

Imperial College London
Department of Surgery and Cancer

PhD Thesis

PROFICIENCY GAIN AND
COMPETENCY ASSESSMENT IN
LAPAROSCOPIC COLORECTAL
SURGERY

by

Danilo Miskovic MD FRCS

Supervisors:

Professor George B Hanna PhD FRCS

and

Professor Ara W Darzi FMedSci HonFREng KBE

DECLARATION

I hereby declare that the work presented in this thesis is original and I confirm that I have performed this research myself. Collaboration and important input from other individuals are declared in the respective chapters.

.....

Danilo Miskovic

LONDON, 17 MAY 2012

ABSTRACT

Aims and objectives

The aim of this thesis was to investigate the proficiency gain curve and competency in laparoscopic colorectal surgery (LCS) for specialist colorectal surgeons. The objectives were (1) to analyse the shortfalls of proficiency gain curve analysis in LCS, (2) to develop methods to describe proficiency gain and competency in LCS and (3) to analyse proficiency gain and competency within the National Training Programme (NTP).

Methods

Objective 1: Two systematic reviews and a meta-analysis of current evidence were performed. Objective 2: Semi-structured interviews and a Delphi method were applied to develop proficiency and competency assessment tools and clinical cases were used for validation. A novel observational clinical human reliability analysis (OCHRA) of clinical cases was developed and evaluated. Objective 3: clinical and educational data from the National Training Programme were used for the analysis of proficiency gain and competency using advanced statistical methods.

Results

Using clinical data alone is insufficient for the description of proficiency gain and competency. A generic assessment scale (GAS form) and a competency assessment tool (CAT) were developed and validated. A combination of CAT and OCHRA has been shown to be highly sensitive and specific to determine competency. Analysis of clinical and educational data revealed a shortening of the proficiency gain curve using the approach of the NTP without the risk of increased rates adverse outcomes.

Conclusion

Tools for proficiency gain and competency assessment of specialist surgeons have been shown to be valid and feasible and are fully implemented in the NTP. Data suggest that a novel technique can be introduced to specialists on a National level using a structured educational approach with safe clinical outcomes.

TABLE OF CONTENT

ACKNOWLEDGMENTS.....	13
PUBLICATIONS AND PRIZES.....	14
ABBREVIATIONS	18
TABLES	20
FIGURES	23
INTRODUCTION AND LITERATURE REVIEW	27
1.1. Introduction of laparoscopic colorectal surgery	28
1.1.1. A revolution?	28
1.1.2. A historical perspective	29
1.2. Definitions and technical classification	31
1.2.1. Definitions	31
1.2.2. Technical classification	32
1.3. Clinical evidence	35
1.3.1. Major randomised controlled trials.....	35
1.3.2. Meta-analyses of randomised controlled trials	38
1.3.3. On-going trials	40
1.4. Proficiency gain and competency assessment of surgical skills.....	41
1.4.1. Observational skills assessment in surgery	41
1.4.2. Proficiency gain curve	42
1.4.2.1. Background.....	42
1.4.2.2. Measuring proficiency	42
1.4.2.3. Challenges in measuring proficiency gain.....	45
1.4.3. Procedural competency.....	46

1.4.3.1. Competence and competency	46
1.4.3.2. Procedural competency.....	46
1.4.3.3. Challenges in measuring procedural competence.....	47
1.5. The National Training Programme in Laparoscopic Colorectal Surgery (NTP).....	49
1.5.1. The impact of NICE	49
1.5.2. The uptake of LCS in England	49
1.5.3. Aim of the National Training Programme.....	51
1.5.4. Administration and training structure.....	52
1.5.4.1. Organisational structure and administration.....	52
1.5.4.2. Training structure	54
1.5.5. Conclusion	57
1.6. Aims of the thesis	58
1.7. Structure of the thesis	59

2. PROFICIENCY GAIN CURVE AND CASE SELECTION IN LAPAROSCOPIC COLORECTAL SURGERY..... 62

2.1. Introduction	63
2.1.1. Background.....	63
2.1.2. Aim	64
2.2. Methods	65
2.2.1. Summary of methods.....	65
2.2.2. Systematic review.....	65
2.2.2.1. Search strategy.....	65
2.2.2.2. Inclusion and exclusion criteria.....	65
2.2.2.3. Data extraction.....	66
2.2.2.4. Parameters	66
2.2.2.5. Study quality assessment.....	67
2.2.3. Proficiency gain curve analysis	69
2.2.3.1. Statistical software.....	69
2.2.3.2. Risk adjusted CUSUM	69
2.2.3.3. Change point analysis.....	70
2.2.4. Risk prediction analysis.....	70
2.2.5. Expert survey and validation	71

2.3. Results	72
2.3.1. Systematic review	72
2.3.2. Proficiency gain curve	75
2.3.3. Risk prediction.....	78
2.3.4. Expert opinion and recommendations	80
2.4. Discussion.....	83
2.4.1. Summary.....	83
2.4.2. Comparison with results from previous studies	83
2.4.3. Ethical implications	84
2.4.4. Value of case selection	84
2.4.5. Limitations.....	85
2.4.6. Conclusion	86
2.5. Disclosure	87
2.6. Acknowledgments	88

3. SYSTEMATIC REVIEW AND META-ANALYSIS ON THE IMPACT OF SUPERVISION ON CLINICAL OUTCOMES IN LAPAROSCOPIC COLORECTAL SURGERY 89

3.1. Introduction	89
3.1.1. What is supervision?.....	90
3.1.2. Aim	91
3.2. Methods	92
3.2.1. Systematic review.....	92
3.2.1.1. Search strategy.....	92
3.2.1.2. Inclusion criteria	93
3.2.1.3. Exclusion criteria.....	94
3.2.1.4. Study quality assessment.....	94
3.2.2. Meta-analysis.....	96
3.2.2.1. Data extraction.....	96
3.2.2.2. Data analysis.....	96
3.3. Results	97
3.3.1. Results from the literature search	97
3.3.2. Study description and publication bias	101

3.3.3. Subgroup comparison and meta-analysis	102
3.4. Discussion.....	105
3.5. Summary.....	105
3.5.1. Are morbidity and mortality parameters useful for performance assessment?	105
3.5.2. The impact of clinical supervision.....	106
3.5.3. Limitations.....	106
3.5.4. Conclusion	107
3.6. Disclosures.....	108
3.6.1. Acknowledgments	109
 4. DEVELOPMENT, VALIDATION AND IMPLEMENTATION OF A MONITORING TOOL FOR TRAINING IN LAPAROSCOPIC COLORECTAL SURGERY	 110
4.1. Introduction	110
4.1.1. Background.....	110
4.1.2. Formative and summative assessment.....	111
4.1.3. Is there a perfect assessment tool?.....	112
4.1.3.1. Classical test theory	112
4.1.3.2. Advanced concepts: generalizability theory	113
4.1.3.3. Feasibility and implementation	113
4.1.4. Aim	114
4.2. Methods	115
4.2.1. Scale development.....	115
4.2.2. Classical test theory	115
4.2.2.1. Validity	115
4.2.2.2. Reliability	116
4.2.3. Feasibility and implementation	116
4.2.4. Statistical analysis.....	116
4.3. Results	118
4.3.1. Scale development.....	118
4.3.2. Construct validity and reliability	121
4.3.3. Feasibility and implementation	123
4.4. Discussion.....	125

4.4.1.	Strengths of the tool in the context of formative assessment	125
4.4.2.	The use of the tool for proficiency gain evaluation.....	126
4.4.3.	Keys for successful implementation.....	126
4.4.4.	Limitations of the study	127
4.4.5.	Conclusion	127
4.5.	Disclosures.....	128
4.6.	Acknowledgments	129

