).

Criterion validity reflects the success of a test used for prediction or estimation of performance. *Concurrent validity* refers to how well the test results correlate with other markers of performance done at the same time. *Predictive validity* refers to how well the test results predict performance at some time in the future.

1.3.3.2 Evaluation

In this thesis, the term evaluation will be used to describe measures looking at the process of training. This usually involves assessment of individuals, but may also include other aspects such as learning needs evaluation, trainer evaluation and performance outcomes evaluation. Kirkpatrick's four levels of programme evaluation[42] provide a framework for evaluating any kind of training programme. They map very closely to Miller's framework of clinical assessment, but focus on the training rather than the individual. The four levels are again hierarchical, with the simplest and easiest level to evaluate at the bottom. With each level, the process becomes more difficult and time-consuming, but potentially more valuable (Figure 8).

Figure 8 Kirkpatrick's four levels of training evaluation



1.3.4 Application of educational theory

Application of these theoretical models underpinned the training and assessment modules that were developed as part of this thesis. Individual learning preferences and styles were taken into account during the design of these training programmes, and feedback was obtained in order to maximise their benefit. A good understanding of adult learning was necessary to apply sound pedagogical principles during development of a module for colonoscopy training on a virtual reality simulator (p.87) and an e-learning module for capsule endoscopy (p.125). The themes of assessment and evaluation run throughout the thesis, with formative and summative assessment embedded within the educational evaluation framework for the courses on simulation training in colonoscopy (p.87), therapeutic endoscopy (p.109), capsule endoscopy (p.125), and virtual colonoscopy (p.139). In each instance, the evaluation aimed to provide the highest level of quality assurance that was possible given the practical limitations.

1.4 Quality Assurance in endoscopy training

1.4.1 Background

Although now over 30 years old, gastrointestinal endoscopy is still a rapidly growing and evolving field, with new techniques and technology being developed and adopted into widespread practice. There has been a steady increase in the number of publications on what is considered to be appropriate training in endoscopy, initially constituting expert opinion[43-47], and more recently recommendations and requirements laid down by national organisations such as the American Society of Gastrointestinal Endoscopy (ASGE)[48] or JAG[32]. However, the first call for the outcomes of training to be evaluated came in 1993 with a description of procedural competence[49], followed by the development of tools for objective evaluation of skills[11,50,51], although it was recognised that “understanding of how endoscopic skills are acquired remains rudimentary” (Baillie, 1993, p.73). This section summarises the previous published literature regarding quality assurance of different aspects of endoscopy training. The following sections identify areas where this quality assurance data are lacking or has the potential to be improved.

1.4.2 Flexible sigmoidoscopy

There has been considerable interest over the years in flexible sigmoidoscopy training, as it was the original technique developed in the 1970's for using flexible endoscopes to investigate the gastrointestinal tract. In the 1980's outcomes from intensive training courses and programmes in flexible sigmoidoscopy were studied[52-54]. As experience increased, the first report of training nurse endoscopists to perform flexible sigmoidoscopy was published in 1998[55], followed by descriptions of training programmes for nurses and other paramedical personnel[56,57], and finally guidelines for training both specialists[58] and non-specialists[59] in screening flexible sigmoidoscopy. The final report is awaited of a large multinational evaluation of flexible sigmoidoscopy in colorectal cancer screening, the FlexiScope trial[60]. This aims to identify training and quality assurance issues, with an interim report suggesting that there is wide variation in performance between endoscopists[61]. As the current UK paradigm for colorectal cancer screening involves faecal occult blood testing plus colonoscopy rather than

flexible sigmoidoscopy, it was decided that until the results of this large trial are published, further investigation of quality assurance issues would not be beneficial.

1.4.3 Colonoscopy

Much of the published literature on colonoscopy training has focussed on the aspects dealing with credentialing, in particular learning curves[62-66] and the number of colonoscopies needed to attain competence[8,12,62,64,67-71]. Most licensing authorities are now moving away from purely numbers done - "how much is enough?" and moving towards other markers that can be used as surrogates of competency such as sedation rate, complication rate, adenoma detection rate[71-73] and caecal intubation rate[74]. Despite the interest in (and medico-legal implications of) competency assessment, there have been only a few efforts to produce objective assessments of technical skill[51,75]. There have been some retrospective surveys of training outcomes[13,76,77], and a substantial number of publications looking at the use of simulation, which is discussed separately here (p.42). One study has looked prospectively at quality assuring training and assessment in colonoscopy and flexible sigmoidoscopy. It focused on designing and validating a novel intensive colonoscopy course[34], including development of a knowledge assessment (MCQ) in colonoscopy[78] and video assessment of withdrawal technique[79], which had previously been ignored in the competency debate. This work demonstrated that intensive, hands on training can have a positive impact on key areas of skill acquisition that is sustained over at least 6 months. It is recognised, however, that intensive courses are only ever going to be adjuncts to regular training[80], and there is work to be done in the future looking at the effect of 'Train the colonoscopy trainers' courses that were developed by JAG during the time-frame of this research.

1.4.4 Upper GI endoscopy

The first publication of the results from a formal programme in upper GI endoscopy training was in 1993[81], although there had been an awareness for over a decade that the results of training programmes were poor[82]. Again, there has been a focus on the number of procedures to attain competency, although as it is a simpler procedure than colonoscopy there is less drive to define what is competent[11]. Efforts have again been made to derive objective measures of technical skill[83] and investigate learning curves[84], and there is one study quality assuring a training

course in upper GI endoscopy, which used various psychological, psychomotor and cognitive tests to determine efficacy of the programme[85].

1.4.5 ERCP and other therapeutics

Complications most often occur when therapeutic intervention is required, and there is therefore an urgent need for competency based assessment and evaluation of training in therapeutic techniques. There have been calls for objective measures and dedicated training in ERCP[86-88]. However, despite the development of magnetic resonance cholangiopancreatography (MRCP), which has decreased the number of diagnostic ERCP's available for training[89], evaluation of training outcomes has been minimal. There has been only one published account of a single-operator learning curve[90], and no publications on quality assurance of training.

There has only been one publication looking at the numbers of Percutaneous Endoscopic Gastrostomies (PEGs) placed during a surgical training programme[91], which demonstrated low numbers but good success rates. There has also only been one paper on training in placing stents in the GI tract[92], which demonstrated that exposure is diminishing and training is limited outside central cancer units.

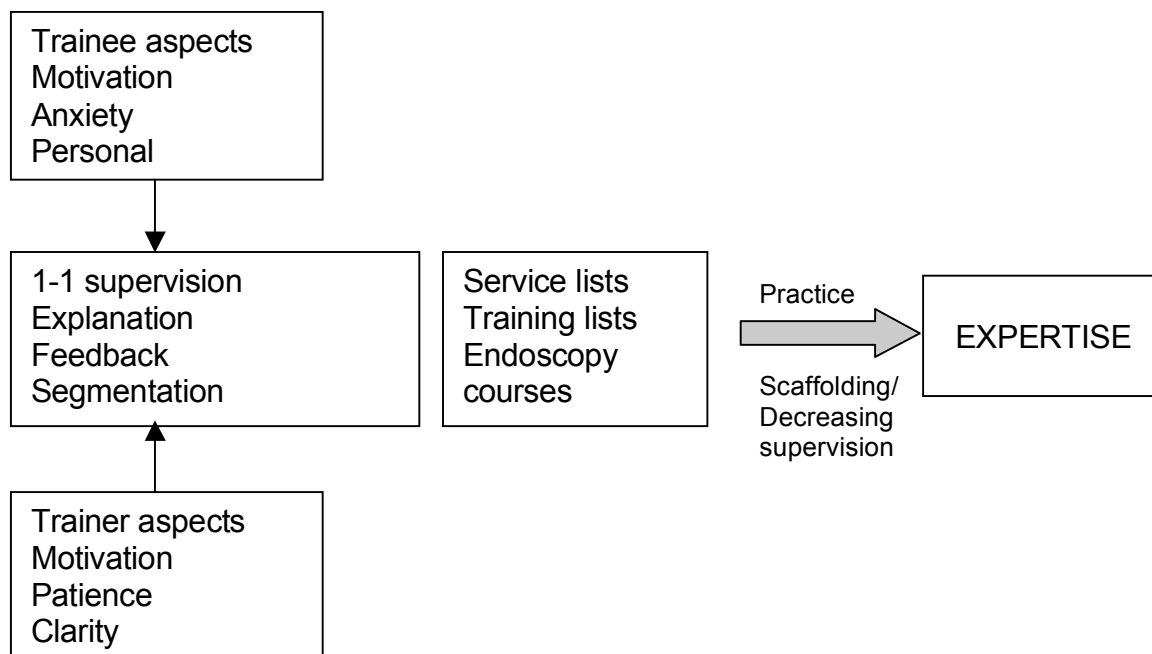
There have been many publications on the technique for polypectomy and the management of complications[93-101], which have increased as newer techniques such as Endoscopic Mucosal Resection (EMR) have been developed[102-104]. However, there are only two published abstracts related to training in polypectomy, both based on work done in the Wolfson Unit. The first was the development of a validated scoring system to assess polypectomy technique[105]. The second was the development and initial validation of an interactive computer-based training tool for polyp detection[106].

There was therefore clearly a need for further work on quality assuring training in virtually all aspects of therapeutic endoscopy, which is addressed in the study described in Chapter 6 (p.59).

1.4.6 Trainee's learning experiences

Although many of the papers highlighted above make reference to the quality of the training outcomes, the evaluation of the trainee's learning experience is mostly limited to cross-sectional surveys at a single point in time[3,10,13,82,92,107], which can only provide a 'snapshot' of the full experience. Only two papers, both by Thuraisingam[108,109], describe a more rigorous investigation of the subject using qualitative methods. Using semistructured interviews with gastroenterology trainees and video of endoscopic teaching encounters, they describe a model of the process of endoscopy learning (Figure 9).

Figure 9 Schematic model of the endoscopy learning experience (from Thuraisingam et al, 2006)



1.4.7 The trainer

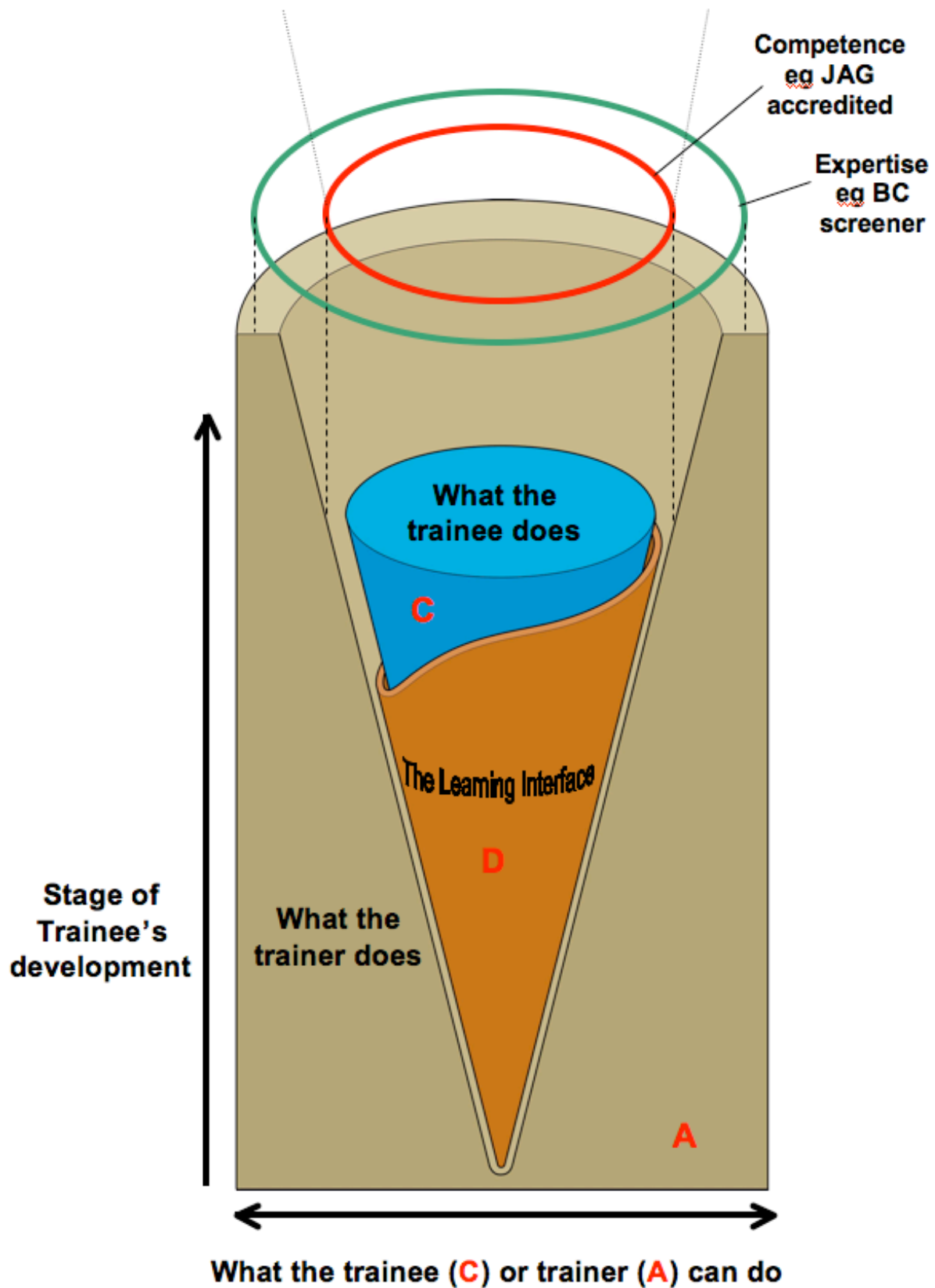
There is a considerable generic literature base on what constitutes a good trainer. A trainer should have expert knowledge and be able to demonstrate competent performance. They should be enthusiastic, open and can play many different roles in the teaching process, as described by Harden[110]:

1. Role model
 - a. On the job role model
 - b. Teaching role model
2. Information provider
 - a. Lecturer
 - b. Clinical teacher
3. Resource developer
 - a. Resource creator
 - b. Study guide producer
4. Planner
 - a. Course organiser
 - b. Curriculum planner
5. Assessor
 - a. Trainee assessor
 - b. Curriculum evaluator
6. Facilitator
 - a. Mentor
 - b. Learning facilitator

1.4.7.1 Qualities of an endoscopy trainer

There have been attempts to define exactly what attributes are desirable for a teacher of endoscopy[5,6], but these have been produced by single or small groups of experts, and so may not reflect the teaching requirements over the whole profession. Recent qualitative research has investigated this aspect of endoscopic training[111]. It builds on the work of Thuraisingam *et al*/ described above (p.39) to provide a more thorough and detailed description of the qualities of an endoscopy trainer and the interaction with the trainee. Although the results of the full study are as yet unpublished, the model of training that was developed (Figure 10) is now being considered for use to improve the training and assessment of the trainers, for example during 'Train the Trainers' courses.

Figure 10 A model of teaching endoscopy (courtesy of C. Wells)



1.5 Use of simulation in endoscopy training and assessment

1.5.1 Background

The traditional 'on-the-job' training paradigm has several inherent problems that make it a less than ideal learning environment. While it is by definition the most 'realistic' of training platforms, it does not allow for standardisation of training, repetition of procedures, practice, or errors. There is also some evidence that the involvement of a trainee can increase the risk to patients[112], and that the risk is increased for therapeutic procedures performed by less experienced practitioners[113-115]. The endoscopy suite is also not the ideal environment for teaching, as it is usually geared towards service provision rather than enhancing learning.

1.5.1.1 Use of simulation for training

The use of simulators can provide a training platform in a suitable learning environment allowing all of the benefits without the initial risk to patients. They have been successfully used for many years for training in aviation and the military, and similar practices can be applied to endoscopy simulation[116]. Although the evidence is limited mainly to skills improvement in novices, the European Society of Gastrointestinal Endoscopy (ESGE) has recommended that simulators, where available, should be used for training[117]. As can be seen in the summary tables in Appendix 1 however, there is relatively little evidence demonstrating transfer of skills from simulators to real-life endoscopy. There is also no literature on the use of simulation to improve skills in more advanced trainees. Learning theory predicts that development of expert performance requires active engagement in deliberate practice focused on improving particular tasks[39], which is ideally suited to simulation rather than ad-hoc clinical experience.

1.5.1.2 Use of simulation for assessment

As has previously been discussed, credentialling for endoscopy based on numbers of procedures performed has produced differing estimates of the numbers required, some as low as 25 procedures[63,64,118]. For colonoscopy, the ASGE currently recommend a minimum of 140 procedures to attain competency[68], although a number of studies have shown that it may take considerably more than this to attain

reasonable standards[11,50,62,67]. In the UK, expert observation of clinical practice is now mandatory to assess competency for trainees, but this is a subjective judgement and vulnerable to bias and human error. For such high-stakes summative assessments, objective and reproducible evaluation of procedural skills on a high-fidelity simulator would increase the reliability and credibility of such a process.

1.5.2 Current types of simulators

There are three main types of gastrointestinal endoscopy simulators currently available; mechanical, computer and biological. The first publication on simulation in endoscopy by Markman in 1969 described a simple mechanical model that allowed practice of the hand-skills required for flexible sigmoidoscopy[119]. Other mechanical models were developed for upper GI endoscopy[120], colonoscopy [121] and for ERCP[122], followed by computer simulations which emerged in the late 1980's[123-127], and the use of biological (ex-vivo) animal models in the last 10 years.

At the start of my research, there were only two computer simulators with published literature commercially available, the GI Mentor (Symbionix Inc, Cleveland, OH, USA), and AccuTouch (Immersion Medical Corporation, Gaithersburg, Maryland, USA), which can provide training in upper GI endoscopy, colonoscopy, ERCP and EUS. Validation studies were also published on a further colonoscopy simulator (KAIST-Ewha) during the period of my research, although it is not yet available commercially. The development and validation of a second-generation colonoscopy simulator (Endo TS-1, Olympus) is described in Chapter 3. Computer simulators have the potential to provide a wide range of learning experiences presented in a fairly realistic fashion, allowing self-instruction and specific feedback. The main obstacle to expanding the use of computer simulators is their cost and the logistics of making them accessible to trainees. At costs ranging from £20,000 to £50,000, most individual departments cannot afford to purchase them.

Anaesthetised, live-animals have previously been used for training, but there is almost no literature base to support this and they are now generally being replaced by ex-vivo models for ethical and practical reasons. There are four ex-vivo models described in the literature, although several more are commercially available. The Erlangen Endo-trainer and the Erlangen Active Simulator for Interventional

Endoscopy (EASIE), both from Erlangen, Germany, feature pig organs inserted into a dummy mannequin. The Compact-Easie is a smaller, portable, lightweight version utilising a flat board and smaller resection specimens. These tend to be used for therapeutic procedures such as ERCP, polypectomy and haemostasis, as they more realistically simulate the tissue elasticity needed than computer or mechanical simulators. One study has looked at a bovine colon model[128] for diagnostic colonoscopy simulation, and there is a dedicated modified ex-vivo porcine model for ERCP, the Neo-Papilla[129]. Whilst more realistic, animal models can be costly to purchase and maintain, and may require specialist facilities, often incurring extra location costs.

1.5.3 Evidence for use of simulation

1.5.3.1 Types of study

There are 23 'review' articles in the literature, which either look at the current state of simulation or expert opinion as to how simulation should be used in endoscopy training[130-155]. Although several of them are very detailed and give insight into the development of simulation in this area, these do not present any novel quantitative evidence for the use of simulation in endoscopy, and are not discussed further here.

Publications providing data that can be used for quality assurance fall into two main categories. The first type, validity studies, represent the first step in assessing a simulator to ascertain how realistically it replicates the procedure. Typically a validity study will include an assessment of face (expert) validity and construct validity. Outcome measures from such studies tend to depend on the metrics recorded by the simulator. These usually include measurements of case completion and speed of completion, but may also include measures of technical skill such as dexterity tasks, ability to keep a luminal view, and simulated patient discomfort scores. Some studies have attempted to validate the simulators by correlating the measurements with expert subjective assessments as well. These studies may also be used as evidence for use of a simulator in an assessment situation. A simulator that has a high degree of construct validity, i.e. is able to distinguish between practitioners of differing ability on a number of different measures, may be considered for use as a tool to assess ability in a procedure. However, as most currently available simulators do not

possess this level of construct validity, there is little evidence for the use of simulation in either low-stakes or high-stakes assessment at present. As construct validity studies require a pre-determined measure to separate the groups, they are by definition non-randomised trials and as such can attain only level 2b evidence (parallel cohort studies).

The second type of study investigates the training effect from use of a simulator. These tend to be parallel comparisons, often in a randomised controlled fashion. The most common studies compare novices trained using a simulator to those who are given no training. These studies aim to show that pre-training with the simulator can shorten the learning curve to competency. Novices are usually chosen as they are on the steepest part of the learning curve, which makes significant changes in performance more likely and easier to measure. Better studies compare training using the simulator with standard bedside training, which offers a much more clinically relevant comparator. There are no studies in the literature that compare training on two different simulators. Outcome measures for these studies are based on completion of simulated tasks or real procedures. Simulated cases use metrics recorded by the simulator to assess performance, but should only use parameters that have previously been shown to exhibit construct validity. Evaluation of performance on real procedures usually uses expert subjective opinion to make an assessment of competency. Most studies use unvalidated scoring sheets to make this assessment, as there are as yet no gold-standard validated clinical assessment measures available for gastrointestinal endoscopy.

The levels of evidence used to rate each study are shown in Appendix 1(i) p.258.

1.5.3.2 Lower GI endoscopy

Details of eleven validation studies done on colonoscopy simulators are given in Appendix 1(ii), p.259. Two studies (Phitayakorn, 2008 and Westman, 2007), both using the GI Mentor II, are descriptive and used experts to look at benchmarking standards and correlate visuospatial coordination between different tasks. The remainder are non-randomised parallel cohort studies. The two largest have both been performed by Koch et al. The first (2007) looked at the validity of the GI Mentor II and had 105 participants in 4 groups. It was able to distinguish novices only on 4

parameters, and no significant differences found between intermediate, experienced and experts on any parameters. A further 3 smaller, less robust studies using the same simulator have also shown validity for only a few outcome measures. The second study by Koch (2008) used the Olympus Endo TS-1, and had 23 experts and 26 novices. It was able to distinguish between the groups only on one parameter, time to caecum. There are three small studies using the Accutouch simulator (Sedlack, 2002; Sedlack, 2003; Mahmood, 2003), each with less than 20 participants in total. All three again found only a few valid outcome measurements that were able to distinguish between absolute novices and experts.

Despite the relative dearth of valid outcome measures, there have been 10 studies using colonoscopy simulators for training (Appendix 1 (iii) p.261). Of the two studies using the Kaist-Ewha colonoscopy simulator (Kim, 2007; Yi, 2008), one used engineering students as the subjects, and the second, using endoscopists, had only 11 subjects. There are three studies using the Accutouch simulator, including the only good-quality randomised controlled trial (Park, 2007). This study randomised 24 novices into 12 subjects who were given 2 hours simulator training and 12 controls who were not. They compared their performance on both simulated procedure and a single real-life colonoscopy and found that those trained on the simulator were rated by experts as better on real patients than the controls, and that the expert ratings of simulated colonoscopy correlated reasonably with two computer outcome metrics, completion time and % time in red-out. Five studies have used the GI Mentor II for training, including the largest trial in the literature. This study by Cohen et al (2006) randomised 45 first year fellows to train for 10 hours on the simulator or to no training. They assessed them on the simulator and then followed them up for their first 200 colonoscopies. They found that subjects had significantly better objective competence than controls for the first 80 cases, but there was no difference in the length of time to 90% completion rate and no difference in patient discomfort between the groups.

There are five studies looking specifically at flexible sigmoidoscopy, 2 validation studies both using the Accutouch simulator, and 3 training studies, one using Accutouch, one using the GI Mentor II, and one using the GastroSim, which is no longer commercially available. The validation studies show some ability to

discriminate novices from experts, but the training studies are variable. The GastroSim study (Tuggy, 1998) showed 6 hours training was superior to no training at all. One study (Sedlack, 2004) showed a slight benefit from 3 hours training prior to 1 week clinical training, but the other study (Gerson, 2003) showed that equivalent bedside training was superior to simulator training in all outcomes.

1.5.3.3 Upper GI endoscopy

There are five validation studies for upper GI endoscopy in the literature. One (Westman, 2007) was descriptive and found no correlations between performance on the GI Mentor with other visuospatial abilities. Four had reasonable but small numbers (20-30 participants), and demonstrated some validity for the GI Mentor in distinguishing novices from experts. One study (Neumann, 2003) used an ex-vivo animal model and an unvalidated score-card to assess trainee performance over 5 days of a course. Although the findings demonstrated validity of the score-card, the assessments were not blinded.

Three of the four training studies (Sedlack, 2007; Neumann, 2003; Ferlitsch, 2002) were continuations of validation studies and the only one of these that was randomised had a mere 8 participants. The only dedicated training study (Di Guilio, 2004) randomised 22 novices to 10 hours simulator training or no training and followed them up for 20 consecutive OGD's. They found the subjects performed more complete procedures, with better overall scores, although this was a subjective, unblinded expert judgement.

1.5.3.4 ERCP

The evidence for ERCP simulation is small, with only three studies. Two studies are essentially descriptive, with one by Sedlack (2003) comparing different types of ERCP simulator using parallel cohorts, with the ex-vivo Erlangen model scoring highest on indices of realism, usefulness, and performance, while the computer was felt to be easiest to incorporate into a training program. The NeoPapilla (Matthes, 2006) was also compared to other simulators and 6/9 experts thought it was the most realistic available. Only one study using the CompactEASIE (Maiss, 2007) compares 2 experts and 2 novices, and found that although the two experts were quicker, both novices managed to complete the test cases.

1.5.3.5 Therapeutic endoscopy

There are five studies in therapeutic endoscopy, all using ex-vivo animal models on control of haemostasis. One was a questionnaire following training looking at preferences (Hochberger, 2004). Three have demonstrated a significant impact on skills from short training sessions (Matthes, 2005; Maiss, 2005; Hochberger, 2005). The most recent study (Maiss, 2007) is the only study given a 1b rating for quality. This randomised 27 novices to receive 19 hours training and then no further training or 12 further training sessions over 7 months. This found significantly higher performance scores ($p < 0.001$) and shorter times ($p < 0.001$) for subjects vs controls. There were no studies looking at any other form of therapeutic endoscopy.

1.5.3.6 Other

Two studies have been done looking at the benefit of simulation for training in endoscopic ultrasound, but both of these were surveys of opinion and did not provide any objective evidence of benefit. One study (Kiesslich, 2005) has shown significantly better endoscopic performance and significantly better crisis management after a standardized training program using two scenarios (GI bleeding with significant blood loss and sedation overdoses). However, this was a non-randomised, non-blinded study and provides only level 3 evidence.

1.5.4 Quality assurance needs

The greatest quantity of evidence exists for the efficacy of computer simulators for colonoscopy, but most of the training studies have inadequate sample sizes or no power calculations and there is only one study which provides level 1 evidence. The validation studies for colonoscopy simulators have only demonstrated a very few metrics that are able to discriminate between novices and experts, and none that will discriminate between intermediate levels of ability. There are fewer, but higher quality studies demonstrating the efficacy of simulation for training in OGD and control of haemostasis, but evidence for other therapeutic procedures such as ERCP, polypectomy or stricture dilatation are negligible or non-existent.

There is clearly scope to improve the quality assurance evaluation of computer simulations by running large, high-quality randomised controlled trials, ideally using a simulator that has a significant number of validated outcome measures and

comparing this with real-life outcome measures. This is described using the Olympus ENDO TS-1 colonoscopy simulator in Chapters 3 (p.73), Chapter 4 (p.87) and Chapter 5 (p.103). There is also a need to demonstrate the efficacy of training in other therapeutic procedures, again ideally using large randomised, controlled trials with blinded assessments. This is addressed in the study on therapeutic endoscopy simulation in Chapter 6 (p.59).

1.6 New technologies for endoscopy training

Alongside the steady increase in the quality and utility of endoscopic technology, there have also been several new technological developments that have the potential to impact on training or for which training standards have yet to be determined.

1.6.1 3-D imager

The first clinical results of a real-time, non-radiation based imaging system were first published in 1995[156]. The current ScopeGuide™ Magnetic Endoscope Imaging (MEI) system uses low-voltage magnetic coils within a colonoscope in combination with a receiver plate outside the patient to produce a 3-dimensional graphical representation of the configuration of the colonoscope shaft. It has been commercially available for 10 years and has the potential to improve colonoscopy by allowing accurate assessment and straightening of loops[157] thus reducing pain and improving completion rates[158]. It is currently used to aid performance assessment during the Bowel Cancer Screening accreditation process, and there is some evidence from single-centre studies that it can enhance the acquisition of colonoscopy skills in trainees[159,160].

A large scale trial of its utility in training is clearly warranted. However the practicalities of this are currently extremely difficult due to the very limited uptake of ScopeGuide around the country and the difficulties in determining end-points for the training. A trial is currently underway looking at the benefit of using a 3-D imager in colonoscopies that are predicted to be difficult and if the results are positive, that may encourage more widespread purchase and subsequently enable a training study to be undertaken. In the meantime, a simulated 3-D imager view is available

on the Olympus ENDO TS-1 colonoscopy simulator, and the studies described in Chapter 3 (p.73) and Chapter 4 (p.87) investigate the training potential further.

1.6.2 E-learning

The most dramatic change in modern living has probably been the rapid development of computers and widespread availability of the Internet. Broadband connections now permit high-speed downloads and access to a vast array of learning and teaching material, which has been termed electronic or e-learning. Over the three years of my research period, e-learning has become a vast industry, with many websites offering online training packages of varying quality and cost. The best websites apply for accreditation with recognised societies to allow Continuing Professional Development (CPD) or Continuing Medical Education (CME) credits.

These sites tend to lend themselves to knowledge-based training rather than the skill-based training, however there are now several websites incorporating or dedicated to endoscopy training. The Wolfson Unit for Endoscopy produced one of the first sites for endoscopy e-learning (www.endoscopylearning.org.uk) with training modules and online video aimed at skills development. The DAVE project (<http://daveproject.org>) is a collection of teaching tools incorporating a video atlas, medical lectures and presentations and has CME accreditation from the American Society of Gastrointestinal Endoscopy (ASGE). The EndoTraining website (www.endotraining.co.uk), produced by the Polish Gastroenterology Foundation (and supported by the UK National Training Lead and Procter & Gamble), has recently gone online. This produces weekly GI clinical cases accompanied by high quality endoscopy videos and has applied for CPD accreditation.

1.6.2.1 Capsule endoscopy training

Since its introduction in 2001, capsule endoscopy (CE)[161,162] has established a role as the investigation of choice for obscure gastrointestinal bleeding[163,164] and for the evaluation of suspected Crohn's disease following a negative initial endoscopic and radiological survey[165,166]. Because of its utility and acceptability to patients, there has been rapid expansion in demand, resulting in devolution of services from specialised tertiary centres to local hospitals. The video capsule

captures up to 60,000 images during the eight hour battery life as it passes through the gastrointestinal tract. Key diagnostic images may be present in only a handful of frames and accurate identification and interpretation of these images requires considerable experience and expertise, with significant inter-observer variation in reporting even in expert hands[167-169].

Despite the difficulties in interpretation and expansion in CE services, there are few formalised training opportunities and it is unlikely that low volume services will have a sufficient caseload to allow for comprehensive training to be performed in-house. Current guidelines regarding accreditation in CE are limited and unvalidated[170]. There is no published literature investigating the effectiveness of different methods of CE training, although a large comparison of expertise in capsule endoscopy by Sidhu et al [171] identified a demand for focussed training to enable trainees to reliably identify and interpret pathology found on CE.

Unlike flexible endoscopy training, there is no practical skills training requirement for CE. Expertise in reporting is purely diagnostic, based on recognition of pathology during video review. The lesion recognition skills required for CE are ideally suited to a web-based e-learning programme which could offer exposure to a wide range of pathology in an interactive learning environment which is independent of geographical location. This approach would allow dissemination of specialist training material and best practice to a wide audience and may help to address the current gaps in CE training provision in the UK and elsewhere. This approach has been adopted by the Joint Advisory Group (JAG) for gastrointestinal endoscopy and a pilot lesion recognition assessment module for upper GI endoscopy is available online at www.lesionrecognition.com. Development and validation of training and assessment modules for lesion recognition in capsule endoscopy are detailed in Chapter 7.

1.6.3 Virtual colonoscopy

Virtual colonoscopy (VC), also known as CT colonography (CTC) is a recently established CT-based investigation for colorectal pathology such as cancer and its precursor, the adenomatous polyp. Like optical colonoscopy, VC is a complex technique whereby performance is closely linked to examination quality, interpretation methods, reader aptitude and experience[172-180]. In response,

experts universally recommend formal training, although with no firm consensus regarding minimum requirements for individuals[181].

In the UK, clinical demand for a more accurate alternative to barium enema and initial enthusiasm amongst both radiologists and radiographers alike has led to rapid and widespread dissemination of VC[182]. However, many UK centres have implemented VC without any kind of formalised training[183]. In addition, radiographers who have been central to the provision of barium enema services in many hospitals have been encouraged to participate in the VC service, often with limited experience.

A potential role for radiographers in VC is evolving, and in several UK centres radiographers now undertake the examination under indirect supervision of radiologists. As a result, they appear well placed to make an initial assessment of examination quality and possibly the presence or absence of colorectal cancer. For example, where distension is deemed inadequate or a colonic segment obscured by fluid in both scan positions, then an additional lateral decubitus scan can be performed (or additional colonic insufflation may be appropriate). Similarly, detecting a cancer might prompt immediate radiologist review, which could precipitate decisions regarding need for same day staging scans and endoscopy, thus obviating requirement for additional bowel preparation, improving diagnostic efficiency and patient care. However, to fulfil such roles, radiographers will require training in both examination technique and interpretation. Chapter 8 (p.139) details the development and quality assurance of a course specifically tailored for radiographers.

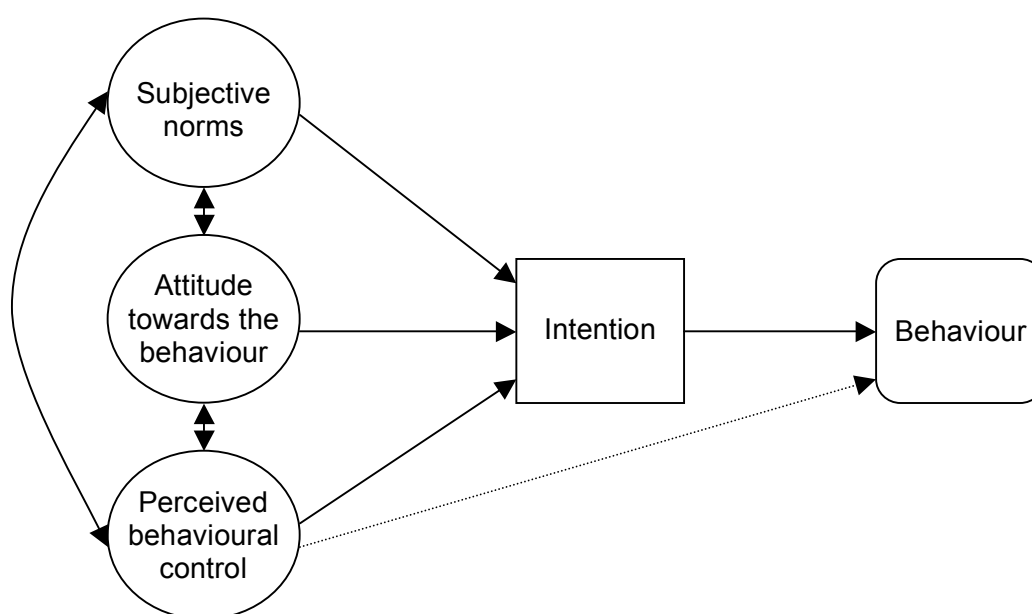
1.7 Attitudes, behaviours and professionalism

1.7.1 Background

Within UK gastrointestinal endoscopy, there are validated assessments of knowledge and skills; a multiple-choice questionnaire on colonoscopy[78] and the JAG Directly Observed Procedural Skills forms for upper and lower endoscopy[184]. These are currently used as both formative and summative assessments for trainees and in the Bowel Cancer Screening accreditation of independent endoscopists. However, according to the learning theories previously described, there is a third domain, the affective or attitudinal domain, which has an important bearing on performance. Attitudes are deemed to be a vital part of professionalism within medicine. The Royal College of Physicians defines medical professionalism as “a collection of attitudes, values, behaviours and relationships that underpins the trust the public has in doctors”[185]. In order to develop training programmes which encourage and enhance professionalism, it is important to be able to define and assess these attitudes, values and behaviours.

Attitudes can be difficult to define and assess, as they are often subconscious, nebulous concepts such as integrity, morality and responsibility. It is possible to make them more concrete by re-defining them as values that form the basis for decisions about appropriate ways to behave, and therefore they can be inferred from more easily observed behaviours. However, the relationship between attitudes and behaviours is likely to be complex. The theory of planned behaviour (TPB) originated in the social psychology literature[186] and states that the best predictor of behaviour is the *intention* to behave, which is in turn influenced by perceived social pressure and subjective norms as well as by attitudes. This model has been shown to be predictive of many behaviours of health-care professionals, including nurses' treatment of HIV-infected patients[187] and doctors' use of clinical guidelines for asthma treatment and antibiotic use[188]. It also provides an explanation for the longstanding awareness of the poor predictive relationship between a given attitude and behaviour[189].

Figure 11 Theory of planned behaviour



It is clear that attitudes and the behaviours that result from them are considered to be important in the training and assessment of professional behaviour. These are often termed ‘human factors’ in the psychological literature.

1.7.2 Human factors research

Research in high-risk industries such as aviation show that errors and adverse outcomes are often not due to failure of knowledge or technical skill, but other ‘human factors’ such as breakdown in communication, poor decision making and fatigue or stress[190]. These can also be categorized as non-technical skills, which are defined as “skills crucial for maintaining safety that are not directly related to technical expertise”. In combination with knowledge and technical skill, they form another definition of professional behaviour; namely safe and competent performance in daily practice[191].

Following investigations into early airline crashes, the aviation industry developed a research base which allowed non-technical skills errors to be identified and understood. These in turn formed the basis for behavioural marker systems such as the Line/LOS Checklist[192] and the NOTECHS system[193,194] which enable evaluation of performance of an individual’s or team’s non-technical skills in their real context, using objective assessment of observable behaviours. This was

subsequently developed into a training methodology aimed at mitigating human factors errors, known as Crew Resource Management[195] (CRM), which has resulted in an excellent safety record over the last three decades. It has been compulsory for all airlines worldwide to participate in such training since 2001.

1.7.3 Importance in medicine

Medical error remains the 8th most common cause of death for hospital inpatients, contributing to one in every 300 deaths per year in the UK. The majority of these errors are, like in the aviation industry, not due to failure of technical skill, but due to other human factor errors. Experts have estimated that nearly 25% of these are possibly preventable and 6% probably or definitely preventable by optimal care[196]. This is also likely to be the case in gastrointestinal endoscopy. In 2004, the National Confidential Enquiry into Perioperative Death (NCEPOD) report, *Scoping our Practice*[197], investigated 1818 deaths within 30 days of a therapeutic gastrointestinal endoscopy. Of the 21 recommendations made as a result of the enquiry, only one mentions technical skill (level of experience required for the management of acute GI bleeding). The other 20 highlight deficiencies in non-technical skills such as organisational issues, patient consent, training and education, pre-procedural review and monitoring.

Similar methodology in identifying and training non-technical skills to that used in the aviation industry has been successfully translated into the medical domain[198], with increasing awareness that the competencies required are different for individual situations and are not applicable 'across the board'. Although many of the principal skills may be generic and transferable across domains, the component behaviours of any assessment system will be specific to the needs and characteristics of a particular environment.

1.7.4 Quality Assurance needs

Although some elements of non-technical skills are included as part of current endoscopic training and assessment, they are not well established or addressed in any detail. It therefore seemed appropriate to start at the very beginning of the process, attempting to define the attitudes and non-technical skills that are important and specific to gastrointestinal endoscopy and arrange them into a taxonomy or skill-

list. This is detailed in Chapter 9 (p.151). Following development of this taxonomy, behavioural markers could then be identified in order to produce an assessment system that can be used for QA and training purposes, which is detailed in Chapter 10 (p.197).

1.8 Summary

Although much more emphasis has recently been placed on attempting to improve endoscopy training in the UK, there are still considerable gaps in the quality assurance evidence base. There are almost no data demonstrating the outcomes from recent changes to both medical and endoscopy training. The use of simulation, although a promising development, is currently limited by several factors. Firstly, the current virtual reality simulators lack realism and there are as yet only a few well-controlled studies showing benefit. Secondly, the evidence for training in therapeutic endoscopy is limited to achieving haemostasis, despite the existence of simulation models for other therapeutic procedures. Thirdly, there are currently no validated tools to evaluate many of the procedures, preventing quality assurance data from being obtained.

New technologies are also now having an impact on training. The incorporation of e-learning into endoscopy training is to be welcomed, but will require educational evaluation to ensure that the outcomes are beneficial. It is by no means certain that making education available over the Internet will result in improved training, much less improved performance. Likewise, competing technologies such as virtual colonoscopy require formalised training and educational evaluation to avoid performance gaps between centres.

Lastly, the focus so far in endoscopy has been purely on training and assessment of knowledge and skills. This neglects the attitudinal domain that the educational literature recognises as being vital for development of professional behaviour. The aviation industry has pioneered training in this area, with translation into the medical field by anaesthetists and surgeons. Gastrointestinal endoscopy needs to follow suit in order to continue to improve performance and reduce error.

1.9 Hypothesis

Training methodologies and assessment of performance within diagnostic and therapeutic gastrointestinal endoscopy can be improved by application of a quality assurance framework.

The aims of this research were:

1. To evaluate the effect of the recent changes in training by closing an audit loop and by surveying trainees nationally about current training
2. To establish the construct validity of a new computer colonoscopy simulator
3. To quality assure the use of a computer simulator for training
4. To establish the criterion validity of a computer simulator
5. To develop and validate new training and assessment tools to quality assure a course on therapeutic endoscopy
6. To design and validate a web-based e-learning course for lesion recognition at capsule endoscopy
7. To design and validate a course on virtual colonoscopy for radiographers
8. To establish a non-technical skills taxonomy for professionalism within endoscopy
9. To establish behavioural markers which map to a non-technical skills taxonomy and develop an assessment tool based on these markers

Chapter 2: Evaluation of current endoscopy training

2.1 Background

There is no recent data regarding the status of endoscopy training in the UK as a whole, despite unprecedented levels of investment. All of the previous surveys in the literature were performed prior to these changes. To establish quality assurance regarding the current standard of training and whether the recent changes have had any impact, two approaches were used. Firstly, a survey performed in 2002 looking at colonoscopy training in a single region was repeated in 2007, allowing a comparison between current conditions against those prior to these changes. The second approach was to take a 'snapshot' of current endoscopy training as a whole by surveying the entire body of gastroenterology trainees nationally.

2.2 Comparative survey on colonoscopy training

2.2.1 Background and Aims

An audit was performed in 2002 looking at the experience of gastroenterology and surgical trainees in the North West Thames region of London with regard to the standard of training in colonoscopy[13]. North West Thames is the second largest post-graduate training region in the UK, containing six teaching hospitals and eight general hospitals, with over 30 specialist gastroenterology and 50 general surgical trainee posts within the region. Results from this audit showed that colonoscopy training was generally poor, with only 17% of the trainees receiving supervision for all of their first 100 colonoscopies and only 39.3% had attended a course in colonoscopy. Quality of the supervision and assessment of competence was also variable. Trainees were routinely supervising more junior trainees in endoscopy without their own competence having been established and trainees were not routinely keeping procedural records. These results were consistent with findings from other studies in the literature [10,14,15].

The aim of this survey was to investigate the changes in provision of training in colonoscopy amongst medical and surgical trainees in the North West Thames region between 2002 and 2007.

2.2.2 Methods

This was a cross-sectional closed survey that was approved by the research and development department of St Mark's hospital as well as the regional training directors for gastroenterology and surgery. The questionnaire used in the previous 2002 study was updated into a Microsoft Word document with 49 questions covering the following aspects of training: endoscopic experience, level of supervision, training methods used by supervisors, assessments of competence, procedural record keeping, and trainees' perception of changes in training. The wording of the questions and non free-text responses was unaltered to allow comparison with the 2002 audit results. Responses were allowed regarding the last three trainers, as trainees may have been exposed to a number of different training styles.

This document was emailed with an invitation for participation to all gastroenterology and surgical trainees in the region as well as those who participated in the previous study who may have completed their specialist training. There were no incentives to respond and participation was entirely voluntary. Trainees could complete the forms electronically and return them using an anonymous web-based secure upload portal. The data was automatically exported to a database on a password-protected secure server. There was a 6-week deadline for response, with two reminder emails sent after 4 weeks and 5 weeks.

2.2.2.1 Statistical analysis

The majority of outcomes were measured on a categorical scale, and for these outcomes Fisher's exact test was used to compare the results between the two cohorts. Continuous measurements that were found to be approximately normally distributed were compared between cohorts using the two-sample t-test. A number of the continuous variables were not normally distributed and the Mann-Whitney test was used. Incomplete questionnaires were included in the analysis, but no weighting or propensity scores were used to adjust for response rates.

2.2.3 Results

The questionnaire was emailed to 113 recipients: 61 surgical trainees, 37 gastroenterology trainees, and 15 of those who participated in the 2002 audit but who were no longer trainees in the region. 26 (70.3%) gastroenterology trainees (specialist registrar grade) with a median (IQR) year of training of 3.5 (2,4.75) but

only 4 (6.6%) surgical trainees (specialist registrar grade) responded. 6 gastroenterology consultants responded; these were previous trainees in the region with a median time of appointment to post of 1 year (1,1). The surgical trainees' responses were not included in the analysis as their low response rate cannot be considered to be representative of the population and are likely to be subject to response bias[19]. For comparison purposes, responses from surgical trainees in the 2002 audit were also excluded from the analysis. Overall, the 2007 cohort included 32 responses, with a response rate from the gastroenterologists of 76.2%.

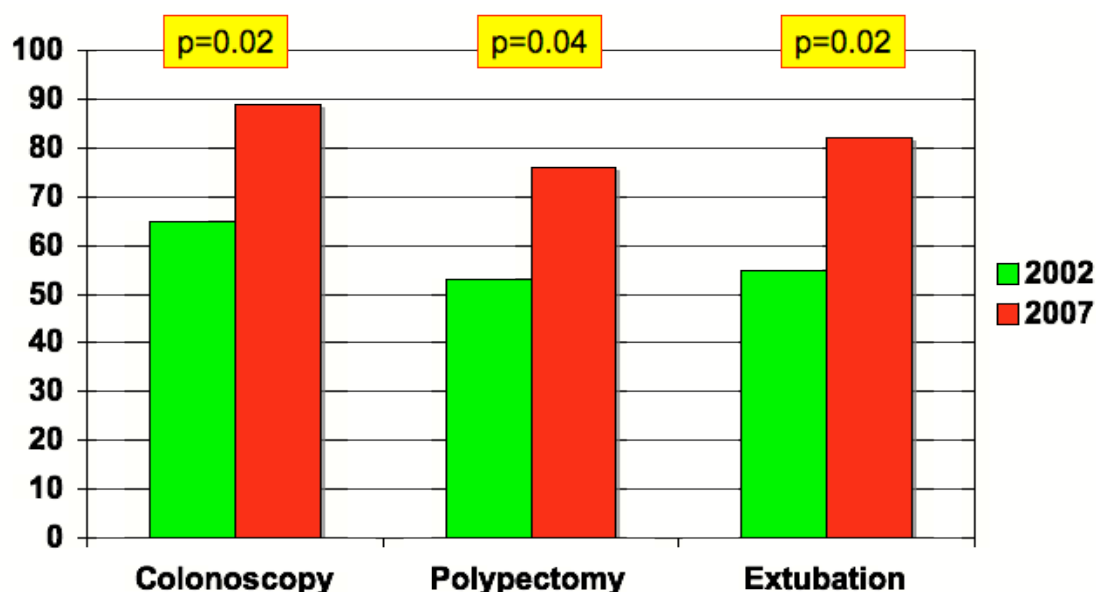
There was no significant difference in the proportion of subjects currently performing colonoscopy between the two cohorts (65% in 2002 vs 78% in 2007, $p=0.27$), and no significant difference between the number of colonoscopies being performed per week ($p=0.08$). If the trainees were performing their own unsupervised lists, trainees had performed significantly more supervised colonoscopies in 2007 than in 2002 ($p=0.05$) (Table 1).

Table 1 Reported number of colonoscopies performed as a trainee prior to performing colonoscopy unsupervised

Number of colonoscopies	2002	2007
<100	10 (45%)	4 (18%)
101-200	11 (50%)	12 (54%)
201-500	1 (5%)	5 (23%)
>500	0 (0%)	1 (5%)

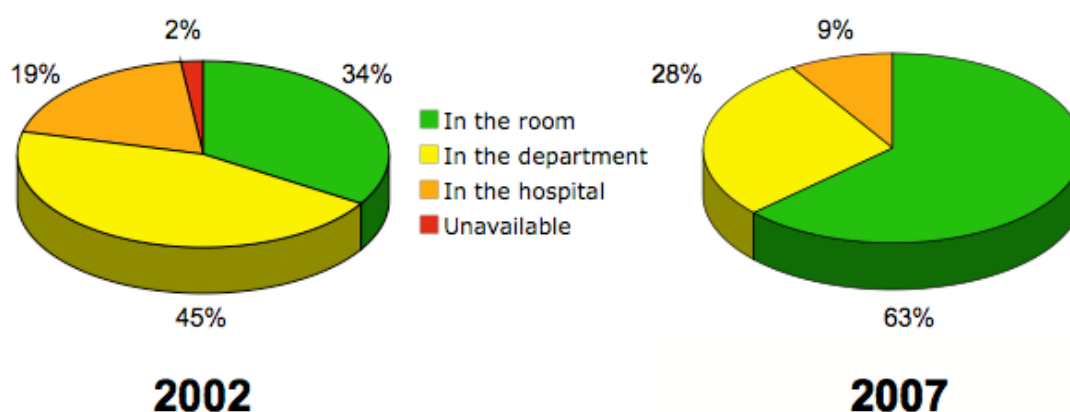
Significantly more trainees reported that they had been formally taught the principles of colonoscopy, polypectomy or extubation technique in 2007 than in 2002 (Figure 12). 59% reported this to be a consistent part of their training, with 31% only receiving intermittent formal training.

Figure 12 Percentage of trainees who had ever received formal training in colonoscopy, polypectomy or extubation technique



Levels of overall supervision for a trainee's first 100 colonoscopies had also significantly improved ($p < 0.001$), with the trainer present in the room for nearly 2/3 of a trainee's first 100 colonoscopies and in no cases was the supervisor completely unavailable (Figure 13).