5. AN OBSERVATIONAL CLINICAL HUMAN RELIABILITY APPROACH (OCHRA) FOR COMPETENCY ASSESSMENT IN LAPAROSCOPIC COLORECTAL SURGERY 130

5.1.	Introduction	130
5.1.1.	Background.....	130
5.1.2.	Human reliability methods in surgery	131
5.1.3.	Aim	132
5.2.	Methods	133
5.2.1.	Surgeons and case selection	133
5.2.2.	Observational Clinical Human Reliability Analysis (OCHRA).....	133
5.2.3.		137
5.2.4.	Dissection-Exposure ratio.....	137
5.2.5.	Expert review.....	138
5.2.6.	Data extraction and statistical analysis	138
5.3.	Results	139
5.3.1.	Case complexity and operating time	139
5.3.2.	Construct validity	139
5.3.3.	D/E ratio	142
5.3.4.	Concurrent validity	142
5.4.	Discussion.....	144
5.4.1.	Summary.....	144
5.4.2.	Interpretation of results.....	144
5.4.3.	Limitations.....	146
5.4.4.	Conclusion	146
5.5.	Disclosures.....	147

6. DEVELOPMENT OF A COMPETENCY ASSESSMENT TOOL (CAT) FOR THE TRANSITION FROM SUPERVISED TO UNSUPERVISED PERFORMANCE	148
6.1. Introduction	149
6.1.1. Background.....	149
6.1.2. Process	149
6.1.3. Competency assessment in laparoscopic colorectal surgery	149
6.1.4. Aim	150
6.2. Methods	151
6.2.1. Development of competency assessment tool (content validity)	151
6.2.1.1. Interview studies.....	151
6.2.1.2. Delphi method	151
6.2.2. Validity and reliability studies.....	152
6.2.2.1. Subjects and materials	152
6.2.2.2. Assessors	152
6.2.2.3. Concurrent validity	152
6.2.2.4. Construct validity	153
6.2.2.5. Reliability	153
6.2.3. Implementation.....	155
6.2.4. Software and analysis	156
6.3. Results	157
6.3.1. Content validity	157
6.3.1.1. Outcomes from the interview study.....	157
6.3.1.2. Outcomes of the Delphi method.....	160
6.3.1.3. Scale development.....	160
6.3.2. Validity and reliability studies.....	163
6.3.2.1. Concurrent validity	163
6.3.2.2. Construct validity	163
6.3.2.3. Reliability	166
6.3.3. Cut-off scores and implementation.....	167
6.4. Discussion.....	168
6.4.1. Summary.....	168
6.4.2. Appraisal.....	168

6.4.3. Limitations.....	169
6.4.4. Clinical implications and conclusions	170
6.5. Disclosures.....	171
6.6. Acknowledgments	172
 7. IMPACT OF THE NATIONAL TRAINING PROGRAMME ON PROFICIENCY GAIN AND COMPETENCY IN LAPAROSCOPIC COLORECTAL SURGERY	 173
7.1. Introduction and aims	173
7.2. Methods	175
7.2.1. Database.....	175
7.2.2. Proficiency gain curve analysis	176
7.2.2.1. Clinical data (RA CUSUM)	176
7.2.2.2. Educational data (GAS forms)	176
7.2.3. Statistical analysis.....	177
7.3. Results	178
7.3.1. Clinical outcomes	178
7.3.2. Proficiency gain curves	181
7.3.2.1. Proficiency gain curve using clinical data.....	181
7.3.2.2. Proficiency gain curve using GAS scores	185
7.3.2.3. Learning rate.....	187
7.3.3. Competency assessment	189
7.3.3.1. Data summary	189
7.3.3.2. Determinants of successful competency assessment.....	190
7.4. Summary of results	193
 8. DISCUSSION.....	 195
8.1. Summary of results.....	195
8.2. Discussion of objectives	197
8.2.1. Objective 1: <i>to identify shortfalls in the evaluation of the proficiency gain curve in laparoscopic colorectal surgery</i>	197

8.2.1.1. Ethical concerns.....	197
8.2.1.2. Reliability of clinical outcome data for the description of the proficiency gain curve 198	
8.2.2. Objective 2: <i>to develop methods to describe the proficiency gain curve in laparoscopic colorectal surgery</i>	199
8.2.3. Objective 3: <i>to analyse proficiency and competency gain for specialists colorectal surgeons within the National Training Programme</i>	201
8.2.3.1. Epidemiological implications	201
8.2.3.2. Educational implications	202
8.3. Future research	204
9. REFERENCES.....	209

ACKNOWLEDGMENTS

First of all I have to thank my supervisor Professor George Hanna for his support towards the completion of this thesis. George is an exceptional surgeon and has become a friend and a true mentor, far beyond his role as a supervisor. His tireless effort to aim for the best, his principle of always placing quality above productivity, precision above speed and choosing the right and not necessarily the easiest way, have all been great inspiration for my research and my surgical career..

I would also like to thank my co-supervisor The Right Honourable Professor The Lord Ara Darzi for giving me the opportunity to work in a unique research environment where there is true academic spirit; a place where opportunities seem limitless.

A number of people have been crucial for the completion of this work and I have acknowledged their input in the relevant chapters. Some have been instrumental in my development as a researcher. I am grateful to Professor Amjad Parvaiz for his outstanding capabilities as a surgical teacher and his immaculate surgery, Mr Robin Kennedy for his vast experience and his stimulus to get me started, Ms Susannah Wyles for being my “research buddy”, Professor Roger Kneebone for pointing me in the right direction, Dr Chris Schlachta for his encouragement that I was on the right track and Dr Melody Ni for opening my eyes to recognise opportunities rather than obstacles in a world of numbers and equations.

My wife Alice is my inspiration, making me proud and happy every single day of my life.

PUBLICATIONS AND PRIZES

The following papers, presentations and academic prizes originated from work for this PhD thesis.

Papers in peer-reviewed journals

D Miskovic, M Ni, SM Wyles, NK Francis, A Parvaiz, RH Kennedy, TA Rockall, AM Gudgeon, MG Coleman, GB Hanna. The English National Training Programme Competency Assessment System for Laparoscopic Colorectal Surgery. *Ann Surg* 2012 (accepted, in print)

D Miskovic, M Ni, SM Wyles, P Tekkis, GB Hanna. Learning curve and case selection in laparoscopic colorectal surgery – an international multi-centre analysis of 4852 cases. *Dis Colon Rectum* (accepted, in print)

D Miskovic, M Ni, SM Wyles, A Parvaiz, GB Hanna. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surg Endosc*. 2012 Mar;26(3):796-803. Epub 2011 Nov 1.

Miskovic D, Wyles SM, Coleman MG, Hanna GB. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. *Surg Endosc* (2011) 25:1136–1142. PMID: 20835723

Miskovic D and Wyles SM, Ni M, Darzi AW, Hanna GB. Simulation and Mentoring in Laparoscopic Colorectal Surgery as an illustration of training in technically advanced procedures. A systematic review. Ann Surg 2010 252(6):943-951. PMID: 21107103

Presentations and academic prizes

D Miskovic, SM Wyles, A Parvaiz, C Cunningham, MG Coleman (presenting), GB Hanna. Educational and Clinical Outcomes of the National Training Programme in Laparoscopic Colorectal Surgery. Tripartite meeting, Cairns/Australia, July 2011

D Miskovic, M Ni, SM Wyles, A Parvaiz, GB Hanna, Validation of observational clinical human reliability assessment (OCHRA) in laparoscopic colorectal surgery performed by specialists. Society of American Gastrointestinal and Endoscopic Surgeons (SAGES), San Antonio TX/ USA, April 2011

D Miskovic, SM Wyles, A Parvaiz, C Cunningham, MG Coleman, GB Hanna. Outcomes of the National Training Programme in Laparoscopic Colorectal Surgery. ALSGBI November 2010, Nottingham/UK, **David Dunn Medal for best annual research paper**

D Miskovic, SM Wyles, NK Francis, TA Rockall, RH Kennedy, GB Hanna. L-CAT – Laparoscopic colorectal competency assessment tool for specialists within the national training programme in England. EAES (European Association of Endoscopic Surgeons) June 2010, Geneva/ Switzerland.

D Miskovic, SM Wyles, A Darzi, GB Hanna. Fifteen years of learning curves in laparoscopic colorectal surgery – what have we learnt? EAES (European Association of Endoscopic Surgeons) June 2010, Geneva, Switzerland.

D Miskovic, SM Wyles, MG Coleman, GB Hanna. The UK National Training Programme in Laparoscopic Colorectal Surgery – can we do it in Switzerland? SGC (Swiss Surgical Association) May 2010, Interlaken, Switzerland.

D Miskovic, SM Wyles, A Darzi, GB Hanna. How to measure the learning curve for laparoscopic colorectal surgery: a systematic review of the literature. SGC (Swiss Surgical Association) May 2010, Interlaken, Switzerland.

D Miskovic, SM Wyles, MG Coleman, GB Hanna. A National Training Program for laparoscopic colorectal surgery in the UK. Society of American Gastrointestinal and Endoscopic Surgeons (SAGES), Washington DC, USA, April 2010

D Miskovic, SM Wyles, M Coleman, BG Hanna. Global Assessment Score (GAS) – A Simple Assessment Tool for Training in Laparoscopic Colorectal Surgery. Section of Coloproctology, Royal Society of Medicine London, November 2009, **John of Arderne Medal for best short paper presented at the Section.**

D Miskovic, SM Wyles, MG Coleman & GB Hanna. The national training programme in laparoscopic colorectal surgery produces expert clinical outcomes during training. Association of Coloproctology of Great Britain and Ireland, Bournemouth, July 2010

D Miskovic, SM Wyles, M Coleman, BG Hanna. Educational Impact of the National Training Programme in Laparoscopic Colorectal Surgery – Preliminary Results after 1 Year. Association of Laparoscopic Surgeons of Great Britain and Ireland, Tonbridge/Kent, November 2009

(Poster) – **Prize for 4th best poster**

ABBREVIATIONS

ABMS	American Board of Medical Specialities
ARCP	Annual review of competency progression
ASA	American Society of Anesthesiologists
BMI	Body mass index
CAT	Competency assessment tool
CL	Confidence level
COLOR	Colon cancer laparoscopic open resection (trial)
COST	Clinical Outcomes of Surgical Therapy Study Group (trial)
CT	Computed tomography
CUSUM	Cumulative Sum
DM	Danilo Miskovic (investigator)
EnROL	Enhanced recovery for open versus laparoscopic surgery (trial)
FAP	Familial adenomatous polyposis
GAS	Global assessment scale
GH	George Hanna (investigator)
HAR	High anterior resection
HALS	Hand assisted laparoscopic surgery
IBD	Inflammatory bowel disease
IMA	Inferior mesenteric artery
JCOG	Japanese Clinical Oncology Group
LAR	Low anterior resection
LCS	Laparoscopic colorectal surgery
LN	Lymph node
LOS	Length of (hospital) stay
MRC CLASICC	Medical Research Council Conventional versus Laparoscopic-Assisted Surgery in Colorectal Cancer (trial)
MRI	Magnetic resonance imaging
N/A	non applicable (also n/a)
NHS	National Health Service
NICE	National Institute for Clinical Excellence
NOTES	Natural orifice endoscopic surgery
NTP	National Training Programme (in laparoscopic colorectal surgery)
OCHRA	Observational Human Reliability Approach
PBA	Procedure based assessment
PRISMA	Transparent reporting of systematic reviews and meta-analyses
RA-CUSUM	Risk adjusted cumulative sum
RCT	Randomised controlled trial
SMW	Susannah M Wyles (investigator)
TA	Technology guidance appraisal
TME	Total mesorectal excision

UC	Ulcerative colitis
UK	United Kingdom
USA	United States of America

TABLES

Table 1.	Summary of evidence from the major multi-centre randomised controlled trials (USA, UK and Europe). The arrows indicate a significant higher (↑) or lower (↓) value for LCS compared to open surgery. Furthermore, “no significant difference” (ns.) and “no available data” (-) are indicated.	37
Table 2.	Summary of the results of meta-analyses on the outcomes of laparoscopic colorectal surgery. After 2006 the main focus was on long-term results. (“ns.” non significance, “-“ no data)	39
Table 3.	Ongoing multicentre trials on the feasibility of laparoscopic colorectal surgery.	40
Table 4.	Evolution of skills based on the Dreyfus and Dreyfus model. From: (Eraut 1994) .	47
Table 5.	Checklist of PRISMA guideline according to Moher 2009.....	68
Table 6.	Quality assessment of included studies. For <u>underlined</u> studies detailed data sets (raw data) were obtained. *The case numbers of these datasets were higher than published in the original papers.	74
Table 7.	Summary of clinical outcomes.	75
Table 8.	Summary of confidence levels and change points for different parameters.	77
Table 9.	Risk predictions for conversion.	79
Table 10.	Risk prediction for complication (diagnosis were not significant).....	79
Table 11.	Expert survey results. The interval scale ranged from 0 (least complexity) to 100 (highest complexity). The list was split into four categories (complexity level) according quartiles calculated from original data (30, 60, 80).	81
Table 12.	Recommendations for case selection during the learning curve.....	81
Table 13.	Search strategy.	93
Table 14.	PRISMA guidelines according to Moher 2009.	95

Table 15.	Summary of included studies.....	100
Table 16.	Comparison of different parameters between the supervised and the control subgroups.....	102
Table 17.	Comparison of conversion rates between subgroups and control group.	103
Table 18.	Classical requirements of a good assessment tool (from REF Ahmed 2010)	114
Table 19.	Generic task analysis for laparoscopic colectomy	119
Table 20.	Generic Juster Scale for a non-judgmental assessment	119
Table 21.	Summary of results.	121
Table 22.	External Error Modes (EEM) according to Joice 1998.	135
Table 23.	Generic task analysis for laparoscopic colectomies	136
Table 24.	Overview of identified errors.....	140
Table 25.	Summary of error counts for different error types and task areas and D/E ratios for different groups.....	141
Table 26.	Prediction accuracy for tissue errors and D/E ratio.	143
Table 27.	Items mentioned for competency skills. Ranks according to final Delphi round (round 2). Number of experts giving the answer (n), average scores after each round and standard deviations (SD) are tabulated. Δ SD demonstrates increasing consistence after round 2.	159
Table 28.	Structure of the CAT form.....	161
Table 29.	Summary of mean scores (standard deviation) for each item of the scoring sheet. 165	
Table 30.	Summary of patient demographics	178
Table 31.	Proportions of indications for surgery (pie chart represents results from both NTP and registrars)	179

Table 32. Types of procedure performed (pie chart represents results from both NTP and registrars)	179
Table 33. Summary of clinical outcomes	180
Table 34. Learning rates according to the power law curve ($y=ab^k$). The numbers represent the variable k (learning rate).....	187
Table 35. Numbers of trainees from different training centres (NW – North West, SW- South West training region).	189
Table 36. Summary of competency assessment outcomes. The pass rate was 71%. There was a significant difference for numbers of training cases performed within the programme (supervised training). There was no difference for the total numbers of cases trainees have performed before entering or at the end of the programme. ..	190
Table 37. Summary of GAS and CAT scores. CAT scores are significantly different between delegates who passed and those who failed the assessment. Overall GAS scores did not differ. However, delegates who failed tended to have a higher discrepancy between the scores given by their trainers and their self-assessment.	191
Table 38. Univariate analysis of factors for prediction of successful sign off.	192
Table 39. Multivariate analysis including covariates that had a $p<0.1$ in the univariate analysis.	192