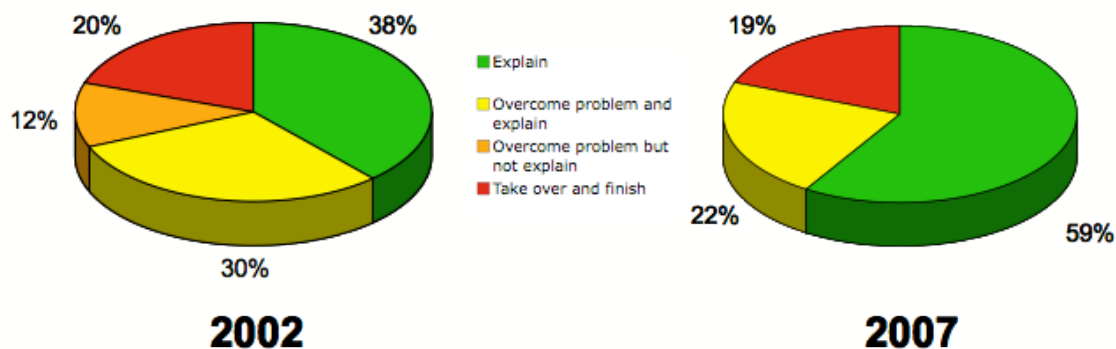
Figure 13 Supervision of the trainees' first 100 colonoscopies



(68 trainer/trainee interactions in 2002, 68 trainer/trainee interactions in 2007)

Training methods used had significantly improved from 2002 to 2007 (Figure 14), with trainers more likely to assist trainees by explaining a problem and attempting to talk through the procedure, rather than taking over and resolving the problem themselves ($p=0.04$).

Figure 14 Level of assistance given to resolve problems during a trainee's first 100 colonoscopies



(67 trainer/trainee interactions in 2002, 70 trainer/trainee interactions in 2007)

Significantly more trainees had been on a colonoscopy course in 2007 than in 2002 (84% vs 48%, $p=0.003$) and although most had to wait between 4-6 months to attend, all trainees were 'quite' or 'very' satisfied with the course.

It was not possible to compare the proportions of those who had been formally assessed as competent prior to independent practice as that question was not asked in 2002. However, in 2007, 82% of those practicing independently had been formally certified as competent at intubation, 59% at polypectomy and 59% at extubation. If informal assessment was included, then there were no significant differences between 2007 and 2002 in the proportion of trainees who had been assessed in competence at intubation (86% vs 69%, $p=0.29$), polypectomy (60% vs 59%, $p=1.00$) or extubation (91% vs 70%, $p=0.14$). The average caecal intubation rate, a surrogate marker of competence, did not change significantly (90% vs 93%; $p=0.20$).

When asked about procedural record keeping, there was an increase in the proportion of trainees who recorded their caecal intubation rate, although this did not reach statistical significance (67% vs 58%, $p=0.60$). There was no significant change

in those who keep a record of their histology results (21% vs 26%, $p=0.76$). The proportion of trainees keeping records of complications was higher in 2007, although not reaching statistical significance (48% vs 29%, $p=0.18$). Reported complication rates were lower, with 3 of 18 (16%) having had a perforation in 2007 compared to 8 of 13 (61%) in 2002.

87% of trainees in 2007 think that their training has been adequate or better than adequate, compared to 25% in 2002. However, 85% of trainees believe that their training can still be improved further. Most report that a greater level of supervision is needed, along with easier access to the JAG accredited courses. New technologies such as simulators and the 3-D imager were considered by the majority to have contributed towards the improvement, but most trainees feel that their main priority was better access to dedicated training lists and courses.

2.2.4 Discussion

This study is one of the first to be able to quantitatively evaluate the impact of recent changes to endoscopy training. It shows that the majority of issues that had caused concerns regarding the provision of colonoscopy training in 2002 have significantly improved by 2007. Trainees report better supervision, more formal training, and greater access to accredited courses. The focus on 'Training the Trainers' has also clearly made an impact by disseminating good practice outside of specialist endoscopy courses. Certain aspects of training still remain issues, particularly those to do with record keeping and formal accreditation of competence. Despite the increase in course attendance, trainees still feel that they require more exposure to colonoscopy, both in terms of dedicated training lists as well as access to more specialised courses.

The main limitation of this study is the relatively small number of trainees who responded to the survey. However, the numbers were equivalent for both cohorts (31 responses in 2002, 32 in 2007), allowing a direct comparison. The low response rate, such as occurred from the surgical trainees, can occur during questionnaire surveys and as such are liable to response bias and sampling errors. However, a response rate of over 50% may be considered to be representative of the population

being surveyed[236]. The 2007 survey is therefore likely to be representative of the experience of the gastroenterology trainees in the region. However, the lack of engagement by the surgical trainees is noteworthy, as this may reflect a disinterest in quality assurance of endoscopy training in independent surgical endoscopists as well.

One difference between the cohorts was that the 2007 questionnaire was administered electronically. Response rates to Internet or electronic questionnaires in general tends to be lower than the traditional pencil and paper version, although in some studies, particularly involving adolescents[237,238] the response rates can be very high. Electronic administration of questionnaires has been shown to have good construct reliability and test-retest reliability in a randomised study of 16 instruments[239] however, so direct comparison with the previous data are valid, particularly as the questions and responses were unchanged.

2.3 National training survey

2.3.1 Background and Aims

Although 'snapshot' surveys of training are not as useful for quality assurance purposes as comparative or prospective studies, they can give some indication of standards at a specific point in time. All previous studies share a common thread of trainee dissatisfaction with their training programmes. A survey of European final year gastroenterology trainees in 2002 found that one-third felt that they did not have sufficient training in endoscopy by the end of their training programmes[10]. In the UK, the colonoscopy audit in 2004 by Bowles et al[2] found that only 17% of practicing endoscopists were felt to have been trained adequately.

A more detailed survey of UK trainees was conducted in 2004[240], but unfortunately the results were not published until 2009 after the survey detailed in this chapter was performed, so a direct comparison of results was not possible. This study surveyed less than half of the gastroenterology trainees in the UK (172 replies from 250 questionnaires), but also demonstrated variation in both the quality and quantity of

endoscopy training. 55% of trainees were being trained less than half the time in procedures in which they still required training, and supervision was often inadequate or absent.

The aim of this survey was to investigate the impact of recent changes to gastroenterology training as a whole and gather data on current standards of teaching, training, and supervision in endoscopy on a national basis.

2.3.2 Methods

This was a cross-sectional closed survey. A web-based questionnaire was developed with questions in eight different categories, including endoscopy. There were nine specific endoscopy questions, with available Yes/No, Likert, numerical and free text responses. An invitation to complete the survey was emailed by the regional Trainees in Gastroenterology (TiG) representatives to all trainees in their region. Several reminder emails to non-responders were sent at weekly intervals, but participation was entirely voluntary. The survey was also advertised on the TiG and BSG websites. A £500 prize was donated by the BSG Endoscopy section which was then allocated by a random draw of all respondents following the close of the survey. Responses were collected during October 2008. Data was automatically compiled by the survey website (www.surveymonkey.com) and exported into a spreadsheet for analysis. No statistics were performed, as this was purely a descriptive rather than comparative survey.

2.3.3 Results

358 out of 581 trainees from 20 different training regions completed the survey, with an overall response rate of 61.6%. Responses by region are shown in Figure 15. There was a good spread of responses across the training years (Figure 16).

Figure 15 Survey responses by region

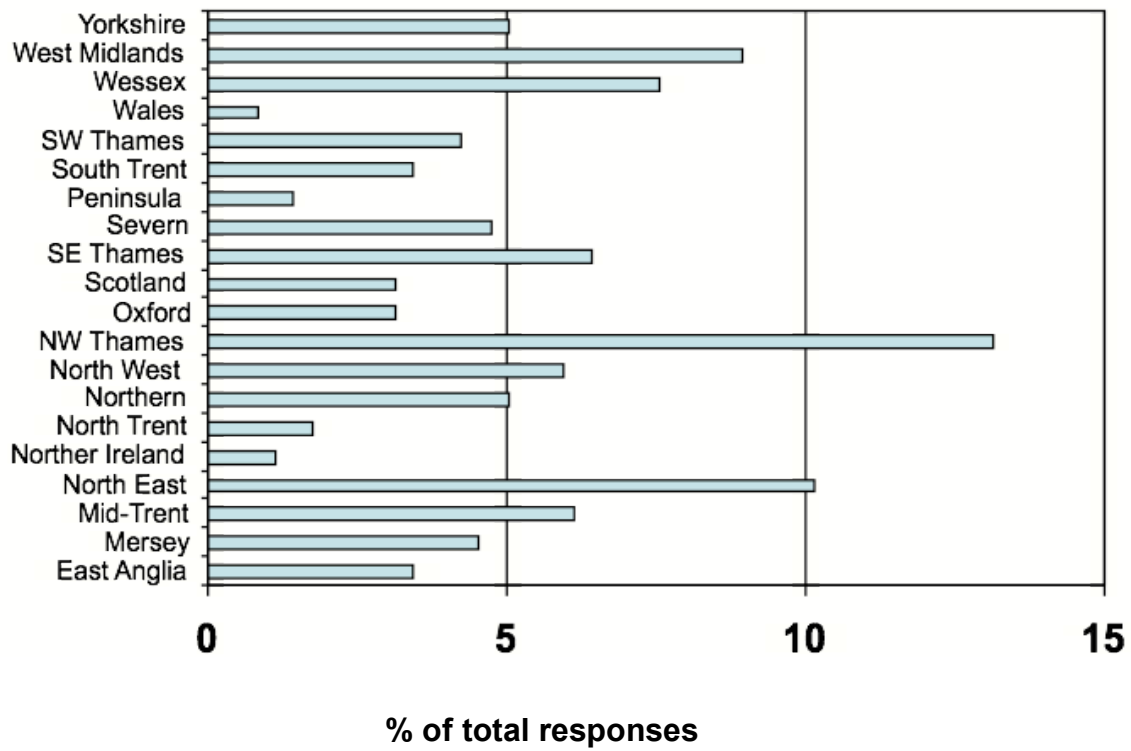
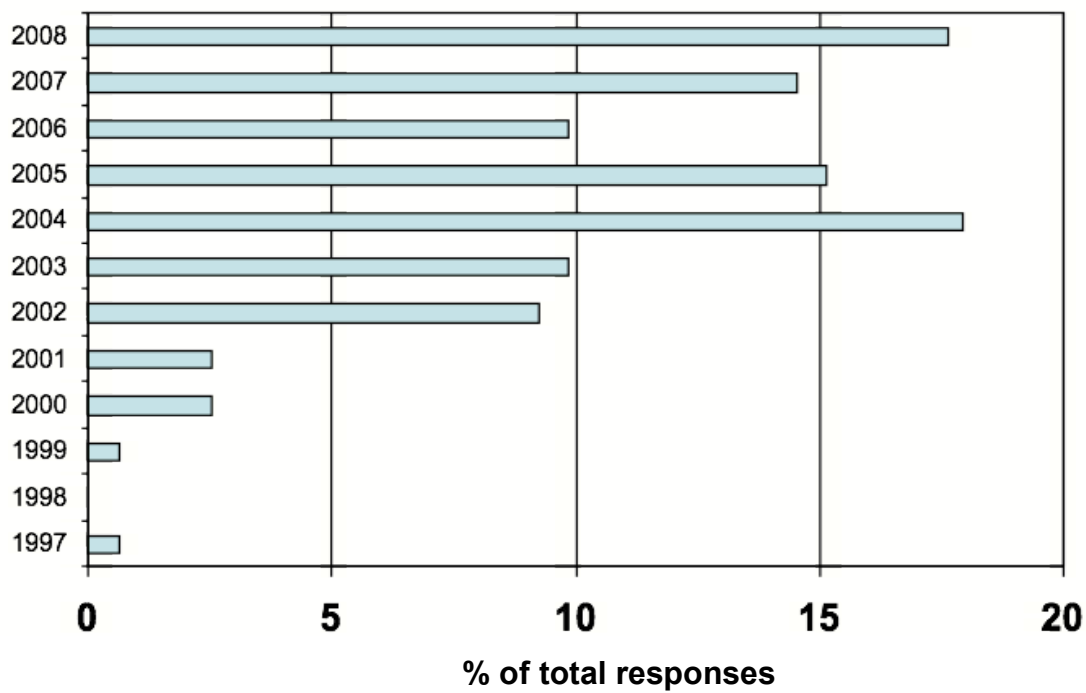
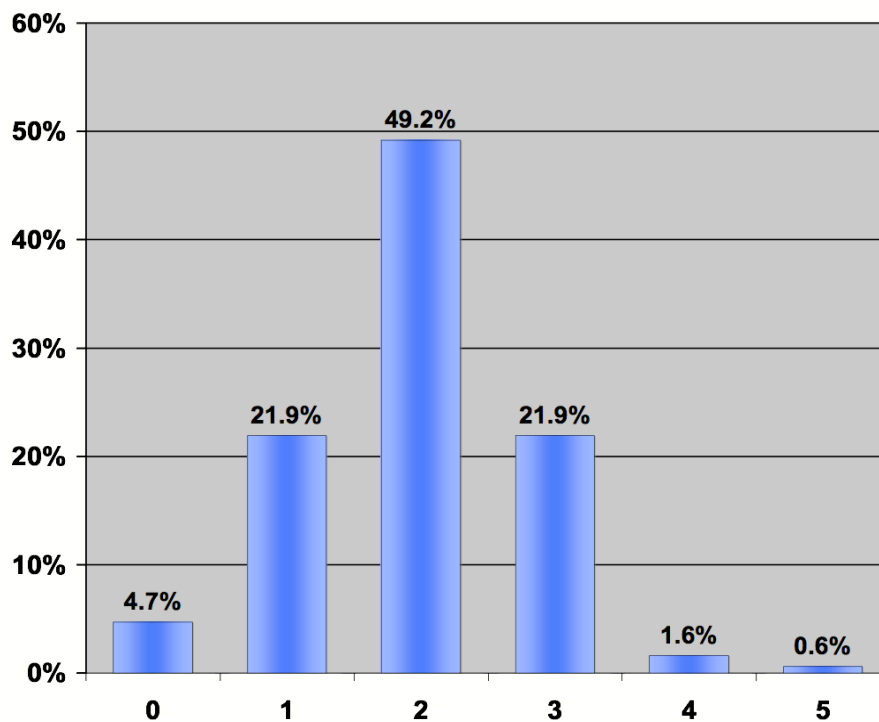


Figure 16 Survey responses by training year (start date)



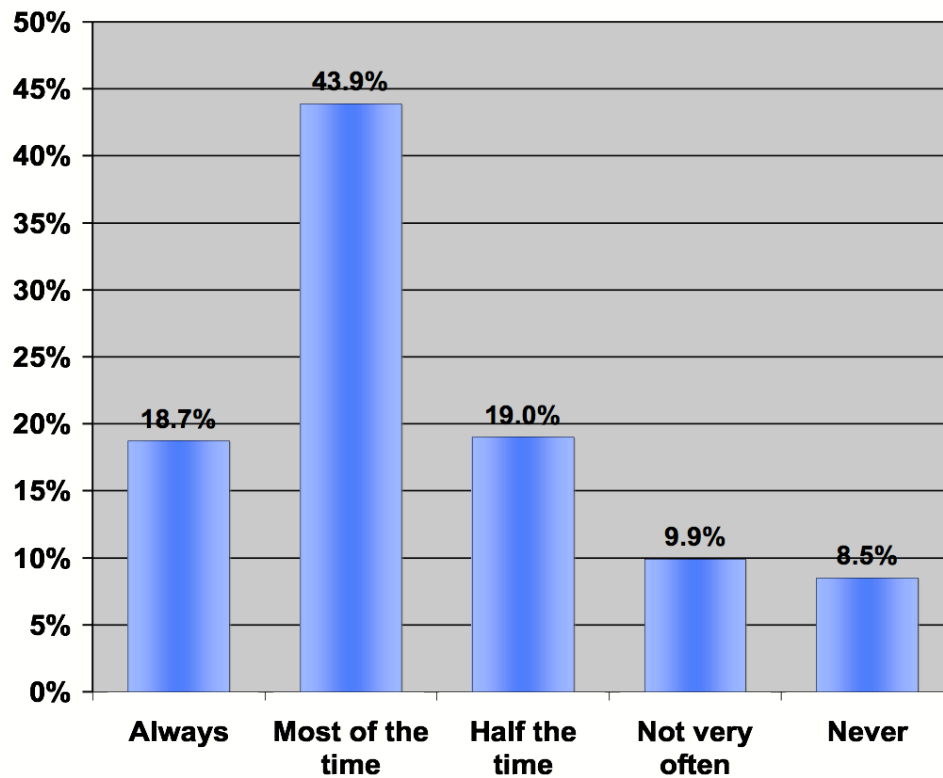
85/319 (26.6%) attend fewer than the recommended minimum of 2 endoscopy sessions per week (2004 guidelines for training) (Figure 17). However, only 49% of these are dedicated training lists and 32% of trainees say they have no dedicated training lists at all.

Figure 17 Number of endoscopy sessions per week



Of those sessions that are allocated as training sessions (rather than service provision), over a third of trainees said they actually happened only half the time or less, with 8.5% saying they never actually had a training list (Figure 18).

Figure 18 Actual occurrence of training lists



However, course attendance was relatively high, with 79% of trainees having been on at least one accredited training course, although only 31% have been on the compulsory basic foundation skills course (Figure 19).

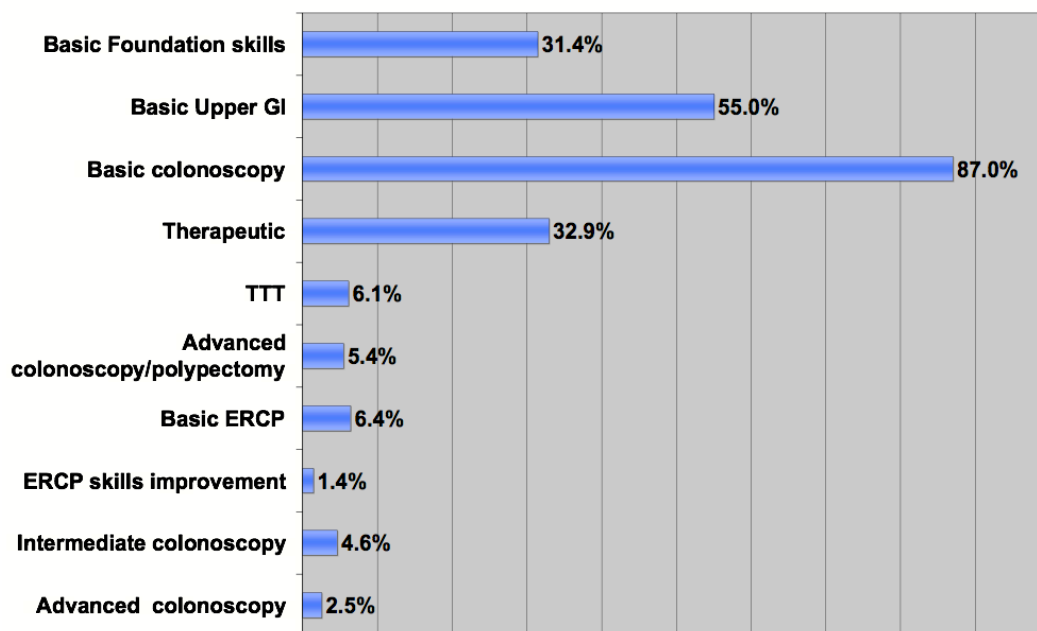


Figure 19 Attendance at endoscopy training courses

Only 41% (126/310) of trainees routinely get formative DOPS filled in for their lists, but just over half (51%) had already completed at least one summative DOPS, with 91% completing a DOPS for OGD and 59% for colonoscopy of those who responded positively. 236/321 (74%) reported that they have had access to an endoscopy simulator at some point in their training. 237/314 (76%) report that they have opportunities to perform emergency endoscopy.

Overall, 258/326 (78.6%) were very or fairly satisfied with their endoscopy training so far, with 19.3% fairly unsatisfied and only 2.1% very unsatisfied. The most common complaints were lack of regular training lists, missing lists due to general medical commitments, and difficulties in getting experience in colonoscopy or therapeutic procedures.

2.3.4 Discussion

In comparison to the data presented from the audit in 2004, the experience of current gastroenterology trainees in endoscopy training is generally positive. Trainees are much more satisfied, and levels of supervision are much improved. Most trainees have access to specialist training in endoscopy, with high levels of course attendance and the majority have had some experience of endoscopic simulation. However, similar areas of improvement were identified, such the conflict between service provision and training and the difficulties in organising dedicated training lists.

The strengths of this study are the large numbers and wide sampling frame that was achieved. We obtained responses from over 60% of all trainees in the country, which gives strong credibility and validity to the results. There was a financial incentive to undertake this survey which may have induced some response bias, although participation was entirely voluntary and the high response rate means that the data can be considered representative of the group as a whole. The results are also very similar to the data that was obtained in the comparative single-region colonoscopy survey, again giving credence to the findings.

2.4 Conclusions

The improvements in training reported by these two surveys are very likely to be due to the impact of the central training programme and funding, as the effects were seen across the country. However, it is difficult, if not impossible to quantify and adjust for potential confounders. Also, as these were both surveys investigating trainees' perceptions of their training, any data are self-reported and potentially biased, so these results may not be present in actuality. Unfortunately, no procedures to prospectively evaluate the effectiveness of the central training programme were put into place so determination of efficacy will have to be made by studies of this kind and to our knowledge these are the most comprehensive assessments yet made.

The data presented here is also limited to gastroenterology trainees. To be comprehensive, it should include other stakeholders such as surgical, nurse and non-medical endoscopists. It is important to engage these groups in the process of quality assuring their training, as failure to do so may result in performance gaps between different practitioners. A pilot study of training nurse endoscopists in polypectomy[241] has been performed by myself and Maggie Vance through the Gastrointestinal Nurse (GiN) training project, but is not covered further in this thesis.

In summary, changes over the last five years have made a significantly positive impact on gastroenterology trainees' perceptions of endoscopy training. Although the national training programme continues and accredited courses continue to be provided, there now have to be paid for out of trainee's individual study budgets. The loss of central investment may have a detrimental effect on future endoscopy training, although the full impact of the continuing 'Train the trainers' programme may not yet have been felt.

Chapter 3: Construct validity of a colonoscopy simulator

3.1 Background and Aims

Simulators potentially have a place in the objective assessment of technical skills in endoscopy. However, my review of the current published literature (p.44) has shown that the majority of studies found only a very few of the metrics recorded were valid measurements of skill, and so their utility as assessment tools remains poor. Two previous reviews of the literature have concluded that until the simulation models improve, they have no place in either assessment or credentialling in gastrointestinal endoscopy[137,141].

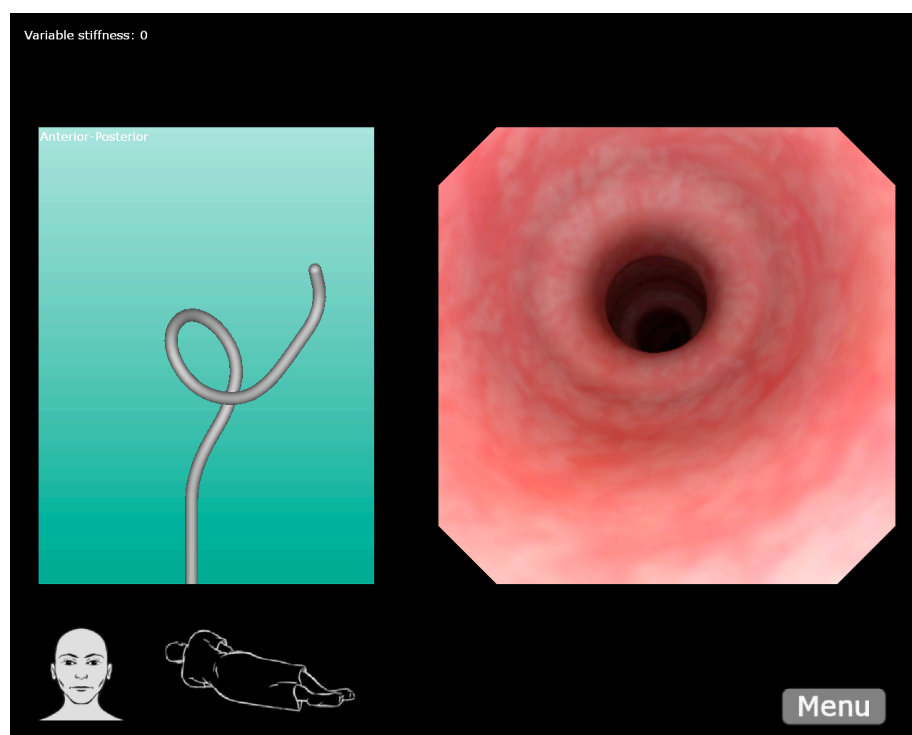
The Olympus colonoscopy simulator (Endo TS-1, Olympus Keymed, Southend, UK; Figure 20) is a second-generation virtual reality simulator for training and assessment of colonoscopy skills. It has been specifically designed to model real-time movements of the colon and provides a much more realistic simulation for colonoscopy than has previously been possible[242]. It utilises a customised Olympus CF180L endoscope and simulates shaft looping, tip contact, variable shaft stiffness, application of abdominal pressure and movement of the patient. It also provides a simulated 3-D endoscope imager view analogous to that provided by ScopeGuide™. It has training material covering insertion and withdrawal technique, including lesion targeting, and the software is currently being developed to simulate therapeutic procedures such as polypectomy.

Figure 20 External view (A) and screen view (B) of the Olympus ENDO TS-1 colonoscopy simulator

A



B



A study was performed by Koch et al in 2007 to establish the initial validity of the simulator[200]. At the time, the simulator only produced six outcome measures, and while the study demonstrated reasonable face validity, it only determined construct validity for one metric, the time taken to reach the caecum. My initial work with the simulator programmer, Dr John Bladen, was therefore in developing further metrics to evaluate performance. Forty-one measurements were developed in four different domains: general measurements, technical aspects, time to landmarks, and looping characteristics. The aim of this study was to further establish **face and construct validity** (see p.34) of the simulator and to establish which of the new assessment measures map to clinical benchmarks of expertise.

3.2 Methods

Participants were recruited from three groups of colonoscopists at a single tertiary referral centre; novices, intermediates and experts. All participants completed a questionnaire on demographics, endoscopy experience, and previous use of simulators. They were given an introduction to the simulator with instructions on how to operate suction, insufflation, variable stiffness, screen wash, and made aware of the patient's facial expression to indicate discomfort. Novices were allowed to familiarise themselves with the endoscope controls before performing their first case. All participants performed three standardised cases which had been chosen to reflect a range of difficulty while keeping to a time of approximately one hour for each subject. The first case was a straightforward sigmoid N-loop combined with a moderate transverse loop. The second case has a straightforward sigmoid N-loop and moderate transverse loop, but in combination with a low pain threshold for insufflation and stretch similar to an 'irritable-bowel type patient'. The third case is a difficult combination of a sigmoid alpha loop and moderate transverse loop. Participants were randomised to availability of the simulated 3-D imager view for all three procedures. None of the participants who were randomised to having the imager view available were given any instruction on the interpretation of the images, although some of the intermediate and expert group were familiar with the use of ScopeGuide™ for real colonoscopy. The participants were instructed to perform the cases to the best of their ability. Cases were considered complete when caecal intubation was achieved or were time-limited to 15 minutes. During the cases, patient position change and abdominal pressure was applied at the direction of the

endoscopist. All cases were recorded by the simulator and automatically exported to a database for analysis. Following each case the participants were asked to rate the level of difficulty, difficulty of loop resolution, and the realism. Following all three cases, participants were asked to rate the simulator in terms of visual appearance, scope movements, force feedback, realism of looping and realism of loop resolution. Ethical approval was obtained from Harrow ethics committee (RD7 062).

3.2.1 Statistical analysis

Analysis was performed by a consultant medical statistician using General Estimating Equations (GEEs). The analysis of continuous outcomes assumed a linear regression approach, with a normal distribution and identity link function. A large number of outcomes were positively skewed, and were given a log transformation before analysis. The analysis of binary outcomes used a logistic regression approach and assumed a binomial distribution with a logit link function. The overall difference between observer groups was examined. If this result was statistically significant, post-hoc tests were used to compare between pairs of groups. To allow for multiple testing, these comparisons were given a Bonferroni adjustment. Where no logistic regression was possible, Fisher's exact test was used to compare categorical data. All p-values presented in the text and tables for the paired comparisons are the adjusted values, and the 'Overall' p-values are unadjusted.

3.3 Results

34 participants were recruited over a six week period; 10 novices (medical students or nurses) with no prior colonoscopy experience, 13 intermediate endoscopists (trainees) with less than 1000 previous colonoscopies (mean 315), and 11 expert endoscopists with over 1000 previous colonoscopies (mean 4772). All participants completed the questionnaire, cases and feedback.

3.3.1 Construct validity

On the general measures (Table 2), the simulator distinguished between the groups for completion rate, maximum tip position and time to completion. There were significant differences between the novices and both the trainees and experts for

completion rate and speed of completion. Although the experts completed the cases in a shorter time than the trainees (299s vs 374s), this was not statistically significant.

For the technical measures, the metrics regarding abdominal pressure indicate whether pressure was applied in an appropriate or an inappropriate zone to prevent the sigmoid looping whilst the tip section is in the region of the splenic flexure, based on previous research[157]. “Correct use of variable stiffness” indicates whether the variable stiffness function was set to a value greater than zero whilst the tip section was at the splenic flexure. There were four discriminatory metrics (Table 3). The first was correct use of the variable stiffness function ($p < 0.001$), which both experts and trainees performed more often than novices. Although expert use of the variable stiffness function was more often correct than the trainees (52% vs 29%), this was not statistically significant ($p = 0.10$). The other metrics were maximum and mean shaft insertion force, demonstrating that novices pushed less than both trainees and experts, and the time spent pushing with an embedded tip, which discriminated novices from both trainees and experts ($p = 0.03$). There were no significant differences in the maximum or mean simulated patient pain scores between the groups.

All of the time-to-landmark measurements were discriminatory (Table 4) for novices compared to both trainees and experts. The hepatic flexure intubation time (time to get round the hepatic flexure from the proximal transverse colon) was not discriminatory, presumably because it was too easy as all three groups managed to pass the flexure in a under 2 seconds. Two metrics did not discriminate between trainees and experts; the time to hepatic flexure, and the time to caecum. Although the experts attained both landmarks more quickly than the trainees, the difference was not significant.

With regard to the looping parameters, three metrics for sigmoid N-looping were discriminatory; the number of loops ($p = 0.02$) and the mean loop size during ($p = 0.01$) and after sigmoid intubation ($p < 0.001$). The more experienced colonoscopists created more loops as they progressed through the sigmoid, but these loops were smaller than for either novices ($p = 0.02$) or trainees ($p = 0.04$). Once through the

sigmoid, they kept the scope straighter for the remainder of the procedure (Table 5). The alpha-looping parameters; time to remove the loop and time with loop after sigmoid intubation, were both discriminatory ($p=0.004$ and $p=0.006$ respectively) (Table 6). Although one comparison for the transverse looping, the number of loops, was statistically different between novices and experts (Table 7), the median number of loops was identical and is not likely to reflect a real difference in performance.

3.3.2 Face validity

The group of experts graded the tasks as expected in terms of difficulty, with the alpha loop proving more difficult to resolve than the simpler sigmoid N-loops (Table 8). The ratings for scope movement, force feedback, looping and loop resolution were good, over 6.0 out of 10, with the scores for realism of appearance slightly lower at 5.8 out of 10.

3.3.3 3-D Imager

The groups had a similar proportion of participants randomised to availability of the imager (novice 50%, intermediate 54% and expert 37%; $p=0.31$, Chi-squared test) and the groups were therefore considered comparable. There were no statistically significant differences between those participants who had the 3-D imager view available and those who did not on any of the measures recorded.

Table 2 General measures

	Trainees / Novices Odds Ratio (95% CI)	Experts / Novices Odds Ratio (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Intubated to caecum	7.20 (2.32, 22.3)	12.1 (3.23, 45.1)	<0.001	0.003	<0.001	1.00
Maximum tip position (max = 40)	3.66 (1.40, 5.92)	4.70 (2.34, 7.06)	<0.001	0.003	<0.001	1.00
	Trainees - Novices Difference (95% CI)	Experts - Novices Difference (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Time taken (s)	-305 (-416, -194)	-405 (-521, -289)	<0.001	<0.001	<0.001	1.00

N – Novice, T – Trainee, E - expert

Table 3 Technical measures

	Trainees / Novices Odds Ratio (95% CI)	Experts / Novices Odds Ratio (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Abdominal pressure used correctly	0.80 (0.31, 2.06)	2.34 (0.73, 7.48)	0.19			
Abdominal pressure used incorrectly	0.34 (0.10, 1.09)	0.82 (0.31, 2.16)	0.14			
Variable stiffness used correctly	11.5 (1.45, 90.5)	32.0 (3.59, 286)	0.004	0.06	0.006	0.10
Excessive inflation used	3.06 (0.81, 11.5)	3.73 (1.08, 12.9)	0.10			
Excessive shaft insertion with embedded tip	0.38 (0.10, 1.39)	0.31 (0.09, 1.11)	0.18			
Time with high shaft force and looped sigmoid (s)	1.26 (0.37, 4.33)	0.98 (0.26, 3.70)	0.91			
	Trainees – Novices Difference (95% CI)	Experts – Novices Difference (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Shaft insertion force – max (N)	1.94 (0.50, 3.38)	2.06 (0.55, 3.57)	0.01	0.02	0.02	1.00
Shaft insertion force – mean (N)	0.35 (-0.05, 0.76)	0.53 (0.11, 0.96)	0.04	0.21	0.03	1.00
Shaft insertion length with embedded tip	0.78 (0.59, 1.01)	0.80 (0.61, 1.04)	0.17			
Shaft insertion time with embedded tip*	0.49 (0.28, 0.84)	0.49 (0.29, 0.84)	0.03	0.03	0.03	1.00
Patient pain – maximum	0.10 (-0.06, 0.27)	0.03 (-0.14, 0.20)	0.43			
Patient pain – mean*	1.13 (0.90, 1.44)	1.13 (0.91, 1.42)	0.40			

N – Novice, T – Trainee, E - expert

*Variables analysed on log scale reported as ratio of values in each group relative to novice group.

Table 4 Time taken to landmarks and time taken to negotiate flexures

	Trainees / Novices Ratio (95% CI)	Experts / Novices Ratio (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Time to mid sigmoid	0.45 (0.31, 0.66)	0.29 (0.20, 0.41)	<0.001	<0.001	<0.001	0.003
Time to sigmoid- descending junction	0.50 (0.33, 0.77)	0.29 (0.19, 0.44)	<0.001	0.009	<0.001	<0.001
Sigmoid intubation time	0.52 (0.38, 0.72)	0.33 (0.24, 0.46)	<0.001	<0.001	<0.001	0.003
Sigmoid intubation and straighten time	0.55 (0.39, 0.77)	0.38 (0.27, 0.54)	<0.001	0.003	<0.001	0.01
Time to splenic flexure	0.53 (0.39, 0.72)	0.35 (0.25, 0.49)	<0.001	<0.001	<0.001	0.02
Splenic flexure intubation time	0.45 (0.32, 0.63)	0.28 (0.19, 0.42)	<0.001	<0.001	<0.001	0.006
Time to mid transverse	0.52 (0.41, 0.66)	0.34 (0.26, 0.43)	<0.001	<0.001	<0.001	<0.001
Time to hepatic flexure	0.55 (0.42, 0.71)	0.43 (0.32, 0.58)	<0.001	<0.001	<0.001	0.13
Hepatic flexure intubation time	0.80 (0.45, 1.41)	0.72 (0.40, 1.29)	0.51			
Time to caecum	0.60 (0.47, 0.75)	0.50 (0.37, 0.68)	<0.001	<0.001	<0.001	0.63

N – Novice, T – Trainee, E - expert

Table 5 Sigmoid N-looping

	Novices N (%)	Trainees N (%)	Experts N (%)	Overall	p-value		
					N vs T	N vs E	T vs E
More than 1 loop formed during sigmoid intubation	0 (0%)	1 (3%)	5 (15%)	0.02	1.00	0.16	0.26
	Trainees - Novices Difference (95% CI)	Experts - Novices Difference (95% CI)		Overall	N vs T	N vs E	T vs E
Time spent looped during sigmoid intubation (s)*	0.89 (0.52, 1.53)	0.80 (0.47, 1.34)		0.68			
Number of loops formed after sigmoid intubation*	1.19 (0.95, 1.48)	1.30 (1.03, 1.65)		0.06			
Max loop size during sigmoid intubation	1.07 (0.37, 3.08)	0.72 (0.25, 2.12)		0.78			
Mean loop size during sigmoid intubation	-0.01 (-0.09, 0.07)	-0.03 (-0.11, 0.05)		0.01	1.00	0.02	0.04
Max loop size after sigmoid intubation	-0.02 (-0.08, 0.04)	-0.09 (-0.16, -0.03)		0.60			
Mean loop size after sigmoid intubation	-0.02 (-0.15, 0.12)	-0.07 (-0.20, 0.07)		<0.001	0.003	<0.001	0.35

N – Novice, T – Trainee, E - expert

*Variables analysed on log scale reported as ratio of values in each group relative to novice group.

Table 6 Sigmoid alpha looping

	Trainees - Novices Difference (95% CI)	Experts - Novices Difference (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Time to remove loop (s)*	0.48 (0.22, 1.06)	0.28 (0.13, 0.60)	0.004	0.20	0.006	0.10
Time with loop after sigmoid intubation (s)*	0.21 (0.06, 0.75)	0.23 (0.09, 0.63)	0.006	0.05	0.02	1.00

N – Novice, T – Trainee, E - expert

*Variables analysed on log scale reported as ratio of values in each group relative to novice group.

Table 7 Transverse looping

	Trainees - Novices Difference (95% CI)	Experts - Novices Difference (95% CI)	Overall	p-value		
				N vs T	N vs E	T vs E
Number of transverse loops*	1.22 (1.03, 1.46)	1.32 (1.10, 1.57)	0.01	0.06	0.01	0.72
Time with transverse loop (s)*	1.79 (0.63, 5.13)	2.08 (0.78, 5.55)	0.32			
Transverse intubation and straighten time (s)*	0.72 (0.45, 1.17)	0.73 (0.36, 1.47)	0.39			
Transverse intubation time (s)*	0.63 (0.39, 1.01)	0.70 (0.37, 1.32)	0.15			
Transverse loop size - Max	0.03 (-0.02, 0.08)	0.06 (0.01, 0.10)	0.07			
Transverse loop size - Mean	-0.03 (-0.09, 0.03)	-0.01 (-0.07, 0.05)	0.67			

N – Novice, T – Trainee, E - expert

*Variables analysed on log scale reported as ratio of values in each group relative to novice group.

Table 8 Expert ratings of the simulator cases

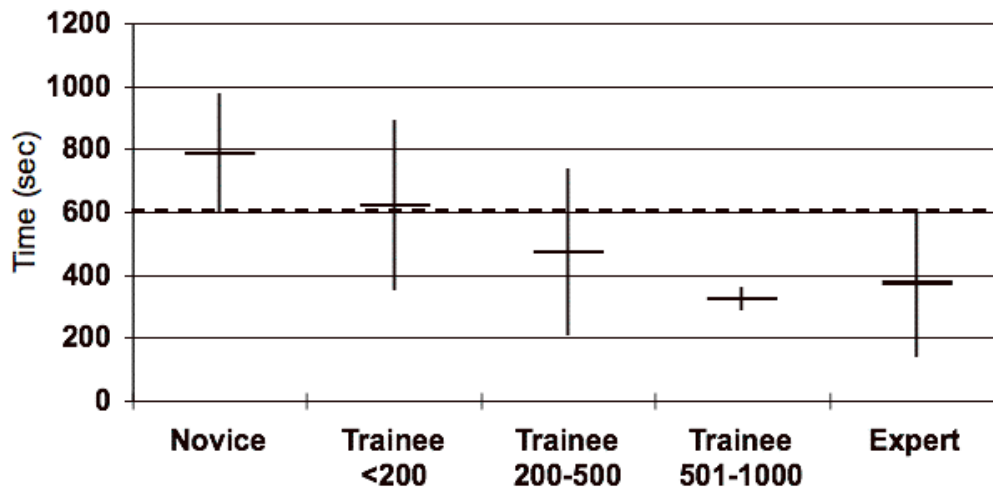
Appearance	Movement	Force feedback	Looping	Loop resolution
5.8 (2.4)	6.4 (2.3)	6.6 (1.5)	6.6 (1.5)	6.8 (1.6)
Case1	Case2	Case3	Overall	
Overall difficulty	3.2 (2.2)	3.3 (2.4)	7.3 (1.9)	4.6 (2.9)
Difficulty of loop resolution	2.8 (1.7)	4.0 (2.4)	6.3 (2.5)	4.4 (2.6)
Overall realism	6.1 (1.7)	6.3 (1.9)	5.8 (2.2)	6.0 (1.9)

Results are mean (SD) out of 10

3.4 Discussion

This study demonstrates that the second-generation Olympus colonoscopy simulator has a high level of fidelity compared to real life colonoscopy. According to the experts, the simulator offers a good representation of colonoscopy in terms of looping and loop removal, force feedback and scope movements. The simulator demonstrates excellent construct validity as an assessment tool in that it can identify and separate subjects by experience level for multiple parameters. Not only does it discriminate on time-to-landmark measurements, which has been demonstrated for a number of previous simulations, but it also discriminates in terms of important clinical endpoints such as completion rate. This study has also established validated metrics for several technical aspects of colonoscopy insertion such as speed of loop resolution, insertion of the shaft with an embedded tip and correct use of the variable stiffness function. These have shown a graded improvement with increasing expertise and may be used as markers of proficiency.

We have used the expert scores on these measurements to benchmark the threshold levels of performance. For example, a cut-off of 10 minutes (1 Standard Deviation below the expert mean) to complete the case could be used as a discriminatory benchmark (see Figure 21). This anchors the assessment to clinically relevant standards based on obtaining optimum patient outcomes, and is considered to be a more appropriate way of setting standards than simply assigning cut-offs based on purely expert opinion[243]. Development of milestone benchmarks can enhance the use of the simulator as a training tool in order to reach the desired performance outcomes. Comparison of contemporary performance abilities can provide formative feedback to trainees and enhance learning. Metrics such as incorrect use of the variable stiffness function could be used to provide feedback to candidates failing an assessment as to how they can improve their technique. This type of information can easily be incorporated into the results screen displayed after task completion, allowing immediate and accurate feedback to be given in the absence of an instructor.

Figure 21 Graph of time to caecum for all cases combined.

The trainee group has been separated into subgroups of <200, 200-500 and 501-1000 previous actual colonoscopies. Measurements are mean $\pm$ 1SD. The dashed line represents a benchmark standard of 10 minutes.

The findings from this study are supported by recent a publication on the use of a novel ex-vivo bovine model for assessment in colonoscopy[128]. This has shown construct validity and results for intubation times that correlated well with median intubation times obtained from clinical practice. The results from this well-constructed study provide evidence that simulators with higher fidelity can be used for skills assessment in colonoscopy. However, most of the performance parameters were based on time-to-landmark measurements, which although objective, do not take into account other measures of skill such as patient pain, looping and insufflation that can be recorded using a computer simulator. Using the bovine model for assessment also suffers from other drawbacks such as the need for expert opinion for some of the assessment metrics and the need for fairly frequent replacement of the bovine colon preparation. However, it does have some advantages over computer simulations, notably the initial purchase cost.

There are several other limitations to our study. Two trainees in the intermediate group had experience of between 500 and 1000 colonoscopies, which in most eyes would deem them to be experienced practitioners. However, as they were still not independent practitioners, they were included in the intermediate group for analysis.

It is important to note that despite their inclusion, the discriminatory metrics demonstrated a graded improvement in performance with experience, and the fact that the simulator could distinguish between the groups even while including colonoscopists of such experience is one of the strongest arguments for its use as a high-stakes assessment tool.

The experienced group only had an 88% overall completion rate, which is lower than might have been expected. This may be because participants were only allowed 15 minutes to intubate to caecum. Although most experts can intubate to the caecum in the majority of real cases in under 15 minutes, difficult cases may take significantly longer to achieve complete colonoscopy. In this study, all four occasions in which the experts did not complete the case were on the difficult alpha loop, again reflecting the high fidelity of the simulation. It may also be that the lack of the 3-D imager view for 7 of the 11 experts (63%) had an effect on loop resolution and therefore completion rates for the difficult case. However, our data did not show any statistically significant benefit for the use of the 3-D imager on any of the measures recorded. The current published literature on the utility of ScopeGuide is limited and the findings inconsistent. Several of the published studies have shown a benefit in terms of completion rate and time to completion[158,159], but these findings have not been reproduced by other studies[244,245]. There is also some other evidence to show a beneficial effect on pain scores[158] and accuracy of localising lesions[246], but again these have not been replicated[247]. Although the study was not specifically powered to investigate the effect of having a 3-D imager view available, our data suggests that there is no significant benefit for untrained users during simulated colonoscopy. It is possible, however, that the imager would have a beneficial effect for colonoscopists who have been trained in its use and interpretation of the images.

The use of the simulator to produce objective metrics does not preclude the addition of expert opinion to assess competence during simulated cases. This has been done in one previous study[210] which demonstrated moderately strong correlations between two out of eight of the computer metrics with faculty ratings. Ideally simulator metrics should correlate with expert ratings of performance on live cases, sometimes termed predictive validity. This is more difficult to demonstrate due to the

inherent variability in live case assessment, and requires a large number of endoscopists to be assessed in a blinded fashion. Chapter 4 details a large multi-centre study on the Olympus simulator investigating training outcomes for novice colonoscopists using such blinded assessments of performance. Chapter 5 uses the data from these training outcomes to investigate the criterion (concurrent) validity for the discriminatory markers validated in this chapter.

3.5 Conclusions

This study has demonstrated that the Olympus Endo TS-1 colonoscopy simulator has good face validity and high construct validity. It can discriminate between colonoscopists with a range of expertise on multiple measures, allowing an objective and repeatable assessment of procedural skill. Benchmarking of the discriminatory metrics has allowed threshold levels of competence to be set, which may allow the simulator to have a place as part of a summative assessment.

Chapter 4: Use of a colonoscopy simulator in training

4.1 Background and Aims

Having determined that the simulator demonstrates face and construct validity, the next step in the evaluation process is to determine its efficacy as a training tool. At the start of this research, the simulator had some training material available including games, videos, full cases and part-task cases (for example 'intubate and straighten the sigmoid colon'). However, these were not arranged in any kind of order, and the case selection only represented a small fraction of the teaching scenarios that the simulator could potentially provide.

The first step was therefore to design an educational package that utilised the simulator's abilities to the fullest extent. A training programme was developed that incorporated all of the necessary knowledge and skills that a colonoscopist requires for safe intubation. Information was given in video and text format which was embedded within the training programme. Knowledge was assessed using multiple choice questionnaires at the end of each segment. To maximise the educational value, each MCQ had a 100% pass-mark requirement, with wrong answers highlighted in the question review, allowing revision and repeated attempts.

Skills training was broken down into component segments and tasks were developed with specific learning objectives to meet these skills requirements. These 'part-tasks' were then arranged in a sequential fashion to enable skills development in the logical order required to progress from rectum to caecum. Dexterity tasks were developed and incorporated into the programme with increasing levels of difficulty to ensure development of hand skills. At the end of the programme, full cases were developed to ensure the trainee was able to integrate the skills. The simulated 3-D imager view was used throughout the training programme to highlight learning points regarding loop development and resolution, including the use of hand pressure and position change.

This programme was felt to provide a thorough grounding in the basics of colonoscopy. After a pilot trial with one novice trainee, it was estimated that it would take approximately 16 hours for a novice to complete the entire course.

The aims of this study were to:

1. Investigate performance outcomes of simulator training using this educational package compared to standard patient-based training
2. Investigate transfer of skills from simulator training to real-life colonoscopy.

4.2 Methods

This was a prospective, multicentre, randomised, blinded evaluation of training at four endoscopy training centres in three countries; St Mark's Hospital, London, UK; Erasmus Medical Centre, Rotterdam and the Academic Medical Centre, Amsterdam, the Netherlands; A. Gemelli University Hospital, Rome, Italy. The study was approved by all four institutional review boards and was given a rating of educational evaluation by the UK National Research Ethics Service and the Amsterdam ethical committee, and approval by the Rotterdam and Rome ethical committees.

4.2.1 Participants

Participants were identified at each centre by a sub-investigator. Trainees could be of any medical background (physicians, surgeons, nursing) and in a position recognised by the training institution as appropriate for training in colonoscopy.

Participants were excluded if they had experience of more than 25 previous colonoscopies or flexible sigmoidoscopies, had previously attended an intensive colonoscopy training course or were previously a subject in a colonoscopy training or simulator training study. Participants who had performed more than 10 laparoscopic surgical procedures were also excluded as it was felt that they may have to 'unlearn' endoscopic skills that may conflict with gaining flexible endoscopy skills. The participants were enrolled by a sub-investigator and randomised into subjects (simulator training) and controls (patient-based training) by the lead investigator using a computer generated block randomisation protocol with 8 per block.

Participants, sub-investigators and trainers in each institution were not blinded to the group allocation.

4.2.2 Pre-training

4.2.2.1 Questionnaire

A questionnaire was completed by all participants providing data on demographics, previous endoscopic experience, self-reported measures of competence and confidence in colonoscopy and expectations regarding training outcomes from the study. All participants received a standardised tutorial on the fundamentals of colonoscopy to ensure a minimum background knowledge with respect to the basic concepts of colonoscopy and scope handling.

4.2.2.2 Simulation pre-training assessment

All participants performed the three previously validated simulator cases to provide a measure of their baseline performance. Each case had a time limit of 20 minutes to intubate to the caecum, and provided a simulated ScopeGuide™ 3-D imager view to standardise it with the patient-based assessments following training. All procedures were recorded automatically by the simulator and evaluated using the computer-generated parameters that had been demonstrated to have construct validity in the previous study.

4.2.3 Training

4.2.3.1 Simulator training

Participants allocated to the subject group were given access to the standardised simulator-training program for up to 16 hours. Trainers were expected to provide minimal tuition and feedback.

4.2.3.2 Patient-based training

Participants allocated to the control group received 16 hours training in patient based colonoscopy (4 half-day sessions) by an expert trainer using a ScopeGuide™ imager. During these four lists the participants were required to perform a minimum of 8 colonoscopies under one-to-one supervision. Recommendations were made for topics to be covered aiming to standardise the training. All trainees were taught to use single-handed, one person technique for colonoscopy, but the instructors were otherwise told to provide the usual training for a novice colonoscopist.

4.2.4 Post training

4.2.4.1 Simulator assessment

All participants subsequently performed the same three standardised simulator cases as for the pre-training assessment to investigate the changes in performance following training. All procedures were recorded and evaluated using the same parameters.