FIGURES

Figure 1. Survey of 1561 US surgeons: surgeons without appropriate training caused bile duct injuries early in their career, suggesting these severe complications to be a direct consequence of lack of training (Archer 2001)	29
Figure 2. Synopsis of different technical approaches for minimally invasive colorectal surgery.	34
Figure 3. Power law of practice.	44
Figure 4. Example for strategy shift. Two groups used different strategy to solve the same task resulting in differences in Response times (RT) and learning rates (α) expressed by the steepness of the logarithmic line. (adapted from Delaney).....	44
Figure 5. Uptake of LCS in England from 2000-2007 (Source: Hospital Episode Statistic (HES))......	51
Figure 6. Organisational structure of the NTP. Delegates are allocated to one of the trainers who belong to one of the 11 training centres, represented by a senior laparoscopic surgeon. The clinical and educational chairmen lead the programme through a steering group.	53
Figure 7. The training pathway for NTP delegates.	54
Figure 8. Flowchart for study selection.....	73
Figure 9. Proficiency gain curves (RA-CUSUM) for different parameters. Dashed lines and numbers indicate the length of the proficiency gain curve (change points). *Lymph node harvest and mortality did not reach sufficient confidence levels.....	76
Figure 10. Agreement of experts responding to the survey regarding the guidelines and its different components. Most disagreement was that BMI limits were too restrictive and that previous abdominal surgery should be considered for inclusion in the guidelines.	82
Figure 11. Flowchart of inclusion and exclusion of studies.....	98

Figure 12. Funnel plot for studies of the supervised subgroup	101
Figure 13. Forrest plot for weighted conversion rates within the supervised subgroup	103
Figure 14. Forrest plot for weighted conversion rates within the non- supervised subgroup .	104
Figure 15. Final version of the paper form of the assessment tool	120
Figure 16. Construct validity for different experience groups.	122
Figure 17. Bland-Altman plot for inter-rater reliability.	122
Figure 18. CUSUM proficiency gain curves for different areas of the procedure.	124
Figure 19. Examples of proficiency gain curves of four individual surgeons.	124
Figure 20. Screenshot of the video rating software used for analysis (Observer XT, Noldus software)	137
Figure 21. Construct validity for different groups for all errors	141
Figure 22. Construct validity using D/E ratio.	142
Figure 23. AUC curves for (a) overall errors, (b) tissue errors only and (c) tissue errors and D/E ratio.	143
Figure 24. Process of competency assessment at the end of supervised training as suggested by the educational committee	155
Figure 25. Finalised version of the CAT form.	162
Figure 26. There was a significant difference of averaged CAT scores between the expert and novice group ($p=0.004$). The main effects could be found between apprentices who passed and those who failed, whereas the difference between experts and apprentices who passed was only moderate.	164
Figure 27. D study for increasing numbers of assessors and observed cases. The critical G coefficient of 0.8 is reached with three assessors observing two cases, or two assessors observing three cases.	166

Figure 28. The scattered plot of individual cases suggests decreasing CAT scores with increasing number of observed errors using the OCHRA method (CAT scores are averaged for two assessors). The solid lines indicate the cut-offs for CAT at 2.5 and for OCHRA at 16.75. The dotted line represents the ideal cut-off for CAT. The grey rectangle represents the “passing” area.	167
Figure 29. Each surgeon has a personalised account to enter the data for each training case.	175
Figure 30. Graphical illustration (histogram) of hospital stay patients operated by NTP delegates and registrars.	180
Figure 31. ROC curves for the prediction model using logistic regression for conversion (left) and complication (right).	181
Figure 32. Number of surgeons submitting data at different stages during the proficiency gain curve.	182
Figure 33. Proficiency gain curve for conversions. The curve remains within the boundaries of 4 standard deviations, indicating that the process within limits.	183
Figure 34. Proficiency gain curve for complications. The curve remains within the safety limits of 4 standard deviations.	183
Figure 35. Subgroup analysis for NTP delegates and registrars for conversions and complications. None of the curves is crossing the boundary of 4 standard deviations (x: procedure number, y: CUSUM scores)	184
Figure 36. Summary of GAS scores for NTP delegates.	185
Figure 37. Summary of GAS scores for registrars	186
Figure 38. Power curve fit model for learning rates for NTP delegates (continuous lines) and registrars (dashed lines) for different subsections of the GAS form (x: procedure number, y: GAS score).	188
Figure 39. Predicted probability using the discrepancy as a sole factor for the prediction of passing the sign off. A cut off at a discrepancy of 0.3 reaches highest sensitivity (72.7%) and specificity (77.8%).	191

Figure 40. ROC curve for multivariate prediction model (Area under the curve (AUC)=0.909)

.....192

INTRODUCTION AND LITERATURE REVIEW

1.1.	Introduction of laparoscopic colorectal surgery.....	28
1.1.1.	A revolution?.....	28
1.1.2.	A historical perspective.....	29
1.2.	Definitions and technical classification.....	31
1.2.1.	Definitions.....	31
1.2.2.	Technical classification.....	32
1.3.	Clinical evidence.....	35
1.3.1.	Major randomised controlled trials.....	35
1.3.2.	Meta-analyses of randomised controlled trials.....	38
1.3.3.	On-going trials.....	40
1.4.	Proficiency gain and competency assessment of surgical skills.....	41
1.4.1.	Observational skills assessment in surgery.....	41
1.4.2.	Proficiency gain curve.....	42
1.4.2.1.	Background.....	42
1.4.2.2.	Measuring proficiency.....	42
1.4.2.3.	Challenges in measuring proficiency gain.....	45
1.4.3.	Procedural competency.....	46
1.4.3.1.	Competence and competency.....	46
1.4.3.2.	Procedural competency.....	46
1.4.3.3.	Challenges in measuring procedural competence.....	47
1.5.	The National Training Programme in Laparoscopic Colorectal Surgery (NTP).....	49
1.5.1.	The impact of NICE.....	49
1.5.2.	The uptake of LCS in England.....	49
1.5.3.	Aim of the National Training Programme.....	51
1.5.4.	Administration and training structure.....	52
1.5.4.1.	Organisational structure and administration.....	52
1.5.4.2.	Training structure.....	54
1.5.5.	Conclusion.....	57
1.6.	Aims of the thesis.....	58
1.7.	Structure of the thesis.....	59

In this thesis the proficiency and competency gain process in laparoscopic colorectal surgery for colorectal specialists (consultants) is investigated. This introductory chapter describes the challenges of the introduction of laparoscopic colorectal surgery, a literature review of the clinical evidence and problems of proficiency and competency assessment. Based on the current evidence, research deficits are identified and aims and objectives of the thesis are defined.

1.1. Introduction of laparoscopic colorectal surgery

“In art there are only fast or slow developments. Essentially it is a matter of evolution, not revolution.”

Béla Bartók, composer, 1881-1945

1.1.1. A revolution?

Laparoscopic colorectal surgery (LCS) was introduced more than 20 years ago. Unlike other minimally invasive procedures, its introduction was slow and is incomplete. Whilst this may be explained by initial concerns about oncologic safety, technical complexity and a long learning curve, it still seems surprising. The potential impact of the minimal invasive concept on the individual patient and its socioeconomic consequences are as ground breaking in colorectal surgery as for any other surgical speciality (Delaney 2003; Schwenk 2005). Twenty years on, a great deal of evidence has been published and the sceptics are getting fewer, but it seems that the “revolution” is still rolling out. This long drawn-out transition period may be an opportunity to avoid the repetition of mistakes made in the past. During the rapid introduction of laparoscopic cholecystectomy in the early 1990s, thousands of patients had complications arising from insufficient or non-existent formal surgical training in the technique (Archer 2001) (**Figure 1**). This demonstrates that the introduction of a new procedure to a wider surgical community is complex and it may not be sufficient to tackle the problem by surgeons simply attending a one-day course.

There is little evidence about the clinical consequences of the introduction of laparoscopic colorectal surgery. A critical view of the COST trial suggests that the quality of laparoscopic surgery at this early stage may have resulted in an increase relative risk for mortality of 16% and 11% for recurrence (Tinmouth 2004). Although data from the COST trial is often

considered to be “historical” it may be representative of the difficulties of the introducing novel technology and adverse events may have been underestimated.

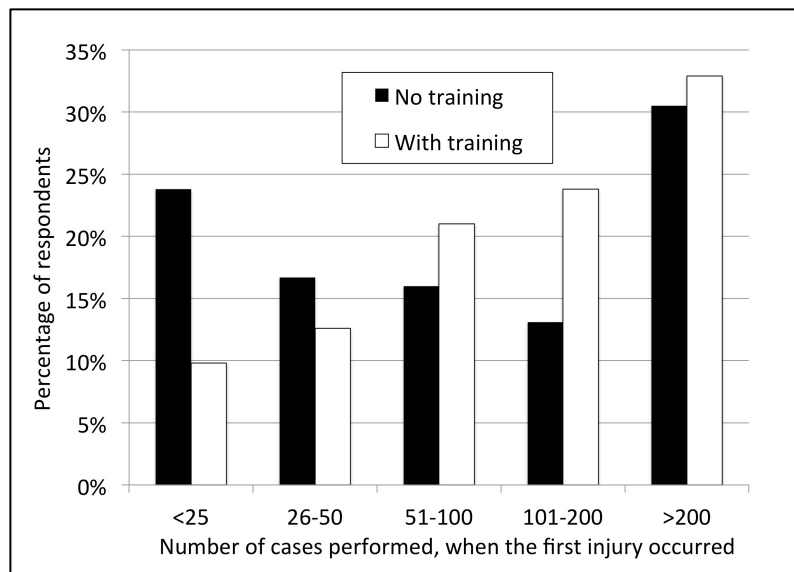


Figure 1. Survey of 1561 US surgeons: surgeons without appropriate training caused bile duct injuries early in their career, suggesting these severe complications to be a direct consequence of lack of training (Archer 2001)

1.1.2. A historical perspective

The first description of laparoscopic colorectal surgery in the medical literature was a series of 20 laparoscopic colonic resections by Moises Jacobs, Juan-Carlos Verdeja and Harold Goldstein from Miami in 1991. In their conclusion they proposed the technique would become “as popular as laparoscopic cholecystectomy” (Jacobs 1991). However, the on-going development of LCS is hardly comparable with the evolution of laparoscopic cholecystectomy. Laparoscopic cholecystectomy almost completely replaced the open technique within a few years after its introduction, and its benefits were so obvious that the range of indications was widened and a rise in the overall rate of cholecystectomies was observed (Legorreta 1993). This rapid development was certainly not reproduced in colorectal surgery. In the 1990s, the two main concerns arising from various reports on LCS were that the technique may not be oncologically safe and there were doubts about any clinical benefit at all. Both issues were

debated intensively. Concerns regarding oncologic risks were dominated by reports on a rise in abdominal wall metastases (port site metastases) (Wexner 1995; Silecchia 2002). There was confusion amongst the surgical community about pathogenesis and the actual risk of this complication (Rae 1995; Tseng 1998). The deeper understanding of aetiology, identification of risk factors and technical innovation gradually led to a decrease in port-site metastases and to an acceptably low level (Curet 2004). Randomised single centre studies provided evidence for clinical benefits, such as shorter hospital stay and faster recovery from the mid 1990s (Lacy 1995; Milsom 1998; Schwenk 1998). However, since no randomised controlled multicentre trial was available until the early 2000s, many of these claims remained controversial until publication of high-level evidence which confirmed the benefits and safety of LCS compared with open surgery (see **chapters 1.3.1-2**).

1.2. Definitions and technical classification

1.2.1. Definitions

There is no official definition for laparoscopic colorectal surgery. In accordance with the definition of laparoscopic surgery, it could be defined as surgical procedures on the colon and rectum using a fibreoptic endoscopic camera and specialised surgical instruments, inserted into the peritoneal cavity through small incisions in the abdominal wall (Merriam-Webster 2011).

For this thesis the term LCS is reserved for laparoscopic surgery used for partial or total resections of the colon and/or rectum. Hence, certain procedures, such as laparoscopic rectopexy or laparoscopic procedures on the small bowel or appendix are not included unless mentioned specifically. The same applies to other endoscopic modalities such as trans-luminal procedures (trans-anal, rectoscopic and colonoscopic procedures) whereby the camera and instruments are introduced through the anus and the procedure is performed on the endoluminal aspect of the bowel.

For each operation there is a laparoscopic phase, starting with the initial insertion of the camera into the peritoneal cavity and an open phase which includes the preparation of the patient, the insertion of the first trocar and closure of abdominal wall and skin. Depending on the technique certain parts of the procedure, e.g. transection of bowel or an extracorporeal anastomosis can be performed outside the abdomen by pulling the specimen through a small incision. These extracorporeal steps are usually planned as part of a routine laparoscopic procedure. Any *unplanned* laparotomy can be defined as conversion to open surgery (Shawki 2009).

1.2.2. Technical classification

There are a variety of laparo-endoscopic techniques for the resection of colon and rectum.

Within the wider context of minimally invasive techniques, they can be classified according to the following characteristics (see also **Figure 2**):

I. Site of resection

- Segmental (wedge) resection (anywhere in the colon)
- Ileocoecal resection
- (Extended) right hemicolectomy
- (Extended) left hemicolectomy
- (Recto-)sigmoid resection
- Anterior/ Abdominoperineal resection (excision) of the rectum
- (Sub-) total colectomy or proctocolectomy

II. Instrumentation

(Fully) laparoscopic: All parts of the procedure are carried out through the laparoscopic approach, with only removal of the specimen at the end of the procedure requiring a larger incision.

Laparoscopically-assisted: The main parts of the operation are performed by a laparoscopic approach, but towards the end of the procedure a planned small incision is performed to exteriorise parts of the colon in order to transect bowel and/or perform an anastomosis. This technique is often also referred to as *laparoscopic*.

Hand-assisted laparoscopic surgery (HALS): in addition to laparoscopic instrument a planned incision is performed in order to allow insertion of the surgeon's hand surgeon into the peritoneal cavity. This (usually non-dominant) hand acts as an additional instrument.

Robotic laparoscopic surgery: the whole or parts of the procedure are performed by introducing one or multiple laparoscopic instruments that are attached to a robotic arm. The instruments can be remotely operated by the surgeon, with a high level of precision.

III. Access

Conventional laparoscopy: A camera port and at least one, but usually two additional laparoscopic instruments are introduced.

Single access techniques: The camera and all instruments are inserted through one access, using a single-port system or conventional trocars that are inserted through the same skin incision.

Natural orifice transluminal endoscopic surgery (NOTES): The access port is introduced through a natural orifice (e.g. vagina, rectum or trans-oesophageally through the stomach wall). Auxiliary ports for working instruments may be introduced through the abdominal wall.

IV. Operative approach

Medial-to-lateral: The dissection begins medial to the colon with the transection of the vascular pedicle (ileo-colic artery on the right and inferior mesenteric artery (IMA) on the left side) and the mobilisation progresses laterally.