4.2.4.2 Patient-based assessment

All participants performed three patient-based colonoscopies following their training. Cases were specifically chosen from patients scheduled to undergo a clinically indicated colonoscopy. In order to reflect the training standard, patients were excluded if they were older than 75 yrs or had pelvic or colonic surgery, and if they had a previous difficult colonoscopy. Standard operating procedures for the unit were followed during all examinations. Patients gave informed consent for participation. Sedation and monitoring were utilised as per standard practice in the unit. A ScopeGuide™ 3-D endoscope imager view was utilised for all the colonoscopies performed.

An expert assessor blinded to the group allocation of the trainee was present during all assessments. Assessors were asked not to provide any assistance, either verbal or practical, during the assessment, but could take over the procedure if they were concerned at any time about the safety of the procedure or the patient's well-being, if the trainee requested assistance or at any other time they considered necessary. If the assessor offered any advice or took over the procedure, then the assessment was terminated at that point. Otherwise the procedure was time limited to 20 minutes or when caecal intubation had been achieved. Caecal intubation was confirmed by the assessor by visualisation of 2 out of 3 of; the ileocaecal valve, the appendix orifice or the triradiate fold, as well as an imager view compatible with the tip of the endoscope in the caecum.

If the trainee did not achieve caecal intubation by 20 minutes, the assessor took over the procedure. The assessor straightened the scope as much as possible to remove any loops, and then recorded depth of insertion in centimetres and estimated the

position of the tip using the ScopeGuide™ image. If a case was terminated due to patient factors (poor bowel preparation, extensive diverticulosis, impassable stricture or poor patient tolerance), the trainee was given the opportunity to repeat the procedure on a second suitable patient. However, no second opportunity was allowed if the termination was due to the trainee experiencing difficulty or placing the patient at risk. At the end of the case, the expert assessor made a judgment as to the difficulty of the case on a 5-point scale.

4.2.4.3 Feedback

All participants completed a questionnaire following the study regarding their training and assessments.

4.2.5 Outcome measures

Assessment of proficiency on real colonoscopy was measured using previously validated structured assessment tools; the UK Joint Advisory Group (JAG) colonoscopy Direct Observation of Procedural Skills (DOPS) assessment form[35] and an expert global rating of performance adapted for colonoscopy, the Global Score[210,248]. Primary outcome measures were the expert assessor's score on the JAG DOPS and the Global Score. Secondary outcome measures were the time to completion, depth of insertion, and improvement in performance parameters at simulated colonoscopy.

Data from the simulator assessments were automatically exported in a coded fashion into a database on a secure server. Data from the patient-based assessments were entered in a coded fashion into a spreadsheet.

4.2.6 Statistical analysis

Similar studies in skills training and transfer have demonstrated effect sizes between 1 and 1.92 standard deviations when comparing treatment and control groups[210,249]. Using the Global Score as the primary measure, it was estimated that the standard deviation of the scores would be 5 units. The study was powered to detect a 5-unit (one standard deviation) difference between the two groups. With a 5% significance level and 80% power, 16 subjects were required for each group, 32

in total. To allow for some drop out of subjects and errors in data collection, recruitment aimed to enrol a total of 36 subjects into the study.

Statistical analysis was performed on an “intention-to-treat” basis. For categorical data, Fisher’s exact test was used to compare the results between the two groups. For ordinal data, a two-sample t-test was used to compare variables that were normally distributed, and the Mann-Whitney and Kruskal-Wallis tests for variables that were not normally distributed. For both the simulator assessments and the patient assessments, each case was analysed separately and then as a combined score from all 3 cases, which is the data presented here. Logistic regression was used to investigate the change in scores from pre to post test for the simulator assessments.

4.3 Results

Forty trainees were randomised, with thirty-six completing the study. Two trainees did not start due to limitations in availability of lists, one trainee completed the simulator pre-training assessment but had to leave for personal reasons before commencing the training, and one trainee completed the training and simulator assessments, but did not complete all three patient-based assessment cases (Figure 22).

The subjects and controls were reasonably well matched in their demographics and previous endoscopy experience (Table 9), with no trainee having any practical colonoscopy experience. Although there were more specialists in training in the control group than the subject group, this was not statistically significant. There were some differences in the performance metrics on the simulator pre-training assessment between the groups, with the controls performing significantly better than the subjects on several measures (Table 10).

Figure 22 Flow diagram of participants through the study

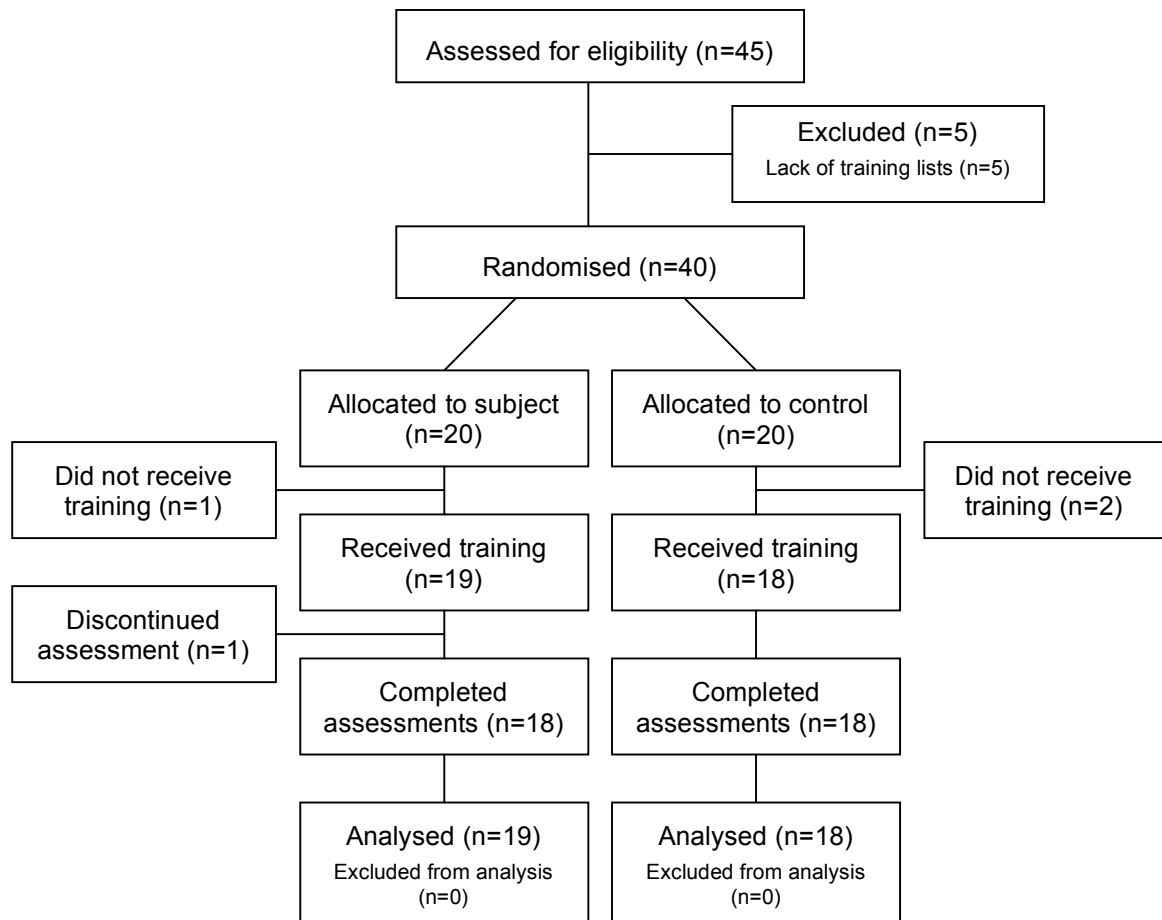


Table 9 Demographics and pre-course endoscopy experience

Variable	Controls (N=18) Median (IQR) or Number (%)	Subjects (N=19) Median (IQR) or Number (%)	p-value
Age	31 (26, 33)	28 (26, 30)	0.32
Male	10 (56%)	6 (32%)	0.19
Grade			
Nurse	3 (17%)	3 (16%)	
General Trainee	6 (33%)	10 (52%)	
Specialist in training	8 (45%)	3 (16%)	0.25
Other	1 (5%)	3 (16%)	
Colonoscopies witnessed	45 (21.25, 137.5)	15 (7.5, 125)	0.15
Colonoscopies assisted	1 (0, 30)	0 (0, 4)	0.21
Colonoscopies performed	0 (0, 0)	0 (0, 0)	0.72
Sigmoidoscopies	0 (0, 1)	0 (0, 0)	0.16
EGD	0 (0, 2)	0 (0, 0)	0.09

Table 10 Simulator pre-training assessment

	Controls	Subjects	
	Median (IQR) or Number (%)	Median (IQR) or Number (%)	p-value
General			
Intubated to caecum	32 (56%)	15 (26%)	0.002
Maximum tip position	40 (30, 40)	29 (26, 40)	<0.001
Time taken	1032 (749, 1200)	1200 (1194, 1200)	<0.001
Technical			
Patient pain – maximum	0.63 (0.38, 0.96)	0.58 (0.45, 0.70)	0.77
Insertion length with embedded tip	0.10 (0.06, 0.22)	0.14 (0.07, 0.29)	0.51
Insertion length with obscured lens	0.04 (0.01, 0.25)	0.13 (0.03, 0.53)	0.05
Insertion force – maximum	13.1 (3.2)	12.2 (4.7)	0.23
Correct use of abdominal pressure	19 (36%)	13 (27%)	0.40
Correct use of variable stiffness	17 (32%)	13 (27%)	0.67
Excessive inflation	16 (28%)	18 (32%)	0.84
Looping			
Number of sigmoid loops during sigmoid intubation	1 (1, 1)	1 (1, 1)	0.19
Time with sigmoid loop during sigmoid intubation (s)	104 (40, 182)	101 (11, 194)	0.90
Number of sigmoid loops after sigmoid intubation	1 (1, 1)	1 (1, 2)	0.05
Time with sigmoid loop after sigmoid intubation (s)	141 (63, 366)	135 (9, 294)	0.20
Number of transverse loops	1 (0, 1)	0 (0, 1)	0.006
Time with transverse loop (s)	64 (0, 184)	0 (0, 69)	0.01
Time to resolve alpha loop (s)	311 (216, 566)	368 (233, 880)	0.34

4.3.1 Post-training simulator assessment

Data from the simulator post-test is shown in Table 11. The subjects were significantly more likely to complete intubation to the caecum (95% vs 70%, $p=0.001$) and took approximately half as long to do so (407 vs 743s, $p<0.001$). They demonstrated superior technical skill in terms of reduced maximum patient pain scores ($p=0.001$), and shorter distances pushing with either an embedded tip ($p<0.001$) or obscured lens ($p<0.001$). They were more likely to use abdominal pressure correctly to assist intubation (79% vs 52%, $p=0.003$), but there were no differences in the correct use of variable stiffness. The subjects straightened sigmoid loops more quickly than the controls (34 vs 68s, $p=0.002$) and kept the scope straighter once the sigmoid had been passed ($p<0.001$). There were no differences in the management of transverse or alpha loops. Logistic regression confirmed that the subjects improved their performance from their pre-training assessment significantly or nearing significance compared to the controls on the majority of measures (Table 12).

4.3.2 Patient-based assessment

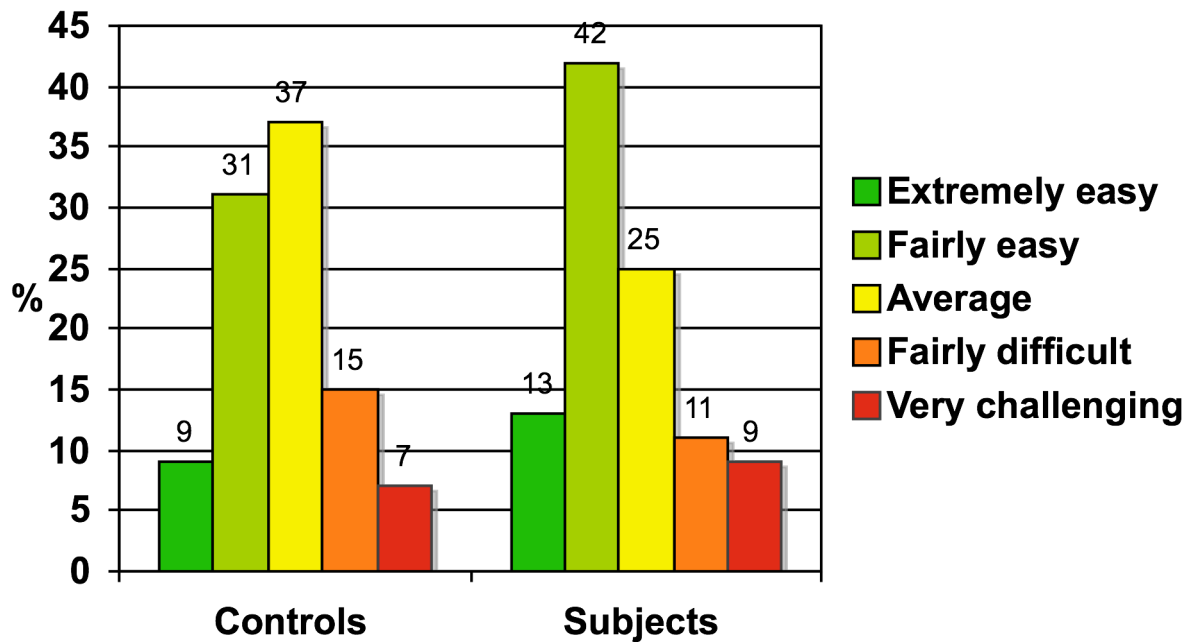
Two cases performed by one subject and one control were deemed to be unsuitable because of patient-related factors and were not included in the analysis. Each of the trainees performed a further case which was included. One subject was stopped by the expert assessor during their first patient-based assessment due to concerns about patient safety and did not perform any further cases. The remainder of the cases were comparable in difficulty between the two groups ($p=0.62$), with a normal distribution for both groups (Figure 23). There were no significant differences between the groups in terms of case completion, maximum tip position achieved, time taken, straight insertion depth, JAG DOPS score or Global Score (Table 13).

Table 11 Simulator post-training assessment

	Controls Median (IQR) or Number (%)	Subjects Median (IQR) or Number (%)	p-value
General			
Intubated to caecum	38 (70%)	54 (95%)	0.001
Maximum tip position	40 (35, 40)	40 (40, 40)	<0.001
Time taken (s)	743 (504, 1200)	407 (327, 504)	<0.001
Technical			
Patient pain – maximum	0.45 (0.19, 0.68)	0.24 (0.05, 0.43)	0.002
Insertion length with embedded tip	0.07 (0.05, 0.12)	0.03 (0.02, 0.08)	<0.001
Insertion length with obscured lens	0.03 (0.00, 0.14)	0.001 (0.00, 0.005)	<0.001
Insertion force – maximum	13.1 (2.8)	11.8 (1.6)	0.003
Correct use of abdominal pressure	28 (52%)	45 (79%)	0.003
Correct use of variable stiffness	21 (39%)	20 (35%)	0.70
Excessive inflation	13 (24%)	7 (12%)	0.14
Looping			
Number of sigmoid loops during sigmoid intubation	1 (1, 1)	1 (1, 1)	0.97
Time with sigmoid loop during sigmoid intubation (s)	68 (31, 104)	34 (9, 61)	0.002
Number of sigmoid loops after sigmoid intubation	1 (1, 2)	1 (1, 1)	<0.001
Time with sigmoid loop after sigmoid intubation (s)	113 (45, 240)	33 (2, 68)	<0.001
Number of transverse loops	1 (1, 1)	1 (1, 1)	0.98
Time with transverse loop (s)	61 (17, 168)	46 (6, 81)	0.12
Time to resolve alpha loop (s)	210 (158, 383)	176 (106, 224)	0.11

Table 12 Logistic regression of scores from pre-training to post-training assessments

	Odds Ratio (95% CI) (Subjects/Controls)	Mean Difference (95% CI) (Subjects - Controls)	p-value
General			
Intubated to caecum	13.3 (3.1, 53.5)		<0.001
Maximum tip position		3 (2, 5)	<0.001
Time taken		-390 (-499, -281)	<0.001
Technical			
Patient pain – maximum		-0.16 (-0.26, -0.06)	0.003
Insertion length with embedded tip		-0.07 (-0.13, -0.02)	0.006
Insertion length with obscured lens		-0.22 (-0.36, -0.08)	0.002
Insertion force – maximum		-1.2 (-2.1, -0.4)	0.005
Correct use of abdominal pressure	5.15 (1.93, 13.7)		0.001
Correct use of variable stiffness	0.94 (0.42, 2.14)		0.89
Excessive inflation	0.42 (0.14, 1.16)		0.09
Looping			
Number of sigmoid loops during sigmoid intubation		-0.1 (-0.5, 0.4)	0.82
Time with sigmoid loop during sigmoid intubation (s)		-49 (-77, -22)	0.001
Number of sigmoid loops after sigmoid intubation		-0.7 (-1.3, 0.0)	0.04
Time with sigmoid loop after sigmoid intubation (s)		-98 (-146, -51)	<0.001
Number of transverse loops		0.5 (-0.8, 1.0)	0.09
Time with transverse loop (s)		-37 (-75, 2)	0.06
Time to resolve alpha loop (s)		-85 (-164, -7)	0.03

Figure 23 Distribution of case difficulty for patient-based assessments**Table 13 Results of patient-based assessments**

	Controls	Subjects	P-value
Completion of case	4 (7%)	6 (11%)	0.51
Maximum tip position			
Sigmoid	28 (52%)	29 (54%)	0.73
Descending	12 (22%)	8 (15%)	
Transverse	8 (15%)	11 (20%)	
Ascending	2 (4%)	0 (0%)	
Caecum	4 (7%)	6 (11%)	
Time taken (min)	20 (20, 20)	20 (19, 20)	0.11
Straight insertion depth (cm)	52 (21)	48 (23)	0.35
JAG DOPS	18 (14, 21)	16 (14, 22)	0.92
Global Score	17 (14, 19)	16 (14, 19)	0.35

Tip position expressed as number (%); Time taken, DOPS score and Global Score as median (IQR) and straight insertion depth as mean (SD)

4.3.3 Feedback

The participants from both groups rated their training experience highly, with a median score of 8.0 out of 10 for both types of training. Of note, the subjects rated training on the simulator as useful (8.0/10) and as enjoyable (8.0/10) as the controls rated training on real patients. Both subjects and controls felt reasonably well prepared for their assessments ($p=0.40$), although both groups found the patient assessments to be difficult and stressful than the simulator assessments (Table 14).

Table 14 Results from the feedback questionnaire

	Control	Subject	P-value
Overall, how good was your training?	8.0 (6.0, 9.75)	8.0 (7.0, 8.5)	0.93
Overall, how useful was your training?	8.5 (8.0, 10.0)	8.0 (7.0, 9.0)	0.13
Overall, how enjoyable was your training?	9.0 (8.3, 10.0)	8.0 (7.0, 10.0)	0.18
How well do you think your training prepared you for the assessments?	7.0 (5.3, 9.0)	6.0 (6.0, 8.0)	0.40
How difficult did you find the simulator assessments?	6.0 (3.3, 8.0)	5.0 (5.0, 6.5)	0.48
How stressful did you find the simulator assessments?	5.0 (4.3, 7.0)	4.0 (3.0, 7.0)	0.25
How difficult did you find the patient assessments?	8.0 (7.3, 8.0)	8.0 (5.5, 9.0)	0.96
How stressful did you find the patient assessments?	7.0 (5.3, 8.5)	6.0 (5.0, 8.0)	0.28

Answers are on a 10-point Visual Analogue Scale (VAS) and expressed as Median (IQR)

4.4 Discussion

In this study we aimed to address weaknesses in previous studies by assessing a large number of trainees from multiple centres in a prospective, randomised, blinded study. The use of patient-based training as a control allowed performance comparisons for 111 simulated and 109 clinical cases. We found that the equivalent time spent on the simulator resulted in significantly greater improvements in

performance on the simulator tasks than traditional patient-based training. We also found that simulator training produced equal performance outcomes on real-life cases, demonstrating a high degree of skills transfer from the simulator to real colonoscopy. Trainees found simulator training to be useful, highly enjoyable and prepared them well for their first patient-based colonoscopies.

One limitation of this study is that despite randomisation, there were some large but non-statistically significant differences in the level of trainee between the groups. 45% of the control group were comprised of specialists in training (defined as trainees who are currently in a formal training programme in gastroenterology), compared to 16% of the subject group. The controls had also seen more colonoscopy previously, although again this difference was not statistically significant. We speculate that although the controls did not have any increased 'hands-on' experience compared to the subjects, a specific interest in gastroenterology and increased exposure to endoscopy may have been enough to produce the differences seen in their initial simulator assessments. The statistical methods used in this study also assume that the results from the three cases performed by each individual are independent of each other, which may not necessarily be the case (their success is likely to be a direct reflection of the skill of the individual, making the three cases from a single individual more similar than three cases done by another individual). However, the ordinal nature and skewness of some of the data made correcting for this possible lack of independence difficult, and would be unlikely to have a significant effect on the results.

We found that the simulator group demonstrated higher performance on the simulated cases but only equivalence on patients. We consider this may be because the assessment tools used (DOPS and Global Score) may not have been sensitive enough to discriminate between relative novices, as they are designed to assess whether a trainee is competent or not. Despite the advanced modelling achievable by the simulator, real colonoscopy is inherently more difficult and unpredictable, so novices would not be expected to achieve competence scores after only 16 hours of training. It may also be that during their training the simulator group had learned to 'play the game' and could complete the simulator cases using techniques that work on the simulator but not on patients. However, if this were so, then we would expect

the subjects to perform less well on real-life colonoscopy, which was not the case. This issue is an inherent problem of even high-fidelity simulators and one of the reasons why simulation is unlikely to ever completely replace bedside training for complex procedures.

This study focused on novice training as it is easier to measure changes in performance on the steep slope of the learning curve. However, there are many potential advantages of simulator training for more experienced trainees. The ability to repetitively perform parts of a procedure as deliberate practice has been recognised as essential to the acquisition of complex skills within medicine[39,250]. The use of high-fidelity simulation may allow this level of practice for colonoscopy, although clinical outcome measures may need to be further validated in order to prove a significant training effect. This study also focused purely on the insertion phase, which forms only part of the skill set needed for colonoscopy. Other skills such as good withdrawal technique to achieve complete visualisation of mucosa, lesion recognition and therapeutic ability are also necessary for competency and the use of simulation to meet these training needs will need to be addressed in future studies.

4.5 Conclusions

The findings of this study provide supportive evidence for the use of the colonoscopy simulator as a training tool for novice colonoscopists. Simulator training and standard patient-based colonoscopy training appear to be equivalent as training approaches, and use of the simulator may therefore shorten the learning curve to competency. It may be considered both as a tool to develop knowledge, skills and confidence prior to clinical practice as well as an adjunct to more traditional training methods. The ability to simulate difficult cases more realistically may be helpful in improving the skills of clinicians who already perform at a more advanced level.

Chapter 5: Criterion validity of a colonoscopy simulator

5.1 Background and Aims

The results of the previous two chapters support the assertion that the Olympus colonoscopy simulator is a significant advance over any of the previous computer models in terms of valid assessment metrics and training impact. Data from the simultaneous performance assessments in the last study now allows an evaluation of the **criterion validity** (see p.34) of the simulator. This is a measure of how well the simulator can be used for prediction of performance.

Ideally an assessment tool should have both concurrent and predictive criterion validity. In this case, a measurement of concurrent validity can be made as the assessments on the simulator and patients were both performed at the end of the training study. **Predictive validity** refers to how well the test results correlate with performance at some point in the future, and would require long-term follow-up and re-assessment.

The aim of this analysis was to examine whether the novice colonoscopists performance on the simulator was predictive of their performance on real patients.

5.2 Methods

Each participant in the previous study was measured on three simulator cases and three patient cases. As there was no link between individual simulator and patient cases, it was not possible to examine the associations at the case level. Instead a single value per person was calculated for both the simulator and patient results, and all analyses were performed at the person level.

The majority of the variables in the analysis were measured on a numerical scale. For these variables, the mean value was taken for the three cases and used in the analyses. Some variables were binary measures (e.g. intubated to caecum), and responses to these variables were categorised into either all negative responses, or one or more positive responses.

The majority of the associations between simulator and patient variables involved two numerical variables. The association between two numerical values was examined using Spearman's Rank correlation. This method was used in preference to Pearson correlation as a large number of variables were ordinal rather than continuous, and due to the highly skewed distribution of a number of variables. When both patient and simulator variables were measured on a categorical scale, Fisher's exact test was used to examine the associations between variables. For situations where there was a numerical patient measure and categorical simulator measure, the Mann-Whitney test was used.

Because there are potentially an extremely large number of analyses examining multiple associations between simulator and patient measures, it is highly likely that some associations will appear to be significant due to chance alone, rather than being genuinely significant relationships. To try and counter this, only selected comparisons were made; the total JAG score, the total Global score, the combined score (JAG + Global Score), and the 'overall score', the expert's global judgement of ability, were examined against all the simulator metrics. Then individual parts of the JAG and Global Score were compared against the relevant simulator metrics only. Also, when multiple comparisons are made, a more stringent significance level is generally employed, with only p-values of less than 0.01 considered to be statistically significant.

5.3 Results

Table 15 shows the associations between the numerical simulator and patient-based variables. Of the 129 comparisons made, only 3 are significant at the $p < 0.01$ level. The first is the time with an alpha loop after sigmoid intubation against the expert's 'overall' judgement of ability, indicating that those trainees who were judged as more proficient also more quickly resolved the alpha loop in the simulator case. The other two significant values are both comparisons of the maximum loop size during sigmoid intubation against the total JAG score and the Total Score. This indicates that those trainees who performed better overall on the patient-based assessments also had smaller sigmoid loops during the simulator-based assessments. There was also a significant association with the 'loop formation' item which forms part of the

JAG score, although at the lower significance level of $p < 0.05$. These findings should be interpreted with caution however, as the Total Score is partly comprised of the JAG score and so may be significant purely because of this.

There were no significant associations between any of the categorical variables (Table 16) or between the case completion rates (Table 17) for simulator and patient-based results.

5.3.1 Discussion

This study has demonstrated that only a few of the variables recorded on the simulator have a significant association with the trainees' actual performance on patients. This is a slightly surprising result considering the strong construct validity previously demonstrated. However, there are several potential reasons why a stronger correlation was not achieved. The first is that this study was not specifically powered for this kind of evaluation, but uses available results from the previous study where the calculations were based on the numbers needed to achieve 80% power to distinguish between the two groups. To examine multiple correlations between the different types of assessment may require a significantly larger number of participants to achieve significant results. Secondly, as was previously discussed, the real-life assessment tools (JAG and Global Scores) may not be able to discriminate between relative novices, as they have been designed to assess competency in more advanced practitioners. If the patient-based assessments are unable to discriminate between performers at the bottom end of the performance spectrum, then it is unsurprising that there were very few significant correlations. Thirdly, the simulator may actually not demonstrate a high level of criterion validity.

Table 15 Associations between numerical simulator and patient-based results

	Time taken	Max tip position	Insertion depth	Luminal view	Atraumatic technique	Loop formation	'Overall' score	Total JAG score	Total GS	Total score
General										
Time taken	0.06						-0.19	-0.23	-0.15	-0.09
Max tip position		0.19	-0.03				0.10	-0.07	-0.11	-0.21
Technical										
Insertion length with embedded tip				-0.05	0.08		0.28	0.11	0.11	0.13
Insertion length with obscured lens				0.17			0.13	-0.04	-0.01	-0.02
Insertion time with embedded tip				0.11	0.19		*0.34	0.24	0.27	0.27
Insertion time with obscured lens				0.22			0.04	-0.05	-0.02	-0.03
Timings										
Time to mid sigmoid	0.29	-0.19					-0.09	-0.02	-0.02	0.00
Time to sigmoid-descending junction	0.24	-0.13					-0.07	-0.01	-0.01	0.02
Sigmoid intubation time	*0.33	-0.17					-0.06	-0.03	-0.01	0.00
Sigmoid intubation & straighten time	0.18	-0.18					-0.10	-0.07	-0.09	-0.07
Time to splenic flexure	0.27	-0.17					-0.05	0.01	-0.03	0.01
Splenic flexure intubation time	-0.05	-0.25					0.06	0.00	0.00	0.03
Time to mid transverse	0.07	*-0.36					-0.07	-0.01	-0.07	-0.01
Time to hepatic flexure	0.09	-0.32					-0.07	-0.04	-0.07	-0.04
Time to caecum	0.01	*-0.39					-0.17	-0.15	-0.20	-0.16
Looping										
Time spent looped during sigmoid intubation						0.06	-0.01	0.02	0.05	0.05
Number of loops after sigmoid intubation						-0.03	-0.03	-0.18	-0.22	-0.19
Time spent looped after sigmoid intubation						-0.06	-0.17	-0.12	-0.19	-0.13
Mean loop size during sigmoid intubation						*-0.37	-0.25	** -0.50	-0.35	** -0.43
Mean loop size after sigmoid intubation						-0.13	-0.20	-0.16	-0.16	-0.15
Number of transverse loops							0.08	-0.07	-0.12	-0.09
Maximum transverse loop size							-0.06	-0.15	-0.26	-0.19
Time to remove alpha loop							-0.20	-0.14	-0.10	-0.12
Time with alpha loop after sigmoid intubation							** -0.42	-0.22	*-0.33	-0.27

Table 16 Associations between categorical simulator and patient-based results

Variable	Group	Overall Performance Median (IQR)	P-value	Total JAG Score Median (IQR)	P-value	Total GRS Score Median (IQR)	P-value	Total Score Median (IQR)	P-value
All 3 simulator cases intubated to caecum	No	2.0 (1.7, 2.3)	0.29	20 (15, 21)	0.24	18 (13, 19)	0.78	37 (31, 40)	0.37
	Yes	2.0 (2.0, 2.7)		15 (14, 21)		16 (14, 19)		32 (28, 39)	
Variable stiffness function used	No	2.0 (1.7, 2.3)	0.95	15 (14, 20)	0.16	16 (13, 17)	0.20	31 (28, 36)	0.17
	Yes	2.0 (2.0, 2.5)		20 (15, 22)		18 (14, 20)		38 (29, 41)	
Number of loops formed during sigmoid intubation	1	2.0 (1.7, 2.3)	0.20	18 (14, 21)	0.48	17 (14, 19)	0.46	32 (28, 40)	0.36
	>1	2.5 (2.3, 2.7)		20 (19, 21)		18 (18, 18)		38 (37, 39)	

Table 17 Association between completion rates for simulator and patient-based results

Simulator results	Patient Results		P-value
	No cases completed N (%)	≥1 cases completed N (%)	
Not all cases completed	12 (86%)	2 (14%)	0.68
All cases completed	17 (74%)	6 (26%)	

5.4 Conclusions

This study has demonstrated only a few simulator metrics that display concurrent criterion validity. However, given the limitations described above, the fact that there were any positive findings is encouraging and further work is indicated to attempt to minimise the limitations, confirm reliability and to establish predictive validity as well. Such a study would require taking colonoscopists of varying levels of ability and then following them up for an extended period of time. It would be almost impossible to standardise the training of these individuals, so would limit the conclusions about the simulator to evaluating purely its assessment rather than training potential. However, given that some assessments now established in the UK are of a 'high-stakes' pass/fail nature, a validated, objective, feasible and acceptable assessment using the colonoscopy simulator would be worth further investigation.

Chapter 6: Therapeutic endoscopy training

6.1 Background and Aims

Therapeutic endoscopic procedures are technically challenging and have higher complication rates than diagnostic procedures[251,252]. Less experienced practitioners are significantly more likely to have a complication than more experienced endoscopists [113-115]. Trainee endoscopists require supervised hands-on experience to develop competence for safe procedural performance, but training in therapeutic endoscopy can be hard to obtain[88,89,92]. The ability to practise therapeutic techniques on endoscopic simulators early on in training would be advantageous without associated risks to the patient. There is evidence that practice on endoscopy simulators can improve technical skills in endoscopic haemostasis (see Appendix 1, p.267). There is, however, no published literature regarding training in other therapeutic skills such as polypectomy, stricture dilation and percutaneous endoscopic gastrostomy (PEG) insertion.

The Basic Therapeutic Endoscopy Course (BTEC) was developed at the Wolfson Unit at St Mark's Hospital in response to demands for a basic (foundation) endoscopy course and a therapeutic endoscopy course, which are now recommended for all trainees wishing to accredit in endoscopy in the United Kingdom (UK)[253]. The hands-on practical session has been run eleven times since 2005 with one hundred and twenty trainees completing the course. Core faculty have undergone a JAG accredited Training the Trainers course.

Feedback from this course has shown it is considered to be an excellent training opportunity (mean overall visual analogue scale rating 9.1 out of 10) and enjoyed by the trainees, but to date there has been no higher level assessment of its efficacy in either improving individual technique or its impact on broader outcomes such as successfully completed procedures. The primary aim of the study was to evaluate the efficacy of the BTEC in delivering training in therapeutic endoscopic techniques using a single blind randomised controlled trial to assess behavioural change.

6.2 Methods

This was a prospective educational evaluation study carried out in a single institution. The study was approved by the St Mark's review board and the Harrow ethics committee (07/H0719/52).

6.2.1 Participants

Participants were drawn from a cohort of endoscopists who had already expressed an interest in attending the course, as well as from applications received following advertisements on three training-related websites and an email to all gastroenterology trainees in the UK.

Participants were eligible for inclusion in the trial if they had not previously attended a therapeutic endoscopy course, were currently having exposure to at least 3 of the 4 therapeutic procedures, and would expect to continue having exposure to these therapeutic techniques following the course. Participants were excluded if they were already rated as independent to safely perform more than 2 of the 4 therapeutic procedures.

The participants were recruited in two blocks of 16, with the study run during two blocks of three days in September 2007 and April 2008. They were enrolled by the lead investigator and then randomised into subjects and controls using a computer generated block randomisation protocol with 8 per block. Participants were informed of their allocation by the lead investigator and the dates of the study intervention, thus neither was blinded to the group allocation.

6.2.2 Interventions

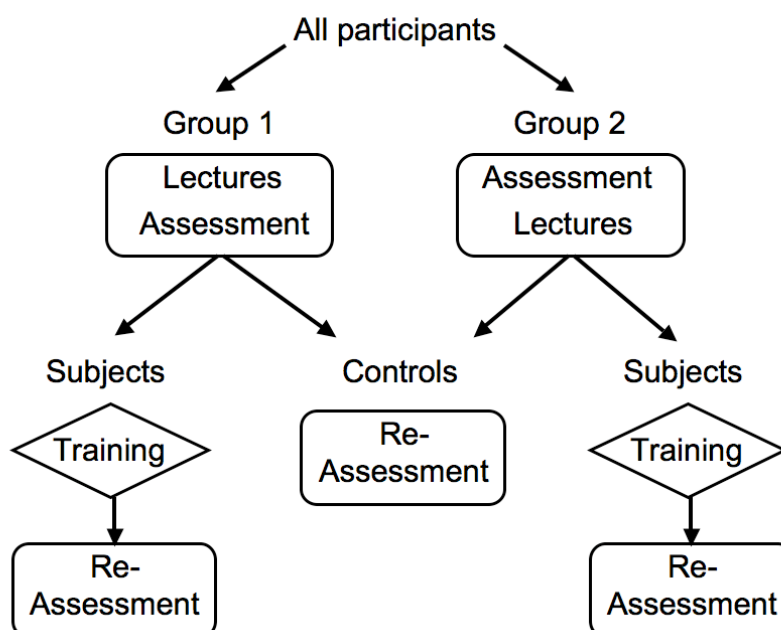
A questionnaire was completed by all the participants to provide data on demographics, previous endoscopic experience, previous therapeutic procedures, self-reported measures of competence and confidence in therapeutic endoscopy and expectations regarding the training outcomes from the course. All participants were given four 40-minute lectures designed to present the knowledge base underlying safe therapeutic endoscopy. This forms a vital part of training and may have an impact on the skills performance on the simulators. To test for the effect of this type of training, participants were to receive either the lectures first then assessments of

knowledge and practical skills (Group 1), or the assessments first then the lectures (Group 2). Randomisation was done using a computer generated blocked allocation protocol with 8 per block.

The participants' knowledge was assessed using a multiple-choice questionnaire (MCQ) containing true-false questions on all four modalities of therapeutic endoscopy. Participants' procedural skills were assessed on each of the four therapeutic simulations. Non-variceal haemostasis, snare polypectomy and percutaneous endoscopic gastrostomy (PEG) insertion were performed on ex-vivo models (Erlangen Endo Trainer, Germany); oesophageal stricture dilation was performed on a modified mechanical upper GI phantom (Adam, Rouilly, UK). The participants were given 10 minutes to complete each task to the best of their ability with an endoscopy assistant who would follow instructions competently, but would not offer instruction or be otherwise proactive. An expert observer and the endoscopy assistant assessed each participant. Both were blinded to the group allocation of participants. Specific skills tested were: use of injection, endoclips and bipolar diathermy probe for non-variceal haemostasis; technique for snare polypectomy and hot biopsy; manipulation of the endoscope for PEG insertion and technique for balloon dilation of an oesophageal stricture.

The subjects then received four 30-minute standardised practical hands-on skills training sessions on each therapeutic simulator taught by experienced endoscopists and senior endoscopy assistants in a one-to-one ratio. The control group were not given any skills training. Both the subjects and controls then had a re-assessment of their practical skills on each simulator in an identical fashion to the initial assessment. Following completion of the study, the control group received training identical to that received by the subject group (see Figure 24 for summary of the study design).

Figure 24 Flow diagram of the study protocol



6.2.3 Outcome measures

The knowledge-based outcome measure was the score obtained on the multiple-choice questionnaire (maximum score 97). The skills-based outcome measures used were station-specific checklists, which were developed based on a previously validated scoring system for polypectomy, and a modified Global Rating Scale termed the Global Score (GS)[254] which assessed competence against previously devised levels of units of competence taken from the UK national endoscopy competence framework[255]. Appendix 2 (p.269) gives details of the scoresheets and Global Score. The expert observer completed both the checklist and GRS and the endoscopy assistant completed just the GS. Each checklist contained 'nodal stops' which were considered to be essential for safe completion of the procedure. Failure to perform a nodal stop satisfactorily incurred a penalty of 3 points. Final checklist score was calculated as the sum of the completed checklist steps and penalty points. The Total score was calculated as sum of the final checklist score, instructor GS score and assistant GS score.

The primary endpoints were the difference in MCQ score between the groups and the change in the skills assessment score from initial assessment to re-assessment.

Secondary endpoints were changes in the individual checklist scores, GS scores, and a subset of the GS score, the overall score, which was the experts rating of competence on a 5 point categorical scale. The data from the questionnaire was correlated with feedback forms completed after the end of the course. Scores from the two types of assessment were correlated in order to validate the assessment tools used.

6.2.4 Statistical analysis

The primary aim of the study was to examine the change in score between the initial skills assessment and the re-assessment. Using the GS as the primary measure, a difference in score of 3 units was considered of clinical importance, and it was estimated that the within-subject standard deviation would be 4 units. With 5% significance and 80% power, 14 participants were required for the subjects group, with an equivalent number for the control group to detect a similar size difference over time. To allow for some drop out of participants and errors in data collection, recruitment aimed to enrol a total of 32 participants into the study in two blocks of 16 participants.

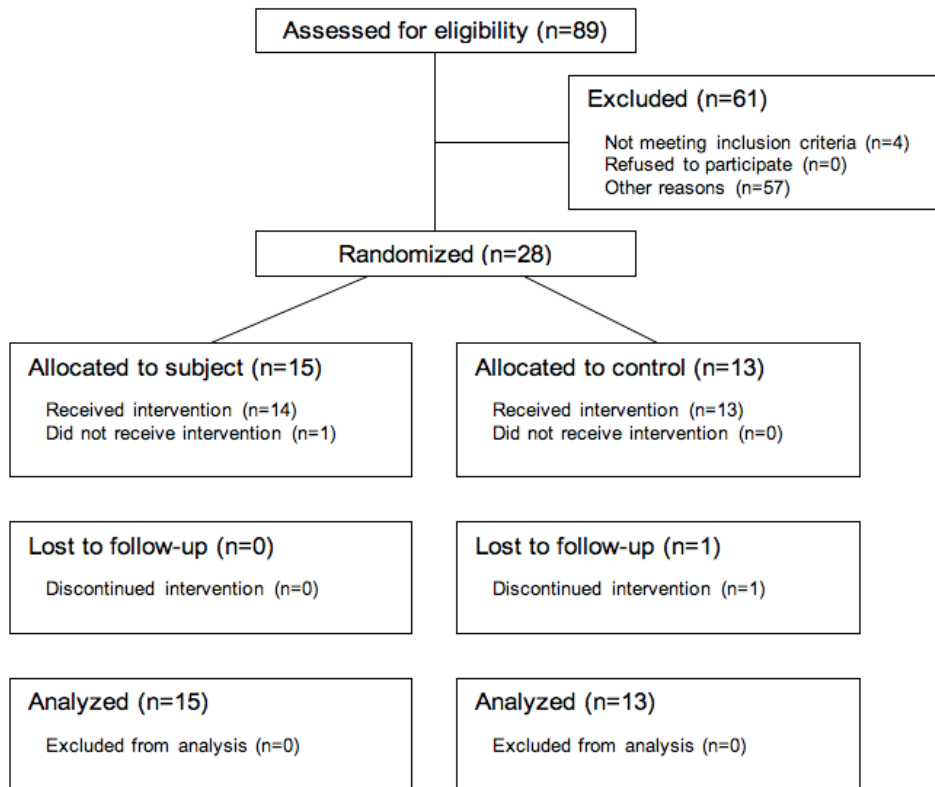
Statistical analysis was performed on an “intention-to-treat” basis. For the outcomes that were measured on a categorical scale, Fisher’s exact test was used to compare between the results between the groups. For continuous measurements that were found to be non-normally distributed, the Mann-Whitney test was used to compare variables between groups. For continuous measurements that were normally distributed, unpaired t-tests were used to compare variables between groups and paired t-tests were used to compare changes within groups over time.

6.3 Results

6.3.1 Study participants

Eighty-nine candidates expressed an interest in attending the course and were considered for inclusion in the study. Four did not meet eligibility criteria and fifty-seven were excluded because of an inability to get study leave or funding. Twenty-eight participants who met the eligibility criteria were randomised, with twenty-six completing the study. One subject did not receive the training because of illness and one control had to leave for personal reasons prior to the final assessments (see Figure 25 for flow diagram of participants through each stage of the study).

There was a spread of experience across the first four training years (4 first year, 6 second year, 4 third year, 4 fourth year) and seven endoscopists currently not in training posts. All had current endoscopy exposure as described in the inclusion criteria. The subjects and controls were well matched with no significant differences in the demographics (Table 18) or previous endoscopy experience (Table 19). They were also well matched in their initial skills assessment scores, with only the Upper GI bleeding checklist score showing the subjects performed significantly better than controls (Table 20). This means that the subjects had a lower chance of improvement on this measure when comparing initial and final scores, which will bias against the subjects.

Figure 25 Flow diagram of participants through each stage of the study**Table 18 Demographics of participants**

Variable	Group	Controls Number (%)	Subjects Number (%)	p-value
Sex	Male	11 (85%)	11 (73%)	0.66
	Female	2 (15%)	4 (27%)	
Handed	Left	0 (0%)	1 (7%)	1.00
	Right	13 (100%)	14 (93%)	
Grade	Specialist Registrar	10 (77%)	11 (73%)	1.00
	Consultant/other	3 (23%)	4 (27%)	
Speciality	Gastroenterology	11 (85%)	11 (73%)	0.66
	Surgery	2 (15%)	4 (27%)	
UK Trained	Yes	11 (85%)	11 (67%)	0.66
	No	2 (15%)	4 (27%)	

Table 19 Age and prior endoscopic experience of participants

	Controls Median (Range)	Subjects Median (Range)	p-value
Age	32 (28, 45)	33 (28, 55)	0.39
Colonoscopy	105 (5, 400)	90 (18, 500)	0.78
Flexible sigmoidoscopy	70 (10, 700)	80 (10, 400)	0.72
Polypectomy	20 (5, 75)	20 (0, 300)	0.55
OGD	400 (75, 4000)	300 (100, 1200)	0.39
Therapeutic OGD	25 (0, 50)	20 (2, 100)	0.93
Oesophageal dilations	2 (0, 20)	1 (0, 10)	0.59
PEG insertion	15 (0, 50)	5 (0, 30)	0.20

Table 20 Initial skills assessment scores for subjects and controls

Group	Variable	Controls Median (Range)	Subjects Median (Range)	p-value
Polypectomy	Checklist score	8 (1, 18)	10 (-3, 18)	0.98
	Number nodal stops	1 (0, 2)	0 (0, 2)	0.56
	Instructor GS total	19 (6, 27)	17 (7, 27)	0.28
	Assistant GS total	18 (7, 30)	16 (6, 28)	0.21
	Overall total	47 (15, 64)	42 (12, 67)	0.45
Dilation	Checklist score	9 (4, 14)	7 (0, 14)	0.44
	Number nodal stops	0 (0, 1)	1 (0, 1)	0.23
	Instructor GS total	15 (9, 27)	16 (6, 23)	0.87
	Assistant GS total	17 (8, 21)	16 (7, 24)	0.61
	Overall total	43 (26, 61)	38 (16, 58)	0.50
UGIB	Checklist score	8 (1, 13)	12 (4, 16)	0.04
	Number nodal stops	0 (0, 1)	0 (0, 1)	0.86
	Instructor GS total	17 (6, 23)	17 (11, 24)	0.71
	Assistant GS total	16 (10, 30)	18 (7, 28)	0.64
	Overall total	40 (17, 59)	47 (24, 63)	0.32
PEG	Checklist score	13 (-2, 18)	14 (4, 18)	0.89
	Number nodal stops	0 (0, 2)	0 (0, 0)	0.50
	Instructor GS total	22 (7, 29)	21 (6, 30)	0.75
	Assistant GS total	21 (7, 30)	20 (6, 30)	0.46
	Overall total	56 (14, 76)	47 (17, 76)	0.60

6.3.2 Expectations of trainees

The expectations of trainees for the course were high, with an 'Overall' pre-course expectation of 8.2 out of 10 on a visual analogue scale (VAS). Expectations of how useful the course would be, how appropriate the learning objectives were for them and how well the course would meet those learning objectives were also high. They also expected the course to be highly enjoyable. Feedback following the course showed that it met or exceeded all of their expectations. Specific feedback for each element of the course was also exceptionally good, with an average score of 8.4 out of 10. The practical hands-on sessions were rated especially highly, with an average score of 8.9 out of 10.

6.3.3 Effect of lectures

Although there was a trend towards higher scores in the skills assessment for the group who had the lectures before assessment (Group 1) than the group who had the lectures after the assessment (Group 2), there was no statistically significant effect of the lectures on the participants' practical skills (Table 21). There was no significant difference found in the initial MCQ score; the average (SD) score for Group 1 was 70.6 (7.1) and Group 2 65.5 (8.7), $p=0.10$, demonstrating no measurable effect of knowledge-based teaching using a knowledge assessment.

6.3.4 Effect of hands-on skills training

There was a significant difference in the final re-assessment scores (Table 22) as well as the median change from baseline (Table 23) for nearly all assessment measures between subjects and controls. There were significant differences between groups for the polypectomy measures, with only the number of nodal stops not showing a difference between groups. For each score, the subject group was more likely to have a greater increase than the control group. The results for the stricture dilation show statistically significant differences between subjects and controls for all measures, with the subject group improving their scores more than the control group, although the difference for assistant GS score was only of borderline statistical significance. For control of upper GI bleeding, there was a significant difference in checklist score, number of nodal stops and overall score, but not in GS. For the checklist score and overall total, the subject group was more likely

to have a greater increase in score than the control group, despite the fact that the subjects originally scored higher than the controls for the checklist score on this station. There was no difference in median values for the number of nodal stops between the two groups, but the range of values indicated that nodal stops were performed correctly more often in the subject group than in the control group. There was no evidence of a difference between the groups for any of the measures for the insertion of a PEG. The change in overall score for each skill is shown in Figure 26.