Lateral-to-medial: The mobilisation starts lateral to the colon (as in open surgery) and progresses medially. The vascular pedicle is usually transected with an open approach after the colon has been exteriorised through a laparotomy.

Inferior-to-superior (right colon only): some surgeons start the dissection inferiorly, around the caecum, progressing towards the hepatic flexure. The vascular pedicle is transected either laparoscopically or by an open technique.

The training programme described in this thesis was designed to teach conventional laparoscopic and laparoscopically-assisted techniques for all types of resection. There was no rule for the type of technical approach (depending on the individual trainer's preference), but additional instrumentation, such as HALS and robotic surgery were not taught. For cancer resections, current technical standards were preconditioned (lymph node count, high tie, total mesorectal excision (TME)).

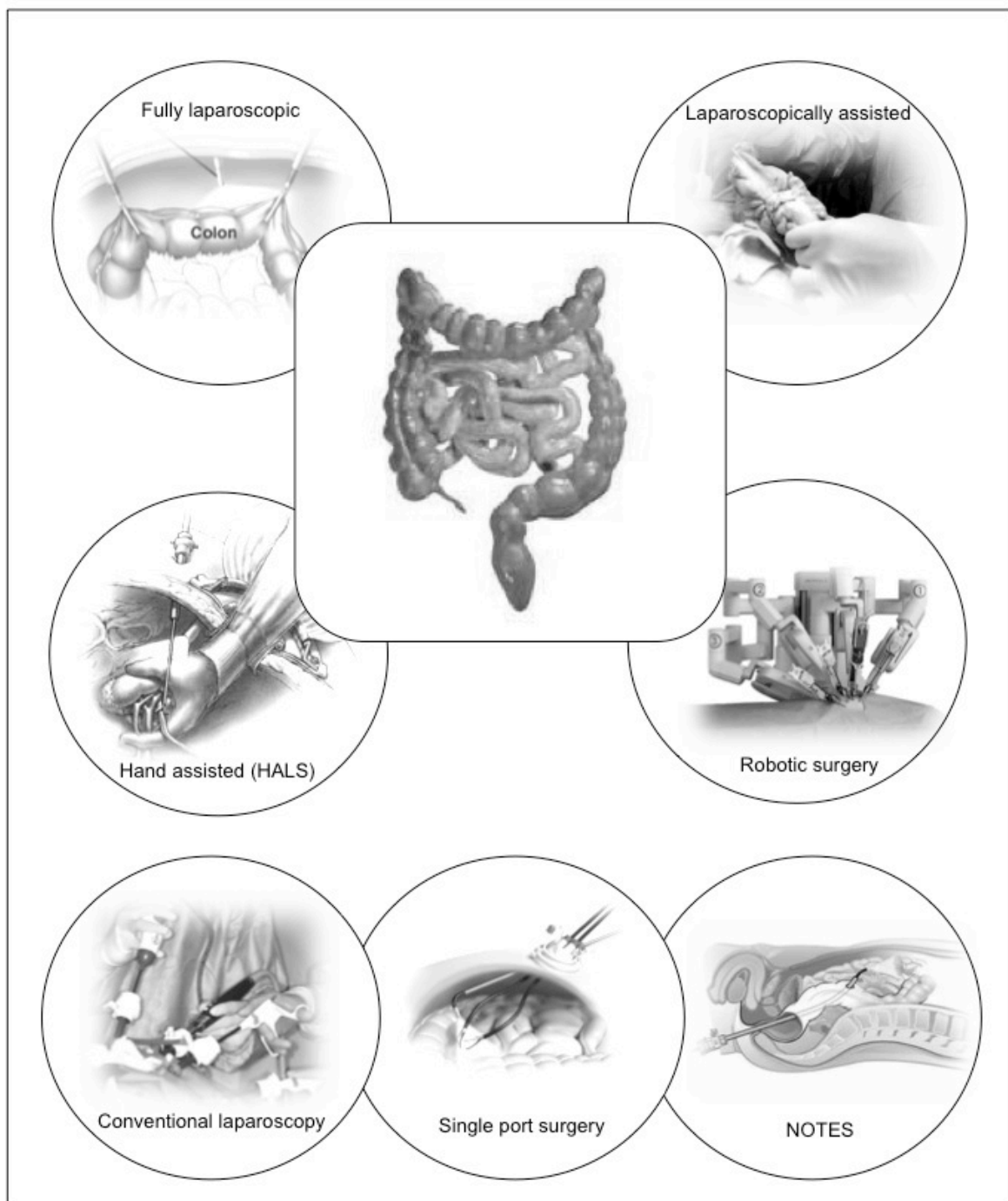


Figure 2. Synopsis of different technical approaches for minimally invasive colorectal surgery.

1.3. Clinical evidence

1.3.1. Major randomised controlled trials

Early randomised controlled trials provided evidence for the short-term benefits of LCS compared to open surgery. Lacy et al. performed the first randomised, prospective study comparing open and laparoscopic colectomies in 1995. The authors of this study produced evidence for earlier oral intake and shorter hospital stay in the laparoscopic group (Lacy 1995). The follow-up data of the series in 2002 suggested that a laparoscopic approach was an independent factor for both lower mortality and tumour recurrence rates for stage III cancer (Lacy 2002). These results however were heavily criticised due to small case numbers, inappropriate statistical methods and a high recurrence rate in the open group. It was suggested that from their data only equivalence with the open group could be claimed (Ceulemans 2003; Evrard 2003). Leung et al. further supported this in their prospective randomised trial showing no difference in overall and disease-free survival rates (Leung 2004).

Solid evidence on the oncologic efficacy and safety of LCS was demonstrated in four major multicentre studies (USA (COST 2004), UK (Guillou 2005; Jayne 2007), Europe (Veldkamp 2005; Buunen 2009b), Germany (Neudecker 2009)). Some trials are still being followed up (see **Chapter 1.3.3.**). Although the common aim of all studies was to prove feasibility of LCS for colonic cancer, the primary endpoints of each study were different. COST (“Clinical Outcomes of Surgical Therapy Study Group”) was set up in 1994 and was designed to test whether LCS is equivalent to open surgery in terms of short-term, postoperative outcome parameters and long-term survival and recurrence rates (Nelson 1995; COST 2004). Data supported evidence of LCS being oncologically safe. The UK trial MRC CLASICC (“Medical Research Council Conventional versus Laparoscopic-Assisted Surgery in Colorectal Cancer”) was designed specifically to look at long-term results, with a special emphasis on quality of life

outcomes. Furthermore this was the first multi-centre trial that included rectal cancer (Jayne 2007; Jayne 2010). The European COLOR study included data from centres in Sweden, Finland, Italy, Spain, Germany, France, UK and the Netherlands. Short-term data was presented separately in an early report (Veldkamp 2005). The main aim of the trial, however, was to establish the 3-year disease-free and overall survival rate after open and laparoscopically assisted colon resections (Buunen 2009b). In the German study (LAPKON II) patients were randomized after excluding cases with tumours deemed unsuitable for laparoscopic resection (Neudecker 2009). The major outcome parameters of all trials are summarised in **Table 1**.

Parameter	COST 2004	MRC CLASICC 2005	MRC CLASICC 2007/2010	COLOR 2005	COLOR 2009	LAPKON II 2009
Country	USA	UK	UK	Europe	Europe	Germany
Main aim	Short-term and long-term results	Short-term results	Long-term results	Short-term results	Long-term results	Short-term results
Patients included	863	794	794	1082	1076	472
Laparoscopic group	435	526	526	536	534	250
Operating time	↑	↑	-	↑	↑	↑
Blood loss	-	-	-	↓	↓	-
Histopathological outcome:						
- Lymph node retrieval	-	ns.	-	-	ns.	↓ ²
- Disease-free margins	ns.	ns.	-	-	ns.	ns.
Hospital stay	↓	↓	-	↓	-	↓
Postoperative complications:						
- overall	ns.	ns. ¹	-	ns.	ns.	ns.
- surgical	ns.	ns.	-	ns.	ns.	ns. ³
- anastomotic leakage	-	ns.	-	ns.	ns.	ns.
- 30-day mortality	ns.	ns.	ns.	ns.	ns.	ns.
Conversion rate % (for laparoscopic group only)	21	17	17	29	29	11.2
Long-term results:						
- Disease-free survival	ns.	ns.	ns.	-	ns.	-
- Mortality	ns.	ns.	ns.	-	ns.	-
- Quality of life scores	ns.	ns.	ns.	-	-	-
¹ Patient with conversion to open surgery had significantly more complications ² Mean values 16 vs. 17 lymph nodes, p=0.049 ³ only surgically treated small bowel obstruction lower in laparoscopic group						

Table 1. Summary of evidence from the major multi-centre randomised controlled trials (USA, UK and Europe). The arrows indicate a significant higher (↑) or lower (↓) value for LCS compared to open surgery. Furthermore, “no significant difference” (ns.) and “no available data” (-) are indicated.

1.3.2. Meta-analyses of randomised controlled trials

Although proving equivalence in terms of oncologic risk is a major achievement, none of the large multi-centre studies were able to show clear superiority of LCS compared with open surgery. A meta-analysis with data from 12 randomised controlled trials of short-term outcomes up to 2002 concluded that there was less morbidity, earlier recovery in terms of bowel function and hospital stay, and less analgesia requirements (Abraham 2004).

Subsequent meta-analyses showed improved postoperative pulmonary function, reduced wound infections, less mechanical ileus, less analgesia requirements, less blood loss and hospital stay was shorter (Schwenk 2005). These results remained unchanged when two further important RCT's (COLOR, MRC-CLASICC) were included in the analysis (Tjandra 2006). After 2006, the majority of meta-analyses focused on long-term results, specifically oncologic outcome, as this data was now available. All studies supported the RCT findings that LCS does not lead to increased oncologic risk, as shown in **Table 2**.

Meta-analysis	RCT	n	Aims	Morbidity			Survival		Recurrence
				Surgical	Medical	30 day Mortality	Overall	Disease free	
Abraham 2004	12	4065	Short-term outcomes	↓	↓	ns.	-	-	-
Schwenk 2005	25	3526	Short-term outcomes	↓	↓	ns.	-	-	-
Tjandra 2006	17	4013	Short-term outcomes	↓	↓	↓	-	-	-
Reza 2006	12	4407	Safety and efficacy of LCS in the treatment of colorectal cancer	↓	-	ns.	ns.	-	ns.
Jackson 2007	10	3830	Oncologic outcomes	-	-	-	ns.	-	ns.
Bonjer 2007	4	1765	Oncologic outcomes (COST, COLOR, and CLASICC)	-	-	-	ns.	ns.	ns.
Laurenco 2008	20	4568	Clinical effectiveness (including HALS)	ns.	ns.	ns.	ns.	ns.	ns.
Liang 2008	10	2474	Recurrence rates	-	-	-	-	-	ns.
Kuhry 2008	12	3346	Long-term outcomes	-	-	-	ns.	ns.	ns.

Table 2. Summary of the results of meta-analyses on the outcomes of laparoscopic colorectal surgery. After 2006 the main focus was on long-term results. (“ns.” non significance, “-“ no data)

1.3.3. On-going trials

There are at least five major randomised controlled trials underway. It is expected that these trials will close current gaps in the literature on short- and long-term outcomes for rectal cancer (COLOR II, JCOG 0404, LAPKON II), the role of LCS in inflammatory bowel disease (LAPDIV-CAMIC, Sigma-trial) and the role of multimodal peri-operative treatment programs (EnROL). An overview of on-going trials is shown in **Table 3**.

Trial	Aim	Country	Centres	Sample size	Results expected	Reference
COLOR II	Short- and long-term outcomes of laparoscopic and open surgery for rectal cancer	Netherlands, Belgium, Germany, Sweden, Spain, Denmark, South Korea and Canada	27	739	Inclusion completed end of 2009	(Buunen 2009a)
JCOG 0404	Short- and long-term results of laparoscopic T3 and T4 colorectal cancer resections	Japan	24	818		(Kitano 2005)
LAPKON II	Long-term results for selected laparoscopic colorectal cancer resections	Germany	27	472	Short-term results published in 2009	(Neudecker 2009)
LAPDIV-CAMIC	Short- and mid-term results after laparoscopic sigmoid resections for diverticulitis	Germany	?	710	Started in 2004	(Schwenk 2004)
Sigma trial	Post-operative morbidity, mortality and quality of life, after elective laparoscopic sigmoid resection	Netherlands, Spain	5	104	Started in 2007	(Klarenbeek 2007)
EnROL	Short-term results with fast-track recovery programme for open and laparoscopic group	UK	6	202	Started in 2008 (168 patients recruited)	(Kennedy 2008)

Table 3. Ongoing multicentre trials on the feasibility of laparoscopic colorectal surgery.

1.4. Proficiency gain and competency assessment of surgical skills

1.4.1. Observational skills assessment in surgery

Research output on the development and validation of observational assessment tools for technical skill in surgery has been steadily increasing over the past two decades. Two recent systematic reviews summarise the current available evidence of assessment tools for technical skills (van Hove ; Ahmed 2011c). Both generic scales and task-specific checklists for the assessment of technical skills in a laboratory and real workplace environments for a variety of surgical specialities were identified. There was only one study describing an assessment procedure at specialist level (Beard 2007), whereas the vast majority of assessment tools was aimed specifically at the trainee surgeon. There are concerns that rating scales aimed at junior surgeons will show a ceiling effect for those at specialist level (Munz 2004). More recently, an increasing amount of published work has focussed on assessment of non-technical skills. Prominent research groups in the field have published similar tools for the assessment of communication, teamwork, leadership, decision making and situational awareness (Sevdalis 2008; Yule 2008). However, a clear distinction between technical and non-technical skills may be challenging. Decision-making is a continuous process with direct impact on the “technical” performance level. The surgeon’s situational awareness, the ability to communicate and work with a team may cause or prevent an operative error (Cuschieri 2006). Hence, using a so-called “technical” skills assessment tool may also measure non-technical skills indirectly.

1.4.2. Proficiency gain curve

1.4.2.1. Background

The term *proficiency gain curve* was introduced by Sir Alfred Cuschieri in 2006, who objected to the use of the term *learning curve* in the surgical literature. He claimed that “proficiency gain’ is a more precise description of improving psychomotor skills as distinct from “learning”, which implies a purely cognitive process. The proficiency gain curve can be defined as the process of increasing levels of technical and non-technical proficiency (Cuschieri 2006). The term is often mistakenly meant to be adapted from economics. Theodore Paul Wright is regularly claimed to have introduced the term “learning curve” in the 1930s in the aircraft industry, in the context of productivity in a factory. However, in the late 19th century, the German psychologist Hermann Ebbinghaus had already used the same term for his research on memory in an educational context. In his experiments, he investigated the “learning curves” of memorizing sequences of “non-sense” syllables (Wozniak 1999). In surgery, the term was used for the first time in 1980 to describe improved graft patency rates with increasing practice in cardiac bypass surgery (Tyras 1980). In this thesis the terms learning curve and proficiency gain curve are used as synonyms.