Figure 26 Graph of change in overall score between initial and re-assessments for each therapeutic modality

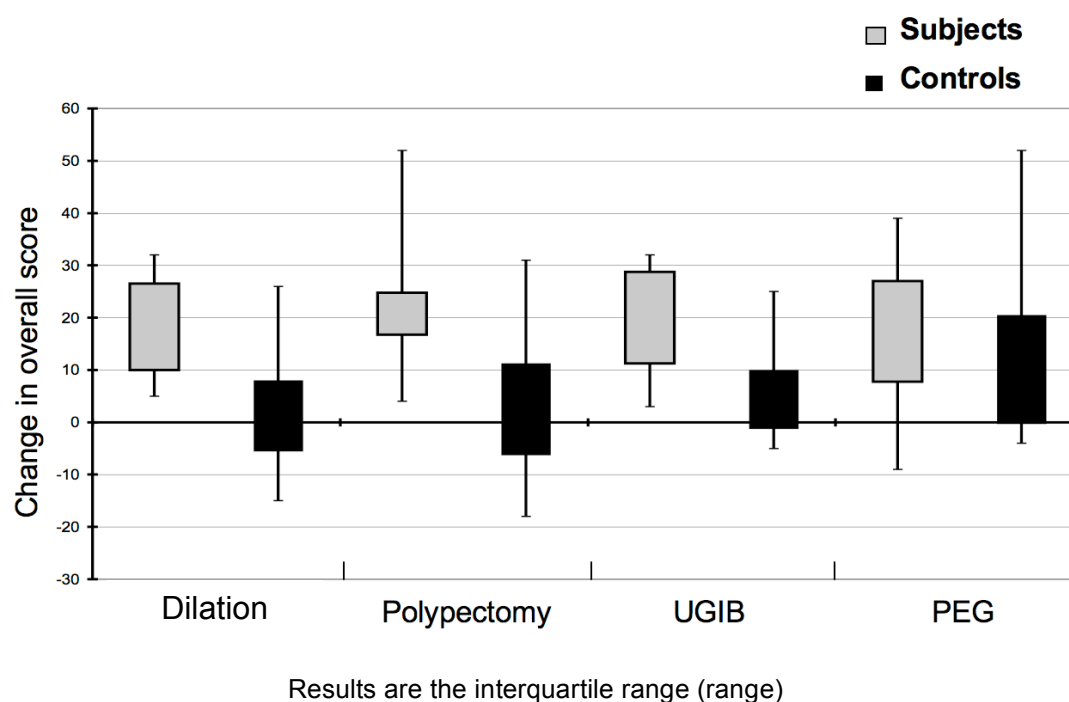


Table 21 Initial skills assessment scores for those who had the lectures before assessment (Group 1) and lectures after assessment (Group 2)

Group	Variable	Group 1 Median (Range)	Group 2 Median (Range)	p-value
Polypectomy	Checklist score	13 (3, 18)	10 (5, 15)	0.13
	Number nodal stops	0.5 (0, 2)	1 (0, 2)	0.52
	Instructor GS total	19 (7, 27)	17 (6, 24)	0.15
	Assistant GS total	18 (8, 30)	16 (6, 28)	0.40
	Overall total	48 (12, 67)	40 (15, 59)	0.08
Dilation	Checklist score	12 (5, 14)	11 (3, 13)	0.51
	Number nodal stops	1 (0, 1)	0.5 (0, 1)	0.25
	Instructor GS total	17 (6, 27)	16 (9, 23)	0.91
	Assistant GS total	17 (7, 21)	16 (7, 24)	1.00
	Overall total	44 (21, 61)	41 (16, 58)	0.76
UGIB	Checklist score	12 (6, 16)	9 (1, 15)	0.30
	Number nodal stops	0 (0, 1)	0 (0, 1)	0.15
	Instructor GS total	18 (11, 24)	17 (6, 24)	0.28
	Assistant GS total	18 (7, 25)	17 (10, 30)	0.58
	Overall total	47 (24, 63)	41 (17, 61)	0.57
PEG	Checklist score	15 (4, 18)	13 (5, 17)	0.55
	Number nodal stops	0 (0, 2)	0 (0, 1)	0.96
	Instructor GS total	24 (6, 29)	20 (8, 30)	0.68
	Assistant GS total	19 (6, 28)	21 (7, 30)	0.30
	Overall total	56 (14, 75)	51 (19, 76)	0.75

Table 22 Final skills re-assessment scores for subjects and controls

Group	Variable	Subject Median (Range)	Control Median (Range)	p-value
Polypectomy	Checklist score	17 (14,19)	12.5 (2,19)	0.011
	Number nodal stops	0 (0,1)	0 (0,2)	0.258
	Instructor GS total	25 (16,30)	16.5 (9,24)	0.001
	Assistant GS total	26 (14,29)	19.5 (11,27)	0.044
	Overall total	67 (44,76)	48.5 (24,67)	0.002
Dilation	Checklist score	14 (8,15)	9 (0,13)	<0.001
	Number nodal stops	0 (0,1)	0.5 (0,1)	0.067
	Instructor GS total	24 (15,30)	21 (10,27)	0.017
	Assistant GS total	20 (13,30)	16.5 (9,22)	0.174
	Overall total	60 (36,75)	48.5 (22,61)	0.005
UGIB	Checklist score	17 (13,22)	8 (3,11)	<0.0001
	Number nodal stops	0 (0,1)	0 (0,1)	0.142
	Instructor GS total	23 (15,29)	18 (10,27)	0.112
	Assistant GS total	24.5 (18,30)	20 (7,29)	0.051
	Overall total	64 (53,75)	46 (24,65)	0.001
PEG	Checklist score	18 (14,19)	16.5 (12,19)	0.044
	Number nodal stops	0 (0,0)	0 (0,0)	1.0
	Instructor GS total	27.5 (21,30)	26 (11,29)	0.077
	Assistant GS total	26 (17,30)	26 (10,29)	0.960
	Overall total	71.5 (52,79)	69 (33,74)	0.129

Table 23 Change in score from initial assessment to re-assessment

Group	Variable	Subject Median (Range)	Control Median (Range)	p-value
Polypectomy	Checklist score	2 (-7, 9)	6 (-2, 15)	0.03
	Number nodal stops	0 (-2, 1)	0 (-2, 1)	0.85
	Instructor GS total	-2 (-11, 11)	7 (1, 15)	<0.001
	Assistant GS total	-1 (-12, 8)	8 (-8, 16)	0.006
	Overall total	-1 (-18, 31)	21 (4, 52)	0.002
Dilation	Checklist score	-2 (-6, 4)	3 (1, 9)	<0.001
	Number nodal stops	0 (-1, 1)	-1 (-1, 0)	0.02
	Instructor GS total	3 (-4, 11)	8 (1, 14)	0.02
	Assistant GS total	2 (-7, 9)	6 (-4, 12)	0.06
	Overall total	3 (-15, 26)	18 (5, 32)	0.003
UGIB	Checklist score	1 (-7, 6)	7 (2, 11)	<0.001
	Number nodal stops	0 (0, 1)	0 (-1, 0)	0.02
	Instructor GS total	3 (-4, 11)	6 (-5, 11)	0.28
	Assistant GS total	3 (-7, 11)	5 (-2, 14)	0.18
	Overall total	3 (-5, 25)	15 (3, 32)	0.004
PEG	Checklist score	4 (-3, 12)	3 (0, 12)	0.68
	Number nodal stops	0 (-2, 0)	0 (0, 0)	0.12
	Instructor GS total	4 (-3, 16)	7 (-5, 22)	0.24
	Assistant GS total	3 (-2, 18)	8 (-4, 13)	0.28
	Overall total	12 (-4, 52)	19 (-9, 39)	0.30

6.4 Discussion

Optimal patient care and outcomes from all types of endoscopic therapeutic procedures is highly dependent on the skill of the operator. How a trainee should safely and reliably gain this skill has frequently been debated. Traditionally, practical skills are gained by practising on patients whilst supervised by an expert, often over a prolonged period of time. Acquisition and refinement of the necessary cognitive and psychomotor skills inevitably results in performance errors by trainees, which may in turn result in poorer outcomes for patients. It may also be difficult for trainees to gain experience in uncommon, emergency or difficult procedures. Surgeons have long recognised that structured training, with the use of simulators is advantageous for these types of procedures to be learned in a safe, controlled manner with no risk to patients[256,257].

With increasing development of endoscope technology, gastrointestinal endoscopy is becoming more and more focused towards the provision of therapy within the GI tract. As with all practical skills, there is a learning curve for these procedures, and the training issues that have concerned surgeons now equally apply to endoscopists. There are many different types of therapeutic endoscopy courses available, ranging from courses offering simple observation of procedures; courses giving hands-on practice using endoscopy simulators; to those offering supervised practice on real patients. While lecture-based or observational courses may offer 'tips and tricks' to aid self-improvement and have the ability to demonstrate advanced techniques without exposing patients to added risk, there is no evidence for their ability to alter the actual skills of attendees. Practice on real patients may offer more expert training than is available at their home hospital, and there is some evidence showing the long-term benefit of intensive hands-on courses in diagnostic colonoscopy[34]. However, courses offering this type of training for therapeutics still run into all the problems of standard training methodologies in terms of limited exposure to procedures for trainees and risk of complications for patients.

This study describes the first evaluation of a training course specifically tailored for the basic therapeutic skills required by independent practitioners. This evaluation has demonstrated the effectiveness of intensive hands-on simulator training for improving practical skills in three different endoscopic therapeutic modalities. This extends the evidence for this type of training from purely endoscopic haemostasis to polypectomy and dilation within the GI tract. It is noteworthy that the training effect was measurable after a relatively short period (30 minutes) of one-to-one instruction. This study has also shown that knowledge-based teaching alone does not have a measurable effect on skills or performance, emphasising the importance of specific hands-on training to improve the practical skills required.

No measurable effect of structured skills training for insertion of a PEG tube was demonstrated. This may be because of the two-person nature of the procedure. Unlike the other three modalities, where the endoscopist performs the majority of the technical procedures and the assistant is in a 'helping mode', PEG insertion requires two operators of more equal skill. The endoscopist's role is limited to examination of the stomach and then snaring the string prior to withdrawal, which effectively

requires competence at a 'diagnostic' level only. The majority of the technical therapeutic aspects of the procedure (localising the site, insertion of the trocha, assembly of the PEG) are performed by the assistant, who must be both knowledgeable and adequately trained in all aspects the procedure. Both the assistant and the endoscopist have input into a successful insertion. The outcomes for this study were based on the change in performance of the endoscopist between two assessments. Both subjects and controls improved on all the scoring systems used, with a large range of scores measured. It is possible that the controls improved their performance merely by observation of a competent assistant during the initial assessment procedure. It may also be that the assessment tools used were additionally measuring the competence of the assistant, which may negate any deficiencies or competencies of the endoscopist. It may also be that this procedure is simpler to learn than the other therapeutic skills, and didactic learning may be enough to enhance performance without the need for hands-on training.

The study did not show any effect of knowledge-based teaching using a knowledge assessment. This is likely to be due to weaknesses of the MCQ used for the assessment, which may not have reflected the lecture material closely enough. The validity of the skills assessment tools could also be questioned. Checklist scores and a Global Score form part of the Objective Structured Assessment of Technical Skill (OSATS) format which has previously been shown to be reliable and valid for use in assessment of surgical training[248,254,258-270]. A checklist approach has also been successfully used in assessment of haemostasis during simulated endoscopy[231]. The polypectomy checklist has been previously validated[271], and the other checklists were created based on this format by the authors, but have not been validated individually. The Global Score was adapted from the surgical platform by the authors for therapeutic endoscopy, but not specifically validated. Analysis of the inter-rater variation using these tools was small, with moderate concordance between the two experts' judgements of ability (Weighted kappa value 0.52, 95% CI 0.43-0.61). The correlation between the checklist score and GS was also highly significant ($p < 0.001$) for all four stations, indicating they have good concurrent criterion validity.

Finally, there may also have been a performance effect due to differing motivations between the groups at the final assessment, with the subjects wanting to demonstrate their skill acquisition and the controls performing less well due to the frustration of repetitive testing. However, the subjects had no additional motivation to improve their performance, whereas the controls were still motivated by the expectation of subsequent training. The results from the PEG station also show that the controls did improve where there was the possibility of experiential learning.

Although this study's results show that hands-on skills training improves performance on simulated procedures, it does not provide evidence that such training translates into improved outcomes in clinical endoscopic practice. Follow-up of trainees' performance in each modality over time using standardised assessment tools would be necessary to show such an effect. One study has shown that trainees receiving simulator as well as clinical training for haemostasis have a higher success rate and a reduction in the frequency of complications in their clinical practice compared to those who are just clinically trained[231]. However, there is as yet no data showing clinical efficacy for the other therapeutic techniques investigated in this study. Ideally trainees would have repeated interval training sessions in therapeutic endoscopy and recurrent evaluations with feedback to enhance learning and skills development. However practical costs and limitations of trainees' study budgets prevent open access to simulator training. Lack of funding was the primary reason for the inability of interested trainees to attend this course and participate in the study.

6.5 Conclusions

The results from this study provide evidence for the efficacy of training using realistic simulation of therapeutic procedures. This evidence should be used to continue improvements in both the simulator platforms and the training programmes and support more widespread provision of affordable training. Additional stations for control of variceal bleeding and of post-polypectomy bleeding are planned for future courses and should undergo similar evaluation. Advanced endoscopy courses including training in endoscopic stent insertion, double-balloon enteroscopy (DBE), and large endoscopic mucosal resection (EMR) will be increasingly in demand for training prior to widespread adoption of hybrid endoscopic/surgical procedures such

as endoscopic submucosal dissection (ESD) and NOTES. Continued evaluation of the training outcomes, especially with correlation to clinical endoscopic practice, will be vital to ensure the competency of practitioners and the safety of patients.

Chapter 7: E-learning in Capsule Endoscopy

7.1 Background and Aims

Capsule endoscopy has been identified as an area which currently lacks any kind of quality assurance of training. Unlike training in other areas of endoscopy which is mainly focused on improvement of technical skill, capsule endoscopy skills are purely diagnostic and may have differing training requirements. To identify key elements of capsule training, a focus group was held with three very experienced capsule readers; one research registrar (AP), one consultant gastroenterologist (CF) and one specialist nurse (AF). The initial question posed was:

“What is it about capsule endoscopy that makes it difficult? Specifically, what issues in capsule endoscopy do you think should be addressed in a training programme?”

Four areas were identified as difficult during capsule endoscopy:

1. Using the equipment
2. Spotting an abnormality
3. Correctly identifying an abnormality
4. Management of the patient once an abnormality is found

Using the equipment was considered to be easily addressed by clear instructional video or reading the manual.

The expert consensus for the best way to spot an abnormality was to have a systematic way of reviewing it to ensure lesions are not missed. The experts felt that ideally this would be done as one-to-one instruction with an expert, but could be incorporated into a self-directed learning programme using currently available DVDs.

Management of the patient and subsequent decision-making was considered very difficult to train. The experts felt that it would require extensive experience in gastroenterology as each case will be different with many individual factors to take into account e.g. clinical history, age, comorbidities, local services, relevance of

findings to presentation. The experts considered that this would require a service led by a consultant gastroenterologist who has extensive experience in the management of all GI conditions. Management experience would therefore be gained in an apprentice-type model.

Correctly identifying an abnormality was also considered to be a very difficult area to train adequately. The key point identified was the ability to recognise a lesion, and this was considered the main requirement for competent performance at capsule endoscopy. Four main ways were identified that may help in training in lesion recognition:

- i. Using an atlas of capsule endoscopy pictures (that is bundled with the software) to learn what pathology looks like at CE
- ii. Self-directed learning using videos of CE in a DVD or computer based format
- iii. One-to-one feedback by expert of results of actual cases (double-reporting)
- iv. Auditing own results of performance (number of lesions found, correlation of findings with other investigations e.g. DBE)

The first three methods would be useful for novices or trainees. The fourth applies to all practitioners as part of continuing professional development. Lesion recognition was considered to be the key feature that would be most amenable to a new approach to training. The experts considered that development of a module in lesion recognition with a built-in self-assessment exercise would be useful and fill a learning gap that is currently not well met.

The primary aim of this study was to develop, validate and pilot a web-based training and assessment module for lesion recognition at capsule endoscopy.

7.2 Methods

All studies were approved by the St Mark's review board and classified as educational evaluation by the UK National Research Ethics Service.

7.2.1 Module development

An initial training needs analysis was carried out using three sources:

- i. Expert consensus (CF, AP and AF).
- ii. Questionnaire survey of study participants regarding current training provision and training needs/preferences (data originally unpublished, courtesy of Sidhu et al)
- iii. Literature review

Learning objectives were derived by two experts (AP, AH) from the needs analysis, and the training module was developed to meet the identified learning objectives. Subject matter (video clips and image stills) were sourced and selected by a CE expert (AP) using the St. Mark's CE database of over 850 CE studies and the Rapid5™ Atlas (Given Imaging Ltd, Yoqneam, Israel). The selected material covered a comprehensive range of CE findings and included normal anatomical appearances, common incidental findings and a wide range of pathological findings. The images were annotated with relevant background information, key references and detailed diagnostic features. The training resource was further developed from its original format (Microsoft Powerpoint 2003™) into a multimedia resource (Illiminatus Opus Pro XE™). The module provided a menu-driven teaching resource incorporating: a keyword index; learning objectives; summary of key learning points; links to related resources and an integrated feedback mechanism ('check your understanding' section). A screenshot is shown in Figure 27.

A test module was developed using sixty video clips which were sourced and selected from the St. Mark's CE database. The clips were 40 seconds in length and played back at 5 frames per second. The clips covered a comprehensive range of normal anatomical appearances, incidental findings and pathological findings. Each clip was presented as a multiple-choice question with one correct answer from five possible options. The test module was further developed from its original format (Microsoft Powerpoint 2003™) to a multimedia resource (Illiminatus Opus Pro XE™). The module allowed manipulation of the real-time video clips such that they could be paused or re-wound to review a selected image. A screenshot is shown in Figure 28.

Figure 27 Screenshot from capsule endoscopy training module

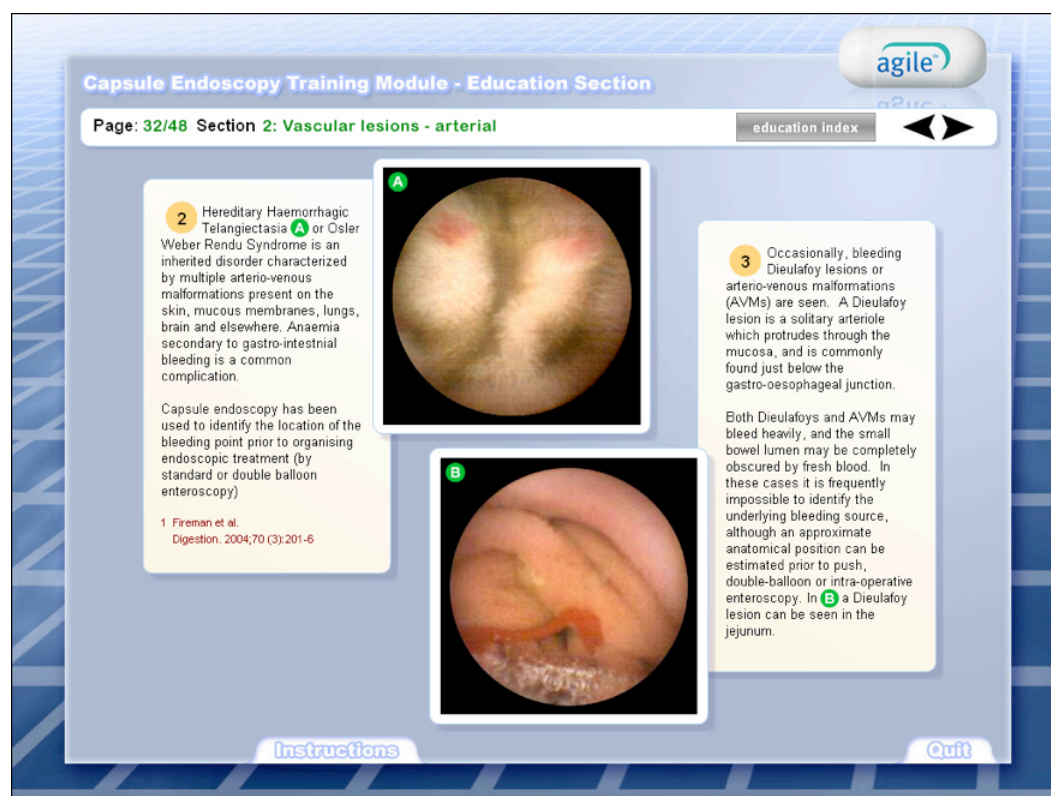
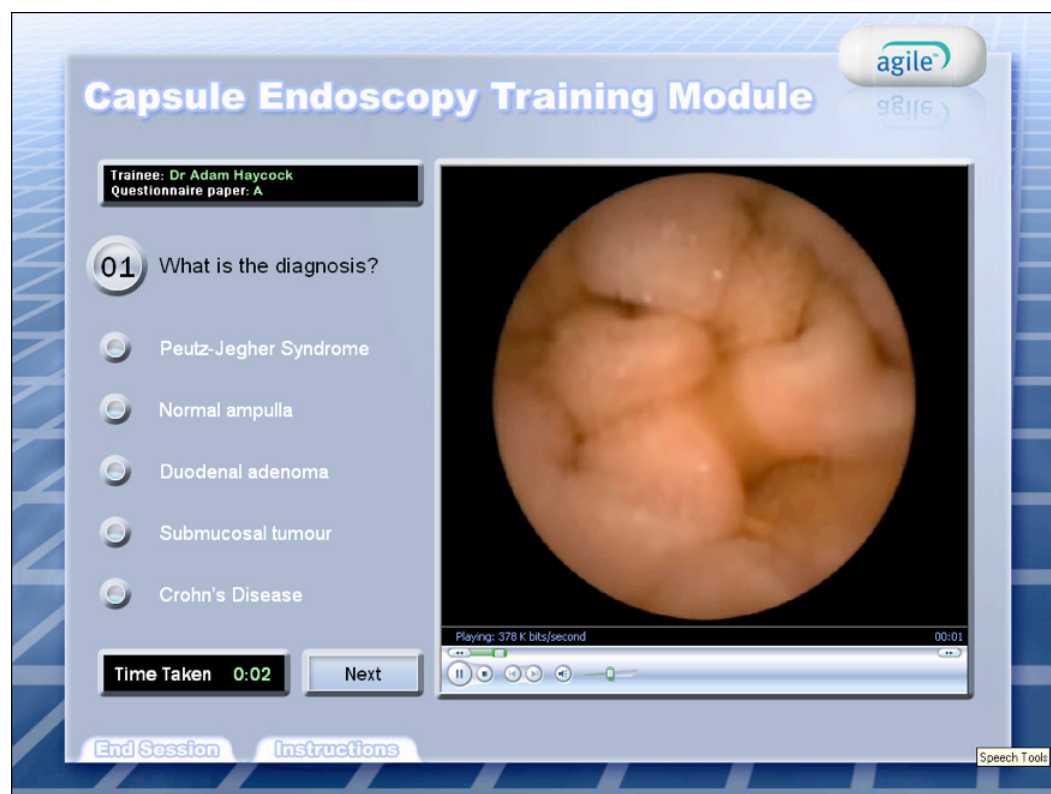


Figure 28 Screenshot from capsule endoscopy test module



7.2.2 Module validation

To assess the construct validity of the assessment questions, we examined the difference in performance between participants with different levels of experience (in medicine, gastroenterology, endoscopy and CE). The content validity of the training module was assessed by examining the change in performance on the test module following training .

Four experienced CE readers (defined as previous experience of ≥ 100 independently reported CE studies) completed the test module in order to benchmark expert performance. Medical students (novices) and gastroenterology specialist registrars (trainees) were recruited prospectively to test the training and evaluation modules. A pre-post test design was used. All trainees completed a questionnaire detailing medical, gastroenterology, endoscopy and CE experience. Baseline performance was assessed using the test module which was completed by the subjects in their own time (approximately 1 hour). No feedback or correct answers were given following the initial assessment. The training module was then completed by the subjects in their own time (approximately 1 hour) until the learning objectives had been achieved and the 'check your understanding' sections could be completed correctly. All subjects were then reassessed on the test module. Individual feedback (correct answers and test performance) was given following completion of the test module. Finally, an assessment of the format and content of the educational module was completed by each subject.

7.2.3 Web-based training pilot

To allow wider dissemination of the module, a web-based version was then developed. Amendments were made to both modules based on the comments made by the participants in the validation study. It was felt that repeating all 60 test questions was tedious and without any feedback did not enhance the educational value of the module. Some of the questions were also deemed to be excessively difficult. To rectify this, 8 questions that demonstrated poor construct validity (defined as $<50\%$ correct answers by the experts) were removed. The remaining 44 questions were stratified in terms of difficulty and separated into two 22 question papers, A and B. These were used in a pre/post test design, with the program

randomly allocating the order of completion (AB or BA) to each individual. This design allows all the questions to be evaluated in both pre and post-training situations, while each individual completing them only sees each question once. An option was included to allow participants to review their own answers to each question, the correct answers, and also a paragraph of explanatory detail including hyperlinks to the relevant section of the teaching module.

The program recorded participant demographics, the answers to each test question, the time taken and the results from the 'test your understanding' sections. A results file was automatically emailed to a dedicated email address once the training module and both test modules had been completed. To encourage completion, participants had 14 days to complete the module from the date of registering their details.

The training and assessment modules were developed into a single downloadable file that was placed on the Wolfson Unit for Endoscopy website. Invitations to Beta-test the module were sent to all gastroenterology trainees registered with TiG, and adverts were placed on the Wolfson Unit for Endoscopy website, the BSG website and the TiG website. Participants interested in completing the module were asked to contact the lead investigator and were sent an email containing download instructions. A reminder was sent after 13 days to each individual reminding them of the 14 day completion requirement. Following completion of the module, an email was sent to each individual asking them to complete a short web-based feedback form.

7.2.4 Statistical analysis

To investigate the initial construct validity, we assumed that a difference in baseline performance scores *between* each of the three groups (novice and trainee, trainee and expert, novice and expert) of 20% would be of clinical importance. With 5% significance and 80% power, a minimum of 4 subjects were required for each of the study groups. Assuming a difference in performance *within* the novice and trainee groups of 10% between pre and post training would be of clinical importance (with 5% significance and 80% power), 14 subjects were required per group for paired analysis to confirm content validity of the training module. For the web-based pilot,

the population was assumed to be much more heterogenous, so a difference in performance of 5% was assumed to be clinically important, requiring 30 participants to complete both pre and post-training assessments.

To determine construct validity, logistic regression analyses were used to compare the baseline performances of the three participant groups (expert, trainee, novice) at the pre-training stage. Robust standard errors were used with the logistic regression procedure to allow for the fact that the data was collected from 32 different observers. In addition to an overall comparison between the three groups, specific comparisons were performed to compare between each pair of participant groups. As this involved multiple comparisons and several statistical tests a Bonferroni correction was applied. For content validity, logistic regression analyses were used to compare the difference between the pre and post training performances of the different participant groups. Robust standard errors were used to allow for the fact that there were multiple observers. This analysis was performed firstly for trainees and students separately and then for the two groups combined. For the web-based pilot, logistic regression analyses were used to compare the difference between participants' pre and post training performance.

7.3 Results

7.3.1 Validation study

7.3.1.1 Participants

4 CE experts (3 consultant gastroenterologists and 1 experienced CE research nurse) were used to benchmark expert performance on the test module. 14 gastroenterology trainees (specialist registrar grade) and 14 medical students were prospectively recruited to test the construct validity of the test module and the content validity of the training module. Prior endoscopy and CE experience for each of the study groups is recorded in Table 24.

Table 24 Previous endoscopy experience of validation study participants

Group	Experience		
	Capsule	OGD	Colonoscopy
CE Experts	300 (175-500)	2250 (1500-4375)	2750 (1500-6125)
Gastroenterology Trainees	0	800 (400-925)	325 (137-462)
Medical Students	0	0	0

Results are Median (IQR)

7.3.1.2 Construct validity

The CE experts answered the most questions correctly, achieving a mean score of 73.8% $\pm$ 8.0. The research nurse expert achieved an individual score of 70%. The gastroenterology trainees achieved a mean score of 49.5% $\pm$ 10.9 and the medical students 29.5% $\pm$ 3.3 (Table 25). Logistic regression analyses demonstrated that there was a significant overall difference between the percentage of correctly answered questions before training for the three study groups ($p < 0.001$). Additional Bonferroni adjusted comparisons were carried out between each pair of groups and indicated that the performance of each study group was significantly different from the other study groups ($p < 0.001$ for each comparison). These findings demonstrate the ability of the test module to distinguish between participants with different levels of experience and knowledge and confirms construct validity.

Table 25 Comparison of baseline (pre-training) test performance

Group	Baseline test performance	P-value
CE Experts	73.8 (8.0)	<0.001
Gastroenterology Trainees	49.5 (10.9)	
Medical Students	29.5 (3.3)	

Results are % correct Mean (SD)

7.3.1.3 Content validity

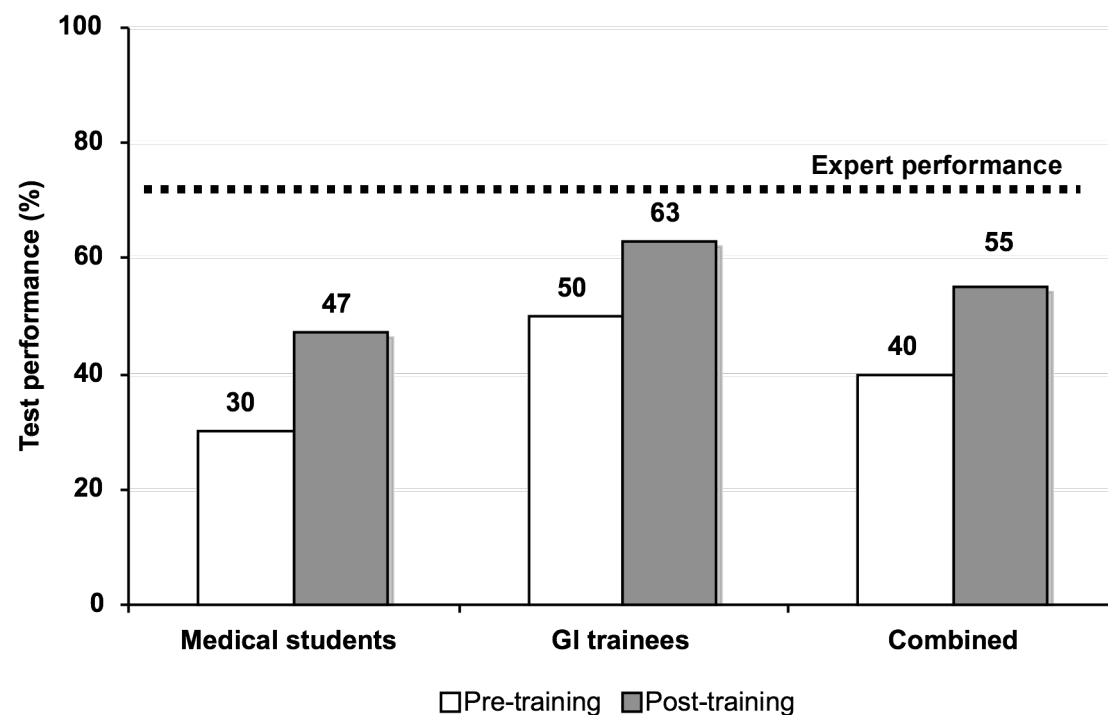
Logistic regression analyses were used to compare the difference between the pre and post training performances on the test module. There was a significant improvement in test performance after training for both the medical students and the gastroenterology trainees, both when considered separately and when considered together (Table 26, Figure 29). These findings demonstrate the ability of the training module for improving lesion recognition skills at CE and confirms content validity.

Table 26 Comparison of pre and post training test performance in validation study

Group	Pre-training test performance	Post-Training test performance	p-value
Gastroenterology Trainees	49.5 (10.9)	62.1 (7.7)	<0.001
Medical Students	29.5 (3.3)	46.7 (6.8)	<0.001
Gastroenterology Trainees and Medical Students	39.5 (12.9)	54.4 (10.6)	<0.001

Results are % correct Mean (SD)

Figure 29 Pre and post training test performance in validation study



7.3.1.4 Feedback

Feedback obtained from the participants was extremely positive regarding both the training and test modules, with median (range) VAS scores out of 10 for their usefulness of 9.1 (6.6, 10) and 8.55 (7.3, 9.9) respectively. When asked how useful they thought an online version would be for training, the median score was 9.55 (7.4, 10), demonstrating that this method of training would have high acceptability for trainees. The most common additional comment was a request to expand the subject matter to include lesion recognition during flexible endoscopy, a gap that is now being addressed with the JAG online e-learning module.

7.3.2 Web-based training pilot

7.3.2.1 Participants

99 individuals have so far expressed an interest in Beta-testing the module and have been sent details on how to access it (as of 8 September 2009). 30 have completed the module, with full results available for 27. The results are not available for 3 participants either because of failure of writing the data to the database or because of difficulties in retrieval. Demographics of the participants are shown in Table 27.

Table 27 Demographics of participants in web pilot

Grade	Number of participants	Experience		
		Colonoscopy	OGD	Capsule
Clinical assistant	1	0	0	300
Staff Grade	1	50	2000	10
Nurse	2	1 (0.25-0.75)	1 (0.25-0.75)	1 (0.25-0.75)
Trainee (Total)	20	245 (89-325)	600 (437-1157)	0 (0-1)
Year 1	3	80 (49-82)	250 (242-875)	0 (0)
Year 2	5	90 (20-120)	400 (180-450)	0 (0)
Year 3	4	245 (210-262)	550 (500-650)	1 (0-16)
Year 4	2	300 (300-300)	1045 (967-1122)	4 (2-6)
Year 5	3	400 (350-700)	1143 (1071-1321)	1 (0.5-20.5)
Year 6+	3	530 (515-615)	2000 (1250-2500)	0 (0-1)
Consultant	3	650 (325-1325)	3000 (1501-3500)	7 (5-11)

Results are Median (IQR)

7.3.2.2 Content validity

Logistic regression analysis was used to compare the difference between the pre and post training performances on the test modules (Table 28). There was a significant improvement in test performance after training ($p=0.037$), although this improvement was slightly smaller than in the validation study (8.9% vs 14.9%). Both papers had similar baseline (pre-training) scores indicating that they were comparable in difficulty ($p=0.86$). However, scores for paper A done after training did not show a significant difference compared to scores done pre-training ($p=0.087$), whereas there was a significant difference for paper B ($p=0.029$). This may indicate that the training material covered the topics tested in paper B better than they did those in paper A.

Table 28 Comparison of pre and post training test performance in web-based pilot study

Group	Pre-training test performance	Post-Training test performance	p-value
Overall	56.9 (15.5)	65.8 (13.1)	0.037
Paper A	56.8 (13.7)	58.4 (14.4)	0.087
Paper B	57.1 (18.4)	69.4 (11.1)	0.029

Results are % correct Mean (SD)

7.3.2.3 Feedback

Feedback obtained from 27 participants was again extremely positive regarding both the training and test modules, with median (range) scores out of 10 for their usefulness of 8.4 (6,10) and 7.2 (4,10) respectively. Trainees found it enjoyable, comprehensive, interactive and easy to use. The most common suggestions for improvement were to do with improved resolution of the videos and requests for a wider variety of normal and pathological findings.

7.4 Discussion

The current expansion in CE use as a primary diagnostic modality for gastrointestinal pathology is likely to continue into the foreseeable future. However, reporting capsule studies is time consuming and due to the increasing workload it is no longer practicable in many specialist centres for capsule reporting to be performed primarily by expert physicians. In some centres GI trainees may report capsule studies independently, in others specialist nurses or technicians are used to pre-read capsule studies prior to expert review of selected images in order to reduce expert reporting time[272-275].

There has also been a devolution of CE services from tertiary centres to local hospitals where it is possible that the lack of specialist experience and a low volume caseload may impact on quality assurance and provision of adequate in-house training. It is increasingly recognised that significant pathology may be missed or significantly underestimated at CE, producing only subtle abnormalities in a limited number of video frames, and that concordance with expert findings varies significantly. These issues highlight the importance of adequate training and experience if diagnostic standards are to be maintained as the service expands.

Our data represent the largest studies in CE training to date, and provide new insights into the need for CE training and how it could be provided at a national level. They are also the first studies to evaluate the effectiveness of an educational tool for CE training. We have shown that a web-based tool can be an effective method of training and assessing lesion recognition skills at CE – a skill which is considered critical for accurate CE reporting. The tool was educationally validated, demonstrating content validity of the training module and construct validity of the test module. The performance of trainees was consistently shown to improve significantly following training, irrespective of level of previous experience and under ‘real-life’ learning conditions. Trainees with previous gastroenterology/endoscopy experience improved to a significantly higher level of performance compared to those without, although even experienced trainees did not achieve the benchmark ‘expert’ performance following training, indicating that further focussed training may be necessary before acquisition of competence and independent practice for CE can be

achieved. Finally, trainees positively endorsed this method of CE training as both valid and appropriate.

It is interesting that the improvement in performance was smaller in the pilot web-based study than in the validation study. This may be because of the much more heterogeneous target population than the selected groups involved in the validation study. It may also be because the trainees in the validation study saw all the test questions twice and a degree of question familiarity may be a confounding factor. However, to counteract this, the web-based trainees were provided with feedback regarding their initial performance, which the trainees in the validation study did not have. However, as the improvement in performance after a relatively short training intervention was consistent in both novices, specialist trainees and a more general population, it appears robust and is consistent with other published studies in the medical education literature[276]. However, it would be inappropriate to make generalised conclusions from our findings as there remains a significant difference between accurate detection and identification of a capsule image in a 30 second CE video and correct interpretation of an 8 hour CE video.

7.5 Conclusions

These studies prove the principle that validated web-based e-learning can be an effective tool for improving CE lesion recognition skills and allow rapid dissemination of quality training material and best practice to a wide audience. This approach could potentially help to address the current issues surrounding provision of CE training in the UK and elsewhere.

Chapter 8: Virtual colonoscopy training for radiographers

8.1 Background and Aims

Optical colonoscopy is an invasive procedure which requires significant time investment for training and to achieve competence. Competing technologies such as virtual colonoscopy (VC) offer a diagnostic capability similar to optical colonoscopy, unlike older, less accurate investigations such as barium enema. In the UK, barium enema services have largely been provided by radiographers with radiologist input. There is now a move to replace barium enema services with virtual colonoscopy, but there is a paucity of data recommending how radiographers are best trained for this new role. The Virtual Colonoscopy Training of Radiographers (ViCToR) course was designed to promote the development of dedicated multi-disciplinary VC teams around the UK by training radiographers and encouraging their supervision by the local lead GI radiologist. It has been running since September 2007 and an application has been approved by the Society of Radiographers (ScOR) for CPD accreditation. The aims of the course are summarised in Table 29.

Table 29 ViCToR Course Aims

1. Development of a nationally recognised radiographer VC training course, validated by formal evaluation methods and accredited by Society and College of Radiography (SCoR).
2. Develop the radiographer's theoretical knowledge, understanding and skills required to perform virtual colonoscopy as part of advanced radiography practice.
3. Provide radiographers with knowledge of anatomy, physiology and pathology of the colon including pathogenesis of colorectal cancer and morphological variants.
4. Provide radiographers with the opportunity to extend their role and critically examine the evidence base for radiographer performed Virtual colonoscopy.
5. Assist radiographers in developing the practical and interpretive skills to enable their participation in a dedicated multi-disciplinary VC team and, where appropriate, to provide initial scan review in routine clinical practice.

The primary aim of this study was to prospectively evaluate the efficacy of the course in delivering training in VC technique and initial interpretation by combining an evaluation of the teaching methodologies with an assessment of behavioural change.

8.2 Methods

This was a prospective study carried out in a single institution (academic radiology department). It was approved by the St Mark's review board and classified as educational evaluation by the UK National Research Ethics Service.

8.2.1 Participant details

Radiographers, either experienced in gastrointestinal imaging (for example barium enema and/or CT), and working in the UK National Health Service (NHS), were invited by national advertising to attend the course in small groups of four delegates (usually two from each department). Their radiologist leads were also invited for the final one and half days of the course to review their radiographers' training and to seek advice and support for developing a VC service from scratch.

Participating radiographers were required to have access to regular training/clinical lists in their base hospital to enable subsequent VC implementation if there was not currently a service developed. Lead radiology consultants from the participating centres were designated as 'mentors' to their radiographer 'trainees' and attended the second half of the course to help promote teamwork and common goals for service implementation or expansion in their own hospitals.

8.2.2 Course design

The course was four days long and incorporated practical technique and 'hands-on' workstation training with active participation in the daily VC service under direct supervision of a highly experienced radiographer and consultant radiologist (with experience of over 1500 and 2000 examinations respectively). Each trainee had access to an individual VC workstation running an up to date version of Virtual colonoscopy software, combining 2D and 3D 'flythrough' displays (Vitrea 4.0, Vital Images, USA) and computer assisted detection software (ColonCAD, Medicsight

PLC, London, UK). Lectures on image interpretation (anatomy, methods and pitfalls), advice on establishing and running a VC service and the interaction with endoscopy were provided (see Appendix 3, p.274 for course details and learning objectives).

8.2.3 Course evaluation

The main aim of the study was to evaluate the course by comparing trainees' knowledge and performance before and after the course. It was not possible to compare against a suitable reference standard as there are currently no validated recommendations for optimal training or minimum requirements for assessment. This study was therefore designed to evaluate the course at multiple levels according to Kirkpatrick's framework for course evaluation[42].

Initial reaction was assessed using a feedback form completed by all the participants at the end of the course. This provided information regarding the perceived quality of the individual course components and allowed identification of areas for improvement. Specific areas were rated on either a visual analogue scale (VAS) or on a categorical Likert scale. Free-text comments were encouraged, and the results were analysed after each course and appropriate adjustments to the course content made.

Evaluation of **learning** was subdivided into assessment of knowledge and skills. To assess knowledge, a two-paper Multiple Choice Questionnaire (MCQ) was used. This consists of fifty true-false questions on ten topics (cancer epidemiology and diagnosis, VC technique, interpretation pitfalls, performance outcomes, faecal tagging, complications, computer aided detection, optical colonoscopy and difficult lesions), split into two papers, A and B. Evaluation of the change in knowledge was by a pre/post test design. The radiographers were given paper A or B (randomly assigned) at the beginning of the course and then the combined paper (AB) at the end of course. This design allowed compensation for possible variation in difficulty of the questions between paper A and B, and for previous experience. It also allowed identification of the level of pre-course knowledge to inform the trainers. Papers were positively marked and a percentage correct score calculated for each trainee.

As indicated previously, there are currently no standards for assessment of procedure performance for VC. Part of this study therefore required development of a training and assessment schema and identification of 'minimum' standards of performance. This was not a summative 'pass/fail' assessment, but rather a formative assessment designed to identify weaknesses in performance and allow directed feedback for future improvement.

The radiographers were assessed on their ability to demonstrate competence in the practical skills required to perform a VC. This was done towards the end of the course following observation of the procedure and specific hands-on training sessions. Assessment of performance was by an expert instructor against a step-by-step checklist of optimum performance as well as a global rating score compiled by expert radiologists and radiographers (Appendix 4, p.277). Previous studies of surgical trainee assessments have shown that a global rating score is a more accurate discriminator of performance, especially in more experienced candidates [254,262]. The checklist is used to provide a training schema and also to identify essential steps (known as nodal points) in the examination. These were highlighted in the checklist, and candidates who failed to complete that step could be stopped by the instructor from proceeding with the examination. The instructor was also permitted to interrupt the procedure at any point if they thought the safety of the patient was being, or had the potential to become, compromised. For the purposes of our analysis, scores were assessed as competent or not if they were within one standard deviation of the mean.

Behaviour was assessed by investigating the change in interpretative performance when reading carefully compiled VC datasets. Each dataset comprised 5 cases incorporating overt significant pathology with a high prevalence of cancers and polyps. These cases had been endoscopically validated and selected by our experienced radiologists as being good examples of significant pathology with a wide range of morphology encountered in day to day practice (Table 30). They were used in a pre/post test design with trainees randomised as to the order of completion; AB or BA. Outcome measures were correct test case classification and large lesion (>10mm polyp or cancer) sensitivity.

Figure 30 gives a summary of the course evaluation process.

Figure 30 Summary of ViCToR course evaluation

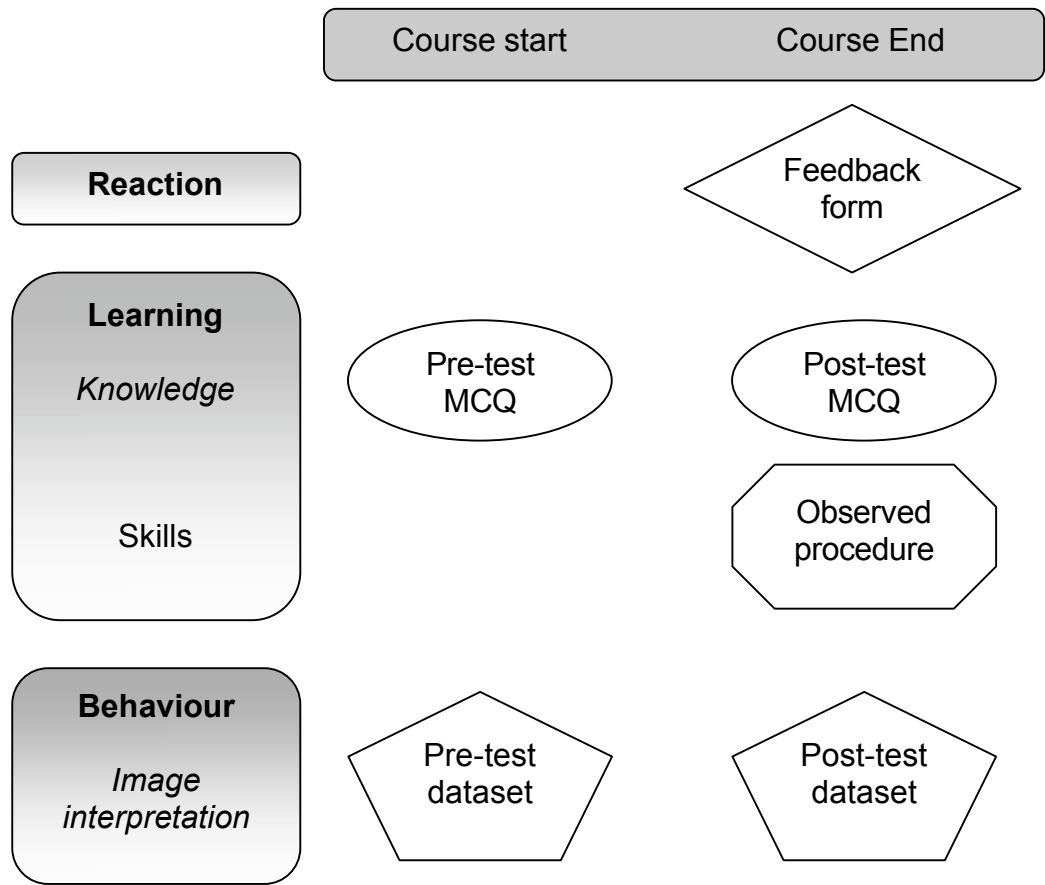


Table 30 VC interpretation test case details

Set A				
Case	Morphology	Size	Location	Classification
1	Polyp	15	Sigmoid	S3
2	Polyp	6	Sigmoid	S2
3	Cancer	31	Sigmoid	S4
4	Cancer	80	Sigmoid	S4
	Cancer	22	Transverse	
	Polyp	10	Rectum	
5	Cancer	28	Rectum	S4
	Polyp	20	Sigmoid	
	Polyp	20	Ascending	
	Polyp	10	Ascending	

Set B				
Case	Morphology	Size	Location	Classification
1	Cancer	11	Sigmoid	S4
2	Polyp	15	Sigmoid	S3
3	Cancer	17	Transverse	S4
	Polyp	6	Descending	
	Polyp	4	Descending	
	Polyp	6	Ascending	
4	Polyp	22	Rectum	S3
	Polyp	10	Sigmoid	
	Polyp	8	Transverse	
	Polyp	27	Ascending	
5	Polyp	7	Sigmoid	S2
	Polyp	5	Descending	
	Polyp	6	Transverse	
	Polyp	7	Ascending	

Size in mm. Classification; S0=Inadequate, S1=Normal, S2=Polyp 6-10mm, S3=Polyp>10mm, S4=Cancer, S5=Diverticular stricture

8.2.4 Statistical analysis

Data was entered in a coded fashion into a database and statistical analysis was performed on an intention-to-treat basis. All continuous variables were found to be normally distributed, so paired t-tests were used to compare pre and post-course measures. Pearson correlation was used to examine whether there were any changes in the results with time as the course developed. A planned analysis six months after the first course was performed to confirm statistical appropriateness, sample size calculations and adjust the course assessments if necessary[277].

8.3 Results

13 courses over 2 years with 49 participants were evaluated. Data was available regarding MCQ scores and image interpretation for all 49, but skills assessment from only 41 candidates, as the assessment evaluation marksheets had to be piloted and were not finalised until after the second course.

8.3.1 Reaction (feedback)

Candidates rated the course very highly, with a mean (SD) score of 8.6 (1.53) out of 10 on a VAS, and it met or exceeded all of their expectations. They also found the course highly enjoyable, with mean scores of 8.9 (1.1) out of 10, and would universally recommend it to a colleague.

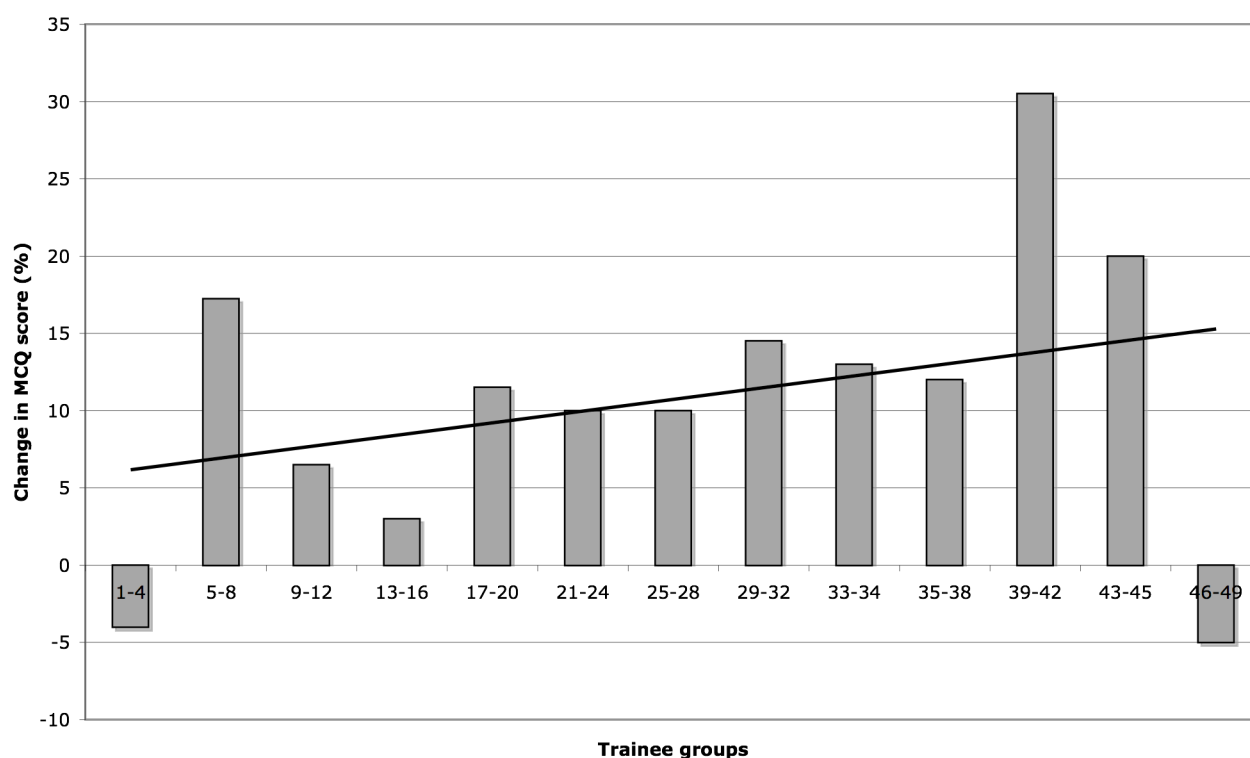
8.3.2 Learning (knowledge)

Interim analysis of the MCQ results (20 delegates from 5 courses) demonstrated that Paper A was significantly easier than Paper B, with mean pre-test scores of 53% and 68% respectively ($P=0.002$). The mean score for the combined paper (AB) post-test was 69%. Answers to individual questions were analysed to determine question difficulty and the question distribution amended between paper A and B in order to match the papers for difficulty. Analysis following the change (29 delegates from 8 courses) demonstrated equivalence of the papers, with mean pre-test scores of 56% and 59% for papers A and B respectively.

Analysis including all 49 delegates showed that trainees significantly improved their scores by a mean of 13%, attaining mean post-test scores of 70% ($p<0.001$ vs pre-

test score). There was a trend towards greater improvements in score with time as the course developed, with a Pearson correlation of 0.32, although the improvement was not significant ($p=0.30$) (Figure 31). The first and last two groups had a mean decrease in MCQ score, which may be the reason the trend was not significant. Reasons for this decrease in score are not clear, although feedback from these groups indicate that they felt rushed for time during their assessments, which may have affected their performance.

Figure 31 Difference in mean MCQ score pre to post training for consecutive blocks of trainees



8.3.3 Learning (skills)

Skills assessment demonstrated that all candidates were competent at performing a VC, with no candidate stopped during the assessment. Mean checklist scores were 14.1 out of 15, and mean Global Score 25.9 out of 30, with no essential steps (nodal points) missed by any candidate.

8.3.4 Behaviour (image interpretation)

Correct classification of test cases improved significantly from a mean of 55% pre-course to 71% post-course, $P < 0.001$. Detection rates for large, biologically significant polyps (>10 mm) and cancers also improved significantly from 49% to 60%, $P = 0.002$ (Table 31).

Table 31 Summary of results for learning and behaviour evaluation

	Pre-training	Post-training	p-value
MCQ (%)	58.6 (10.6)	69.0 (9.5)	<0.001
Correct test case classification (%)	54.7 (24.7)	71.0 (22.7)	<0.001
Large lesion sensitivity	48.5 (19.3)	60.4 (17.1)	0.002

Results are mean (SD)

8.4 Discussion

Virtual colonoscopy is now established as the preferred radiological test for examining the colon for colonic neoplasia [278,279] and in several countries, radiographers are likely to play a central role in its wider dissemination, including conversion of current services from barium enema. This study is the first to attempt to define standards for radiographer training in virtual colonoscopy. We present evidence for the effectiveness of a high-intensity, structured training program in virtual colonoscopy for radiographers by formally demonstrating improved knowledge and skills in both performance and initial interpretation of VC. Most courses are evaluated only at the very lowest level, with changes based purely on candidate feedback. By evaluating the course at multiple levels, we can reassure delegates and funding authorities of its effectiveness.

Our findings demonstrate that radiographers can safely perform virtual colonoscopy and with suitable training can potentially provide an initial interpretation to detect significant pathology. As a result, radiographers could be well placed to request immediate radiologist review when a cancer is found at the time of examination. If cancer is confirmed and resources allow, patients could benefit from same visit CT

staging and also same day endoscopy, improving diagnostic efficiency, potentially reducing patient anxiety and avoiding the need for an additional bowel preparation, thus improving patient care.

However, as the standard deviations in Table 31 demonstrate, there is a range of performance and overall detection rates are relatively low compared to experienced radiologists, albeit similar to previously published performance for non-academic sub-specialist radiologists[280]. The course can provide the initial foundations of core knowledge and skills, but there is clearly a requirement for further training and additional experience to attain competency and support an active role in examination review. As can be seen in Figure 31, there was also a learning curve for development of the training delivered in the course, which was particularly noticeable when adjustments were being made in order to tailor it to the feedback received from the radiographers. Different professional groups clearly have differences in learning needs and learning styles that need addressing.

The course was designed to be multi-disciplinary in nature to encourage continued training and development for both radiographers and radiologists alike. We did require 'buy-in' from all involved parties in order to enhance the learning benefit from the course. Our findings support the introduction of intensive courses as a 'minimum standard' for training, as has been consistently recommended by experts[281]. Individual trusts wishing to institute a VC service should consider investing in training programmes as part of a local business case for service development. When assessing the costs of funding external courses, including decreased capacity in their own departments to cover absence for study leave, it is important to also recognise the wider benefits the course provides, such as supporting team building, role extension and clinical risk management. It is also vital to ensure PMETB standards are met for quality assurance of such courses.

Variability in uptake of training may worsen the performance gap that already exists between centres, and perhaps all radiographers wishing to do so should have access to such training. This may require a national training initiative and we await the reaction of the Department of Health and the Bowel Cancer Screening Programme to the publication of SIGGAR1[282] and the international VC standards

document providing recommendations for practice[283]. Creation of a nationally recognised qualification in the training, performance and interpretation of VC examinations would provide a level of quality assurance for both trusts and individuals alike. Minimum standards could be set to enable both formative and summative assessments of performance, which could then be combined with internal audit to ensure ongoing high quality service provision. Implementation of such a programme has the potential to improve performance on a national level.

The radiographers attending the course have been encouraged to audit their own practice in terms of polyp/cancer detection rates, using radiologist second reads and endoscopy findings as comparators, to try to evaluate the efficacy of the training at the highest Kirkpatrick level (Results). Until such follow-up data are available, there are limits to the conclusions that can be made regarding the long-term efficacy of our training program. This is made more difficult by the lack of current recommendations against which to make a comparison of performance. It is also acknowledged that some of the data presented in this study may be skewed towards better performance in the later groups, due to ongoing changes to both the course and the assessments as a result of feedback and the interim analysis. However, although it may affect the study data, if these changes result in improved quality of the training, this should be seen as a positive outcome rather than a limitation.

8.5 Conclusions

A dedicated training course for radiographers can ensure competent technical performance of VC and may enable an initial review by radiographer to expedite the patient pathway for colorectal cancer and improve diagnostic efficiency. However, a national training initiative and accreditation may be required similar to that developed for bowel cancer screening colonoscopy in order to reduce performance gaps between centres and ensure the quality assurance process.

Chapter 9: Professionalism in endoscopy – non-technical skills taxonomy development

9.1 Background and Aims

The current emphasis in endoscopy training is on improving knowledge, technical skill and assessment of competency, and the previous chapters in this thesis have specifically addressed these issues. However, there is an omission in these educational aims, which ignore the attitudes and non-technical skills which form an integral part of learning. As discussed in the introduction (p.53), they also form the basis of what is considered to be professionalism within any field of medicine. In the absence of any previous work in gastrointestinal endoscopy to base any research on, this chapter details the first attempt to determine the non-technical skills which are specific to and underpin professionalism in this field. The next chapter will detail the development of a marker system that is intended to enable training and assessment of these skills.

The aims of this study were:

1. To review the current literature regarding attitudes, non-technical skills and the development of behavioural marker systems in medicine
2. To identify what non-technical skills are important in gastrointestinal endoscopy
3. To develop a taxonomy of non-technical skills

9.2 Literature review

9.2.1 Current non-technical skills taxonomies and behavioural marker systems

A literature search was performed to establish the current state of theoretical and applied knowledge on non-technical skills in medicine and in gastrointestinal endoscopy in particular. Seven existing non-technical skills and behavioural marker systems were identified as currently in use within medicine. There are three behavioural marker systems of non-technical skills relating to surgery (NOTSS[284],

HFRS[285], NOTECHS[193,286,287]), one surgical attitude questionnaire (ORMAQ[288]), one basic surgical training assessment tool (EBSTAF[289]), one observational surgical teamwork assessment tool (OTAS[290]), and one taxonomy of non-technical skills relating to anaesthesia (ANTS[291]). These were analysed to establish their component skills and structure.

The four behavioural marker systems all had a hierarchical structure, with ratings at both category and element level. The teamwork assessment tool had a linear structure based on a timeline of pre-during-post surgery skills. The questionnaire and basic surgical training assessment tool both had a simple categorical framework. Many skills were included in virtually all of the systems, for example, situation awareness and communication skills. This would be compatible with the psychological literature which indicates that many non-technical skills are generic and applicable to multiple situations. However, there are clearly some skills that may be specific or particularly important to an individual discipline. For example, prioritising tasks is only listed as a separate skill in the anaesthetic marker system, although it is implied in many of the categories for the surgical ones as well (see Appendix 5, p.279 for details of the component analysis).

No existing taxonomies or behavioural marker systems were found relating to gastrointestinal endoscopy. However three related documents were analysed to identify the non-technical skills contained; the Endoscopy National Competency Framework[255], a list of competencies derived for accreditation of competence for nurse endoscopists; the Global Rating Scale[292], a national endoscopy unit audit document; and the JAG guidelines for the training, accreditation and assessment of trainees in gastrointestinal endoscopy[32]. Requirement for many of the same skills and behaviours could be identified in these documents, including communication, monitoring, safety, feedback, risk identification and maintaining standards. However these were often either implied as necessary or, if specified, were generally subsumed as part of the technical requirements.

9.2.2 Non-technical skills identified

From the literature search, a list of 24 non-technical skills were identified that are currently in use in medicine. These are shown in Table 32.

Table 32 Non-technical skills identified from literature search

Situation awareness	Teaching	Monitoring
Communication style	Maintaining team climate	Reviewing situation
Leadership	Decision making	Crisis management
Workload management	Team building	Declaring emergency
Confidence/assertion	Planning	Organisational climate
Feedback	Problem solving	Error reporting
Prioritisation	Information sharing	Stress/fatigue management
Recognition	Preparation	Documentation

9.2.3 Development of behavioural marker systems

Flin and Martin (2001)[293] identified five key areas to be considered when developing behavioural marker systems. These are:

1. Unit of assessment
2. Identification of skills and their associated behaviours
3. Assessment method
4. Rater reliability
5. Rater training

For the purposes of this study, the unit of assessment was deemed to be at the individual rather than the team level. This is because although endoscopists do not work in isolation, the teams are usually small and fluid, and at the training level

individuals need to develop the necessary skills to allow them to work in different teams and in different situations. Moreover, final assessment of competency in endoscopy is currently at the level of the individual (MCQ/DOPS) or the unit (GRS) rather than the team.

To identify skills and associated behaviours, there are a number of techniques that can be considered, all of which lend themselves to qualitative rather than quantitative analysis. It is well recognised that in qualitative studies, using one single method of enquiry is likely to result in inadequate data collection, and using multiple methods is much more likely to produce an accurate representation of the important human factors in individual disciplines[294]. Qualitative methods include questionnaires, experimental studies, focus groups, semi-structured interviews, direct observation and video analysis.

Detailed workpackages have been published of the process by which two of the systems in anaesthesia (ANTS)[295] and surgery (NOTSS)[296] have been developed. Multiple methods of investigation were used in these studies, under the generic term of cognitive task analysis (CTA)[297]. This methodology seemed to be the most applicable to the development of a taxonomy for endoscopy, and was adopted for use in this study. The details of this will be discussed in the Methods section.

To determine the behavioural markers, many of the systems adopted the Delphi process using repeated revision of expert opinion to arrive at a consensus opinion, usually taking place during specially organised workshops. However, several of the systems, particularly those involving teamwork, made direct observations of performance. This has the advantage of being able to identify very specific behaviours, although it comes with the trade-off of producing ever larger assessment schemas. However, using previously identified constructs or taxonomies can successfully avoid this problem by relating exemplar behaviours to previously identified dimensions[298]. As this project aimed to develop a taxonomy prior to identifying the behavioural markers, direct observation seemed to provide the most valid and systematic method of identification and was adopted for use in this study.

The assessment method relates to the way in which direct and measurable output of performance is recorded, usually involving observer judgement of performance utilising a scale based on behavioural markers. There are a wide range of scales in use, with different number of points and anchors. Assessment can also be of a summative type, with direct ratings of 'Pass' or 'Fail'. Alternatively it can be formative, usually with graded performance levels and then if needed, criteria can be set to determine a summative pass/fail divide.

Currently the technical skills for endoscopy (DOPS) are rated on a four point scale[35]. The DOPS can be used both formatively and summatively, with major and minor criteria set in order to determine competency. The existing behavioural marker systems for anaesthesia and surgery are also rated on a four point scale, with specific emphasis on behaviours that enhance or endanger patient safety. However, they are currently not used summatively outside a research environment.

Flin and Martin's fourth and fifth points relate to validation of the tool in terms of reliability and in terms of training assessors to use the tool. As this study focussed on development of the tool, these points will be discussed in the section on future work.

9.3 Methods

This study was approved by the St Mark's review board and given ethical approval by the UK National Research Ethics Service.

9.3.1 Triangulation

A combination of techniques were adapted for use in this study to try to minimise the limitations of collecting qualitative data and the difficulties in accessing implicit knowledge. This concept is known as triangulation and involves the use of multiple sources to enhance the rigour of qualitative research. Depending on epistemological viewpoint, it has been seen either as a method of getting nearer the 'truth' of a phenomenon[299], or as a means of increasing the understanding of the phenomenon through a process of 'crystallisation'[300].

Denzin (cited in Stake, 1995) described four types of triangulation:

1. Data source triangulation
2. Methodological triangulation
3. Investigator (observer) triangulation
4. Theory triangulation

I used the first three types to increase the rigour of this research. Firstly, I have used two different groups of participants, a focus group and an interview group, to triangulate the data sources. Secondly, although I was the main investigator for the majority of the analysis, the data was also critically analysed using the same methods by my supervisor, Dr Thomas-Gibson, and also by a clinical psychologist, Dr Katherine Wolff. Thirdly, for methodological triangulation, I used multiple methods to elicit findings; the main method to obtain data was the critical incident analysis, but three further tasks were also used to give corroborative evidence. The four different tasks used in this study to elicit details regarding non-technical skills are detailed below:

- i. Knowledge audit
- ii. Sorting task
- iii. Rating task
- iv. Critical incident analysis

9.3.1.1 Knowledge audit

A knowledge audit[301] seeks to investigate aspects of expertise by asking for specific examples and using probe questions to elicit further details. It is typically run as a 'brainstorming' exercise. The question asked in this study was:

What kinds of non-technical skills are important or make a good endoscopist?

The follow-up probe questions then aimed to elicit details regarding skills development and use of non-technical skills in crisis situations (see Appendix 6, p.283 for details of schedule). A record of the suggestions (either on paper or a

white-board) was kept in real-time to facilitate the subsequent discussion. The time allocated for this was approximately 15 minutes.

9.3.1.2 Sorting task

A sorting task can be useful to show how different concepts are grouped together, and so identify how the understanding of the task is structured[302]. In this task, the aim was to identify how endoscopists view the relationships between non-technical skills and to identify recurrent themes that may provide over-arching domains. The skills identified from the literature search (Table 32) were put on to individual cards and the participants asked to arrange them however they thought appropriate. There were no definitions provided at the start of the task, and participants were asked to make decisions based purely on what they understood by the terms. Having sorted them as they felt appropriate, participants were asked to talk through the reasons for their choices and to suggest what the groups might be called.

9.3.1.3 Rating task

A rating task can be useful to stratify items in terms of importance and identify those that are considered to be more integral to a particular field than others[302]. This can identify representational skills and shared 'mental models' between experts. The participants were given a list of the previously identified skills and asked to rate each on a scale of 1-5, 1 being not important at all, and 5 being very important. The time allocated for this and the sorting task was approximately 15 minutes.

9.3.1.4 Critical incident analysis

This task aimed to provide concrete evidence of the non-technical skills that are important for safe endoscopic practice by identifying those skills that were involved in actual critical incidents. The critical incident technique[303] and the critical decision method[304] were adapted for use as they have been shown to be effective in eliciting knowledge on non-technical skills in anaesthesia[295]. Participants were asked to recount an actual critical incident or near-miss to provide examples of where human factors were implicated in situations that have adversely affected patient safety. There were three iterations of the account: (i) participants initially gave an unstructured free-recall of the event, helping the interviewee remember the case and for the interviewer to check appropriateness. Then (ii) the interviewer repeated

the case back to check that it was understood, clarifying details if needed. The final iteration (iii) involved running through the account point-by-point to elicit further details regarding the non-technical skills identifiable, using a structured set of probe questions where more information was needed (see Appendix 6, p.283). The time allocated for this was between 30-45 minutes.

9.3.2 Phases of data collection

9.3.2.1 Focus group

A group of endoscopists and endoscopy staff from the Wolfson Unit for Endoscopy (doctors, nurses, and support staff) were asked to participate in an initial focus group to investigate the degree of congruence between those skills identified in the literature search to those that are considered important in endoscopy. This was a semi-structured group discussion, with all staff from the unit invited to attend in order to encompass the subset of professionals with an interest in the subject topic. Although all participants were from a single institution (convenience sampling) and thus may not be representative of the profession as a whole, it was considered a suitable initial starting point for a focus group. The knowledge audit was performed with the group as a whole to identify which non-technical skills are applicable to endoscopy. The sorting task was done in two groups, one comprising endoscopists and the other comprising nursing and secretarial staff. The rating task was completed by each individual. To avoid disruption, the focus group took place during a regular weekly meeting for all unit staff and lasted approximately 40 minutes. Staff were given a participant information sheet prior to the meeting and verbal consent for participation and voice recording was obtained.

9.3.2.2 Semi-structured interviews

For the semi-structured interviews, consultant gastroenterologists and surgeons were chosen as experts in their field likely to have insight into the non-technical skills involved in gastrointestinal endoscopy and had been practicing long enough to have a significant or memorable incident to recount. For practical purposes, all of the consultants selected were personally known to me and based in London and the South East of England. My assumption is that the experiences related by this group

are representative of the experiences of other endoscopists throughout the UK, and that the results can therefore be used to identify generic themes that are applicable to that wider context.

There was no predetermined number of interviews. Kvale (1996) states that one should simply “interview as many subjects as necessary to find out what you need to know”[305]. In framework analysis, the data collection can occur alongside the initial data analysis (during the familiarisation process), and it is possible to cease data collection when thematic saturation occurs. As the data collected was extremely dense and information rich, we followed recommendations to keep the sample size small in order to keep the data at a manageable level[306]. Seven initial interviews were carried out and a further two interviews were conducted following the initial analysis to increase the sampling frame.

Each consultant was invited by email to participate, and each interview was arranged at a suitable time and in a private place convenient for them. Each consultant was sent an information sheet prior to the interview and also asked to think of a performance example that they would be willing to recount for the critical incident analysis. Verbal consent was obtained at the time of the interview, and all participants agreed to the use of a voice-recording device.

9.4 Analysis

The data from the knowledge audit, sorting and rating tasks was transcribed to a database and tables created to investigate similarities between how the participants grouped and rated the skills. The interviews were transcribed into documents, and extracts from the transcripts were managed using spreadsheets to analyse the data using the framework approach[294]. The framework approach was chosen over other methodologies (such as grounded theory, discourse analysis or content analysis) as it can “meet specific information needs and produce actionable outcomes” (Ritchie and Spencer, 1993). It provides a pragmatic and informative approach in qualitative research where the research question has been identified in advance, rather than emerging from the data. It is therefore particularly suited to, and has been used previously in studies of this type investigating non-technical skills in

anaesthetics and surgery. The steps used in the analysis of my data are detailed below.

9.4.1 Familiarisation

This first step involved becoming immersed in the data in order to become familiar with it prior to formal analysis. This began with me carrying out the interviews and then personally transcribing and proof-reading each transcript. This gave me an extremely thorough awareness of the richness of the data.

9.4.2 Identifying a thematic framework

This next step involved 'coding' the transcripts to extract units of meaning within the data[307]. In this methodology, codes emerge from the data as it is read rather than allocating examples to predetermined groups. Whenever a non-technical issue was identified, a notation was made of the page and line within each transcript where it was encountered and a brief descriptor was generated, which could include references to knowledge, skills or behaviours. As each code was developed, further references to different aspects of the code were allocated and gradually a framework of codes was generated. The evolving thematic framework was applied to each sequential transcript to look for further data extracts that were compatible with the code and to identify new concepts in order to create new codes.

This process was performed by myself and Dr Thomas-Gibson independently on seven interview transcripts and each individual's codes were collated. These were categorised and compared, with each incidence of a skill listed as being identified as the same by both investigators (full agreement), or by both investigators but in different places (partial agreement) or by one investigator alone (no agreement). The codes in the collated framework were then discussed individually and the key ideas and recurrent themes were identified and developed into an initial taxonomy (skill list). Two further interview transcripts were then coded by both of us using the preliminary taxonomy. Three interviews, two of the previous transcripts and one of the new transcripts, were also coded by a clinical psychologist using the taxonomy. This was to look at agreement between readers and to ensure that all of the skills identified could be coded using the taxonomy. This will increase both the confirmability and the credibility of the analysis.

9.4.3 Indexing

The taxonomy of skills was then applied back to all the transcripts and verbatim examples of each of the identified skills were extracted. A further discussion was then held between the two main investigators (AH and STG), a clinical psychologist (KW) and the National Endoscopy Lead (RV) in order to refine the taxonomy.

9.4.4 Charting

The skills and verbatim examples were then arranged by code in order to produce a chart which shows the range of descriptive detail regarding that item. This aims to provide an overview of each code and examination of the depth and breadth of the examples attached to it.

9.5 Results

9.5.1 Participants

The focus group comprised two consultant gastroenterologists, one nurse consultant, three gastroenterology trainees (specialist registrars), four endoscopy nurses, and three secretarial/administrative support staff.

Nine consultants participated in the semi-structured interviews. Table 33 gives details of their demographics.

Table 33 Demographics of consultants for semi-structured interviews

Participant	Years as consultant	Sex	Speciality	Endoscopy experience		
				OGD	Colon	ERCP/other
1	1	M	Gastroenterology	2000	1200	130
2	8	M	Gastroenterology	3000	>3000	
3	4	M	Gastroenterology	3000	2500	
4	11	M	Gastroenterology	3500	1500	1500
5	18	M	Gastroenterology	10000	5000	2500
6	17	M	Surgery	5000	6000	1000
7	5	F	Surgery	500	1000	
8	2	F	Gastroenterology	2000	1500	<50
9	5	M	Gastroenterology	8000	6000	2000

9.5.2 Triangulation tasks

9.5.2.1 Knowledge audit

The knowledge audit performed during the focus group produced a list of 30 different non-technical skills (Table 34). Twelve of the skills were deemed to be different aspects of four 'categorical' skills; communication, empathy, infrastructure and professionalism. The remainder were simply statements that seemed to stand on their own. Although the phraseology is different, it was notable that the list produced was very similar to the list derived from the literature search (Table 32), providing some credibility for the knowledge audit as a process.

Table 34 Non-technical skills elicited from the focus group knowledge audit

Communication	Own limitations (knowing when to call for help)	Decision making
	Assessing communication	Sense of control
	Listening	Background knowledge
	Giving information	Coping with stress
Empathy	Patient/bedside manner	Time management
	Appropriate sense of humour	Confidence
Infrastructure	Unit organisation	Organisation
	Outside environment	Team builder
	Budgeting	Appearance/ Body language
Professionalism	Manners	Data protection/recording
	Politeness	Report writing
	Respect for colleagues	Dictation

Participants were then asked how they thought these skills are currently developed. The following were listed:

Role modelling – learning from observation

From mistakes – informally or formally via critical incident analysis or audit

Formal degree (such as an MSc)

General environment/ethos of the unit

Peer review (e.g. 360 degree assessment)

Communication skills training

Several of the participants identified ways in which they had personally developed some of their own non-technical skills, although the only method that was recognised as designed specifically to improve non-technical skills was communication skills training. There was a consensus that some of the skills may not be very amenable to training, for example 'sense of humour'.

Performing the knowledge audit with individual consultants in the semi-structured interviews did not elicit any new skills from those previously identified by the focus group.

9.5.2.2 Sorting and rating tasks

The sorting task provided some interesting insights into the interactions between the skills identified from the literature search. In the focus group, the endoscopists created a hierarchical arrangement with cascading skills arranged into three main sections, which they called 'Daily Skills', 'Core Skills' and 'Crisis Skills' (Figure 32). The non-endoscopists segregated the skills more definitely into three groups, one of which was also called 'Crisis Skills'. They sorted the rest of the skills into 'General Management Skills' and then an 'in-between' category of 'Overlap Skills' (Figure 33). When asked if the skills needed for crisis or emergency situation were different to those required in normal or ordinary ones, two points arose. Firstly, the majority thought that the skills needed were the same, but that some might become more important in a difficult situation, for example assertiveness and leadership. Secondly, there was a suggestion that different 'ranges' might be appropriate for different

situations, for example there was general agreement that it was important to remain calm and not panic, but that you didn't want to be too laid back, as you needed a certain amount of 'adrenaline' in order to get the problem solved.

Figure 32 Focus group sorting task results - endoscopists

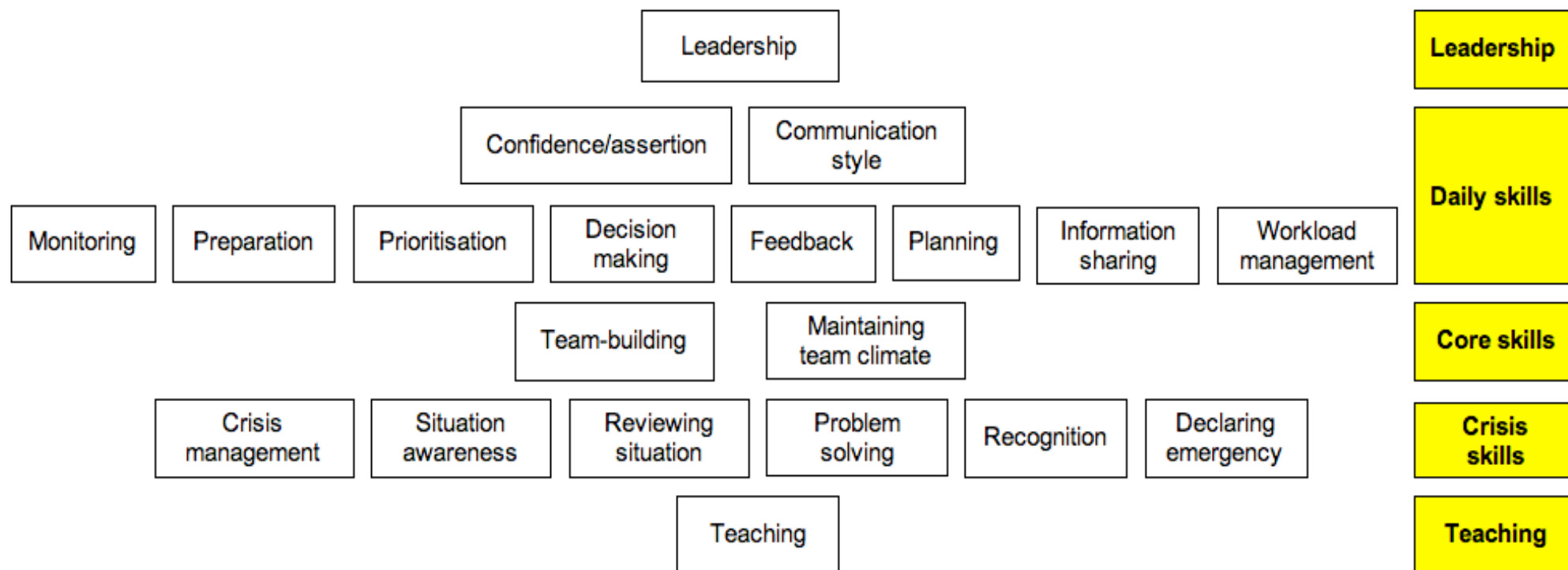
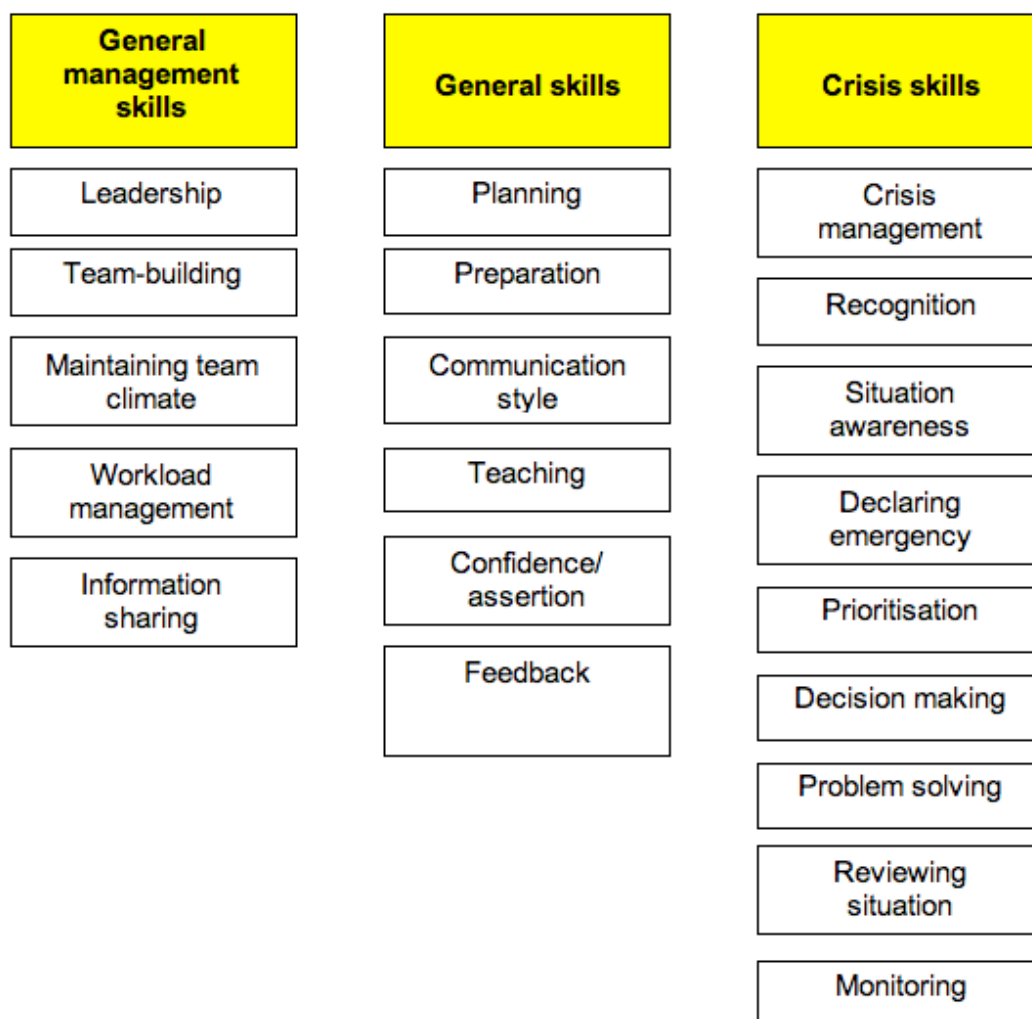


Figure 33 Focus group sorting task results – non-endoscopists



When the sorting task was performed by individual consultants, nine different categories were used (Table 35). However, five of these categories were only ever used by one consultant, and two by only two consultants.

Table 35 Category names given during the sorting task by individual consultant endoscopists

- 1=Organisation/System/Environment
- 2=Individual/inherent/character
- 3=Negative
- 4=Core attributes
- 5=Emergency/crisis
- 6=Bridging qualities/feedback
- 7=Routine/task skills/misc
- 8=Professionalism
- 9=Team functions
- 10=Preparation
- 11=Teachable skills

Analysing the skills by which categories they were most often placed in showed that each skill was most often allocated to one of three categories; 'Organisational skills', 'Individual skills' and 'Crisis management skills'. One skill, 'Preparation' was placed in both 'Organisational' and the 'Individual' categories equally, and Prioritisation was placed in the 'Individual' and 'Crisis Management' categories equally (see Table 36).

Table 36 Arrangement of sorting task skills by most common allocation

Organisational skills	Individual Skills	Crisis Management Skills
Preparation		Crisis management
Workload management	Situation awareness	Declaring emergency
Teaching	Prioritisation	
Maintaining team climate	Communication style	
Team building	Confidence/assertion	
Planning	Leadership	
Information sharing	Recognition	
Monitoring	Decision making	
Organisational climate	Problem solving	
Error reporting	Reviewing situation	
Documentation	Professionalism	
	Stress/fatigue management	

The rating task did not add much insight to the analysis, as all the skills were rated as being important or very important by all individuals in the focus group and by all the consultants in the semi-structured interviews.

9.5.3 Critical incident analysis

9.5.3.1 Familiarisation

During the transcription and initial read-throughs of the first seven interviews, it was apparent that all interviewees had selected an incident with a negative experience to recount, despite the question not being framed in that way (they were asked to recount an incident that was difficult for them). This is understandable in terms of negative incidents being more memorable than positive incidents (regret theory[308]) but may introduce bias to the skill list by underrepresenting the non-technical skills used when successfully dealing with a difficult situation. In the two subsequent interviews I therefore emphasised the chance to recount a difficult situation which resulted in a successful outcome, rather than a complication. This resulted in one 'near miss' type critical incident and one very complex case where there was a complication which eventually had a successful outcome, but that is now currently undergoing litigation.

The other notable fact was that nearly half (4 out of 9) of the incidents involved a trainee. Two cases occurred when the consultant were themselves trainees, and in the other two, a trainee caused the problem and the consultant was needed to resolve it. In three of the cases, the trainee was not being supervised by a consultant in the room. Two of these incidents occurred in the days when there was no requirement for competency accreditation, so the trainee was perhaps inappropriately left alone during the procedure. One incident occurred in a trainee who had only recently been ‘signed-off’ as competent, but who attempted “the largest polypectomy he’d ever seen”, again perhaps inappropriately as the current summative assessment for competency does not have a specific assessment of polypectomy technique.

The two types of complication recounted were bleeding and perforation, which are the two most significant complications of gastrointestinal endoscopy, giving further face validity to the critical incident analysis process. The range of incidents recounted is shown in Table 37.

Table 37 Range of incidents recounted for the critical incident analysis

Post-polypectomy bleeding	
1	Trainee performing unsupervised large polypectomy in the colon
2	Endoscopic Mucosal Resection (EMR) of oesophageal lesion
3	Large gastric polypectomy
Perforation	
4	Trainee performing unsupervised flexible sigmoidoscopy
5	Complication of colonic stricture dilatation
6	Trainee performing unsupervised diagnostic colonoscopy
7	Diagnostic colonoscopy
8	Trainee performing supervised ERCP
‘Near-miss’	
9	Planned oesophageal dilatation moved to inappropriate environment with inadequate equipment

9.5.3.2 Identifying a thematic framework

The initial 7 transcripts were coded independently by the two investigators (AH and STG) to create their own lists of non-technical skills. 33 codes were identified in total. 18 skills were identified using the same code at the same point in the transcripts by both investigators (full agreement). 14 skills were identified by both investigators using the same code, but in different places in the transcripts (partial agreement), of which only 1 skill (training/supervision) was not also used with full agreement. 14 codes were identified by one investigator alone (no agreement). The full list of codes is shown in Table 38.

Table 38 Codes developed during initial analysis of critical incident transcripts by two independent readers

Full agreement	Partial agreement	No agreement
Communication	Communication	Following rules
Confidence	Confidence	Information gathering
Crisis management	Crisis management	Problem solving
Emotional control	Emotional control	Learning
Reflection	Reflection	Modesty
Responsibility	Responsibility	Risk taking
Judgement	Judgement	Cautiousness
Decision Making	Decision Making	Clear thinking
Crisis recognition		Monitoring
Team organisation		Preparation
Assessing situation		Declaring emergency
Seeking Help		Conflict resolution
Leadership	Leadership	Delegation
Teamwork	Teamwork	Professionalism
Focus		
Planning	Planning	
Awareness	Awareness	
Unit organisation	Unit organisation	
	Training/supervision	

Each code was then discussed by the two investigators in order to determine more precisely what aspects of each non-technical skill were being recounted, frequently referring back to the transcripts to elicit further details as to the exact circumstances referred to by the interviewees.

A thematic framework was then developed detailing 16 codes and subordinate aspects (see Table 39). A number was allocated to each code and a letter to each subordinate aspect for ease of reference. These will be discussed individually in turn, with details from transcripts provided as corroborative detail. 17 codes were either deemed to be covered by other codes and not included in the framework, or were deemed to be sub-sections of another code and included as a subordinate aspect.

Table 39 Initial thematic framework of codes

Skill		Aspects	
1	Communication	A	With nursing staff (in the endoscopy room)
		B	With trainee
		C	With colleagues
		D	With the patient
		E	With relatives
2	Confidence	A	Lack of confidence in own abilities or judgement
		B	Overconfidence – assumption that problems would not occur
3	Crisis recognition	A	Ability to recognise when an untoward event has occurred
4	Crisis management	A	Ability to 'switch mode' to deal with a crisis
		B	Ability to institute appropriate management
		C	Knowing when to ask for help
5	Emotional control	A	Awareness of own reaction
		B	Awareness of other's reaction
		C	Control of own reaction
		D	Control of other's reactions
6	Reflection	A	Post procedure of what went wrong
		B	Post procedure of what went right
		C	Feedback of reflection to other members of the team
		D	Making changes based on reflection to improve practice
7	Responsibility	A	For own behaviour and actions during the procedure
		B	For follow-up post procedure
8	Judgement	A	Assessment of pre-procedural risk
		B	Self-awareness of own abilities and limitations
		C	Assessment of situation
9	Decision Making	A	Decision whether to perform or stop a procedure
		B	Ability to put an action plan in place to manage problems
10	Team organisation	A	Pre-procedure organisation and briefing of team
11	Assessing situation	A	Collecting data, establishing facts, risk assessment
12	Leadership	A	Towards a trainee – appropriate supervision and role model
		B	Of the team
		C	Of the unit
13	Teamwork	A	Good working relationship/communication with team
		B	Seeks opinion of team
14	Focus	A	Concentration
		B	Control of the environment appropriate to match the situation
15	Planning	A	Ensure appropriate situation before the start of the procedure
		B	Ensure appropriate environment at the start of the procedure
16	Awareness	A	Of patient's condition
		B	Of team
		C	Of procedural progression

1. Communication

- a. *With nursing staff***
- b. *With trainee***
- c. *With colleagues***
- d. *With the patient***
- e. *With relatives***

Communication was the most commonly utilised code, with every single transcript eliciting multiple examples. Due to the number and variety of these examples, it was considered necessary to develop multiple sub-categories that were independent but fell within the same overlying theme of communication. This category could have been subdivided into communication skills that were particularly relevant to specific times, for example before, during and after the procedure, but in general it was thought to be more useful to identify specific individuals to whom good communication was important/essential as this is where it broke down in the critical incidents.

A. *With nursing staff (in the endoscopy room)*

Communication with staff in the room was coded in terms of the manner in which it was done, in this example the interviewee is describing the need for clear and calm instructions:

“we need some adrenaline and can somebody go and get the APC please, in that manner”

There was also an aspect relating to difficulties arising from lack of familiarity with staff:

“communication with the nursing staff in the endoscopy room, who at the time didn’t know me desperately well, which made it a bit harder”

B. With trainee

There were clearly issues specifically with respect to communication with trainees, both in terms of setting boundaries:

“I don’t know whether anyone’s actually sat him back and said, and expressed it, that these are the rules of your solo colonoscopy. If you get to a really big polyp that you think is, you know, one of the biggest things you’ve ever seen, or you’re at all worried, then you call someone else.”

and in terms of direct communication difficulties as a trainee with more senior colleagues:

“So I phoned him back and told him what had happened, and he wouldn’t believe me”

C. With colleagues

Communication with other colleagues was coded identically on 6 occasions. One type related to seeking help for correct decision-making:

“I actually called in my colleague and said, “Now, what setting would you do for this and what would you do?”

“And I phoned the boss and said “I don’t think that we should, I should proceed”, because I thought it was hazardous and I didn’t really see what it was going to achieve anyway”

The other main area that communication issues arose was post-complication management. These clearly overlap with crisis management and reflection, but were felt to be specific issues related to direct communication:

“I told the consultant and rang the surgeon straight away.”

“I would speak to the most senior person, not the registrar”

D. With the patient and relatives

Communication with patient and relatives was clearly important in management of complications:

[I had] “spoken to the patient and explained to her that there had been a complication.”

However, it was considered that this was a vital skill for an endoscopist in all situations, and warranted inclusion in the more general skill of ‘communication’ as well as in crisis management.

2. Confidence

a. Lack of confidence in own abilities or judgement

b. Overconfidence – assumption that complications would not occur

This code had clear positive and negative elements. On several occasions it was clear that an incident could have been avoided if the endoscopist had had more confidence in their own abilities or judgement to say no or to back off.

“I persisted even though I knew I shouldn’t be. And I suppose I could have actually put my foot down and said “No, I’m not going to continue”

There were also clear examples of procedures performed at the limit of or beyond the endoscopists’ capabilities, representing overconfidence in their own abilities.

“I was going to have to do it sometime and now I’m going to have to do it. Here’s a big polyp, you know, I know what I should be doing and [they] kind of get on with it”

There was also an example of failure to recognise a complication because of overconfidence that it would be extremely unlikely to occur.

“The boss then came down, reinserted the colonoscope, went through the perforation again, accepted there had been a perforation,”

3. Crisis recognition

a. *Ability to recognise when an untoward event has occurred*

This came across as an important skill. Examples given were phrases such as “*the penny dropped*” or mentally “*pushing the emergency button*”, and played an important role when situations were not recognised appropriately:

“He didn’t really recognise the seriousness of the bleeding until the situation was completely out of hand.”

It was considered that there are likely to be different levels of ‘crisis’, from simple incidents e.g. a problem with equipment, to recognition of a more significant complication such as a perforation. Several examples of uncertainty or puzzlement prior to recognition were also noted, particularly in inexperienced practitioners or for medical endoscopists vs surgical endoscopists in recognising a perforation.

“I found I was looking at an area – one hadn’t a clue what one was seeing. And after a little bit twigged that this had probably gone through a hole”

4. Crisis management

a. *Ability to ‘switch mode’*

b. *Ability to institute appropriate management*

c. *Knowing when to ask for help*

This code potentially had many different aspects to it. It could be broken down into a variety of different components such as procedural management, communication, emotional control, teamwork factors. However, most of these were covered in other codes and although they would apply in a crisis situation, there seemed to be three key areas that were essential in management. Firstly, the ability to ‘switch’ to ‘emergency mode’, which allowed a focus on the urgency and importance of the task without resorting to panic:

“You switch from registering that as a fear to focusing on work... the problem”

Secondly, the ability to logically institute appropriate management, which clearly includes the technical aspects of the procedure, but may also include the non-technical aspects such as emotional control, management of the team etc.

“And, you know, pulled out and so from then on I guess my management was, even in retrospect, was as it should be...”

Thirdly, a recognition of when extra help was required to safely manage a situation, either because the endoscopist has reached the limit of their own capabilities, or because of a need for additional hands/kit/advice.

“that actually if the polyp was in the sigmoid it may have been easy just to come out, phone you, and if somebody’s not around to bring him back on another list another day”

5. Emotional control

- a. Awareness of your own reaction**
- b. Awareness of other’s reaction**
- c. Control of your own reaction**
- d. Control of other’s reactions**

Awareness of the endoscopist’s own reaction was coded with full agreement on eight occasions. Emotions identified included shock, horror, anxiety, guilt and calmness:

“you know... I was shocked really”

“I was just so horrified as to what I’d done. There was something really horrible about seeing the inside of the peritoneal cavity from the end of the colonoscope, that you are slightly attenuated in your thought processes, by just being so horrified with what you’ve done.”

“Yes, I was clearly worried about it, oh yeah”

“then I felt all the things that I suppose anybody would feel. Anxious and you know, guilty.”

“So my first reaction was, um, the usual one which was - Don’t panic”

There was clearly a difference between the awareness or experience of the emotion and then the subsequent ability to control this reaction.

“my reaction although was “oh shit” was, right ... I felt that I did not show any obvious [panic].”

It also occurred several times in the context of teamwork, with other’s reactions noted by the endoscopist:

“But their reaction to seeing the screen fill up with blood was clearly um, one of, um slightly more panic actually”

“two nurses with me, um both very experienced nurses, um were more, were visibly uh, taken aback and agitated”

It also occurred in terms of leadership with the endoscopist’s reaction (of calmness or control) providing a basis to encourage the other members of the team not to panic:

“I think it’s important actually to radiate um, calmness and assurity because I think they feed off you. If you panic and show panic, and start, um throwing things or getting agitated, it can infect them and make things worse.”

“my demeanour, I think would have encouraged a non-panicking response”

Discussion identified the need to debrief team members following an incident to identify both one's own and other's reactions to the event and improve control in future episodes.

6. Reflection

- a. *Post procedure of what went wrong***
- b. *Post procedure of what went right***
- c. *Feedback to other members of the team***
- d. *Making changes based on reflection to improve practice***

This was coded with full agreement on seven occasions. On several occasions the use of the word "obvious" or "obviously" came up in terms of the need to reflect back on an incident. However, this is not necessarily the case, as not all transcripts (2 out of the 7) elicited this code. It was clear that it was necessary to elicit not just what went wrong:

"And in fact, we went on under the same general anaesthetic and removed her colon. Again, in retrospect, I am not sure of the wisdom of doing actually"

but also learn from what went right, feed this back to the wider team and then to act on this.

"So I guess I learnt... there are different attitudes towards training. That was one of the things I learnt from that"

"One thing that I definitely felt, as a response to this was that if I ever do this again, I would want an adrenaline-filled needle ready at the side there and then and the APC in the room with me, which we did not have"

7. Responsibility

- a. *For own behaviour and actions during the procedure***
- b. *Post procedure for follow-up***

There were several instances where people took or failed to take responsibility for actions during the procedure:

“So I felt that I was responsible because I was thinking through what was happening.”

or for follow-up after a complication:

“So following-up and... Yeah, I think I would be now much more involved in, whereas then I suppose I handed over responsibility”

During discussion, it was felt that the ability to take responsibility is a prime requirement of all endoscopists (or indeed all doctors to their patients) and should apply to all procedures, whether there was a complication or not. It was considered to be clearly one of the fundamentals of ‘Professionalism’.

8. Judgement

- a. *Assessment of pre-procedural risk***
- b. *Self-awareness of own abilities and limitations***
- c. *Assessment of situation***

These were coded with full agreement 14 times in the transcripts and were considered to be extremely important skills for an endoscopist. Examples of risk assessment and decision-making were:

“The technical questions were really, you know, would this bleed. Is this an unusual position and therefore could we expect a complication, either bleed or perforate?”

“We were already in a situation where we were thinking about surgery and decision-making around surgery. And it was a high risk of perforation, as colonoscopies go. And I think we should have had the thinking process around, you know, what would have happened if we had not been able to continue with the colonoscopy”

The combination of failure of judgement of risk with failure to recognise limitations in skill occurred in multiple transcripts and was clearly the direct cause of several complications:

“basically he said to me it was the biggest polyp he’d ever seen, but... uh... he didn’t call a senior..”

“when he started getting the bleeding... he started trying to recover the thing [polyp].”

9. Decision Making

a. Decision when to stop a procedure

b. Ability to put an action plan in place to manage problems

This occurred eleven times with full agreement in the transcripts, in both positive and negative ways. The most common occurrence was the decision to perform a procedure that resulted in a complication which, with hindsight, probably should not have been done:

“my initial surprise was that he’d gone ahead and done what he’d done without actually calling someone before, or having backed away.”

“I would do a colectomy on the emergency list, either the same day or the following morning. She had a defunctioned bowel. Give her antibiotics and do this in a planned way. Not wake up from a colonoscopy and realise you’ve had your colon removed”

However, it also occurred several times in terms of getting a confirmatory opinion prior to a procedure.

“Yep. So I thought, let’s ask somebody else. Who’s probably not done very much either, and let’s get to a collective decision”

“We said, let’s go ahead and do it, and I talked to the patient and talked him through it”

There were also several instances involving the decision-making ability to take the correct action once a complication had occurred.

“You have to then make a decision that you’re either going to manage this conservatively or you’re going to do an operation and you’re either going to do the operation right now or you’re going to wake the patient up. They’re very clear decisions.”

10. Team organisation

a. Pre-procedure organisation and briefing

This occurred 5 times with full agreement, with two different themes. The first was a deliberate action on the part of the endoscopist to ensure that the appropriate staff were on hand for a particular procedure, and that they were aware of the case and the potential for problems:

“I had 2 good nurses in the room and in fact, I deliberately sought them out because... before this procedure, knowing that it was the first time I was doing it, and I wanted people who I felt confident would react appropriately for my needs”

“I go to the nurse in charge and say “look, I’m doing this. I would like so-and-so or even you to come and assist” and I’ve never had a refusal on that. So I would try and adjust the team appropriately”

The second was the reverse, when the staffing level or experience of the nursing staff was not adequate for the type of case on the list:

“I think there was probably, um, probably only one nurse present, which is uh, how a lot of the lists get run I think, unfortunately”

“it didn’t bring over an air of confidence that the environment was suitable for this”

11. Assessing situation

a. *Collecting data, establishing facts, risk assessment*

This was only coded once with full agreement, but was also deemed to be the overarching code for another term used by an individual coder; “Information gathering”. Risk assessment was coded in this category, but was often subsumed into the decision-making category where the risk assessment was implicit in the decision as to whether to perform a procedure or not.

“when I got in there, you know, I was confident that the patient was fine, told him who I was, what was going on etc. etc. “

12. Leadership

a. *Towards a trainee*

b. *Of the team*

c. *Of the unit*

This was a recurring theme, occurring 11 times with full agreement. Three separate sub-categories became clear, with supervision of a trainee being the commonest occurrence. This occurred in a positive sense when addressing training needs and supervision of trainees:

“although it’s very difficult, we tried to organise appropriate training opportunities for that person”

It also occurred in a negative manner where trainees were not being given the supervision that they required:

“Well, I think that I felt the consultant was responsible, but I didn’t think that he was actually engaged in the process”

“I suppose, um, I don’t know whether anyone’s actually sat him back and said, and expressed it, that these are the rules of your solo colonoscopy”

Leadership of the team often involved emotional control, such as ensuring that the team performed in a calm and controlled manner in a crisis, or to inject a sense of urgency into the management of a problem:

“I think it’s important actually to radiate um, calmness and assurity because I think they feed off you. If you panic and show panic, and start, um throwing things or getting agitated, it can infect them and make things worse”

“you might have to wake them [the trainee] up and say “ok, you’ve got a real problem”

Leadership of a unit was also coded in terms of ensuring that trainers and endoscopy colleagues were aware of unit guidelines.

“all the trainers, all the list supervisors, have been sent a letter which they have returned signed, know that the unit policy is: you have to be... to scope unsupervised you have to have been accredited and that the unit will not take any responsibility for any complications “

13. Teamwork

- a. *Good working relationship/communication with team***
- b. *Seeks opinion of team***

This code was felt to overlap with several other codes, including leadership, team organisation and communication. However, two aspects were coded identically on 7 occasions and were felt to be important in isolation. There was an element of familiarity within a team that worked together regularly that contributed to a good working relationship and enhanced communication.

“You know when you work with a team and you’ve done it a hundred times”

There were also clearly situations where the opinion of team members were sought in order to clarify specific issues. The opposite was felt to contribute to complications in several cases, including those in which the team had not worked together before, or when opinions were not sought or not given because the team members were considered too junior or felt themselves unable to contribute.

“They were not really involved particularly, or forthcoming with any input really. We had, I think it was a pretty junior endoscopy nurse and there was an anaesthetist there...”

14. Focus

- a. *Concentration***
- b. *Control of the environment appropriate to match the situation***

There were two occasions in which this was coded with full agreement, but there were other occasions which were coded as “clear thinking” which were considered to be the same. The main thrust of this code was the ability to resist distractions and concentrate on the job in hand, which involved control of self as well as the environment in the room. Examples were:

“I’m gonna deal with it and you just get on and deal with it. You don’t let that, uh, distract you from what is the job in hand”

“I suppose in my own mind, what needed to come together is that we had the equipment, that we had the room, the staff, and a patient who was happy to go ahead with that after 2 hours of waiting.”

15. Planning

- a. Ensure appropriate situation pre-procedure e.g. notes, radiology available, appropriate case for endoscopist level***
- b. Ensure appropriate environment at the start of the procedure e.g. appropriate equipment available***

This was coded with full agreement on six occasions. There were two main elements that were identified – ensuring the procedure was well planned, with appropriate staff, case type etc.

“I deliberately sought them out because... before this procedure, knowing that it was the first time I was doing it, and I wanted people who I felt confident would react appropriately for my needs”

The second was ensuring that the room environment was appropriate. This was usually in the context of equipment that was not available when it should have been, either causing problems managing a complication or delaying a procedure.

“patient basically waited... for most of the day while most of the equipment was being found”

16. Awareness

- a. Of patient's condition**
- b. Of team**
- c. Of procedural progression**

This was clearly important in both the prevention and management of complications. Awareness (or lack of) of the patient's condition resulted in either good decision making:

"And was looking pale, blood pressure, you know, wasn't going down, but I think it was clear that it was going to... Best get in early"

Awareness was also vital in recognition and management of a problem when it did occur:

"You have that sort of awful sinking feeling, you realise what's done and try and minimise the harm you can do. And basically sort of, uh, get out of it and try and assess roughly what damage you've done."

Awareness of the team's response was considered to overlap with both teamwork and focus, as well as emotional control. While the majority of codes identified in the transcripts related to management of complications, awareness of other aspects of the team was considered an important skill for an endoscopist in non-crisis situations, so was included as a sub-section of this code rather than just crisis management.

There was also a clear awareness of how the procedure was progressing, in order to spot any problems that may arise:

"And so we did that, we applied it very well and very slowly and it all went very well"

Codes not included in the framework

Several codes were considered on discussion to be either the same as another code (just using a different name) or formed part of a subset of another code. The following were subsumed into the above codes or omitted.

Training/supervision	Learning	Preparation
Unit organisation	Modesty	Declaring emergency
Seeking Help	Risk-taking	Conflict resolution
Following rules	Cautiousness	Delegation
Information gathering	Clear thinking	Monitoring
Problem solving	Professionalism	

No new codes were developed following analysis of the two further interviews by AH and STG or the three interviews by the clinical psychologist, KW, although more examples of each of the current codes were identified. A further discussion was then held to refine the thematic framework into a taxonomy (STG, RV, AVH, KW).

Although the triangulation tasks had identified three main 'categories' of organisation, individual and crisis skills, it actually seemed much more appropriate to arrange the skills into a time-line (i.e. pre, during, post-procedure) in order to be more specific and user-friendly.

Five skills seemed to be 'overarching' or generic skills that would either be required during all parts of a procedure, or were not specific to any particular part. These were:

1. Communication
2. Teamwork
3. Leadership
4. Confidence
5. Emotional control

The remainder of the skills were classified in terms of which part of the procedure they referred to, for example 'Reflection' is clearly a post-procedural skill. In order to simplify the taxonomy, some of the sub-categories were combined in the descriptions. For example, awareness and control of emotions were combined in the

category 'Emotional control'. Where there were clear opposites in sub-categories, these were combined and the descriptor changed to 'appropriate level' for simplicity.

The categories of 'Crisis recognition' and 'Crisis management' were changed to 'Problem recognition' and 'Problem management' because it was felt that these skills were applicable to a much wider range of situations than what might be described as a 'crisis', and would be utilised in dealing with many types of problem. A refined taxonomy was produced, with the skills incorporated into a time-line (Table 40).

Table 40 Non-technical skills taxonomy

Skill		General Skills			
1	Communication	A	With nursing staff		
		B	With trainee		
		C	With colleagues		
		D	With the patient		
		E	With relatives		
2	Teamwork	A	Good working relationship/communication with team		
		B	Seeks opinion of team		
3	Leadership	A	Towards a trainee – appropriate supervision and role model		
		B	Of the team		
		C	Of the unit		
4	Confidence	Has appropriate level of confidence in own abilities, judgement and in the capabilities of the team.			
5	Emotional control	A	Awareness of own emotions and emotional reaction, and ability to control these appropriately		
		B	Awareness of other's emotions and emotional reaction, and adjusts own behaviour appropriately		
	Skill		Pre-procedure	During procedure	Post-Procedure
6	Planning	A	Makes assessment of pre-procedural risk		
		B	Ensures appropriate environment at the start of the procedure		
		C	Ensures pre-procedure organisation and briefing of team		
7	Assessing situation	Continually re-assesses situation; collecting data, establishing facts, and re-evaluating risk assessment			
8	Judgement & Decision Making	A	Has self-awareness of own abilities and limitations		
		B	Recognises when own abilities or the environment are not appropriate for the situation and makes appropriate decisions		
9	Focus	A		Has appropriate level of concentration for the situation	
		B		Keeps control of the environment appropriate to match the situation	
10	Awareness	A		Of patient's condition & wishes	
		B		Of team	
		C		Of procedural progression	
11	Problem recognition			Ability to recognise when an untoward event has occurred	
12	Problem management	A		Ability to 'switch mode' to deal with a problem	
		B		Ability to institute appropriate management	
		C		Knowing when to ask for help	
13	Responsibility	A		For own behaviour and actions during procedure	
		B		For follow-up post procedure	
14	Reflection	A		Post procedure of what went right or wrong	
		B		Feedback of reflection to other members of the team	
		C		Making changes based on reflection to improve practice	

9.5.3.3 Indexing

The taxonomy was then applied back to all of the transcripts, as the refined framework had not yet been applied to the entire dataset and the initial coding may not have incorporated aspects that emerged later in the analysis. Excerpts that had been originally labelled with codes that formed part of the final taxonomy were merely rearranged to fit in with the final framework. However, excerpts that had been coded with a description that was not used in the final framework were recoded to fit in with the new taxonomy. For example, in the initial coding, the following extract was coded as Delegation:

“he knew he needed help and got help, fortunately. And was very happy to hand it over to me”

Following indexing, this was coded as ‘Problem management – knowing when to ask for help’. This process led to enrichment of the data in each code and subordinate aspects.

9.5.3.4 Charting

Having indexed all the transcripts with the refined taxonomy, each example relating to an individual code was then extracted and put into a spreadsheet with the examples arranged by code. This permits an ‘overview’ of each code and examination of the depth and breadth of the examples attached to it. Using one of the smallest sub-categories as an example, ‘Seeks opinion of team’ in the category ‘Teamwork’ contains the following excerpts:

“I think there are people that I would feel less comfortable about doing something complex like this than others. So I would probably be looking for a team response”

“I actually called in my colleague and said, “Now, ..., who was next door, what setting would you do for this and what would you do?” So he looked at it and said “Yeah, I think we should remove that, don’t you?” and I said “Yes”.”

“Who’s probably not done very much either, and let’s get to a collective decision. Which is what I would do as a surgeon anyway if I’ve got a particularly difficult case.”

“They were not really involved particularly, or forthcoming with any input really. We had, I think it was a pretty junior endoscopy nurse and there was an anaesthetist there...”

“We discussed it with the anaesthetist, who was a registrar and who was just happy to do whatever we wanted to do really”

An attempt was made to derive ‘definitions’ about each code using the examples as a basis for the statements. For example, for the subcategory above, the statement derived was:

Feels able to seek opinions from other staff in order to make correct decisions

However, the statements tended to be of one of two types. The first type, as in the example given above, tended to be quite subjective and involve emotive verbs such as “feels able” or cognitive verbs such as “understands”. Although such statements may help to define the category, they are of limited usefulness in terms of either training or assessment. The second type of statement actually described behaviours attributable to the skills rather than the skills themselves. For example, the statement for ‘Leadership towards a trainee’ was:

Sets ground-rules for training and ensures appropriate training opportunities for skills development

In view of the plan to develop behavioural markers based on actual observation in the second part of this project, it was thought that these statements may become superfluous. It was also recognised that the descriptions were heavily biased towards the non-technical skills utilised in a crisis or critical incident, rather than day-to-day skills, as a direct reflection of the process which was used to develop them.

They were, however, used where examples of skills were not seen during the video recording analysis in the behavioural marker system development (p.206). In that event, it was felt that these statements should form the basis for the behavioural markers as they represent direct descriptors of where non-technical skills may break down and contribute to adverse events.

9.6 Discussion

This section has detailed the development of a non-technical skills taxonomy specific to gastrointestinal endoscopy. Although it is recognised that professionalism in endoscopy requires more than just knowledge or technical skill, this is the first time a structured methodology has been used to determine what other skills are needed for safe, competent practice.

The skills identified in the final taxonomy are strikingly similar to those developed for other taxonomies, including, as expected, very generic domains such as communication, planning and teamwork. However, the process has identified several aspects which may be particular to endoscopy. The skill which I have labelled 'Focus' recognises that the majority of the work in endoscopy may be 'day-to-day' or 'routine', which should be straightforward for a competent practitioner. If the procedure is relatively routine, an experienced endoscopist is able to perform multiple tasks simultaneously, both technical and non-technical, without much cognitive interference. However, if there are difficult moments during a procedure or if a problem occurs, the endoscopist should concentrate their attention on those specific tasks which need to be completed, if necessary ignoring other, less important tasks, until the difficulty is resolved. Our analysis showed that the lack of ability or failure to do this contributed to or directly resulted in several adverse events.

9.6.1 Strengths of the study

By utilising multiple qualitative research methods, a large quantity of data was collected for analysis, providing a very rich pool of detailed examples on which the framework of the taxonomy is hung. Although interviews lack the objectivity that comes with experimental research, the flexibility they provide makes them well suited

for investigations of this type. Triangulation of methods, observers and data sources was used in order to increase the credibility of the results. This has resulted in a robust, empirically based taxonomy of skills specific to gastrointestinal endoscopy.

9.6.2 Methodological issues

The results from this study are based on incidents recounted by a small sample of individuals. Ideally, a 'purposeful sampling' approach would be used to select as wide a range of individuals as possible within the sampling frame, with consultants from all parts of the UK invited to participate in order to allow identification of as broad a breadth of skills as possible ('maximum variation' or 'heterogenous' sampling). However, for practical purposes, we used a convenience sampling approach. Although this has been argued to be neither purposeful nor strategic[309], we felt that given the sensitivity of the information being discussed, it was important to have volunteers who felt comfortable taking part.

It is well established that memory is subject to various effects such as selective recall and hindsight bias, and that using recalled events as the basis for critical incident analysis may not be truly representative of the phenomenon[310]. A review of the cognitive task analysis approach considers that, although not immunising from memory loss effects, "it involves techniques that are intended to aid in overcoming memory loss"[311]. In particular, the minimal interruption by the interviewer and separation of the initial recall of the incident and the questioning should allow the person to re-immersing himself or herself in the situation. Despite best efforts, however, it is certain that bias will have been introduced. For example, as the interviewer I clearly had an influence on the data generated and in some instances may have imposed my own views onto the interviewee. In general, however, the transcripts of the interviews substantiate my belief that I remained neutral and impartial during majority of the interviews, allowing the participants to articulate their own feelings, views and interpretations of the recounted events.

When presenting examples of the data, I give each quote in isolation and do not always provide the context of the preceding or following comments. This does mean that the quotes could potentially be misinterpreted, but I felt that considering the extent of the data, to put each quote into full context would be excessive and would

not particularly help the reader's understanding. I accept that by not including the context that I am asking the reader to assume that the concepts expressed in the data extracts are not influenced by myself as an interviewer.

A further issue is that the majority of the results and much of the analysis was performed by me as the primary researcher, and therefore the findings could be regarded as my own construction of reality, having no bearing on the actual non-technical skills relevant to gastrointestinal endoscopy. I have attempted to offset this through the triangulation processes I have described, including involvement of multiple reviewers, although ultimately the results were the product of my own interaction with the data. I have, however, provided a detailed description of the research process during this chapter in order to provide transparency and credibility to the results. I believe that another researcher would produce a similar data set and a similar set of results if they were to follow the same methods of data collection and analysis.

9.6.3 Validation of results

The main aim of this study was to identify the non-technical skills specific to gastrointestinal endoscopy. However, as was alluded to in the introduction (p54) non-technical skills are often quite nebulous and difficult to define. Although the taxonomy was produced using a variety of qualitative research techniques, it will be difficult, if not impossible, to confirm that the skills contained within it are a "correct" or complete interpretation of the skill set without replicating the study in its entirety. This is particularly true considering that the number of interviews performed was quite limited, despite producing a large amount of data. The next chapter details the development of behavioural markers which aims to have a much stronger basis in observed evidence. Mapping the taxonomy to these behavioural markers will go some way to validating the results.

9.7 Conclusions

This project has investigated the non-technical skills required for gastrointestinal endoscopy and produced a taxonomy which can be used to inform the development of a behavioural marker scoring system.

Chapter 10: Professionalism in endoscopy - behavioural marker system development

10.1 Background and Aims

Having identified and produced a non-technical skills taxonomy as detailed in the previous chapter, this study aimed to produce a tool that can be used for training and assessment of these skills. It is difficult, if not impossible to adequately evaluate or assess a cognitive process, but it is possible to do so for behaviours that result from these. Using an aviation example, it is hard to tell if a pilot has good situational awareness, but it is possible to get an indication from observation of their information gathering skills, questioning style, discussions regarding the status of the aircraft. As a result, guidance for evaluation of Crew Resource Management (CRM)[312] has stated that evaluation should “focus on clearly defined, observable behaviour rather than inferences about underlying cognitive processes”. Previous work has shown that non-technical skills such as communication and planning have directly observable outcomes that may be used as behavioural markers in high-demand professions[193]. These behavioural markers have been defined as ‘observable, non-technical behaviours that contribute to superior or substandard performance within a work environment’ (Klampfer et al. 2001, p.10). They can be used to support the measurement of non-technical skills based on observation of behaviour either in a live setting or from video recordings.

Performance assessment is often hampered by the lack of sound measurement tools, with the risk that raters may assess performance in different ways and against ill-defined or non-standardised criteria. Systematic development of observational rating tools is therefore essential to ensure they have sufficient validity (does it measure what it is supposed to measure) and reliability (does it measure consistently) for clinical use. During development of previous behavioural marker systems, design criteria have been suggested to try to maximise their utility. These are (adapted from Fletcher et al, 2004):

1. The behaviours should be observable or inferred through communication.

2. The system should be easy and versatile to use with minimal training.
3. The system should be complementary to the competency-based approach being adopted in medical education in the UK.

Once developed, the tool needs to be evaluated in terms of rater reliability and rater training (see p.153). Although this is considered beyond the scope of this study, validation is discussed at the end of this chapter (p.195) and in the section on future work (p.230).

This study therefore has two main aims:

1. To identify behavioural markers which map to the non-technical skills previously identified
2. To develop a behavioural rating system based on the behavioural markers

10.2 Methods

This study was approved by the St Mark's review board and given ethical approval by the UK National Research Ethics Service.

10.2.1 Selection of cases

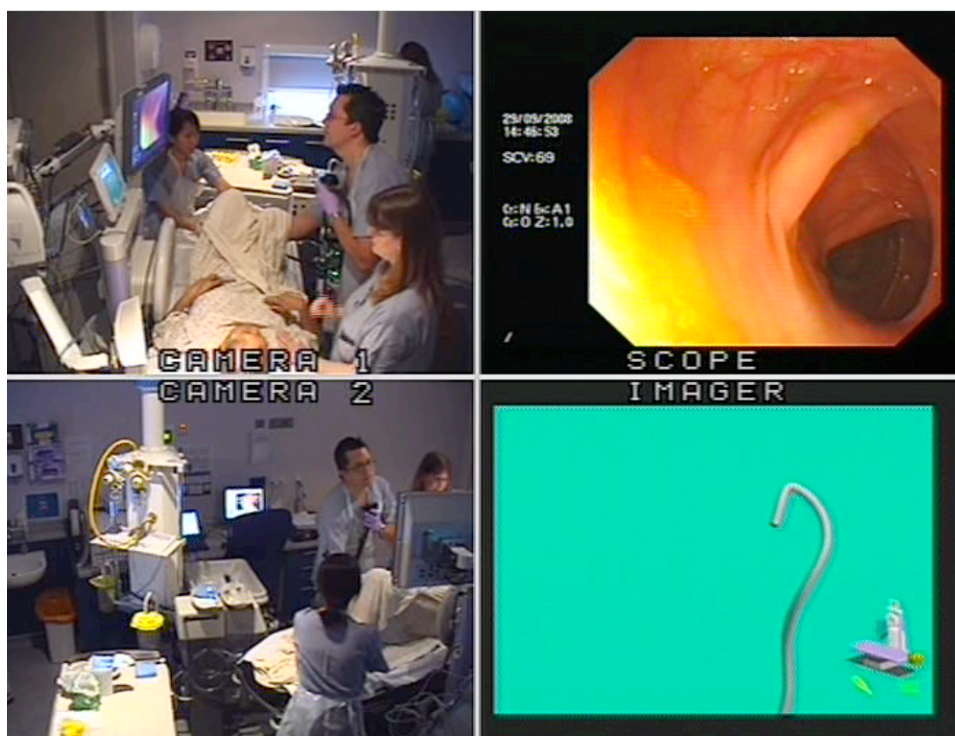
It is patently impossible to observe every variation of endoscopic performance, particularly of poor performance as the very nature of observing is likely to prevent many such behaviours. However, it is possible to identify the most expert practitioners and observe examples of good practice and then develop markers of both good and poor performance based on this. For this reason, the subset of colonoscopy performed for bowel cancer screening was chosen for this study. Bowel cancer screening in England is currently now only performed by colonoscopists who have already undergone a stringent accreditation procedure to ensure their expertise according to valid, nationally agreed criteria[313]. This involves presentation of data demonstrating their competence, a knowledge-based assessment, and a directly observed assessment of two live cases by expert assessors. This is the only cohort of current endoscopists that have objectively documented expertise. The patient

cohort also has a very high incidence of therapeutic intervention, which is likely to elicit the full range of non-technical skill behaviours that are required for safe, expert endoscopy in often challenging clinical cases. Written consent for recording was obtained from all patients and staff present in the room during the procedures.

10.2.2 Video recordings

Video recordings of colonoscopy procedures were used in this study rather than live observation for two reasons. Firstly, it is difficult for an observer to be present in the room for live cases due to the space limitations of the endoscopy suite. Secondly, video provides a permanent record of the procedure, allowing repeated review by many different observers. Evaluation of behavioural markers from video recordings is a recognised methodology and has been used successfully in both surgery[314] and anaesthesia[315]. Two suites of the endoscopy unit are equipped with state-of-the-art recording equipment that is able to produce a 'Quad-split' view combining two camera views of the room, the endoscopic view, a view of the 3-D imaging device (ScopeGuide) and sound, as shown in Figure 34. This can be recorded onto a DVD and played back on any computer or DVD player. This provides an extremely rich media resource that has been successfully used for many years during training in gastrointestinal endoscopy.

Figure 34 Example of video recording 'Quad-split' view



10.2.3 Analysis

10.2.3.1 Coding the video

The video was analysed using the taxonomy by the lead investigator (AH). Using the domains and skills from the taxonomy, specific behaviours were identified from actual observations or inferred from communication. Each behaviour was labelled by a video code, time-stamp and a description of either the behaviour or verbatim quotes if the behaviour was inferred from communication. Each behaviour was then mapped to the appropriate taxonomy skill. Behaviours could be mapped to more than one skill if it was thought multiple skills were being demonstrated.

10.2.3.2 Interpreting the behaviours

During the mapping process, behaviours were initially mapped directly to the skills and aspects of the taxonomy and then arranged to place examples of similar behaviours together within each skill. During this process, further categorisation and rearrangement was deemed necessary in order to simplify and refine the mapping process. The rationale and detailed explanation for this is described in the results section.

10.2.4 Descriptor development

Behaviours that provided examples of good performance were taken directly from the video analysis. Examples of poor performance were developed either by taking the opposite of the good behaviour or were drawn from the statements derived from actual examples elicited during the critical incident analysis described in the previous chapter. Particular care was taken with the phrasing of each descriptor, using clear, jargon-free language and active verbs to emphasise the observable aspects of each behaviour.

10.3 Results

10.3.1 Participants

Procedures performed by four different screening endoscopists were recorded. Each endoscopist was a consultant at the Wolfson Unit for Endoscopy and has performed

over 2000 colonoscopies. Three were consultant gastroenterologists (two male, one female) and one was a nurse consultant (female). All four consultants were bowel cancer screening accredited and three were accreditors themselves. Several procedures were recorded for each endoscopist in case of technical errors and to capture a range of behaviours.

10.3.2 Initial coding

Four procedures (one from each endoscopist) were randomly selected and coded by a single investigator (AH). Over 400 behaviours were identified from actual observations or inferred from communication and allocated to the taxonomy categories.

There were no behaviours identified in the videos for the following categories:

- 1C Communication with colleagues
- 1E Communication with relatives
- 3C Leadership of the unit
- 12C Asking for help

This is likely to be a direct result of the methodology used. The video is unable to record anything outside the endoscopy room, which is presumably where communication with colleagues and relatives would take place. Similarly, unit leadership skills are likely to be utilised outside of the endoscopy room in the majority of cases. None of the videos contained problems that the endoscopist was not able to resolve without help, so there were no behavioural markers for this category.

10.3.3 Re-coding based on ‘purpose’

While allocating the behaviours to the taxonomy categories, many behaviours were repeated. It became clear that the same behaviours could be allocated to several different non-technical skills. For example, in one video the endoscopist saw a small polyp. The observed behaviours were inferred from their comments at the time:

“We’ll do that one on the way back”

This was coded as 'Communication with nursing staff' as well as 'Planning'. However, it was fairly clear that the *purpose* of the behaviour was to ensure that the nursing staff were aware that there was a polyp that would need to be resected on the way back out, in effect developing an understanding of what needed to be done in the future.

The following comment was also coded in three categories. This was during a difficult polypectomy, where the position was unstable and there was a problem with the diathermy machine which wasn't working. There was clearly some urgency to get the problem sorted:

"Right, so it might be 'all hands on deck' now please, because I don't want to lose this position or the polyp."

This was coded under 'Leadership of the team', 'Focus – keeping control of the environment appropriate to match the situation' and 'Teamwork – good working relationship with the team'. However, the aim of the comment was clearly to ensure that the team understood that there was some urgency to deal with the problem and that everyone was going to have to pull together to resolve it. Again, the *purpose* was to ensure that the team had a shared understanding of what the current situation was and the need to resolve a problem.

It also became clear that different behaviours might be grouped together as they were serving the same purpose, although utilising different skills. For example, during several situations that involved dealing with problems, behaviours that had been coded as six different types of non-technical skill were utilised:

1. **Problem recognition:** The most straightforward and obvious behaviour was an observation that equipment was not working, for example a problem with the inflation button:

"It doesn't seem to be blowing"

2. **Leadership of the team:** When the diathermy was not working, the endoscopist took charge to try to diagnose the problem:

“Right, so it’s something to do with the contact on his leg...”

3. **Emotional control:** During an exceptionally difficult sigmoid intubation, the endoscopist admits out loud that they were finding it hard, and then gathered themselves together to continue the procedure:

“Gosh, dearie me that was tough. OK...”

4. **Assessing situation:** During insertion, the imager view was lost and the endoscopist quickly notices this:

“Why have we not got an imager view?”

5. **Judgement & Decision Making- Self awareness of abilities and limitations:** Prior to a polypectomy, the endoscopist makes a judgement about the risk of post-polypectomy bleeding and opts to put a clip on to prevent this:

“I tend to put a clip on when they are twitching like that”

6. **Awareness of the team:** Having had to get a second diathermy machine in (as the first one was not working), the endoscopist recognises that the other rooms may not now have the equipment they need and pre-empt the problem:

“If somebody needs the diathermy back, it’s there”

Rearranging the behaviours into what the *purpose* of the behaviour was thought to be seemed to be a much more natural way of classifying them. This resulted in thirteen different ‘purposes’ or elements, with multiple skills being used in each.

These elements again then naturally sorted into four different categories, as shown in Table 41.

Table 41 Classification of behavioural elements and categories

Communication & Teamwork	Situation Awareness	Leadership	Judgement & Decision Making
Exchanging information	Preparation	Supporting others	Considering options
Maintaining a shared understanding	Continuous assessment	Maintaining standards	Making decisions
Maintaining a patient-centred approach	Problem recognition	Dealing with problems	Reviewing the situation
	Focus		

10.3.3.1 Map to taxonomy

In order to relate these behaviours back to the non-technical skills taxonomy, a chart was produced to map the skills onto the new behavioural elements and categories. The four skills that did not have any behavioural markers identified in the video analysis were placed into what were felt to be the appropriate elements. The chart is shown in Figure 35 below.

Figure 35 Map of non-technical skills to behavioural elements and categories

Category	Element	Skills	
Communication and Teamwork	Exchanging information	1A 2A/B 3A 1B 7	Communication with nurses Teamwork Leadership towards a trainee Communication with trainee Assessing situation
	Maintaining a shared understanding	1A 2A/B 3B 10A	Communication with nurses Teamwork Leadership of the team Awareness of patient's condition and wishes
	Maintaining a patient-centred approach	1A 1D 1E 2A/B 4 5A/B 7 10A	Communication with nurses Communication with the patient Communication with relatives Teamwork Confidence Emotional control Assessing situation Awareness of patient's condition and wishes
Situation Awareness	Preparation	2A/B 6A/B/C 7 8A 10C	Teamwork Planning Assessing situation Self-awareness of own abilities Awareness of procedural progression
	Continuous assessment	7 10A	Assessing situation Awareness of patient's condition and wishes
	Problem recognition	11 3B 5A/B 7 8A/B 9B 10B	Problem recognition Leadership of the team Emotional control Assessing situation Judgement Control of the environment Awareness of team
	Focus	9A 9B	Appropriate level of concentration Control of the environment

Figure 35 continued

Leadership	Supporting others	2A/B 3A 3B 5B 10B	Teamwork Leadership towards a trainee Leadership of the team Awareness of other's emotions Awareness of team
	Maintaining standards	3B 9B 13A/B	Leadership of the unit Keeping control of the environment Responsibility
	Dealing with problems	1A 2A/B 3B 5A/B 9B 11 12A 12B	Communication with nurses Teamwork Leadership of the team Emotional control Control of environment Problem recognition Ability to 'switch mode' Ability to institute appropriate management

Judgement & Decision Making	Considering options	7 8A 10C 12C	Assessing situation Self awareness Awareness of procedural progression Knowing when to ask for help
	Making decisions	1A 3B 4 8A 10C 12A	Communication with nurses Leadership of the team Confidence Self-awareness Awareness of procedural progression Ability to 'switch mode'
	Reviewing the situation	7 8A 14A/B/C	Assessing situation Self-awareness Reflection

10.3.4 Behavioural marker system

A behavioural marker system was then developed comprising a three-level hierarchy consisting of categories, elements and behaviours. Each behaviour may reflect several of the non-technical skills identified in the taxonomy that mapped to that particular element. For each element, behaviours that provided examples of good

performance were taken directly from the video analysis. Examples of poor performance were developed either by taking the opposite of the good behaviour or were drawn from actual examples elicited during the critical incident analysis.

Figure 36 demonstrates how the system might be used to assess an individual at both the category and element level. Descriptions summarising each category and element are provided below the score sheet, with examples of good and poor behaviours in order to aid understanding and assessment.

Figure 36 Behavioural marker system structure at category and element level

Category	Rating	Element	Rating
Communication and Teamwork		Exchanging information	
		Maintaining a shared understanding	
		Maintaining a patient-centred approach	
Situation Awareness		Preparation	
		Continuous assessment	
		Problem recognition	
		Focus	
Leadership		Supporting others	
		Maintaining standards	
		Dealing with problems	
Judgement & Decision Making		Considering options	
		Making decisions	
		Reviewing the situation	

10.3.5 Descriptors and example behaviours

10.3.5.1 Communication and Teamwork

Skills for working within a team to ensure that knowledge is shared and understanding is reached to provide an effective patient-centred approach to safe procedural completion.

i. Exchanging information

Giving and receiving knowledge and information in a clear and timely fashion.

Good behaviours	Poor behaviours
Gives clear, specific instructions to staff and patient	Fails to give clear instructions
Seeks further information to aid understanding e.g. previous endoscopy reports	Does not seek further information or makes inappropriate assumptions
Listens and responds to team input	Does not listen to or acknowledge team members
Confirms team preparation including equipment availability	Does not check if team ready or if equipment available

ii. Maintaining a shared understanding

Ensuring that both the team and the endoscopist are working together from the same information and understand the 'big picture' of the case.

Good behaviours	Poor behaviours
Clarifies indication and objectives with team	Does not discuss case beforehand with team members
Confirms shared information with team e.g. medication doses, patient parameters, therapeutic efficacy	Does not check information with team
Talks about progress of procedure, including difficulties and concerns	Fails to keep team informed about progression or problems
Gives notice prior to therapeutic intervention to allow preparation time	Does not anticipate need for therapy
Explains unusual findings to team or trainee to increase understanding	Does not discuss findings within the team
Calmly indicates when situation requires urgency	Fails to convey need for urgency when required

iii. Maintaining a patient-centred approach

Ensuring that the patient is at the centre of the procedure, emphasising safety, comfort and giving information in a clear and understandable fashion.

Good behaviours	Poor behaviours
Greets patient and introduces self and team	Does not introduce self
Allays patient anxiety and maintains a relaxed atmosphere	Makes no attempt to reassure patient or maintain relaxed atmosphere
Gives clear instructions to patient	Does not give clear instructions
Regularly checks patient comfort	Does not check or ignores patient discomfort
Warns patient prior to uncomfortable event e.g. PR examination	Makes no effort to warn patient prior to uncomfortable events
Keeps patient informed about procedural progression (if appropriate)	Does not attempt to involve patient in the procedure
Explains findings to patient and/or relatives in clear, understandable language	Does not explain findings to patient and/or relatives or uses complex language or jargon

10.3.5.2 Situation Awareness

Creating and maintaining a dynamic awareness of procedural progression by good preparation, continuous assessment of the situation, recognition of problems or potential issues and maintaining clear focus in order to deal with these.

i. Preparation

Ensuring that the patient is fit, the procedure is appropriate, and that it is being done by an endoscopist with the necessary skills, equipment and assistants for safe and successful completion.

Good behaviours	Poor behaviours
Checks indications are appropriate	Does not make a pre-procedural review of notes or patient
Checks patient is fit for the procedure, including comorbidities and allergies	Fails to question indications and proceeds with inappropriate procedure or unfit patient
Checks unfamiliar assistants are adequately trained and experienced for the procedures	Fails to appreciate limitations of staff experience or views
Ensures equipment present and functioning correctly	Makes no effort to check equipment supplies or functioning
Optimises environmental conditions before starting e.g. bed height, equipment positioning	Proceeds with procedure in inadequately set-up or inappropriate surroundings
Does not perform procedure beyond own level of skill or experience	Proceeds with procedure beyond own limitations

ii. Continuous assessment

Maintaining a continuous evaluation of the patient's condition and updating the shared understanding to identify any mismatch between the current situation and expected state.

Good behaviours	Poor behaviours
Regularly checks patient response to sedation	Does not monitor patient or over-relies on assistants to identify problems
Articulates findings clearly	Overlooks or ignores findings
Uses all available techniques to inform decision-making process	Fails to adequately assess for pathology
Monitors results from therapy e.g. bleeding, patient pain	Discards results or findings that are not expected
Re-evaluates risk regularly depending on findings	Ignores results or findings that may increase risk

iii. Problem recognition

Recognising a mismatch between the current situation and the expected state and anticipating what may happen as a result of possible actions, interventions or non-intervention.

Good behaviours	Poor behaviours
Identifies issues quickly and highlights them to the team	Fails to identify problems
Articulates difficulties in procedural progression	Fails to discuss potential problems
Recognises increased risk due to unexpected finding	Proceeds with overconfidence with no regard for what may go wrong
Reflects and discusses significance of issues with team	

iv. Focus

Ensuring lack of distractions and maintaining concentration, particularly during difficult situations.

Good behaviours	Poor behaviours
Minimises interruptions (e.g. by locking door)	Fails to limit distractions (e.g. not turning mobile phone off)
Stops inappropriate discussions or distracting behaviour by staff	Tolerates inappropriate discussion or distracting behaviour
Keeps focus on screen at all times	Allows attention to be diverted easily
Maintains silence if needed during technically difficult manoeuvres	

10.3.5.3 Leadership

Demonstrating leadership by supporting team members, following procedures to maintain high quality clinical care and providing direction when dealing with problems.

i. Supporting others

Providing emotional and cognitive support to team members and trainees by tailoring leadership and teaching style appropriately.

Good behaviours	Poor behaviours
Maintains a relaxed atmosphere	Shows hostility or negativity to other team members
Gives praise for tasks done well	Fails to provide recognition for tasks done well or criticises inappropriately
Uses varied teaching techniques according to trainee needs	Fails to recognise needs of trainee or other staff
Does not rush staff when not necessary	Fails to recognise needs of others, requiring task reallocation

ii. Maintaining standards

Supporting safety and quality by adhering to current protocols and codes of clinical practice.

Good behaviours	Poor behaviours
Clearly follows unit procedures and protocols	Fails to observe protocols and standards
Ensures privacy and patient dignity	Shows disrespect to the patient
Adequately documents procedure immediately afterwards	Fails to adequately document procedure

iii. Dealing with problems

Adopting a calm and controlled demeanour when under pressure. Utilising all resources to maintain control of the situation and taking responsibility for patient outcome.

Good behaviours	Poor behaviours
Emphasises urgency of the situation if needed	Suppresses or dismisses concerns over problems
Gives clear directions to team to help resolve problem	Fails to assume leadership role
Delegates tasks in order to achieve goals	Fails to use team effectively to address situation
Remains calm under pressure	Panics or loses temper when under pressure
Maintains control and assumes responsibility for the patient	Blames others for errors and does not take personal responsibility

10.3.5.4 Judgement & Decision Making

Utilising all resources for dealing with issues and making a judgement in order to choose an appropriate course of action.

i. Considering options

Generating possible courses of action to solve an issue or problem, including assessment of risk and benefit.

Good behaviours	Poor behaviours
Generates options to resolve problems	Does not discuss options
Initiates discussion of options	Does not solicit views of team members
Weighs up pros and cons	Makes no evaluation of risk
Seeks help or opinion of colleagues	Fails to seek help when needed

ii. Making decisions

Choosing a solution to a problem, communicating this to team members and implementing it.

Good behaviours	Poor behaviours
Reaches and clearly communicates decisions	Hesitates or fails to reach a decision when time critical
Implements plan effectively	Selects inappropriate option that leads to increased risk or complication
Makes provision for alternate options	Does not develop provisional plan if option is unsuccessful

iii. Reviewing situation

Reviewing outcomes of procedure or options for dealing with problems. Reflecting on issues and instituting changes to improve practice.

Good behaviours	Poor behaviours
Re-evaluates outcomes and checks for complications	Does not review the impact of actions
Asks for opinion of team members	Fails to seek alternate opinions
Debriefs team and reflects on procedural difficulties and alternate solutions	Makes no effort to discuss problems or successes
Ensures appropriate follow-up for patient	Fails to arrange suitable follow-up
Makes changes based on reflection to improve practice	Makes the same error repeatedly

10.3.6 Rating scale

In order to assess individuals using this system, a four point scale was selected as is currently used in the endoscopic technical skills assessment (DOPS)[35] and in other behavioural marker systems. This was chosen as it has a wide enough range to differentiate levels of skill, yet is not too cumbersome or difficult to use. The scale emphasises patient safety as the primary outcome, and can be used to rate both category and element-level skills. It also recognises that not all behaviours may be observed or relevant in each particular case, in which case N/A should be used.

1	Poor	Performance endangered or potentially endangered patient safety. Serious remediation is required
2	Marginal	Performance indicated some cause for concern. Considerable improvement is needed
3	Acceptable	Performance was of a satisfactory standard, but could be improved
4	Good	Performance was of a consistently high standard, enhancing patient safety. It could be used as a positive example for others
	N/A	Not applicable

10.4 Discussion

This study has produced a tool to allow assessment and feedback regarding key observable non-technical skills needed by gastrointestinal endoscopists. This can potentially be used alongside a knowledge assessment and technical skills assessment to provide a more complete evaluation of professional behaviour within this field.

10.4.1 Strengths of the study

By using video analysis, this study has been able to identify specific behaviours and create a classification system based on direct observation. This provides an advantage over a system developed utilising expert opinion in that it should be less

subjective and contain examples that are more specific to the domain being investigated. Using recordings of procedures performed by endoscopists who have already been formally accredited as high-quality practitioners should also enhance this benefit by providing markers of behavioural excellence.

10.4.2 Methodological issues

It is recognised that the methods of data collection and analysis used in this study may not have identified behavioural markers for the entire range of non-technical skills, particularly those involving crisis situations or emergency procedures. These events are rare and unpredictable and therefore difficult to observe. However, when mapped to the skills taxonomy derived from actual critical incidents, the behavioural markers did cover virtually all of the skills categories. I believe that the absence of skills was a result of logistical issues regarding scope of the video recording and not inherent methodological ones, although this surmise will need to be proven by evaluating the system in both routine and non-routine situations.

One of the main difficulties in this project was to provide enough detail to make the system useful for skills development while avoiding the production of a long, unwieldy tool that is too cumbersome for use by endoscopists. A fairly pragmatic approach had to be taken in terms of grouping the behaviours into elements and categories in order to make the system easier to use[316]. Behaviours were identified that could fit into several of the skills taxonomy categories, and many categories contained different behaviours. It was therefore necessary to redefine the classification system in order to more easily group the large number of behaviours together into something that 'made sense'.

The work was done mainly by one individual and it is likely that this has influenced both the analysis and the creation of the marker system in terms of beliefs, judgements and prejudices. However, the observation of actual lists was an attempt to improve the construct validity of the marker system by gathering empirical data, and as such is an improvement on any of the behavioural marker systems previously developed. It also does not preclude the use of expert opinion to further refine the system, as will be discussed under face validity below.

The behavioural marker system can sit alongside the taxonomy, enhancing both the understanding of the skills and assessment of the behaviours that result from them. This is particularly important if they are to be applied for training purposes. The behavioural marker system is important in order to observe performance outcomes, but the taxonomy can provide an understanding and appreciation of the non-technical skills involved in producing the behaviours.

10.4.3 Validation

Having developed the marker system, it is imperative that it is validated prior to clinical use. As previously described, it should be evaluated for face, content and construct validity, and then assessed for reliability. The first step will be to ask experts in the field to evaluate the system for face validity. One comment made by Dr Roland Valori, the national endoscopy lead, was:

"I have just had a look and my first reaction is wow. It makes immediate sense and there is a deceptive: 'well that is obvious'."

If other experts have the same reaction, then that will be a measure of its face validity. Refinements can be made using a Delphi-type approach (sequential revisions according to comments made by a group of experts).

Further studies will need to be done to look at other types of validity and reliability. These are discussed in the section on future work (p.230).

10.5 Conclusions

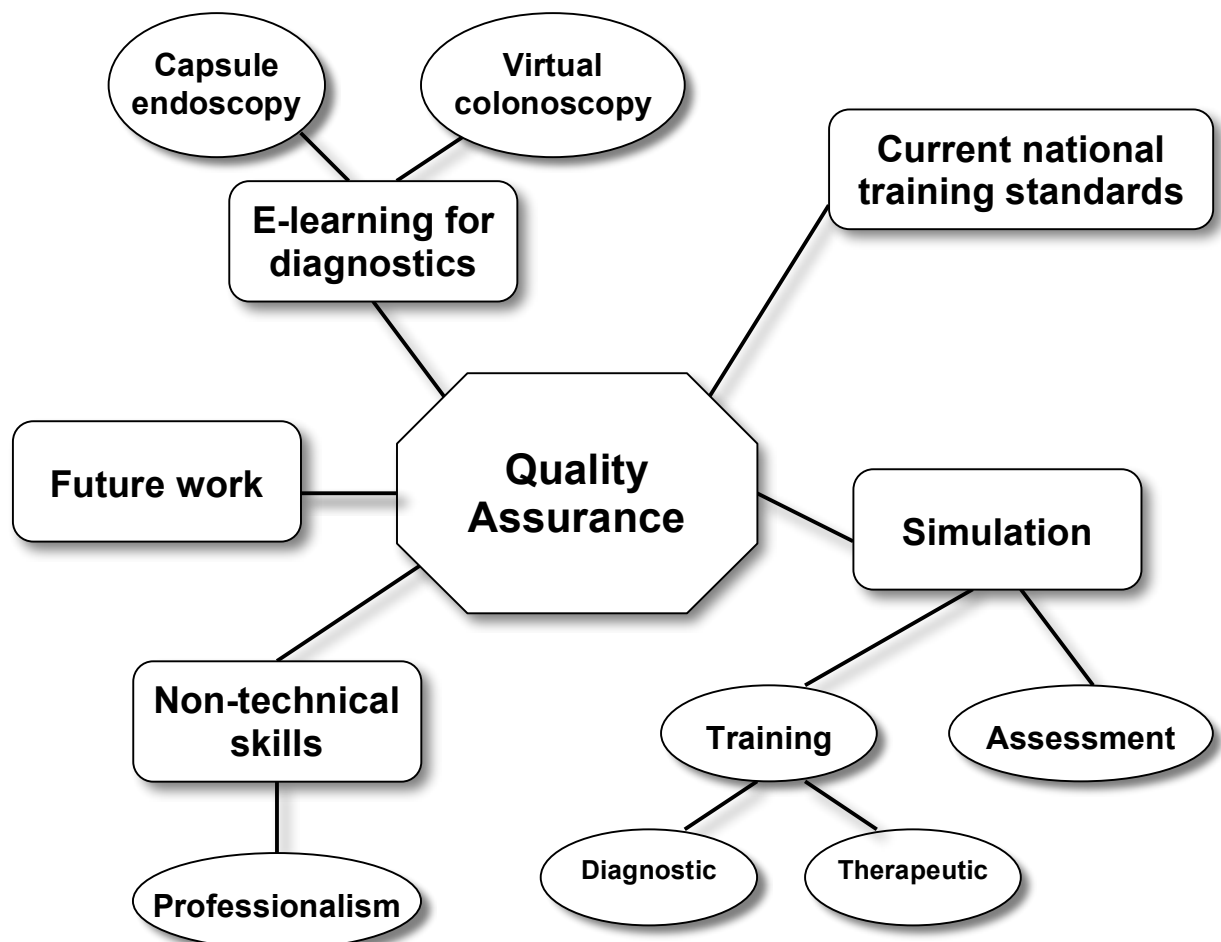
This study has described the development of an objective scoring system for endoscopic non-technical skills based on observable behavioural markers. This has the potential to widen the current training and assessment of endoscopy to allow a more complete evaluation and development of professionalism within this field. Further work needs to be done to refine and validate this tool.

Chapter 11: Discussion

11.1 Summary

The main hypothesis of this research was that application of the quality assurance process to training within gastrointestinal endoscopy can develop training methodologies and demonstrate improvements in learning outcomes and performance. A diagram of the discussion points in the following section is shown in Figure 37 below.

Figure 37 Overview of Discussion



The first set of projects described in Chapter 2 aimed to evaluate whether recent changes to the training methodologies within endoscopy in the UK have had a measurable effect on the current status of training. This was achieved by closing an audit loop on colonoscopy training within a single region, which provided the first quantitative evaluation of the changes to colonoscopy training over a 5 year period. This study demonstrated that the majority of issues previously identified as problems have significantly improved. There are still some areas of concern for trainees, particularly regarding service provision and the time allotted to endoscopy training. A national survey of all gastroenterology trainees in the UK about their current endoscopy training supported these findings, with high levels of satisfaction and greatly improved levels of supervision compared to 5 years previously, but with concerns about the ongoing effects of the reduction in hours for training.

The use of simulation in endoscopy training has the potential to offset some of the trainee's concerns regarding difficulties gaining access to high quality, properly supervised training. Simulators are becoming increasingly commonplace in medical training, and have great potential to provide repeatable, structured skills training with no risk to patients. They may also provide a platform for objective assessment of clinical skills, provided the simulation is a valid reproduction of the clinical task. The second set of projects aimed to evaluate a novel computer simulator for colonoscopy, the Olympus ENDO TS-1. Assessment of construct validity in Chapter 3 demonstrated that this 'second-generation' simulator has a high level of fidelity compared to real life colonoscopy, with 20 different metrics providing excellent construct validity. This was also the first study to show that a computer simulator can distinguish between novices, trainees of intermediate experience and experts. The advanced computer modelling that underlies the simulator's ability to more accurately replicate colonoscopy also allows many technical aspects of colonoscopy insertion to be benchmarked. It has the potential to be used for both formative and summative objective assessments of ability, which has previously not been possible due to the poor validity of other simulators. However, further work will need to be done to further validate it prior to widespread adoption as an assessment tool. Data presented in Chapter 5 does not support concurrent validity for the simulator, although the study was not specifically designed for this and the tools used to assess clinical performance are likely to lack sensitivity to discriminate between novices.

The multi-centre randomised controlled trial described in Chapter 4 demonstrates high utility of the simulator for teaching the necessary knowledge and skills for successful intubation at colonoscopy. The results showed that a dedicated simulator training programme resulted in significantly greater improvements in performance on simulated tasks than traditional patient-based training. It also found that simulator training produced equal performance outcomes on real-life cases, demonstrating a high degree of skills transfer from the simulator to real colonoscopy. This study provides the most robust evidence for the use of simulation in colonoscopy training in the literature, prospectively comparing a large number of trainees against standard training rather than no training.

One of the limitations of this study was that it used novices and focused on the early part of the learning curve. However, quality assurance of other areas of simulation require assessment of trainees with more ability and experience. One such area is training in therapeutic endoscopy. It can be difficult for trainees to gain ‘on-the-job’ experience in such procedures, and exposing patients to inexperienced practitioners may actually cause harm. Training in therapeutics using simulators may therefore have huge potential advantages. The randomised controlled trial on therapeutic training described in Chapter 6 demonstrated the efficacy of this approach for three of the four procedures investigated; control of haemorrhage, polypectomy and stricture dilatation. The training effect on procedural skills was measurable after a short period of one-to-one hands-on instruction, whereas no effect from knowledge-based training could be demonstrated. Training in the fourth procedure, insertion of a PEG tube, was shown to have no significant difference in skills improvement from controls, which may be due to the relative simplicity and two-person nature of the procedure. Such quality assurance data are vital as both diagnostic and therapeutic procedures become more complex and the time for training is reduced. The need to establish competency of practitioners requires continued evaluation of training outcomes to ensure the safety of patients.

Capsule endoscopy is a new diagnostic technique for gastrointestinal pathology with rapidly expanding indications and utility, but with no formalised training recommendations or validated assessment methods. Chapter 7 described the development, validation and initial pilot of web-based training and assessment in

lesion recognition at capsule endoscopy. The assessment module demonstrated good construct validity, successfully discriminating between novices, trainees and experts. The training module demonstrated content validity, with consistently significant performance improvements following training, irrespective of level of previous experience and under 'real-life' learning conditions. This study has established that validated web-based e-learning can be an effective tool for improving lesion recognition skills and allow rapid dissemination of quality training material and best practice to a wide audience. This approach could potentially help to address the current issues surrounding provision of both capsule endoscopy training and lesion recognition skills in endoscopy as a whole.

Chapter 8 described the evaluation of a training course for another new diagnostic technique for colonic pathology, virtual colonoscopy. This CT-based procedure is now replacing barium enema as the radiological test of choice, and like capsule endoscopy, has no formalised training or accreditation process. The course set out to define standards for training radiographers in technique and interpretation of virtual colonoscopy examinations. Accreditation of radiologists to provide initial reviews may potentially expedite the patient pathway for colorectal cancer and improve diagnostic efficiency. The study demonstrated the effectiveness of the course in providing the initial foundations of core knowledge and skill, although further training and experience would be required for competency.

The final two chapters described the initial attempt to expand the current training and assessment paradigm from merely knowledge and technical skill to include those attitudes, non-technical skills and behaviours that underlie professionalism in gastrointestinal endoscopy. Chapter 9 described the development of a non-technical skills taxonomy using qualitative techniques including critical incident analysis, sorting and rating tasks. These skills form the framework underlying the behavioural marker scoring system described in Chapter 10, which provides an objective, observable means of training and assessing non-technical skills. It can potentially be used alongside a knowledge and technical skills assessment to establish a more complete evaluation of professional behaviour within gastrointestinal endoscopy.

11.2 Quality assurance process

When applied to educational activities, the quality assurance process aims to help understand the training process and therefore improve the outcomes. However, this may require repeated cycles of evaluation as part of a continuous educational evaluation strategy. This thesis has focused on achieving high-quality educational evaluation at multiple levels, but at the expense of obtaining 'snapshot' results rather than longer term outcome data. This type of long-term follow-up data represents the highest level of the Kirkpatrick evaluation pyramid, 'Results', which is ultimately the purpose of training in medicine; good clinical outcomes for patients. This thesis has, however, been able to use the quality assurance process to demonstrate successful training outcomes at other levels, demonstrating that intensive specialist training can provide individuals with the necessary knowledge and skills for safe technical performance. It has also taken the first step towards the incorporation of attitudes, non-technical skills and professionalism into gastrointestinal endoscopy training and assessment, which should now be evaluated using the same quality assurance process.

11.3 Simulation in endoscopy

This thesis has provided evidence for the effectiveness of several types of simulation training within gastrointestinal endoscopy. However, there are inherent issues with the use of simulators that should be addressed before incorporation into any training programme. Simulation is, by definition, not real and there is potential to learn bad habits or not to learn at all if adequate feedback is not provided as part of the training process[217]. The training programmes developed as part of the Olympus simulator evaluation and the capsule endoscopy lesion recognition module did incorporate some degree of automatic feedback. However, it is unlikely that this will ever match the quality of feedback given by an expert trainer, which should form part of the learning process. However, the advantages that simulation can bring in terms of repeatability, learner-centredness and allowance for errors probably outweighs this consideration.

During simulation training, particularly when considering computer simulation, there is also potential to focus purely on technical skill, with a resulting lack of non-

technical skills development. The non-technical skills taxonomy and behavioural marker systems developed in this thesis now make it possible to rectify this omission. It is possible to combine high-fidelity simulators, real patients and a convincing simulated environment to produce contextualised simulation or 'rehearsal'[317]. This may be a safe way of providing training which can reflect the uncertainties and complexity of real-world clinical situations, and may therefore avoid the narrow focus on technical skill.

The place of simulation in a training programme should be carefully considered. It may be useful for novice or inexperienced practitioners to gain initial skills. It also clearly has a place for training in technically difficult, rare or complex cases such as therapeutics. It may be used for deliberate practice by a reasonably experienced individual in order to attain mastery of a particular skill. It can also be combined with real or simulated patients to elicit the full range of technical and non-technical behaviours[318]. Each situation may require a different mix of individual and instructor time and effort to maximise learning. Part of the quality assurance process must therefore focus on such considerations to determine how to gain maximum educational value from the use of simulation in training.

The use of simulation for assessment still remains controversial. Most simulators for endoscopy demonstrate low construct validity, which effectively precludes their use as assessment tools. However, higher-fidelity simulators such as the Olympus colonoscopy simulator and the ex-vivo models described in this thesis may have a role, certainly as low-stakes formative assessment tools. The data presented here suggests that they are potentially more discriminatory for trainees lower down the learning curve compared to current clinical assessment tools, which aim to determine competency. The possibility of using them for high-stakes summative assessment is attractive, as it would provide an objective, repeatable, standardised method of ascertaining competence. It is important to note that this would not have to be divorced from the current live-case, subjective assessments, but could form part of an accreditation process including both types to support the validity and robustness of the process. Clinical audit could then be used to ensure that standards are met over time. The use of simulation can potentially also combine technical and non-technical skills to enable an integrated assessment of performance. This would

be a remarkable step towards embracing the idea that endoscopists are not merely technicians, but professionals who require many other skills to perform at an expert level.

11.4 E-learning for diagnostic procedures

Web-based learning is now becoming accepted as a valid educational method of postgraduate training in a wide range of medical specialities[319]. It has been shown to be at least as effective at knowledge improvement as traditional methods in occupational medicine[320] and ambulatory medicine[321]. Non or minimally invasive tests such as capsule endoscopy and virtual colonoscopy provide an ideal forum for utilising this new training methodology, as there is little technical skills requirement. Such tests are also inherently more acceptable to patients, and may therefore eventually exceed the demand for more invasive flexible endoscopic procedures for diagnostic purposes.

This thesis reports data supporting improved lesion recognition skills for capsule endoscopy. However, the level of improvement was relatively small for the 'real-life' web-based learning, with mean performance below expert levels even after training. The assessment module was also not representative of the real-life scenario, using 40 second video clips compared to the 6 hour videos that are reviewed in reality. The module could potentially include several long case (15-20 minute) assessments for more accurate simulation of real-life reporting, and could ultimately form part of an accreditation process for CE privileges in the UK. However, robust data are currently lacking in terms of improvements in patient outcomes from the training, and the quality assurance process would have to be applied to any new assessment modules that are developed. The principle, however, remains sound and other modules in lesion recognition at flexible endoscopy are now being developed by national societies such as JAG, potentially providing more widespread access to high quality training material than is currently available.

This thesis also reports improvements in both knowledge and skill in performance and interpretation of virtual colonoscopy. This project extended the application of the quality assurance process to an entirely different approach to diagnostics within the GI tract. It also confirms the utility of this approach towards training in a

multidisciplinary fashion. The positive results from evaluating a course incorporating both radiologists and radiographers provides robust evidence to support the hypothesis that well-constructed educational evaluation can quality assure training outcomes.

11.5 Attitudes, non-technical skills and professionalism

Non-technical skills taxonomies and behavioural marker systems are a relatively recent innovation to the medical field. As yet there is no requirement for their implementation in either training or assessment, and current specialist trainees are unlikely to have had any experience of them. However, they have many potential uses, from providing a framework for reflection of day-to-day problems through to high-fidelity crisis-management simulation and teamwork training. Current medical students and Foundation Year doctors in London now receive non-technical skills training as part of their educational programme. It is likely that there will be an increase in this type of training with the development of flagship projects like the Simulation and Technology-enhanced Learning Initiative (STeLI) funded by the NHS and the London Deanery (www.simulation.londondeanery.ac.uk).

Both non-technical skills and behavioural markers are integral to the concept of professionalism. There is now an increasing awareness of the need to teach professionalism explicitly to ensure that the doctors of the future avoid some of the failures of the profession to meet legitimate expectations of society. The growing literature base emphasises the sociological and bioethical aspects of professionalism, indicating the importance of the professional relationships of doctors with both patients and society. Teaching programmes must address both issues, which is likely to require multidisciplinary involvement from all medical teaching staff. This may include classroom teaching with medical ethics support, assessment of interactions with patients, their families and other workers involved in the delivery of healthcare.

The assessment and evaluation of attitudinal and judgement domains has been recognised by the English national training and endoscopy leads as important for the continued development of competency-based assessment for gastrointestinal

endoscopy[322]. The taxonomy and behavioural marker system developed as a result of this research are planned to be used initially to support the accreditation system for the national bowel cancer screening programme. If this proves successful, it could then provide the basis for training in endoscopic professionalism and also formative and summative assessment for all endoscopists and trainees.

11.6 Future work

The main conclusion from this thesis is that application of the quality assurance process to training and assessment in endoscopy can have a beneficial effect on the both the learning process and outcomes. The corollary to this is that this process should be applied across the board in order to quality assure training in all aspects of endoscopy, not just those focused on for this thesis. However, there are a number of specific issues raised by this work that should be addressed in future research.

As competing technologies such as capsule and virtual colonoscopy become more widespread for diagnosis of pathology, flexible endoscopy is likely to become more focused on provision of therapy. There is clearly a need to improve both the training and assessment of therapeutic endoscopy. While this thesis has presented data regarding the training and assessment of basic polypectomy, the ex-vivo model used is currently fairly basic and the assessment tools demonstrate only moderate interrater reliability. Work is currently in progress to develop better models that more realistically simulate both flat, sessile and pedunculated polyps in order to expand the training potential to include endoscopic mucosal resection (EMR) and piecemeal polypectomy. Construct validity of these models will need to be demonstrated to ensure their efficacy in training. In order to do this, the assessment tools will need to be refined and validated in order to ensure that they can reliably discriminate between trainees. A Directly Observed Polypectomy Skills (DOPyS) tool is currently under development, with the aim of stratifying polypectomy into skill levels in order to provide better training and further define the competency assessment criteria. More advanced therapeutic techniques such as stent insertion, variceal banding and NOTES will also require the same development and evaluation to ensure training efficacy as the boundaries of the medical/surgical divide become increasingly blurred.

The impact of the Train the Trainers courses has been perceived by trainees to have improved the quality of their training at their base hospitals. However, there is likely to be value in applying the same quality assurance process to these courses as to those providing technical skills training. Indeed, this may be more important, as improvements in technical skill are relatively easy to observe, whereas changes in an individual's teaching skill or their training methodology may be much more difficult to ascertain. Evaluation at multiple levels will be required to assess the impact of these courses and to provide insights into the learning process and hence improve the training. The recent body of research by Wells et al has explored this topic in detail and may provide the basis for evaluation standards in the future.

The use of e-learning clearly has the potential to revolutionise some aspects of training in endoscopy by providing rapid and widespread availability to learning material. However, it is not the case that traditional teaching material can be just transported into a multimedia environment, as learning is likely to occur in a different fashion. Despite the recent explosion of online learning material available, high-quality web-based training is actually quite rare. New material will have to be developed using sound educational principles and appropriate evaluation conducted to establish the efficacy of individual modules. There is now a national initiative, E-learning for Healthcare (www.e-lfh.org.uk) which aims to provide quality assured free online training content for the healthcare profession in the UK. It incorporates traditional proven teaching models with state of the art e-learning techniques and technology, including validated e-assessments. The endoscopy module will consist of ten modules, including advanced therapeutic endoscopy and a trainers section.

This thesis has presented the first step towards incorporation of non-technical skills into the training and assessment of gastrointestinal endoscopy. The provisional behavioural marker system described here seems to meet the criteria set by Flin and Martin (2001) regarding design and development. However, a significant amount of work remains to be done to assess its validity and reliability. Studies to investigate construct and content validity will be needed to ensure that the tool is accurate and to check that no vital skills have been omitted and that the skills present can be observed. Assessments of its utility in realistic situations including both routine and emergency endoscopy are required. Using observational tools is likely to require

training, as has been recognised by other behavioural marker systems such as ANTS and NOTSS. This should include provision of background knowledge on human performance, error management, and the principles of using psychometric tools for performance rating to potential assessors. Subsequent investigation of interrater reliability will be needed to ensure it can be used reliably, consistently and accurately by different endoscopists scoring the equivalent performances. The combination of this tool with technical skills DOPS provides the exciting potential for an integrated assessment of performance, providing a much more complete evaluation of professionalism within endoscopy than is currently possible.

Gastrointestinal endoscopy involves a complex set of knowledge, skills and attitudes. Enormous change has occurred in the development of training in the UK, but little done to evaluate the effectiveness. Processes should be put in place to establish QA outcomes as the technology develops to ensure that the workforce can keep up with the technological advances.

Bibliography

1. Office for National Statistics. Cancer survival 1992-1999. 2001
2. Bowles CJ, Leicester R, Romaya C, et al. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow? *Gut* 2004;53:277-283
3. Sharma VK, Coppola AG, Jr., Raufman JP. A survey of credentialing practices of gastrointestinal endoscopy centers in the United States. *Journal of clinical gastroenterology* 2005;39:501-507
4. Cullinane M, Gray AJG, Hargraves CMK, et al. Scoping Our Practice. In: National Confidential Enquiry into Patient Outcome and Death; 2004
5. Teague R, Soehendra N, Carr-Locke D, et al. Setting standards for colonoscopic teaching and training. *Journal of gastroenterology and hepatology* 2002;17 Suppl:S50-53
6. Balfour TW. Training for colonoscopy. *Journal of the Royal Society of Medicine* 2001;94:160-161
7. Cass OW. Training to competence in gastrointestinal endoscopy: a plea for continuous measuring of objective end points. *Endoscopy* 1999;31:751-754
8. Bond JH, Frakes JT. Who should perform colonoscopy? How much training is needed? *Gastrointest Endosc* 1999;49:657-659
9. Pearl JP, Marks JM. The future of teaching surgical endoscopy. *Surgical innovation* 2006;13:280-282
10. Bisschops R, Wilmer A, Tack J. A survey on gastroenterology training in Europe. *Gut* 2002;50:724-729
11. Cass OW. Objective evaluation of competence: technical skills in gastrointestinal endoscopy. *Endoscopy* 1995;27:86-89
12. Wexner SD, Garbus JE, Singh JJ. A prospective analysis of 13,580 colonoscopies. Reevaluation of credentialing guidelines. *Surg Endosc* 2001;15:251-261
13. Thomas-Gibson S, Saunders BP. Colonoscopy training in the UK, the Trainees' Perspective. *CME Journal Gastroenterology, Hepatology and Nutrition* 2004;6:44-47
14. Freeman ML. Training and competence in gastrointestinal endoscopy. *Rev Gastroenterol Disord* 2001;1:73-86

Bibliography

15. Viiala CH, Zimmerman M, Cullen DJ, Hoffman NE. Complication rates of colonoscopy in an Australian teaching hospital environment. *Internal medicine journal* 2003;33:355-359
16. Bennett JR. Prospective audit of upper gastrointestinal endoscopy in two regions of England: safety, staffing, and sedation methods. *Gastrointest Endosc* 1995;42:383-384
17. Fasih T, Varma JS, Tabaqchali MA. Prospective audit of quality of colonoscopy in a surgical coloproctology unit. *Surgeon* 2004;2:107-111
18. Imperiali G, Minoli G, Meucci GM, et al. Effectiveness of a continuous quality improvement program on colonoscopy practice. *Endoscopy* 2007;39:314-318
19. Calman K. Hospital doctors: training for the future. Report of the working group on specialist medical training [Calman Report]. Heywood Heath: Department of Health Working Group on Specialist Medical Training; 1993
20. Aojanpong T. The new medical training program in UK: modernising medical careers. *Journal of the Medical Association of Thailand = Chotmai het thangphaet* 2005;88:1741-1744
21. Delamothe T. Modernising Medical Careers: final report. *BMJ Clinical research ed* 2008;336:54-55
22. Chikwe J, de Souza AC, Pepper JR. No time to train the surgeons. *BMJ (Clinical research ed)* 2004;328:418-419
23. Standards for curricula and assessment systems - July 2008. Postgraduate Medical Education and Training Board. Available from <http://www.pmetb.org.uk/index.php?id=scas>; Accessed August 2009
24. Aggarwal R, Hance J, Darzi A. Surgical education and training in the new millennium. *Surg Endosc* 2004;18:1409-1410
25. HMSO. The NHS Cancer Plan. 2000
26. Sant M, Capocaccia R, Coleman MP, et al. Cancer survival increases in Europe, but international differences remain wide. *Eur J Cancer* 2001;37:1659-1667
27. HMSO. Department of Health Modernisation Agency Meeting The Challenges. 2004
28. Weller D, Coleman D, Robertson R, et al. The UK colorectal cancer screening pilot: results of the second round of screening in England. *British journal of cancer* 2007;97:1601-1605
29. Goodyear SJ, Stallard N, Gaunt A, et al. Local impact of the English arm of the UK Bowel Cancer Screening Pilot study. *The British journal of surgery* 2008;95:1172-1179

30. Improving Endoscopy Services. Meeting the Challenges. NHS Modernisation Agency 2004
31. Improving Endoscopy Services. Meeting the Challenges. Working towards a sustainable future. NHS Modernisation Agency 2007
32. Joint Advisory Group for Gastrointestinal Endoscopy. Guidelines for the Training, Appraisal and Assessment of Trainees in Gastrointestinal Endoscopy. 2004
33. Hussein AM, Bartram CI, Williams CB. Carbon dioxide insufflation for more comfortable colonoscopy. *Gastrointest Endosc* 1984;30:68-70
34. Thomas-Gibson S, Bassett P, Suzuki N, et al. Intensive training over 5 days improves colonoscopy skills long-term. *Endoscopy* 2007;39:818-824
35. JAG DOPS forms. Available from <http://www.thejag.org.uk/Forms/tabid/72/Default.aspx>. Accessed August 2009
36. Holman JR, Marshall RC, Jordan B, Vogelmann L. Technical competency in flexible sigmoidoscopy. *The Journal of the American Board of Family Practice / American Board of Family Practice* 2001;14:424-429
37. Peyton JWR. *Teaching and Learning in Medical Practice*: Manticore; 1998
38. Bloom B, Englehart MD, Furst EJ, Hill WH, Krathwohl D. *The Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook I: Cognitive Domain*; 1956
39. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med* 2004;79:S70-81
40. Patrick J. *Training: research and practice*. London: Academic Press; 1992
41. Miller GE. The assessment of clinical skills/competence/performance. *Acad Med* 1990;65:S63-67
42. Kirkpatrick DL. *Evaluating Training Programs: The Four Levels*. Second Edition ed: Berrett-Koehler Publishers; 1998
43. Lindner H, Probst P. Education and training in gastroenterological endoscopy--an educational program (author's transl). *Leber, Magen, Darm* 1974;4:177-180
44. Bowman MA, Wherry DC. Training for flexible sigmoidoscopy. *Gastrointestinal endoscopy* 1985;31:309-312
45. Hogan WJ. Uniform standards for gastrointestinal endoscopic training in the U.S.--a need for evaluation and definition. *Gastrointestinal endoscopy* 1988;34:362-363

Bibliography

46. Myren J. Education and training in gastrointestinal endoscopy. *The Italian journal of gastroenterology* 1991;23:461-462
47. Bond JH. Advanced endoscopic training: teaching us older dogs some new tricks. *Gastrointestinal endoscopy* 1992;38:631-633
48. Vennes JA, Ament M, Boyce HW, Jr., et al. Principles of training in gastrointestinal endoscopy. American Society for Gastrointestinal Endoscopy. Standards of Training Committees. 1989-1990. *Gastrointestinal endoscopy* 1992;38:743-746
49. Baillie J, Ravich WJ. On endoscopic training and procedural competence. *Annals of internal medicine* 1993;118:73-74
50. Cass O, Freeman ML, Cohen J, et al. Acquisition of competency in endoscopic skills (ACES) during training; a multicenter study. *Gastrointest Endosc* 1996;43:308
51. Cass OW, Freeman ML, Peine CJ, Zera RT, Onstad GR. Objective evaluation of endoscopy skills during training. *Annals of internal medicine* 1993;118:40-44
52. Rodney WM, Albers G. Flexible sigmoidoscopy: primary care outcomes after two types of continuing medical education. *The American journal of gastroenterology* 1986;81:133-137
53. Rodney WM, Ruggiero C. Outcomes following continuing medical education on flexible sigmoidoscopy. *Family practice* 1987;4:306-310
54. Lewan RB, Baskin WN, Sharon MW, et al. A supervised training program in flexible sigmoidoscopy: evaluating skills from residency training to clinical practice. *Family medicine* 1989;21:25-29
55. Duthie GS, Drew PJ, Hughes MA, et al. A UK training programme for nurse practitioner flexible sigmoidoscopy and a prospective evaluation of the practice of the first UK trained nurse flexible sigmoidoscopist. *Gut* 1998;43:711-714
56. Levine DF, Teague RH. A UK training programme for nurse practitioner flexible sigmoidoscopy. *Gut* 1999;44:897
57. Schoenfeld P. Flexible sigmoidoscopy by paramedical personnel. *Journal of clinical gastroenterology* 1999;28:110-116
58. Levin TR, Farraye FA, Schoen RE, et al. Quality in the technical performance of screening flexible sigmoidoscopy: recommendations of an international multi-society task group. *Gut* 2005;54:807-813
59. Guidelines for training non-specialists in screening flexible sigmoidoscopy. *Gastrointest Endosc* 2000;51:783-785

60. Atkin WS, Edwards R, Wardle J, et al. Design of a multicentre randomised trial to evaluate flexible sigmoidoscopy in colorectal cancer screening. *Journal of medical screening* 2001;8:137-144
61. Atkin W, Rogers P, Cardwell C, et al. Wide variation in adenoma detection rates at screening flexible sigmoidoscopy. *Gastroenterology* 2004;126:1247-1256
62. Parry BR, Williams SM. Competency and the colonoscopist: a learning curve. *Aust N Z J Surg* 1991;61:419-422
63. Marshall JB. Technical proficiency of trainees performing colonoscopy: a learning curve. *Gastrointest Endosc* 1995;42:287-291
64. Tassios PS, Ladas SD, Grammenos I, Demertzis K, Raptis SA. Acquisition of competence in colonoscopy: the learning curve of trainees. *Endoscopy* 1999;31:702-706
65. Dafnis G, Granath F, Pahlman L, et al. The impact of endoscopists' experience and learning curves and interendoscopist variation on colonoscopy completion rates. *Endoscopy* 2001;33:511-517
66. Kruglikova I, Grantcharov TP, Drewes AM, Funch-Jensen P. Assessment of early learning curves among nurses and physicians using a high-fidelity virtual-reality colonoscopy simulator. *Surg Endosc* 2009
67. Church J, Oakley J, Milsom J, Strong S, Hull T. Colonoscopy training: the need for patience (patients). *ANZ journal of surgery* 2002;72:89-91
68. Wexner SD, Eisen GM, Simmang C. Principles of privileging and credentialing for endoscopy and colonoscopy. *Surg Endosc* 2002;16:367-369
69. Thuler FP, Di Sena VO, Hervoso CM, Della Libera E, Ferrari AP. [How many procedures are necessary to achieve competency in colonoscopy?]. *Arq Gastroenterol* 2004;41:211-214
70. Lee SH, Chung IK, Kim SJ, et al. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. *Gastrointest Endosc* 2008;67:683-689
71. Spier BJ, Benson M, Pfau PR, et al. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc* 2009
72. Rex DK. Colonoscopic withdrawal technique is associated with adenoma miss rates. *Gastrointest Endosc* 2000;51:33-36
73. Bretagne JF, Ponchon T. Do we need to embrace adenoma detection rate as the main quality control parameter during colonoscopy? *Endoscopy* 2008;40:523-528

Bibliography

74. Gorard DA, McIntyre AS. Completion rate to caecum as a quality measure of colonoscopy in a district general hospital. *Colorectal Dis* 2004;6:243-249
75. Chak A, Cooper GS, Blades EW, Canto M, Sivak MV, Jr. Prospective assessment of colonoscopic intubation skills in trainees. *Gastrointest Endosc* 1996;44:54-57
76. Cullado MJ, Porter JA, Slezak FA. The evolution of surgical endoscopic training. Meeting the American Board of Surgery requirements. *The American surgeon* 1991;57:250-253
77. Larson GM, Mullins RJ, Wieman TJ, Polk HC, Jr. Evaluation of endoscopy training in a general surgery residency. *The American surgeon* 1988;54:64-67
78. Thomas-Gibson S, Saunders BP. Development and validation of a multiple-choice question paper in basic colonoscopy. *Endoscopy* 2005;37:821-826
79. Thomas-Gibson S, Rogers PA, Suzuki N, et al. Development of a video assessment scoring method to determine the accuracy of endoscopist performance at screening flexible sigmoidoscopy. *Endoscopy* 2006;38:218-225
80. ASGE guidelines for clinical application. Statement on role of short courses in endoscopic training. American Society for Gastrointestinal Endoscopy. *Gastrointest Endosc* 1999;50:913-914
81. Cosgrove JM, Margolis IB, Riou JP, Wait RB. Formal endoscopy training for senior and junior house staff. *Journal of laparoendoscopic surgery* 1993;3:525-529
82. Max MH, Polk HC, Jr. Perceived needs for gastrointestinal endoscopic training in surgical residencies. *American journal of surgery* 1982;143:150-154
83. Neumann M, Friedl S, Meining A, et al. A score card for upper GI endoscopy: Evaluation of interobserver variability in examiners with various levels of experience. *Zeitschrift fur Gastroenterologie* 2002;40:857-862
84. Neumann M, Siebert T, Rausch J, et al. Scorecard endoscopy: a pilot study to assess basic skills in trainees for upper gastrointestinal endoscopy. *Langenbecks Arch Surg* 2003;387:386-391
85. Neumann M, Meining A, Buerschaper C, et al. Training in GI endoscopy: can we objectively predict the performance of trainees? A prospective correlation study. *Zeitschrift fur Gastroenterologie* 2005;43:445-450
86. Cotton PB. How many times have you done this procedure, doctor? *Am J Gastroenterol* 2002;97:522-523
87. Hellier MD, Morris AI. ERCP training--time for change. *Gut* 2000;47:459-460
88. Baillie J. ERCP training: for the few, not for all. *Gut* 1999;45:9-10

89. Jenkins JT, Glass G, Ballantyne S, Fullarton GM. Effect of MRCP introduction on ERCP practice: are there implications for service and training? *Gut* 2006;55:1365-1366
90. Verma D, Gostout CJ, Petersen BT, et al. Establishing a true assessment of endoscopic competence in ERCP during training and beyond: a single-operator learning curve for deep biliary cannulation in patients with native papillary anatomy. *Gastrointest Endosc* 2007;65:394-400
91. Lowe JB, Page CP, Schwesinger WH, Gaskill HV, Stauffer JS. Percutaneous endoscopic gastrostomy tube placement in a surgical training program. *American journal of surgery* 1997;174:624-627; discussion 627-628
92. Atkinson RJ, Donnelly MT. Oesophageal Stenting: Trends in training in the UK. *Endoscopy* 2006;38:A162
93. Waye JD. Techniques of colonoscopy, hot biopsy forceps, and snare polypectomy. *Progress in clinical and biological research* 1988;279:61-69
94. Waye JD. Saline injection colonoscopic polypectomy. *Am J Gastroenterol* 1994;89:305-306
95. Waye JD. New methods of polypectomy. *Gastrointestinal endoscopy clinics of North America* 1997;7:413-422
96. Zlatanich J, Waye JD, Kim PS, Baiocco PJ, Gleim GW. Large sessile colonic adenomas: use of argon plasma coagulator to supplement piecemeal snare polypectomy. *Gastrointest Endosc* 1999;49:731-735
97. Williams CB, Saunders BP, Talbot IC. Endoscopic management of polypoid early colon cancer. *World journal of surgery* 2000;24:1047-1051
98. Miller K, Waye JD. Polyp retrieval after colonoscopic polypectomy: use of the Roth Retrieval Net. *Gastrointest Endosc* 2001;54:505-507
99. Rutter M, Saunders B, Emmanuel A, Price A. Endoscopic snare polypectomy for bleeding postinflammatory polyps. *Endoscopy* 2003;35:788-790
100. Fraser C, Saunders B. Preventing postpolypectomy bleeding: obligatory and optional steps. *Endoscopy* 2004;36:898-900
101. Waye J. It ain't over 'til it's over: retrieval of polyps after colonoscopic polypectomy. *Gastrointest Endosc* 2005;62:257-259
102. Waye JD. Endoscopic mucosal resection of colon polyps. *Gastrointestinal endoscopy clinics of North America* 2001;11:537-548, vii
103. Waye JD. Advanced polypectomy. *Gastrointestinal endoscopy clinics of North America* 2005;15:733-756

Bibliography

104. Rubin PH, Waye JD. Colonoscopic polypectomy: a critical review of recent literature. *Current gastroenterology reports* 2006;8:430-433
105. Haycock A, Suzuki N, Thomas-Gibson S. Development of a novel video scoring system for basic polypectomy: an initial validation study. *Gastrointest Endosc* 2007;65:AB329
106. Youd P, Haycock A, Preston S, Thomas-Gibson S. An interactive polyp detection training tool: development and initial validation. *Gut* 2008;57(Suppl 1):A044
107. Speer AG, Eaves ER, Gijsbers AJ, St John DJ. Training in sigmoidoscopy. An assessment in three teaching hospitals. *The Medical journal of Australia* 1984;141:223-225
108. Thuraisingam A, Shenoy AH, Anderson JT. What are the key components of hands-on colonoscopy training? *Gut* 2007;56:A103
109. Thuraisingam AI, MacDonald J, Shaw IS. Insights into endoscopy training: a qualitative study of learning experience. *Medical teacher* 2006;28:453-459
110. Harden RM, Crosby J. AMEE Guide No 20: The good teacher is more than a lecturer-the twelve roles of the teacher. *Medical teacher* 2000;22:334-347
111. Wells CW, Corbett S, Welfare MR, Barton JR. A model of teaching endoscopy. *Gut* 2008;57(Suppl 1):A272
112. Galandiuk S, Ahmad P. Impact of sedation and resident teaching on complications of colonoscopy. *Digestive surgery* 1998;15:60-63
113. Hernandez LV, Jacobson JW, Harris MS. Comparison among the perforation rates of Maloney, balloon, and savary dilation of esophageal strictures. *Gastrointest Endosc* 2000;51:460-462
114. Anderson PE, Cook A, Amery AH. A review of the practice of fiberoptic endoscopic dilatation of oesophageal stricture. *Ann R Coll Surg Engl* 1989;71:124-127
115. Riley SA, Attwood SE. Guidelines on the use of oesophageal dilatation in clinical practice. *Gut* 2004;53 Suppl 1:i1-6
116. Cohen J, Nuckolls L, Maurant RR. Endoscopy simulators: lessons from the aviation and automobile industries. *Gastrointestinal endoscopy clinics of North America* 2006;16:407-423
117. Axon AT, Aabakken L, Malfertheiner P, et al. Recommendations of the ESGE workshop on Ethics in Teaching and Learning Endoscopy. First European Symposium on Ethics in Gastroenterology and Digestive Endoscopy, Kos, Greece, June 2003. *Endoscopy* 2003;35:761-764

118. Wigton RS, Nicolas JA, Blank LL. Procedural skills of the general internist. A survey of 2500 physicians. *Annals of internal medicine* 1989;111:1023-1034
119. Markman HD. A new system for teaching proctosigmoidoscopic morphology. *Am J Gastroenterol* 1969;52:65-69
120. Classen M, Ruppin H. Practical training using a new gastrointestinal phantom. *Endoscopy* 1974;6:127-131
121. Williams CB, Saunders BP, Bladen JS. Development of colonoscopy teaching simulation. *Endoscopy* 2000;32:901-905
122. Leung JW, Chung RS. Training in ERCP. *Gastrointest Endosc* 1992;38:517-518
123. Barde CJ. Endoscopic simulation for training. *Gastrointestinal endoscopy* 1989;35:462-463
124. Gholson CF, Provenza JM, Silver RC, Bacon BR. Endoscopic retrograde cholangiography in the swine: a new model for endoscopic training and hepatobiliary research. *Gastrointestinal endoscopy* 1990;36:600-603
125. Williams CB, Baillie J, Gillies DF, Borislow D, Cotton PB. Teaching gastrointestinal endoscopy by computer simulation: a prototype for colonoscopy and ERCP. *Gastrointestinal endoscopy* 1990;36:49-54
126. Baillie J, Jowell P, Evangelou H, Bickel W, Cotton P. Use of computer graphics simulation for teaching of flexible sigmoidoscopy. *Endoscopy* 1991;23:126-129
127. Baillie J, Jowell P, Evangelou H, Bickel W, Cotton P. Teaching by endoscopy simulation. *Endoscopy* 1991;23:239-240
128. Sedlack RE, Baron TH, Downing SM, Schwartz AJ. Validation of a colonoscopy simulation model for skills assessment. *Am J Gastroenterol* 2007;102:64-74
129. Matthes K, Cohen J. The Neo-Papilla: a new modification of porcine ex vivo simulators for ERCP training (with videos). *Gastrointest Endosc* 2006;64:570-576
130. Adamsen S. Simulators and gastrointestinal endoscopy training. *Endoscopy* 2000;32:895-897
131. Bar-Meir S. A new endoscopic simulator. *Endoscopy* 2000;32:898-900
132. Bar-Meir S. Endoscopy simulators: the state of the art, 2000. *Gastrointest Endosc* 2000;52:701-703
133. Bittner JGt, Marks JM, Dunkin BJ, et al. Resident training in flexible gastrointestinal endoscopy: a review of current issues and options. *Journal of surgical education* 2007;64:399-409

Bibliography

134. Cisler JJ, Martin JA. Logistical considerations for endoscopy simulators. *Gastrointestinal endoscopy clinics of North America* 2006;16:565-575
135. Dunkin B, Adrales GL, Apelgren K, Mellinger JD. Surgical simulation: a current review. *Surg Endosc* 2007;21:357-366
136. Dunkin BJ. Flexible endoscopy simulators. *Semin Laparosc Surg* 2003;10:29-35
137. Fitzgerald TN, Duffy AJ, Bell RL, et al. Computer-based endoscopy simulation: emerging roles in teaching and professional skills assessment. *Journal of surgical education* 2008;65:229-235
138. Gerson LB. Can colonoscopy simulators enhance the learning curve for trainees? *Gastrointest Endosc* 2006;64:369-374
139. Gerson LB. Evidence-based assessment of endoscopic simulators for training. *Gastrointestinal endoscopy clinics of North America* 2006;16:489-509, vii-viii
140. Gerson LB, Van Dam J. The future of simulators in GI endoscopy: an unlikely possibility or a virtual reality? *Gastrointest Endosc* 2002;55:608-611
141. Gerson LB, Van Dam J. Technology review: the use of simulators for training in GI endoscopy. *Gastrointest Endosc* 2004;60:992-1001
142. Greenwald D, Cohen J. Evolution of endoscopy simulators and their application. *Gastrointestinal endoscopy clinics of North America* 2006;16:389-406
143. Hochberger J, Maiss J. Currently available simulators: ex vivo models. *Gastrointestinal endoscopy clinics of North America* 2006;16:435-449
144. Hochberger J, Maiss J, Hahn EG. The use of simulators for training in GI endoscopy. *Endoscopy* 2002;34:727-729
145. Hochberger J, Maiss J, Magdeburg B, Cohen J, Hahn EG. Training simulators and education in gastrointestinal endoscopy: current status and perspectives in 2001. *Endoscopy* 2001;33:541-549
146. Kefalides PT, Gress F. Simulator training for endoscopic ultrasound. *Gastrointestinal endoscopy clinics of North America* 2006;16:543-552, viii
147. Lucero RS, Zarate JO, Espiniella F, et al. Introducing digestive endoscopy with the "SimPrac-EDF y VEE" simulator, other organ models, and mannequins: teaching experience in 21 courses attended by 422 physicians. *Endoscopy* 1995;27:93-100
148. Matthes K. Simulator training in endoscopic hemostasis. *Gastrointestinal endoscopy clinics of North America* 2006;16:511-527, viii

149. Nelson DB, Bosco JJ, Curtis WD, et al. Technology status evaluation report: endoscopy simulators: May 1999. *Gastrointest Endosc* 2000;51:790-792
150. Sedlack RE. Endoscopic simulation: where we have been and where we are going. *Gastrointest Endosc* 2005;61:216-218
151. Sedlack RE. Simulators in training: defining the optimal role for various simulation models in the training environment. *Gastrointestinal endoscopy clinics of North America* 2006;16:553-563
152. Soehendra N, Binmoeller KF. Overview of interactive endoscopy simulators. *Endoscopy* 1992;24 Suppl 2:549-550
153. Sturm LP, Windsor JA, Cosman PH, et al. A systematic review of skills transfer after surgical simulation training. *Annals of surgery* 2008;248:166-179
154. Trifan A, Stanciu C. Computer-based simulator for training in gastrointestinal endoscopy. *Revista medico-chirurgicala a Societatii de Medici si Naturalisti din Iasi* 2007;111:567-574
155. Wagh MS, Waxman I. Animal models for endoscopic simulation. *Gastrointestinal endoscopy clinics of North America* 2006;16:451-456
156. Saunders BP, Bell GD, Williams CB, Bladen JS, Anderson AP. First clinical results with a real time, electronic imager as an aid to colonoscopy. *Gut* 1995;36:913-917
157. Shah SG, Saunders BP, Brooker JC, Williams CB. Magnetic imaging of colonoscopy: an audit of looping, accuracy and ancillary maneuvers. *Gastrointest Endosc* 2000;52:1-8
158. Hoff G, Bretthauer M, Dahler S, et al. Improvement in caecal intubation rate and pain reduction by using 3-dimensional magnetic imaging for unsedated colonoscopy: A randomized trial of patients referred for colonoscopy. *Scandinavian journal of gastroenterology* 2007;42:885-889
159. Shah SG, Brooker JC, Williams CB, Thapar C, Saunders BP. Effect of magnetic endoscope imaging on colonoscopy performance: a randomised controlled trial. *Lancet* 2000;356:1718-1722
160. Shah SG, Thomas-Gibson S, Lockett M, et al. Effect of real-time magnetic endoscope imaging on the teaching and acquisition of colonoscopy skills: results from a single trainee. *Endoscopy* 2003;35:421-425
161. Swain P. Wireless capsule endoscopy. *Gut* 2003;52 Suppl 4:iv48-50
162. Iddan G, Meron G, Glukhovsky A, Swain P. Wireless capsule endoscopy. *Nature* 2000;405:417

Bibliography

163. Appleyard M, Glukhovsky A, Swain P. Wireless-capsule diagnostic endoscopy for recurrent small-bowel bleeding. *The New England journal of medicine* 2001;344:232-233
164. Pennazio M, Santucci R, Rondonotti E, et al. Outcome of patients with obscure gastrointestinal bleeding after capsule endoscopy: report of 100 consecutive cases. *Gastroenterology* 2004;126:643-653
165. Mason M, Swain J, Matthews BD, Harold KL. Use of video capsule endoscopy in the setting of recurrent subacute small-bowel obstruction. *Journal of laparoendoscopic & advanced surgical techniques* 2008;18:713-716
166. Swain P. Wireless capsule endoscopy and Crohn's disease. *Gut* 2005;54:323-326
167. Lai LH, Wong GL, Chow DK, et al. Inter-observer variations on interpretation of capsule endoscopies. *European journal of gastroenterology & hepatology* 2006;18:283-286
168. Christodoulou DK, Haber G, Beejay U, et al. Reproducibility of wireless capsule endoscopy in the investigation of chronic obscure gastrointestinal bleeding. *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* 2007;21:707-714
169. Chen GC, Enayati P, Tran T, et al. Sensitivity and inter-observer variability for capsule endoscopy image analysis in a cohort of novice readers. *World J Gastroenterol* 2006;12:1249-1254
170. Faigel DO, Baron TH, Adler DG, et al. ASGE guideline: guidelines for credentialing and granting privileges for capsule endoscopy. *Gastrointest Endosc* 2005;61:503-505
171. Sidhu R, Sakellariou P, McAlindon ME, et al. Is formal training necessary for capsule endoscopy? The largest gastroenterology trainee study with controls. *Dig Liver Dis* 2008;40:298-302
172. Burling D, Taylor SA, Halligan S. Virtual colonoscopy: current status and future directions. *Gastrointestinal endoscopy clinics of North America* 2005;15:773-795
173. Cotton PB, Durkalski VL, Pineau BC. Computed tomographic colonography (virtual colonoscopy): a multicenter comparison with standard colonoscopy for detection of colorectal neoplasia. *JAMA* 2004;291:1713-1719
174. Ferrucci JT. Colonoscopy: virtual and optical--another look, another view. *Radiology* 2005;235:13-16
175. Halligan S, Altman DG, Mallett S, et al. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. *Gastroenterology* 2006;131:1690-1699

176. Pickhardt PJ, Choi JR, Hwang I, et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. *N Engl J Med* 2003;349:2191-2200
177. Slater A, Taylor SA, Tam E, et al. Reader error during CT colonography: causes and implications for training. *Eur Radiol* 2006;16:2275-2283
178. Spinzi G, Belloni G, Martegani A, et al. Computed tomographic colonography and conventional colonoscopy for colon diseases: a prospective, blinded study. *Am J Gastroenterol* 2001;96:394-400
179. Fidler JL, Fletcher JG, Johnson CD, et al. Understanding interpretive errors in radiologists learning computed tomography colonography. *Academic radiology* 2004;11:750-756
180. Taylor SA, Halligan S, Burling D. CT colonography: effect of experience and training on reader performance. *Eur Radiol* 2004;14:1025-1033
181. Taylor SA, Laghi A, Lefere P, Halligan S, Stoker J. European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. *Eur Radiol* 2007;17:575-579
182. Burling D, Halligan S, Taylor SA, Usiskin S, Bartram CI. CT colonography practice in the UK: a national survey. *Clinical radiology* 2004;59:39-43
183. Burling D, Moore A, Taylor S, La Porte S, Marshall M. Virtual colonoscopy training and accreditation: a national survey of radiologist experience and attitudes in the UK. *Clinical radiology* 2007;62:651-659
184. Yi SY, Ryu KH, Woo HS, et al. Sectional Analysis of Learning on the KAIST-Ewha Colonoscopy Simulation II. *Studies in health technology and informatics* 2009;142:432-434
185. Doctors in Society: Medical Professionalism in a Changing World. Report of a Working Party of the Royal College of Physicians of London. 2005:7-48
186. Eagly AH, Chaiken S. Attitude structure and function. In: Gilbert DT, Fiske ST, Lindzey G eds, *Handbook of Social Psychology*. 4th ed. Boston: McGraw-Hill; 1998:269-322
187. Tessaro I, Highriter M. HIV and the work intentions of public health nurses. *Public health nursing (Boston, Mass)* 1994;11:273-280
188. Limbert C, Lamb R. Doctors' use of clinical guidelines: two applications of the theory of planned behaviour. *Psychol Health Med* 2002;7:301-310
189. Ajzen I. From intentions to actions: a theory of planned behaviour. In: Kuhl J, Beckmann J eds, *Action-Control: from Cognition to Behaviour*. Heidelberg: Springer; 1985:11-39

Bibliography

190. Helmreich RL. Managing human error in aviation. *Scientific American* 1997;276:62-67
191. Callaghan K, Hunt G, Windsor J. Issues in implementing a real competency-based training and assessment system. *The New Zealand medical journal* 2007;120:U2510
192. Helmreich RL, Klinect JR, Wilhelm JA. System safety and threat and error management: The line operations safety audit (LOSA). *Proceedings of the Eleventh International Symposium on Aviation Psychology*. Colombus, OH: The Ohio State University; 2001
193. Flin R, Martin L, Goeters K, et al. The development of the NOTECHS system for evaluating pilots' CRM skills. *Human Factors and Aerospace Safety* 2003;3:95-117
194. Avermaete van JAG, Kruijsen EAC. NOTECHS. The evaluation of non-technical skills of multi-pilot aircrew in relation to the JAR-FCL requirements. Amsterdam: EC NOTECHS Project Final Report: NLR-CR-98443; 1998
195. Helmreich RL, Merritt AC, Wilhelm JA. The evolution of Crew Resource Management training in commercial aviation. *The International journal of aviation psychology* 1999;9:19-32
196. Hayward RA, Hofer TP. Estimating hospital deaths due to medical errors: preventability is in the eye of the reviewer. *Jama* 2001;286:415-420
197. Cullinane M, Gray AJG, Hargraves CMK, et al. National Confidential Enquiry into Patient Outcome and Death: Scoping Our Practice. 2004
198. Helmreich RL. On error management: lessons from aviation. *BMJ (Clinical research ed)* 2000;320:781-785
199. Phitayakorn R, Marks JM, Reynolds HL, Delaney CP. Expert benchmark for the GI Mentor II(TM). *Surg Endosc* 2008
200. Koch AD, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ. A second-generation virtual reality simulator for colonoscopy: validation and initial experience. *Endoscopy* 2008;40:735-738
201. Koch AD, Buzink SN, Heemskerk J, et al. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. *Surg Endosc* 2007
202. Westman B, Ritter EM, Kjellin A, et al. Visuospatial abilities correlate with performance of senior endoscopy specialist in simulated colonoscopy. *J Gastrointest Surg* 2006;10:593-599
203. Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *Jsls* 2005;9:130-133

204. Eversbusch A, Grantcharov TP. Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. *Surg Endosc* 2004;18:1514-1518
205. Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. *Gastrointest Endosc* 2003;57:214-218
206. Mahmood T, Darzi A. A study to validate the colonoscopy simulator. *Surg Endosc* 2003;17:1583-1589
207. Ferlitsch A, Glauninger P, Gupper A, et al. Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. *Endoscopy* 2002;34:698-702
208. Sedlack RE, Kolars JC. Colonoscopy curriculum development and performance-based assessment criteria on a computer-based endoscopy simulator. *Acad Med* 2002;77:750-751
209. Yi SY, Ryu KH, Na YJ, et al. Improvement of colonoscopy skills through simulation-based training. *Studies in health technology and informatics* 2008;132:565-567
210. Park J, MacRae H, Musselman LJ, et al. Randomized controlled trial of virtual reality simulator training: transfer to live patients. *American journal of surgery* 2007;194:205-211
211. Buzink SN, Koch AD, Heemskerk J, et al. Acquiring basic endoscopy skills by training on the GI Mentor II. *Surg Endosc* 2007;21:1996-2003
212. Kim WS, Woo HS, Ahn W, et al. Non-clinical evaluation of the KAIST-Ewha Colonoscopy Simulator II. *Studies in health technology and informatics* 2007;125:214-216
213. Cohen J, Cohen SA, Vora KC, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointest Endosc* 2006;64:361-368
214. Clark JA, Volchok JA, Hazey JW, Sadighi PJ, Fanelli RD. Initial experience using an endoscopic simulator to train surgical residents in flexible endoscopy in a community medical center residency program. *Current surgery* 2005;62:59-63
215. Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D. Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? *Endoscopy* 2005;37:1198-1204
216. Sedlack RE, Kolars JC. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *Am J Gastroenterol* 2004;99:33-37
217. Mahmood T, Darzi A. The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. *Surg Endosc* 2004;18:1224-1230

Bibliography

218. Sedlack RE, Kolars JC, Alexander JA. Computer simulation training enhances patient comfort during endoscopy. *Clin Gastroenterol Hepatol* 2004;2:348-352
219. MacDonald J, Ketchum J, Williams RG, Rogers LQ. A lay person versus a trained endoscopist: can the preop endoscopy simulator detect a difference? *Surg Endosc* 2003;17:896-898
220. Gerson LB, Van Dam J. A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy* 2003;35:569-575
221. 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;16:1459-1463
222. Tuggy ML. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *The Journal of the American Board of Family Practice / American Board of Family Practice* 1998;11:426-433
223. Sedlack RE. Validation of computer simulation training for esophagogastroduodenoscopy: Pilot study. *Journal of gastroenterology and hepatology* 2007
224. Moorthy K, Munz Y, Orchard TR, et al. An innovative method for the assessment of skills in lower gastrointestinal endoscopy. *Surg Endosc* 2004;18:1613-1619
225. Di Giulio E, Fregonese D, Casetti T, et al. Training with a computer-based simulator achieves basic manual skills required for upper endoscopy: a randomized controlled trial. *Gastrointest Endosc* 2004;60:196-200
226. Maiss J, Diebel H, Naegel A, et al. A novel model for training in ERCP with double-balloon enteroscopy after abdominal surgery. *Endoscopy* 2007;39:1072-1075
227. Sedlack R, Petersen B, Binmoeller K, Kolars J. A direct comparison of ERCP teaching models. *Gastrointest Endosc* 2003;57:886-890
228. Maiss J, Millermann L, Heinemann K, et al. The compactEASIE is a feasible training model for endoscopic novices: a prospective randomised trial. *Dig Liver Dis* 2007;39:70-78; discussion 79-80
229. Matthes K, Cohen J, Kochman ML, et al. Efficacy and costs of a one-day hands-on EASIE endoscopy simulator train-the-trainer workshop. *Gastrointest Endosc* 2005;62:921-927
230. Maiss J, Wiesnet J, Proeschel A, et al. Objective benefit of a 1-day training course in endoscopic hemostasis using the "compactEASIE" endoscopy simulator. *Endoscopy* 2005;37:552-558

231. Hochberger J, Matthes K, Maiss J, et al. Training with the compactEASIE biologic endoscopy simulator significantly improves hemostatic technical skill of gastroenterology fellows: a randomized controlled comparison with clinical endoscopy training alone. *Gastrointest Endosc* 2005;61:204-215
232. Hochberger J, Euler K, Naegel A, Hahn EG, Maiss J. The compact Erlangen Active Simulator for Interventional Endoscopy: a prospective comparison in structured team-training courses on "endoscopic hemostasis" for doctors and nurses to the "Endo-Trainer" model. *Scandinavian journal of gastroenterology* 2004;39:895-902
233. Matsuda K, Hawes RH, Sahai AV, Tajiri H. The role of simulators, models, phantoms. Where's the evidence? *Endoscopy* 2006;38 Suppl 1:S61-64
234. Bhutani MS, Wong RF, Hoffman BJ. Training facilities in gastrointestinal endoscopy: an animal model as an aid to learning endoscopic ultrasound. *Endoscopy* 2006;38:932-934
235. Kiesslich R, Moenk S, Reinhardt K, et al. Combined simulation training: a new concept and workshop is useful for crisis management in gastrointestinal endoscopy. *Zeitschrift fur Gastroenterologie* 2005;43:1031-1039
236. Fink A, Kosecoff J. How to conduct surveys: A step-by-step guide. Third Ed. ed. Thousand Oaks, California: Sage Publications; 2005
237. Mangunkusumo RT, Moorman PW, Van Den Berg-de Ruitter AE, et al. Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. *J Adolesc Health* 2005;36:70 e71-76
238. Raat H, Mangunkusumo RT, Mohangoo AD, Juniper EF, Van Der Lei J. Internet and written respiratory questionnaires yield equivalent results for adolescents. *Pediatric pulmonology* 2007;42:357-361
239. Ritter P, Lorig K, Laurent D, Matthews K. Internet versus mailed questionnaires: a randomized comparison. *Journal of medical Internet research* 2004;6:e29
240. Wells CW, Inglis S, Barton R. Trainees in gastroenterology views on teaching in clinical gastroenterology and endoscopy. *Medical teacher* 2009;31:138-144
241. Haycock A, Thomas-Gibson S, Vance M. Results from the national nurse endoscopy project: hands-on training improves and maintains polypectomy skills in the long-term. *Gut* 2009;58(Suppl 1):A10
242. Williams CB, Thomas-Gibson S. Rational colonoscopy, realistic simulation, and accelerated teaching. *Gastrointestinal endoscopy clinics of North America* 2006;16:457-470
243. Michelson JD, Manning L. Competency assessment in simulation-based procedural education. *American journal of surgery* 2008;196:609-615

Bibliography

244. Wehrmann K, Fruhmorgen P. Evaluation of a new three-dimensional magnetic imaging system for use during colonoscopy. *Endoscopy* 2002;34:905-908
245. Cheung HY, Chung CC, Kwok SY, Tsang WW, Li MK. Improvement in colonoscopy performance with adjunctive magnetic endoscope imaging: a randomized controlled trial. *Endoscopy* 2006;38:214-217
246. Shah SG, Pearson HJ, Moss S, et al. Magnetic endoscope imaging: a new technique for localizing colonic lesions. *Endoscopy* 2002;34:900-904
247. Shah SG, Brooker JC, Thapar C, et al. Effect of magnetic endoscope imaging on patient tolerance and sedation requirements during colonoscopy: a randomized controlled trial. *Gastrointest Endosc* 2002;55:832-837
248. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing technical skill via an innovative "bench station" examination. *American journal of surgery* 1997;173:226-230
249. Naik VN, Matsumoto ED, Houston PL, et al. Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? *Anesthesiology* 2001;95:343-348
250. Ericsson KA. Deliberate practice and acquisition of expert performance: a general overview. *Acad Emerg Med* 2008;15:988-994
251. Dafnis G, Ekbohm A, Pahlman L, Blomqvist P. Complications of diagnostic and therapeutic colonoscopy within a defined population in Sweden. *Gastrointest Endosc* 2001;54:302-309
252. Green J. Guidelines on complications of gastrointestinal endoscopy. In: *British Society of Gastroenterology*; 2006
253. General Recommendations on Training in Gastrointestinal Endoscopy. In: *The Joint Advisory Group on GI Endoscopy*; 2008
254. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *The British journal of surgery* 1997;84:273-278
255. Vance M, Baulff M. Endoscopy Competence Framework. Available from <https://tools.skillsforhealth.org.uk/suite/show/id/36>. Accessed September 2009
256. Emken JL, McDougall EM, Clayman RV. Training and assessment of laparoscopic skills. *Jsls* 2004;8:195-199
257. Fried GM. Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. *Gastrointestinal endoscopy clinics of North America* 2006;16:425-434

258. Goff B, Mandel L, Lentz G, et al. Assessment of resident surgical skills: is testing feasible? *American journal of obstetrics and gynecology* 2005;192:1331-1338; discussion 1338-1340
259. Ault G, Reznick R, MacRae H, et al. Exporting a technical skills evaluation technology to other sites. *American journal of surgery* 2001;182:254-256
260. Bann S, Khan M, Datta V, Darzi A. Surgical skill is predicted by the ability to detect errors. *American journal of surgery* 2005;189:412-415
261. Broe D, Ridgway PF, Johnson S, Tierney S, Conlon KC. Construct validation of a novel hybrid surgical simulator. *Surg Endosc* 2006
262. Datta V, Bann S, Mandalia M, Darzi A. The surgical efficiency score: a feasible, reliable, and valid method of skills assessment. *American journal of surgery* 2006;192:372-378
263. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. *American journal of surgery* 2002;184:70-73
264. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Acad Med* 1996;71:1363-1365
265. Goff BA, VanBlaricom A, Mandel L, Chinn M, Nielsen P. Comparison of objective, structured assessment of technical skills with a virtual reality hysteroscopy trainer and standard latex hysteroscopy model. *The Journal of reproductive medicine* 2007;52:407-412
266. Hernandez JD, Bann SD, Munz Y, et al. Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system. *Surg Endosc* 2004;18:372-378
267. Immenroth M, Burger T, Brenner J, et al. Mental training in surgical education: a randomized controlled trial. *Annals of surgery* 2007;245:385-391
268. Kohls-Gatzoulis JA, Regehr G, Hutchison C. Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. *Canadian journal of surgery* 2004;47:277-283
269. Lentz GM, Mandel LS, Goff BA. A six-year study of surgical teaching and skills evaluation for obstetric/gynecologic residents in porcine and inanimate surgical models. *American journal of obstetrics and gynecology* 2005;193:2056-2061
270. Nielsen PE, Foglia LM, Mandel LS, Chow GE. Objective structured assessment of technical skills for episiotomy repair. *American journal of obstetrics and gynecology* 2003;189:1257-1260

271. Haycock A, Suzuki N, Thomas-Gibson S. Development of a novel video scoring system for basic polypectomy: an initial validation study. In, Digestive Disease Week. Washington D.C.; 2006
272. Sidhu R, Sanders DS, Kapur K, et al. Capsule endoscopy: is there a role for nurses as physician extenders? *Gastroenterol Nurs* 2007;30:45-48
273. Ewertsen C, Svendsen CB, Svendsen LB, et al. [Is screening of wireless capsule endoscopies by non-physicians feasible?]. *Ugeskr Laeger* 2006;168:3530-3533
274. Bossa F, Cocomazzi G, Valvano MR, Andriulli A, Annese V. Detection of abnormal lesions recorded by capsule endoscopy. A prospective study comparing endoscopist's and nurse's accuracy. *Dig Liver Dis* 2006;38:599-602
275. Levinthal GN, Burke CA, Santisi JM. The accuracy of an endoscopy nurse in interpreting capsule endoscopy. *Am J Gastroenterol* 2003;98:2669-2671
276. Kulier R, Hadley J, Weinbrenner S, et al. Harmonising evidence-based medicine teaching: a study of the outcomes of e-learning in five European countries. *BMC medical education* 2008;8:27
277. Wylie P, Muckian J, Haycock A, et al. Validating virtual colonoscopy training in the United Kingdom. ESGAR. Istanbul, Turkey; 2008
278. Bartram C. Observer variation in the detection of colorectal neoplasia on double-contrast barium enema: implications for colorectal cancer screening and training. *Clinical radiology* 2004;59:641-642
279. Halligan S, Altman DG, Taylor SA, et al. CT colonography in the detection of colorectal polyps and cancer: systematic review, meta-analysis, and proposed minimum data set for study level reporting. *Radiology* 2005;237:893-904
280. Burling D, Halligan S, Atchley J, et al. CT colonography: interpretative performance in a non-academic environment. *Clinical radiology* 2007;62:424-429; discussion 430-421
281. Soto JA, Barish, M.A., Yee, J. Reader training in CT colonography: how much is enough? *Radiology* 2005;237:26-27
282. Halligan S, Lilford RJ, Wardle J, et al. Design of a multicentre randomized trial to evaluate CT colonography versus colonoscopy or barium enema for diagnosis of colonic cancer in older symptomatic patients: The SIGGAR study. *Trials* 2007;8:32
283. Burling D. CT Colonography and CAD: promoting excellence and ensuring minimal standards. *European Radiology* 2009;19 (supp 2):637
284. Flin R, Yule S, Paterson-Brown S, et al. Teaching surgeons about non-technical skills. *Surgeon* 2007;5:86-89

285. Undre S, Koutantji M, Sevdalis N, et al. Multidisciplinary crisis simulations: the way forward for training surgical teams. *World journal of surgery* 2007;31:1843-1853
286. Mishra A, Catchpole K, Dale T, McCulloch P. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2007
287. Catchpole K, Giddings AEB, de Leval M, et al. Identifying and reducing systems failures through non-technical skills. *Surgeon* 2005;3
288. Schaeffer H, Helmreich RL. The operating room management attitudes questionnaire. (ORMAQ). Austin, Texas: NASA/University of Texas; 1993:93-98
289. Baldwin PJ, Paisley AM, Brown SP. Consultant surgeons' opinion of the skills required of basic surgical trainees. *The British journal of surgery* 1999;86:1078-1082
290. Undre S, Healey AN, Darzi A, Vincent CA. Observational assessment of surgical teamwork: a feasibility study. *World journal of surgery* 2006;30:1774-1783
291. Fletcher G, Flin R, McGeorge P, et al. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *British journal of anaesthesia* 2003;90:580-588
292. Endoscopy global rating scale (GRS). www.grs.nhs.uk.
293. Flin R, Martin L. Behavioural Markers for CRM: A survey of current practice. *International Journal of Aviation Psychology* 2001;11:95-118
294. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess R eds, *Analysing qualitative data*. London: Routledge; 1993:173-194
295. Fletcher G, Flin R, McGeorge P, et al. Rating non-technical skills: developing a behavioural marker system for use in anaesthesia. *Cogn Tech Work* 2004;6:165-171
296. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D. Development of a rating system for surgeons' non-technical skills. *Medical education* 2006;40:1098-1104
297. Klein G, Militello LG. Some guidelines for conducting a cognitive task analysis. In: Salas E ed, *Advances in human performance and cognitive engineering research*. New York: JAI Press; 2001
298. Dickinson TL, R.M. M. A conceptual framework for teamwork measurement. In: Brannick MT, Salas E, Prince C eds, *Team performance assessment and measurement: theory, methods, and applications*. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 1997
299. Stake RE. *The art of case study research*. Thousand Oaks, Ca: Sage; 1995

Bibliography

300. Tobin GA, Begley CM. Methodological rigour within a qualitative framework. *Journal of advanced nursing* 2004;48:388-396
301. Militello LG, Hutton RJ. Applied cognitive task analysis (ACTA): a practitioner's toolkit for understanding cognitive task demands. *Ergonomics* 1998;41:1618-1641
302. Rugg G, McGeorge P. The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. *Expert Systems* 1997;14:80-93
303. Flanagan JC. The critical incident technique. *Psychological bulletin* 1954;51:327-358
304. Klein G, Calderwood R, MacGregor D. Critical decision method for eliciting knowledge. *IEEE Trans Syst Man Cybern* 1989;19:462-472
305. Kvale S. *InterViews. An introduction to qualitative research writing*. Thousand Oaks, CA: Sage Publications; 1996
306. Ritchie J, Lewis J, Elan G. Designing and selecting samples. In: Ritchie J, Lewis J eds, *Qualitative research practice: A guide for social science students and researchers*. London: Sage; 2003:77-108
307. Miles MB, Huberman AM. *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: CA: Sage; 1994
308. Le Minor M, Alperovitch A, Knill-Jones RP. Applying decision theory to medical decision-making--concept of regret and error of diagnosis. *Methods of information in medicine* 1982;21:3-8
309. Patton MQ. Designing qualitative studies. In: Patton MQ ed, *Qualitative evaluation and research methods*. Newbury Park: Sage; 1980:145-198
310. Chell E. Critical incident technique. In: Cassell C, Symon G eds, *Essential guide to qualitative methods in organizational research*. London: Sage; 2004:45-60
311. Hoffman RR, Crandall B, Shadbolt N. Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis *Human Factors: The Journal of the Human Factors and Ergonomics Society* 1998;40:254-276
312. Birnbach RA, Longridge TM. The regulatory perspective. In: Weiner EL, Kanki BG, Helmreich RL eds, *Cockpit Resource Management*. London: Academic Press; 1993
313. Barton R. Outcomes and perceptions of a high-stakes colonoscopy skills assessment: the Joint Advisory Group/Bowel Cancer Screening Programme screening colonoscopy assessment. *Gut* 2008;57:A53
314. Yule S, Flin R, Maran N, et al. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World journal of surgery* 2008;32:548-556

315. Weller JM, Bloch M, Young S, et al. Evaluation of high fidelity patient simulator in assessment of performance of anaesthetists. *British journal of anaesthesia* 2003;90:43-47
316. Seamster T, Edens E. Cognitive modelling of CRM assessment expertise: identification of the primary assessors. In, *Proceedings of the human factors and ergonomics society, 37th annual meeting*. San Diego: HFES; 1993
317. Kneebone R, Baillie S. Contextualized simulation and procedural skills: a view from medical education. *Journal of veterinary medical education* 2008;35:595-598
318. Kneebone R, Bello F, Nestel D, Yadollahi F, Darzi A. Training and assessment of procedural skills in context using an Integrated Procedural Performance Instrument (IPPI). *Studies in health technology and informatics* 2007;125:229-231
319. Cook DA. Web-based learning: pros, cons and controversies. *Clinical medicine (London, England)* 2007;7:37-42
320. Hugenholtz NI, de Croon EM, Smits PB, van Dijk FJ, Nieuwenhuijsen K. Effectiveness of e-learning in continuing medical education for occupational physicians. *Occupational medicine (Oxford, England)* 2008
321. Cook DA, Dupras DM, Thompson WG, Pankratz VS. Web-based learning in residents' continuity clinics: a randomized, controlled trial. *Acad Med* 2005;80:90-97
322. Anderson J. Assessment in endoscopy. Available from www.grs.nhs.uk. Accessed 31 July 2009.

Publications arising from this research

1. CT colonography training for radiographers – a formal evaluation. Haycock A, Burling D, Wylie P, Muckian J, Ilangoan R, Thomas-Gibson S. Accepted for publication in Clinical Radiology.
2. Training and transfer of colonoscopy skills: a multinational randomized blinded controlled trial of simulator versus bedside training. AV Haycock, AD Koch, P Familiari, F van Delft, JS Bladen, E Dekker, L Petruziello, J Haringsma, S Thomas-Gibson. Gastrointest Endosc 2010. Feb;71(2):298-307. Epub 2009 Nov 3.
3. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. AV Haycock, P Bassett, J Bladen, S Thomas-Gibson. Endoscopy 2009 Nov;41(11):952-8. Epub 2009 Oct 2.
4. Evaluating changes in gastrointestinal endoscopy training over 5 years: closing the audit loop. Haycock AV, Patel JH, Tekkis PP, Thomas-Gibson S. Eur J Gastroenterol Hepatol. 2010. Mar;22(3):368-73. Epub 2009 July 17.
5. Simulator training improves practical skills in therapeutic gastrointestinal endoscopy: results from a randomized, blinded controlled study. AV Haycock, P Youd, P Bassett, P Tekkis, S Thomas-Gibson. Gastrointest Endosc. 2009 Nov;70(5):835-45. Epub 2009 Jun 25.
6. Development of a Novel Esophageal Stricture Simulation. Haycock AV, East JE, Swain D, Thomas-Gibson S. Dig Dis Sci. 2010 Feb;55(2):321-7. Epub 2009 Mar 13.
7. Computer-aided learning in capsule endoscopy leads to improvement in lesion recognition ability. Postgate A, Haycock A, Thomas-Gibson S, Fitzpatrick A, Bassett P, Preston S, Saunders BP, Fraser C. Gastrointest Endosc. 2009 Aug;70(2):310-6. Epub 2009 Apr 21.
8. How should we train capsule endoscopy? A pilot study of performance changes during a structured capsule endoscopy training programme. AJ Postgate, AV Haycock, A Fitzpatrick, G Schofield, P Bassett, S Thomas-Gibson, C Fraser. Dig Dis Sci. 2009 Aug;54(8):1672-9. Epub 2008 Nov 26.

Abstracts presented to learned societies

*(*oral presentation)*

1. *Acceptability of simulator training in colonoscopy: results from a multinational randomised blinded controlled trial. AV Haycock, AD Koch, P Familiari, F van Delft, JS Bladen, E Dekker, L Petruziello, J Haringsma, S Thomas-Gibson. Gut 2009;58(Suppl 1):A11.
2. *Results from the national nurse endoscopy project: hands-on training improves and maintains polypectomy skills in the long-term. A Haycock, S Thomas-Gibson, M Vance. Gut 2009;58(Suppl 1):A10.
3. *Gastroenterology training in 2008: results from the tig/bsg national training survey. AV Haycock, P Flanagan, A Ignjatovic, S Kendrick, D Mathias, R Mead, P Mizen, M Mottershead, V Patel, K Pollock, T Valliani, J Williams. Gut 2009;58(Suppl 1):A11.

4. Endoscopy training in 2008: results from the TiG/BSG National Training Survey. AV Haycock, P Flanagan, A Ignjatovic, S Kendrick, R Mead, P Mizen, M Mottershead, V Patel, K Pollock, T Valliani, J Williams. *Gut* 2009;58(Suppl 1):A50.
5. Colonoscopy skills transfer from a second-generation virtual reality simulator to patients: a multinational randomised blinded controlled trial. AV Haycock, AD Koch, P Familiari, F van Delft, JS Bladen, E Dekker, L Petruziello, J Haringsma, S Thomas-Gibson. *Gut* 2009;58(Suppl 1):A50.
6. *Validating virtual colonoscopy training in the United Kingdom. P Wylie, J Muckian, A Haycock, A Gupta, M Marshall, D Burling. Oral presentation ESGAR June 2008. SS6.03.
7. *Efficacy of hands-on skills training in therapeutic endoscopy: results from a randomised blinded controlled study. AV Haycock, S Thomas-Gibson. *Endoscopy* 2008;40(Suppl 1):A5
8. *An interactive polyp detection training tool: development and initial validation. P Youd, AV Haycock, S Preston, S Thomas-Gibson. *Gut* 2008;57(Suppl 1):A18.
9. Mid-point results of a randomised blinded controlled study of hands-on therapeutic endoscopy training. A Haycock, S Thomas-Gibson. *Gut* 2008;57(Suppl 1):A59.
10. The UK National Endoscopy Training Programme has improved colonoscopy training over a 5 year period: Closing the audit loop. AV Haycock, JH Patel, S Thomas-Gibson. *Gut* 2008;57(Suppl 1):A60.
11. Development of a web-based e-learning module to improve lesion recognition at capsule endoscopy. AJ Postgate, AV Haycock, S Thomas-Gibson, A Fitzpatrick, S Prestion, C Fraser. *Gut* 2008;57(Suppl 1):A109.
12. JAG colonoscopy training guidelines are not being met following a one-to-one hands-on colonoscopy training course. AV Haycock, P Youd, N Suzuki, S Thomas-Gibson, BP Saunders. *Gut* 2007;56(Suppl 2):A91.
13. Trainees continue to improve in colonoscopy despite low procedural experience following an intensive one-to-one hands-on colonoscopy training course. A Haycock, P Youd, N Suzuki, S Thomas-Gibson, BP Saunders. *Gastroenterology* 2007;132:4(Suppl S1).
14. Development of a novel video scoring system for basic polypectomy: an initial validation study. A Haycock, N Suzuki, S Thomas-Gibson. *Gastrointestinal Endoscopy* 2007;65:5,AB329.
15. Achieving realistic simulation: Initial validation of the Olympus Colonoscopy Simulator (ENDO TS-1) AV Haycock, CB Williams, BP Saunders, S Thomas-Gibson. *Endoscopy* 2007;39(Suppl I):A364.

Appendix 1 Literature search on simulation in GI endoscopy

i. Levels of evidence

- 1a Systematic reviews (metaanalysis) containing at least some trials of level 1b evidence, in which results of separate, independently conducted trials are consistent
- 1b Randomised controlled trial of good quality and of adequate sample size (power calculation)
- 2a Randomised trials of reasonable quality and/or of inadequate sample size
- 2b Nonrandomised trials, comparative research (parallel cohort)
- 2c Nonrandomised trial, comparative research (historical cohort, literature controls)
- 3 Nonrandomised, noncomparative trials, descriptive research
- 4 Expert opinions, including the opinion of Work Group members

ii. Colonoscopy: Validation studies

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Phitayakorn (2008)[199]	GI Mentor II	Benchmarking	23 expert surgeons (>1000)	Single case (Module 1, case 5)		Time to caecum, time to completion, % mucosa visualised, % time looped	Wide range of variability for all measurements	3
Koch (2008)[200]	Olympus Endo TS-1	Validation	23 experts (>1000), 26 novices (0)	Dexterity task, single case	Survey	Dexterity score, time to caecum, max shaft insertion force, shaft torque, tip section force, max pain Expert judgement of face validity	Time to caecum only discriminated All scores 6.9/10 or higher	2b
Sedlack (2007)[128]	Bovine ex-vivo	Validation	13 experts (?), 13 intermediate (100-150), 13 novices (0)	Single case	Survey	unvalidated blueprint - time to markers Expert judgement of face validity	Completion rate, completion time, time to markers, % mucosa visualised, quality of mucosa exam all discriminate All except one (haustra realism) better than neutral	2b
Koch (2007)[201]	GI Mentor II	Validation	35 experts (>1000), 20 experienced (200-1000), 15 intermediate (<200), 35 novices (0)	Dexterity task (Endobubble level 1), 2 cases (Module 1, cases 1 & 3)	Survey	Dexterity score, time to caecum, % time with clear view, lost view of lumen, excessive local pressure, % time in pain, excessive loop formed.	Novices distinguished on dexterity task, completion time, visibility and patient pain. No significant differences found between intermediate, experienced and experts on any parameters	2b
Westman (2007)[202]	GI Mentor II	Visuo-spatial correlation	11 experts	Single case (Module 1, case 3), PicSOr, card rotation, cube comparison	?	?	The endoscopists who performed better in the visuospatial tests also were better at maintaining visualization of the colon lumen. Those who performed better in the PicSOr test formed fewer loops during colonoscopy	3
Grantcharov (2005)[203]	GI Mentor II	Validation	8 experienced (>200), 10 residents (<50), 10 novices (0)	?	?	Time to caecum, % time with clear view, lost view of lumen, excessive local pressure, % time in pain, excessive loop formed.	Novices distinguished on completion time, % mucosa seen, efficiency of screening, patient pain, loop formation, time with loop, excessive local pressure. No sig differences found between experienced and experts	2b

Appendix 1: Literature search on simulation

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Eversbusch (2004)[204]	GI Mentor	Learning curve	8 experts (>200), 10 residents (<50), 10 novices (0)	Endobubble. Single case.	10 repetitions of Endobubble	Time to caecum, % mucosa visualised, % time with clear view, efficiency of screening, excessive local pressure, pain, time in pain, loop formation, time with loop.	Learning curve plateau on 2nd repetition for experts, 5th for residents, 7th for novices.	2b
Sedlack (2003)[205]	Accutouch	Validation	10 experts (?), 6 fellows (150), 6 residents (0)	2 cases (3 & 4)	Survey	Total time, insertion time, withdrawal time, time in red-out, max insertion depth, total path length, % mucosa visualised, % discomfort, volume air insufflated, max scope force, complications, extra sedation, identification of pathology Expert judgement of face validity	Time, insertion time and red-out distinguished novices from residents. No significant differences found between residents and experts. All better than neutral	2b
Mahmood (2003)[206]	HT Immersion (Accutouch)	Validation	5 experts (>100), 7 intermediate (11-100), 10 novices (<10)	2 cases (3 & 4) - average of 5 attempts		Total time, % mucosa seen, path length, perforation	Total time, % mucosa seen, perforation discriminated overall, but no direct comparisons made between the three groups	2b
Ferlitsch (2002)[207]	GI Mentor	Validation (Part 1)	11 experts (>1000), 13 novices (0)	Endobubble, Endobasket, 2 cases		Endobasket, endobubble scores, correctly identified pathology, insertion time, successful retroflexion, adverse events, excessive wall pressure, impaired view	Experts significantly better than novices	2b
Sedlack (2002)[208]	Accutouch	Validation	10 experts (?), 5 partially trained (?), 2 novices (0)	2 cases		Time to complete, distance intubated, % mucosa seen, complications, pain	Average and minimal performance standards established. Simulator likely most effective early in training	2b

iii. Colonoscopy: training studies

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Yi (2008)[209]	KAIST-Ewha	Training and transfer	6 fellows, 5 residents randomised into subjects and controls	2 cases	Training to pre-determined goals	?	?	?
Park (2007)[210]	Accutouch	Training and transfer	24 novices (<3) randomised into 12 subjects and 12 controls	Pre-test single case (Module 1) Post-test single patient colonoscopy	2-3 hours on simulator	Computer: 4 validated (completion time, % time in red-out, % lumen visualised, total scope path length), 4 unvalidated (max scope insertion depth, max scope force, amount air insufflated, time with patient discomfort). Expert opinion using Global Rating Score. Completion rate, critical flaws, Global Rating Score	Procedure completion time and % time in red-out had reasonable correlation with GRS. Subjects sig better than controls on GRS (effect size 0.8, p=0.04). Not sig for completion rate or critical flaws.	1b
Buzink (2007)[211]	GI Mentor II	Training	5 experts (>1000), 30 novices (0). Also compared against data from 20 experienced (200-1000) and 35 experts (>1000)	Dexterity task (Endobubble level 1). Repetition of case (Module1, case 3).	15 tasks, with repetitive task performed 4 times	Dexterity score, time to caecum, % time with clear view, lost view of lumen, excessive local pressure, % time in pain, excessive loop formed.	Novices significantly improved on all parameters whereas experts did not. Simulator could distinguish between novices and both experienced and experts, but not experienced from experts.	2b/2c
Kim (2007)[212]	Kaist-Ewha	Training	31 engineering students (0)	?	Training to predetermined goals	?	?	?
Cohen (2006)[213]	GI Mentor II	Training	45 first year fellows (<10) randomised to 23 subjects and 22 controls	Follow-up for 200 colonoscopies	10 hours training on simulator	Total time, time to caecum, % mucosa examined, excessive pressure, efficiency score. Ability to reach caecum independently, correct identification of abnormalities, overall competency, patient discomfort (expert rating)	Subjects significantly better objective competence than controls for first 80 cases. No difference in length of time to 90% completion rate. No difference in patient discomfort	2a
Clark (2005)[214]	GI Mentor	Training	5 first year fellows, 8 surgical residents	Follow-up over 2 years	Monthly training	?	?	?

Appendix 1: Literature search on simulation

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Ahlberg (2005)[215]	Accutouch	Training	10 novices (0) - 8 surgical, 2 medical residents randomised to 6 subjects and 4 controls	Follow-up for 10 colonoscopies	Training to pre-determined goals (mean 20 hours)	Completion rate, time to completion, patient discomfort	Subjects had significantly higher completion rates with shorter time and less patient discomfort.	2a
Sedlack (2004)[216]	GI Mentor	Training	8 novices (0) randomised to 4 subjects and 4 controls	Follow-up for 15 colonoscopies	6 hours training on simulator	Insertion time, depth of unassisted insertion, independent procedure completion, ability to identify endoscopic landmarks, inserts in a safe manner, adequately visualizes mucosa on withdrawal, responds appropriately to patient discomfort	Subjects significantly better than controls for all except insertion time. Three parameters (depth of insertion, independent completion, ability to identify landmarks) demonstrated a continued advantage out to 30 colonoscopies, but no further.	2a
Mahmood (2004)[217]	HT Immersion (Accutouch)	Training	26 subjects - 9 research fellows, 5 PRHO's, 6 SpR's, 6 consultants. No details given of experience.	5 consecutive cases from Modules 3 & 4	14 did one trial (5 cases), 12 did two trials (10 cases)	Time taken, % mucosa visualised, path length, depth	With no feedback, there was no improvement in skills	2a
Ferlitsch (2002)[207]	GI Mentor	Training (Part 2)	13 novices (0) randomised to 7 subjects and 6 controls	Endobubble, Endobasket, 2 cases pre and 1 case post	30 hours simulator training vs no training	Endobasket, endobubble scores, correctly identified pathology, insertion time, successful retroflexion, adverse events, excessive wall pressure, impaired view	Subjects better on number of bubble hits and insertion time only.	2a

iv. Flexible sigmoidoscopy studies

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Sedlack (2004)[218]	GI Mentor	Training and transfer	38 residents randomised to 19 subjects and 19 controls	<10 Patient-based FS	3h simulator training prior to 1 week patient based training vs 1 week patient based training	Ability to reach flexure, perform biopsy, patient discomfort	Less patient discomfort for subjects. No difference in skill level	2a
MacDonald (2003)[219]	Pre-op (Accutouch)	Validation	10 clerical staff (0), 19 residents (mean <15), 5 experts (750)	Module 1 (task 1,2, and 3)		Time, insertion length, % mucosa viewed, % pain, red-out, air insufflated, air not removed, max force, % pathology visualised, perforation	Clerical staff worse than residents or experts. Experts not significantly better than residents and had higher perforation rate	2b
Gerson (2003)[220]	Accutouch	Training and transfer	16 residents (<10) randomised to 9 subjects and 7 controls	5 patient-based FS	Unlimited time on simulator over 2 weeks vs 10 patient-based FS over 2 weeks	Time, independent completion (to splenic flexure), required assistance, flexure recognition, retroflex completed, patient satisfaction, discomfort	Bedside training superior in all outcomes	2a
Datta (2002)[221]	Pre-op (Accutouch)	Validation	15 expert (>200), 15 intermediate (5-50), 15 novice (0)	3 cases		Time, insertion length, % mucosa viewed, % pain, red-out, air insufflated, air not removed, max force, efficiency score, path length, % pathology visualised, perforation	% mucosa visualised and efficiency ratio discriminated between all 3 groups. Novices were slower and had lower pathlength.	2b
Tuggy (1998)[222]	Gastro-Sim	Training and transfer	10 family medicine residents randomised to 5 subjects & 5 controls	1 patient-based FS	6-10 hours training on simulator vs no training	Insertion time, length of examination, directional errors, % mucosa visualised, viewing quality	Subjects superior to controls for all measures	2a

v. OGD validation studies

Study	Simulator	Type	Subjects	Assessment Module/task	Intervention	Outcomes	Findings	Level of evidence
Sedlack (2007)[223]	GI Mentor II	Validation (Part 1)	7 experts, 7 intermediate, 7 novice	2 cases	survey	7 measures	Only 2 had good fidelity - anatomy and scope maneuverability	2b
					computer-based parameters		Only novices could be distinguished	
Westman (2006)[202]	GI Mentor II	Visuo-spatial correlation		Single case (Module 1, case 3), PicSOr, card rotation, cube comparison	11 experts	?	No significant correlations between visuospatial abilities and performance parameters	3
Moorthy (2004)[224]	GI Mentor	Validation	10 experienced (>200), 11 intermediate (10-50), 11 novices (0)	2 cases	Comparison with 2 blinded observers using video endoscopic (VES) score	Time, % mucosa visualised, % pathology visualised, inappropriate retroflexions	Able to distinguish novice from expert on computer parameters and VES.	2b
Neumann (2003)[84]	Erlangen	Validation (Part 1)	5 experienced (?), 10 medical students (0), 10 physicians (0)	Single case daily over 6 days of the course	10 hours practice over 5 days of course	Scorecard (unvalidated, not blinded)	Scorecard distinguished experts from novices	2b
Ferlitsch (2002)[207]	GI Mentor	Validation (Part 1)	11 experts (>1000), 13 novices (0)	Endobubble, Endobasket, 2 cases		Endobasket, endobubble scores, correctly identified pathology, insertion time, successful retroflexion, adverse events, excessive wall pressure, impaired view	Experts significantly better than novices	2b

vi. OGD Training studies

Study	Simulator	Type	Subjects	Assessment Module/task	Intervention	Outcomes	Findings	Level of evidence
Sedlack (2007)[223]	GI Mentor II	Training and transfer (Part 2)	4 novice subjects, 4 novice controls		6h sim training + 1/12 patient training vs 1/12 patient training	Unvalidated scoring: Patient discomfort, sedation, independence, competence	Patient training alone superior	2a
Di Giulio (2004)[225]	GI Mentor	Training	22 novices (0) randomised to 11 subjects and 11 controls	20 consecutive OGD	10h simulator training	completeness, the number of identified or missed lesions, overall subjective judgment, complications (not blinded)	Subjects performed more complete procedures with better overall scores	2a
Neumann (2003)[84]	Erlangen	Training (Part 2)	5 experienced (?), 10 medical students (0), 10 physicians (0)	Single case daily over 6 days of the course	10 hours practice over 5 days of course	Scorecard (unvalidated, not blinded)	Scores improved daily and were significantly higher at the end of the course for novices	2b
Ferlitsch (2002)[207]	GI Mentor	Training (Part 2)	13 novices (0) randomised to 7 subjects and 6 controls	Endobubble, Endobasket, 2 cases pre and 1 case post	30 hours simulator training vs no training	Endobasket, endobubble scores, correctly identified pathology, insertion time, successful retroflexion, adverse events, excessive wall pressure, impaired view	Subjects better on number of bubble hits and adverse events only.	2a

vii. ERCP studies

Study	Simulator	Type	Subjects	Assessment Module/task	Intervention	Outcomes	Findings	Level of evidence
Maiss (2007)[226]	CompactE ASIE	Initial report	2 experts (experienced at ERCP and DBE). 2 novices (0 DBE, 0 ERCP)	ERCP with DBE	Beginners were assisted	Time to papilla, successful stent exchange, sphincterotomy	Experts were quicker, but beginners managed to complete.	2b
Matthes (2006)[129]	Neo-Papilla	Validation	9 experts (?)	Cannulate duct, sphincterotomy, stent placement		Tissue pliability, papillary anatomy, visual realism, cannulation realism, overall ERCP experience.	Scores all > average.	4
						Comparison with 3 other models (Erlangen, live pig, computer)	6/9 thought it was the most realistic simulator available.	2c
Sedlack (2003)[227]	Live pig, Erlangen, computer	Comparison	20 randomly selected volunteers from course workshop	Biliary cannulation and interventions		Tissue pliability, papillary anatomy, visual realism, cannulation realism, overall ERCP experience.	Erlangen model scored highest on indices of realism, usefulness, and performance. Computer was felt to be easiest to incorporate into a training program.	2b

viii. Therapeutic endoscopy studies

Study	Simulator	Type	Module/task	Intervention	Subjects	Outcomes	Findings	Level of evidence
Maiss (2007)[228]	Compact EASIE	Training	Haemostasis	All participants had 19 hours practical pre-training. Subjects then received 12 training sessions every 2 weeks over 7 months in manual skills, injection therapy, haemoclip, band ligation vs no training for controls	27 novices randomised to 14 subjects and 13 controls	Previously used checklists (0-10) for each skill (manual skills, injection therapy, haemoclip, band ligation) and total time	Significantly higher performance scores ($p<0.001$) and shorter times ($p<0.001$) for subjects vs controls	1b
Matthes (2005)[229]	EASIE	Training the Trainers	Haemostasis	6 hour training session for new tutors. Fellows had 45mins training on each of manual skills, injection therapy, haemoclip, band ligation.	7 tutors (no previous simulator teaching experience). 8 fellows (minimal experience - no haemoclip prior to course)	Previously used checklists (0-10) for each skill (manual skills, injection therapy, haemoclip, band ligation) and total time. Results compared with historical cohort.	Trainees significantly improved their scores on all 4 skills. No differences from historical controls.	2c
Maiss (2005)[230]	Compact EASIE	Training	Haemostasis	60 mins training on each of manual skills, injection therapy, haemoclip, band ligation	14 US fellows (year 1-3). 18 French fellows (final year).	Previously used checklists (0-10) for each skill (manual skills, injection therapy, haemoclip, band ligation) and total time	Significant improvement in performance for all 4 skills. Time improvement for all but injection.	2b
Hochberger (2005)[231]	Compact EASIE	Training	Haemostasis	3 sets of 45 min training sessions over 7 months on each of manual skills, injection therapy, haemoclip, band ligation	28 fellows randomised to 13 subjects (7 first year, 5 second year, 2 third year) and 14 controls (6 first year, 4 second year, 4 third year). 4 drop-outs not analysed.	Previously used checklists (0-10) for each skill (manual skills, injection therapy, haemoclip, band ligation) and total time	Significant improvement for all skills in intensive group. Controls improved in variceal ligation alone.	2a
Hochberger (2004)[232]	EndoTrainer (4 courses), Compact EASIE (7 courses)	Training	Haemostasis	30 min theory then 4h training	207/291 workshop participants	Questionnaire	No significant difference in preferences between models	2b

ix. Other studies

Study	Simulator	Type	Subjects	Assessment module/task	Intervention	Outcome measures	Findings	Level of evidence
Matsuda (2006)[233]	EUS FNA-box, EUS Mentor, EUS RK model, live pig	Evaluation	8 EUS experts	EUS	Survey	Realism scores, utility as educational tool, ordered rankings	Experts considered all four simulations useful in varying degrees throughout training	4
Bhutani (2006)[234]	live pig	Evaluation	20/38 participants attending an ASGE EUS training course in 1997 and 34 participants in 2000	Upper EUS	Retrospective survey		Majority rated course as excellent and had increased their EUS practice following the course	4

Appendix 2 Therapeutic checklists and Global Score sheet

Polypectomy & Hot Biopsy Checklist

Please tick the appropriate boxes to the right of the tasks when the candidate has completed them. Any boxes left blank will indicate that the candidate did not complete that task successfully.

Section 1: Checking basic function	
Checks endoscope functions (air, water, suction, control wheels)	☐
Makes appropriate adjustments	☐
Checks polypectomy equipment (forceps, snare, diathermy available)	☐
Section 2: Optimising view of polyp:	
Identifies polyp and lines up appropriately (5 o'clock position)	☐
Ensures adequate view by sufficient insufflation, wash, suction as appropriate	☐
Section 3: Selection of appropriate method of polypectomy	
Hot biopsy if <5mm, snare polypectomy if 5mm or greater	☐
Checks diathermy settings (NB for pig model settings different)	☐
Section 4: For hot biopsy	
Checks forceps open & close prior to insertion into therapeutic channel	☐
Forceps guided accurately onto polyp	☐
Appropriate tissue captured with jaws	☐
Insulating portion of forceps visible	☐
Mucosa safely tented	☐
NODAL POINT	DO NOT LET CANDIDATE PROCEED UNTIL IT IS SAFE FOR DIATHERMY TO BE APPLIED
Appropriate degree of coagulation prior to withdrawing forceps	☐
Section 5: For snare polypectomy	
Checks snare opens & closes and handle marked prior to insertion into therapeutic channel	☐
Snare loop passed accurately over polyp head	☐
Snare sheath advanced towards polyp as snare closed	☐
Snare closure at appropriate site (eg on stalk, or correct amount tissue, tenting)	☐
Asks assistant to close 'to the mark'	☐
NODAL POINT	DO NOT LET CANDIDATE PROCEED UNTIL IT IS SAFE FOR DIATHERMY TO BE APPLIED
Endoscopist performs snare closure and coagulation	☐
Appropriate degree of coagulation at polyp base	☐
Retrieval of specimen by re-snaring polyp or suctioning into polyp trap	☐
Checks retrieved specimen is representative of whole polyp	☐
Section 6: Post polypectomy	
Examination of polypectomy base to confirm haemostasis	☐
No contralateral burns	☐
Section 8: Documentation	
Asks to document procedure and gives instructions for after care	☐

Therapy to Bleeding Lesion Checklist

Please tick the appropriate boxes to the right of the tasks when the candidate has completed them. Any boxes left blank will indicate that the candidate did not complete that task successfully.

Section 1: Checking basic function Checks endoscope functions (air, water, suction, control wheels) ☐ Makes appropriate adjustments ☐ Checks equipment (injection needle, diathermy, clips) available ☐	
Section 2: Optimising view of lesion: Identifies lesion and lines up appropriately (7 o'clock position) ☐ Ensures adequate view by sufficient insufflation, wash, suction as appropriate ☐ Stabilises position (asks assistant to hold scope if necessary) ☐	
Section 3: Selection of appropriate method of haemostasis Performs at least two of the three techniques below (one <u>must</u> be injection) ☐	
Section 4: Injection Checks needle opens & closes prior to insertion into therapeutic channel ☐ Checks adrenaline concentration ☐ Checks line flushed through prior to injection ☐ Injection performed at 4 quadrants 2-3mm around bleeding vessel ☐ Evidence of sufficient volume injected (raised bleb, blanching etc) ☐	
Section 5: Diathermy Checks leads and diathermy settings prior to insertion into therapeutic channel ☐ Probe guided accurately onto lesion ☐	
NODAL POINT	Insulating portion of probe visible ☐ DO NOT LET CANDIDATE PROCEED UNTIL IT IS SAFE FOR DIATHERMY TO BE APPLIED
Appropriate degree of coagulation applied to central vessel ☐ Use of flush to disengage probe if necessary ☐	
Section 6: Haemoclip application Checks assistant aware of procedure to open and fire clip prior to insertion into therapeutic channel ☐ Optimises position for clip application (rotation of scope or clip) ☐ Clip opened to full extent prior to deployment ☐ Clip applied accurately to vessel ☐	
Section 7: Post therapy Checks haemostasis achieved at end of procedure ☐	
Section 8: Documentation Asks to document procedure and gives instructions for after care ☐	

Oesophageal Stricture Balloon Dilatation Checklist

Please tick the appropriate boxes to the right of the tasks when the candidate has completed them. Any boxes left blank will indicate that the candidate did not complete that task successfully.

Section 1: Checking basic function		
	Checks endoscope functions (air, water, suction, control wheels)	☐
	Makes appropriate adjustments	☐
	Checks dilatation equipment (balloon dilator, sizes) available	☐
Section 2: Identification of stricture		
	Identifies stricture and does not attempt to pass scope through if narrow	☐
	Maintains position 2-3 cm above stricture during procedure	☐
Section 3: Dilatation of stricture		
	Passes guidewire (if available) accurately through lumen	☐
	Ensures balloon dilator functions before insertion into working channel	☐
NODAL POINT	Balloon passed through stricture and positioned with 1cm cuff visible	☐
	DO NOT LET CANDIDATE PROCEED UNTIL BALLOON POSITIONED APPROPRIATELY	
	Maintains position of balloon within stricture during inflation	☐
	Maintains inflation for between 20-60 seconds	☐
	Uses incremental balloon inflation x 3 as appropriate	☐
Section 4: Post dilatation		
	Does not withdraw balloon until completely deflated	☐
	Attempts to pass stricture with the endoscope (either with balloon still inflated or after deflation)	☐
	Withdraws endoscope in a controlled manner	☐
Section 5: Documentation		
	Asks to document procedure and gives instructions for after care	☐

Percutaneous Endoscopic Gastrostomy (PEG) Checklist

Please tick the appropriate boxes to the right of the tasks when the candidate has completed them. Any boxes left blank will indicate that the candidate did not complete that task successfully.

Section 1: Checking basic function	
Checks endoscope functions (air, water, suction, control wheels)	☐
Makes appropriate adjustments	☐
Checks PEG kit (PEG, gloves, Betadyne, lignocaine) available	☐
Section 2: Identification of site:	
Chooses appropriate site (no scarring, polyps, pathology)	☐
Checks finger indentation	☐
Uses transillumination	☐
Visualisation of green needle	☐
NODAL POINT	DO NOT LET CANDIDATE PROCEED UNTIL APPROPRIATE SITE CHOSEN AND 2 OF 3 CHECKS COMPLETED
Section 3: Insertion of trocha	
Trocha insertion visualised clearly	☐
NODAL POINT	DO NOT LET CANDIDATE PROCEED UNLESS ABOVE BOX COMPLETE
Section 4: Thread snaring	
Instrument check before insertion into working channel	☐
Thread snared or caught successfully	☐
Scope withdrawn successfully without loss of thread	☐
Section 5: Placement of PEG	
Checks exit site suitably sized for PEG (may require further incision)	☐
PEG attached securely to thread	☐
PEG pulled through in a controlled manner	☐
Section 6: Check PEG placement	
Scope re-inserted in a controlled manner	☐
Bumper visualised in situ	☐
Skin to flange distance noted	☐
PEG secured correctly using attachments	☐
Section 8: Documentation	
Asks to document procedure and gives instructions for PEG care afterwards	☐
Asks to test aspirate for pH	☐

Global Rating

Please circle the appropriate number (1 to 5) corresponding to the candidate's performance in each section. Guidelines are given in the text below each section.

GLOBAL RATING EVALUATION OF ENDOSCOPIC PERFORMANCE				
Knowledge of procedure				
1	2	3	4	5
<i>Insufficient knowledge Looked unsure and hesitant (Had to be stopped at one or more nodal points)</i>		<i>Knew all important steps of procedure (did not have to be stopped at any nodal points), but some mistakes or omissions</i>		<i>Demonstrated familiarity with all steps of procedure. No mistakes or omissions</i>
Communication with assistant				
1	2	3	4	5
<i>Poor communication. No evidence of teamwork.</i>		<i>Reasonable dialogue but occasional communication difficulty or uncertainty</i>		<i>Clear instructions and dialogue with assistant. Clear evidence of teamwork.</i>
Handling of scope and accessories				
1	2	3	4	5
<i>Repeatedly makes awkward or tentative moves with instruments through inappropriate use</i>		<i>Competent use of instruments but occasionally appeared stiff or awkward</i>		<i>Fluid movements with instruments and no stiffness or awkwardness</i>
Flow of procedure				
1	2	3	4	5
<i>Many unnecessary moves. Disjointed or very slow progression of procedure</i>		<i>Demonstrated some forward planning and reasonable progression of procedure</i>		<i>Obviously planned procedure with efficiency from one move to another</i>
Care for tissue				
1	2	3	4	5
<i>Frequently used unnecessary force on tissue or cause damage by inappropriate instrument use</i>		<i>Careful handling of tissue but occasionally caused inadvertent damage</i>		<i>Consistently handled appropriately with minimal damage to tissue</i>
Overall performance				
1	2	3	4	5
<i>Very poor. Requires full supervision and continued basic training.</i>	<i>Adequate but requires continued supervision & further training</i>	<i>Reasonably competent. Capable of performing procedure with minimal supervision</i>	<i>Competent. Capable of performing procedure unsupervised</i>	<i>Clearly superior</i>

Appendix 3 ViCToR course details and learning outcomes

Monday

- Welcome
- VC - a Radiographer's experience
 - To include what VC is, why we do it etc
- Intro to workstations, how CAD works and how to review a case with and without CAD. Including basic segmental anatomy
- Review of test cases

Tuesday

- Observe and assist with VC session
- Radiation dose- how low can we go?
- Anatomy and technique. Including examples of tortuous colon, spasm post surgical colon
- Pitfalls and how to avoid them
 - To include identification of IC valve, faecoliths, residues, complex fold patterns and an appreciation of the mobility of sections of the colon.
Also include reference to tagging
- Bowel prep alternatives. (Delegates already have info sheets in course pack)
 - When and why we use faecal tagging.
- Review of cases – cancers and large polyps

Wednesday

- Teams split:
 - Team 1 visit to Endoscopy to observe cases, interact with staff and appreciate their requirements with regard to prep, information about localisation of polyps for removal etc
 - Team 2 interpretation
- Importance of good communication. Including patient information, consent and compliance with preparation regime.
- Medico-legal aspects of consent
- Interpretation with faecal tagging and standard prep

- VC from a patient's perspective
 - An informal presentation encouraging the delegates to understand the patient's experiences of VC

Thursday

- Teams split:
 - Team 2 visit to Endoscopy to observe cases, interact with staff and appreciate their requirements with regard to prep, information about localisation of polyps for removal etc
 - Team 1 interpretation
- Interface between OC and VC
 - Building upon what was learned in the observation sessions, e.g. why OC is not suitable for all patients, why suitable prep is vital, how our teams support each other and, crucially, why it is essential to maintain excellent links between the 2 departments
- Setting up a VC service
 - To highlight what support is needed e.g. support and mentorship from an enthusiastic Consultant, financial support, staffing, dedicated CT time, support from referring clinicians and Endoscopy etc.
- Final test case review. 5 cases, submitted for assessment

Friday

- Assessment of practical skills and knowledge of technique
- Presentation of certificates.

Learning outcomes of the course

On completion of the course the radiographer will be able to:

- Demonstrate knowledge of the range of bowel preparation regimes available.
- Demonstrate knowledge of the function and safe operation of the manual insufflator.
- Demonstrate competence in the practical skills required to perform a VC.
- Critically evaluate the diagnostic quality of a VC scan.
- Demonstrate knowledge of the radiological appearance at VC of a range of pathologies.
- Identify key concepts of the role of the VC radiographer and evaluate their contribution to advanced practice.
- Critically evaluate the legal, ethical and professional issues relevant to radiographer performed VC.

Appendix 4 VC checklist and Global Score sheet

VC Checklist

Pre-procedure:

Patient assessment

1. Indication for procedure checked ☐
2. Contraindications checked (e.g. glaucoma, contrast allergy) ☐

Consenting

1. Sequence of events during procedure explained fully ☐
2. Risks and potential complications of procedure explained fully ☐
3. Potential additional procedures explained fully ☐
4. Checked patient understanding ☐
5. Radiographer to sign ☐
6. Patient to sign ☐

General

Poor

Excellent

7. Quality of verbal information
given to patient

Procedure

8. Correctly position the patient in the scanner, including movement between prone and supine scans ☐
9. Safely insert the rectal tube ☐
10. Safely inflate the retaining balloon ☐
11. Identify and perform the correct scanner protocol ☐
12. Safely remove the rectal tube ☐

Post-procedure

13. Assess the quality and adequacy of the acquired images ☐
14. Give the patient the correct information on completion of the study ☐

NODAL STOPS ARE IN RED: DO NOT LET TRAINEE CONTINUE UNTIL IT IS SAFE TO DO SO. IF A TRAINEE HAS TO BE STOPPED AT THESE POINTS, PLEASE CIRCLE THE BOX.

VC Global Rating Evaluation

GLOBAL RATING EVALUATION OF VC PERFORMANCE				
Care for patient				
1	2	3	4	5
Little awareness or consideration for patient's needs or comfort during procedure		Demonstrates basic awareness of patient needs and responds appropriately to specific requests		Consistently dealt with patient in a caring and considerate manner with high level of awareness
Pre-procedure				
1	2	3	4	5
Poor interaction with patient with obvious problems in communication		Adequate assessment and explanation of procedure, but with some errors or omissions		High quality information given to patient clearly. Superior assessment of patient understanding.
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 or awkwardness
Flow of procedure				
1	2	3	4	5
Frequently stopped and seemed unsure of next move		Demonstrated some forward planning and reasonable progression of procedure		Obviously planned procedure 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 procedure		Demonstrated familiarity with all steps of procedure
Overall performance				
1	2	3	4	5
Very poor		Competent		Clearly superior

Overall Global Score
(out of 30)

Appendix 5 Literature search for non-technical skills

Non-technical Skills for Surgeons (NOTSS) - Yule et al

Situation Awareness	Gathering information Understanding information Projecting and Anticipating future state
Decision making	Considering options Selecting and communicating option Implementing and reviewing decisions
Communication and Teamwork	Exchanging information Establishing a shared understanding Co-ordinating team
Leadership	Setting and maintaining standards Supporting others Coping with pressure

Non-technical Skills (NOTECHS) – Mishra et al

Leadership & Management (LM)	Leadership Maintenance of standards Planning and preparation Workload management Authority & assertiveness
Teamwork & cooperation (TC)	Team building/maintaining Support of others Understanding team needs Conflict solving
Problem-solving + decision making	Definition and diagnosis Opinion generation Risk assessment Outcome review
Situation awareness (SA)	
Patient	Notice
Procedure	Understand
People	Think ahead

Anaesthetists Non-Technical Skills – Fletcher et al

Task management	Preparation and planning Prioritising Identifying resources Maintaining standards
Teamwork	Coordination with team members Exchange information Use authority/assertiveness Assess capabilities Support others
Situation awareness	Gathering info Recognising and understanding Anticipating
Decision making	Identifying options Selecting options Re-evaluating
Human factors rating scale (HFRS) - Undre	
Communication and interaction	Instructions to assistant - clear + polite Waits for acknowledgment from assistant Instructions to nurse - clear and polite Waits for acknowledgment from nurse
Vigilance/situation awareness	Monitored patient's parameters Awareness of anaesthetist Actively initiates communication during a crisis situation
Team skills	Maintains positive rapport with team Open to opinions from others Acknowledges contribution by others Supportive of other team members Conflict handling: focus on what's right
Leadership & management skills	Adherence to best practice Time management Resource utilization Debriefing the team Authority/assertiveness
Decision making - crisis	Prompt identification of the problem Informed team members Outlines strategy/institutes a plan Anticipates potential problems and prepares contingency plan Option generation

Observational Teamwork Assessment for Surgery (OTAS) - Undre

Task checklist	Patient
	Equipment
	Communications
Team behaviour	Communication
	Co-operation
	Co-ordination
	Shared-leadership
	Monitoring

Divided into Pre, Op and Post phases and 3 stages per phase.

Operating Room Management Attitudes Questionnaire (ORMAQ) Sexton

Leadership-structure
Confidence-Assertion
Information Sharing
Stress and Fatigue
Teamwork
Work Values
Error
Organizational climate

Edinburgh Basic Surgical Training Assessment Form (EBSTAFF)

Technical skills	Baldwin & Paisley
Clinical skills	
Communication	
Teamwork	
Application of knowledge	

Endoscopy National Competency Framework

Communication
Provide information
Keep records
Take responsibility
Review
Planning
Preparation
Coordination across services
Monitoring
Safety awareness
Teamwork
Reporting abnormalities
Decision making
Identify complications

Global Rating Scale
Clinical quality

Consent
Safety
Monitoring
Review
Monitoring
Communication
Feedback

Workforce

Staff support
Career planning
feedback
Risk identification
Adverse incident reporting
Staff valued and contribute
Sharing knowledge

Training

Feedback
Maintaining standards

Appendix 6 Interview schedule for non-technical skills

Introduction

Description of project

- Aim to identify non-technical skills in GI endoscopy and develop a list of behavioural markers
- Liase with National Endoscopy Project to inform development of Training and GRS
- No funding declarations

Use of tape recorder

- To avoid copious note-taking
- Will be transcribed, de-identified then deleted.
- Sound check

Format

Part 1: performance example. 30 mins

Describe a case from your own endoscopic experience that you consider was difficult for you as an endoscopist.

Part 2: Skill identification exercise. 15 mins

Identification of important NTS characteristic of an experienced endoscopist.

Part 3: Sorting task. 15 mins

Rate and group a number of non-technical skills

Focus of interview

- Interested in NTS and behaviours
- No judgement about performance
- No right/wrong answers

Any questions?

Personal information:

- Endoscopic experience (approx)
 - OGD
 - Colon
 - ERCP/other
- Years as a consultant

Part 1: Performance example

You were asked to think of a case from your own experience in endoscopy that you found particularly difficult or challenging.

I will ask you to walk through the case a number of times

1. A brief description of the case
2. I will repeat back the key aspects to check
3. Describe the case again in more detail to focus on the NTS

Please give as much information as possible, but not personal details about the patient or members of staff.

Description of case and development of timeline

I will now repeat the case back to you to check the details and identify the key management points

Repeat back case

I would now like you to go through the case and give a description of what you were thinking, decisions you had to make, communications with colleagues, planning the tasks etc. If I think anything is particularly important, I may ask you further questions to identify particular points.

Go through case

Specific points to identify:

- What teamwork issues arose during the case?
- Who was in the leadership role?
- What information were you using to make your decisions?
- Did you have previous experience for you to draw on?
- How did you decide what option to take? What factors affected your decision?
- What resources did you have to support you? How did you use them?
- Were there any communication issues specific to this case?

Additional questions:

1. Can you tell me why you picked this case? What was so challenging for you?
2. Where do you think things otherwise could have broken down in this scenario?
3. How do you think someone with less experience i.e. a trainee, might have handled this situation?

Part 2: Skill identification exercise

What kinds of non-technical skills are important or make a good endoscopist?

Think about the skills that make a good effective endoscopist (in terms of non-technical skills) and what might distinguish a really experienced endoscopist from a novice. It may help to think about a colleague you have observed or work with.

Question: How do you think these skills are currently developed?

Question: How might a trainee gain these skills?

Question: Do you think there are any differences between the skills needed for normal situations and crisis situations?

Part 3: Sorting task

I have identified some non-technical skills from a literature review. There are no definitions so make your decisions based on what you understand by the term.

Please group them together however you feel appropriate.

Please explain why you put those skills together and what the groups might be called

Please rank each of the skills in terms of importance

Use a scale of 1-5

1= not important at all

2= slightly important

3= fairly important

4= quite important

5= very important