1.4.2.2. Measuring proficiency

In classical experiments on learning curves, reaction or solution times were used as the primary outcome parameter, e.g. the time needed to solve an arithmetic problem or the reaction time after a stimulus. In such standardized experiments, time seems to be a reliable parameter to describe increasing skill levels. A surgical operation is a non-standardized, complicated complex of tasks, requiring cognitive, psychomotor and non-technical skills. The time required

to perform the task depends on multiple factors, such as case complexity, morbidity of the patient, team performance etc. Clinical outcome parameters, such as complications, mortality or conversion to open surgery have also been used to describe the proficiency gain process, but again, these factors naturally depend on various factors. Hence, using parameters to describe skill is always only an approximation. Each parameter comes with a certain level of “noise” which can also be referred to as “performance shaping factors” (Kirwan 1992b).

Experiments on learning behaviour have shown that this relationship is unlikely to be linear. Proficiencies accumulate quickly at the beginning and slower at the end of the “learning curve”. The curve approaches an asymptote (“plateau”) which can be defined as the hypothetical maximum level of performance. The time required to reach the asymptote is called the length of the “proficiency gain curve”. The shape of the curve is defined by the learning rate which is represented by the slope of the curve (Feldman 2009). Depending on the outcome parameter the curve can become more negative (e.g. operating time) or positive (e.g. performance score) with increasing training.

A variety of curve shapes have been proposed to describe learning behaviour, including power law, exponential or hyperbolic functions. Among those, the power function has become popular as the power law of practice and it has been argued that this is the best fitting curve for most learning behaviours (Newell 1981). Ramsay et al. used operating times for 190 consecutive laparoscopic funduplications by a single surgeon and applied twelve different curve models. Logarithmic, power law and log linear curves showed the best fit, whereas log-log-linear, linear and inverse curves were less accurate (Ramsay 2001).

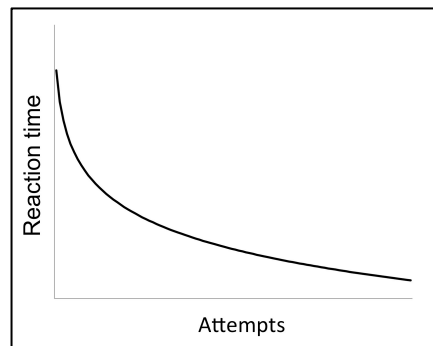


Figure 3. Power law of practice.

Although the power law of practice is widely accepted, it has been argued that it only fits when observing the average performance of a group of individuals. When observing individuals, exponential curves often fit better and the curves are less smooth and demonstrate discontinuities at multiple levels (Heathcote 2000). It has been argued that these discontinuities express strategy shifts in the learning approach. Applying different learning strategies during training can cause acceleration or deceleration of proficiency gain (Delaney PF 1998) (**Figure 4**).

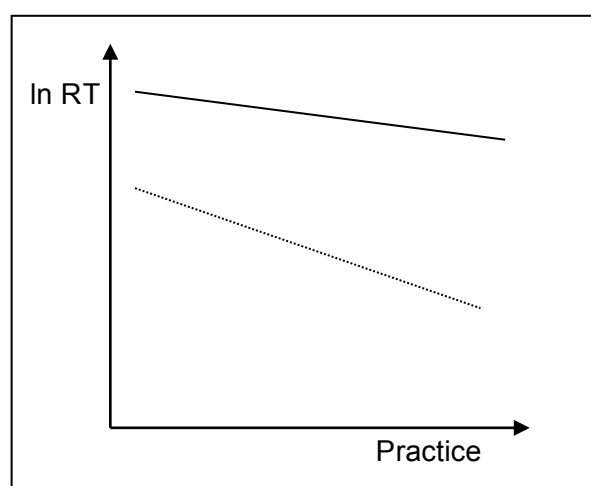


Figure 4. Example for strategy shift. Two groups used different strategy to solve the same task resulting in differences in Response times (RT) and learning rates (α) expressed by the steepness of the logarithmic line. (adapted from Delaney).

1.4.2.3. Challenges in measuring proficiency gain

As shown above, describing the proficiency gain curve is only an approximation of reality.

This approximation depends on the parameter used to measure proficiency and the method used to analyse proficiency gain. Although certain curve patterns have been shown to be applicable in different experiments (e.g. the power law of practice) there are situations or individuals insufficiently represented by the generalised model (Feldman 2009). Despite numerous studies on the “learning curve” in laparoscopic colorectal surgery, the proficiency gain process is poorly understood. There is no consensus, either on the length of the proficiency gain curve, or how it should be measured. Also, those factors influencing (i.e. shortening) the proficiency gain curve are unknown.

1.4.3. Procedural competency

1.4.3.1. Competence and competency

Competence is defined as “the ability to do something successfully or efficiently” (Oxford 2005). However, professional competence should be regarded as a normative rather than a descriptive concept (Short 1984). Hence, it is important to define competence in its professional context (Eraut 1994). Although competence and competency are often used as synonyms in dictionaries, the educational literature makes a distinction between the two. Competence is used as a generic term to describe a person who is competent in his or her realm, without knowing what exactly he or she is doing. Competency, on the other hand, is used to describe the ability of performing a specific skill (Short 1984).

In a clinical context, competence is often defined through the description of different professional facets. In the UK clinical competencies are defined in the Good Medical Practice framework for competency and include (1) knowledge, skills and performance, (2) safety and quality, (3) communication, partnership and teamwork, and (4) maintaining trust (GMC 2012). The American Board of Medical Specialties (ABMS) defines core competencies of a clinician in the USA: medical knowledge, patient care, interpersonal and communication skills, professionalism, system based practice, and practice based learning and improvement (ABMS 2012).

1.4.3.2. Procedural competency

Technical competence should not be confused with dexterity. Dexterity merely describes manual nimbleness, whereas technical skills also include the cognitive process behind it, the strategy on how to approach an operation and the flexibility to adapt (situational awareness).

Procedural competencies for specific operations involve both non-technical and technical skills.

However, procedure-specific non-technical skills are predominantly required during the perioperative phase (e.g. consenting the patient, case selection etc.), whereas intraoperative non-technical skills are mainly generic.

In order to understand competency in the context of skill evolution the Dreyfus and Dreyfus model is helpful. **Table 4** summarizes its adaption by Eraut (Eraut 1994). According to this model, a competency is a step on the ladder to expertise. A competent person copes well with multiple activities, and accumulation of information, so called “coping with crowdedness”, which is determined through deliberate planning and formulation of goals, and having insight into the necessary actions for their achievement.

Novice	"rigid adherence to taught rules or plans" no exercise of "discretionary judgment"
Advanced beginner	limited "situational perception" all aspects of work treated separately with equal importance
Competent	"coping with crowdedness" (multiple activities, accumulation of information) some perception of actions in relation to goals deliberate planning formulates routines
Proficient	holistic view of situation prioritizes importance of aspects "perceives deviations from the normal pattern" employs maxims for guidance, with meanings that adapt to the situation at hand
Expert	transcends reliance on rules, guidelines, and maxims "intuitive grasp of situations based on deep, tacit understanding" has "vision of what is possible" uses "analytical approaches" in new situations or in case of problems

Table 4. Evolution of skills based on the Dreyfus and Dreyfus model. From: (Eraut 1994)

1.4.3.3. Challenges in measuring procedural competence

A number of reports on procedure-specific and generic assessment tools for technical skills in surgery have been published. Recent systematic literature reviews reveal that the published evidence on assessment of procedural competency in laparoscopic colorectal surgery is sparse (van Hove ; Miskovic 2010b; Ahmed 2011c).

However we have attempted to define and measure competence, the ultimate aim being ability to perform in the real-world environment. Messick distinguished between competence and performance. He argued that “competence refers to what a person can do under ideal circumstances whereas performance refers to what is actually done under real circumstances” (Messick 1984). It remains a major challenge to establish a link between these interpretations of competence and performance.

1.5. The National Training Programme in Laparoscopic Colorectal Surgery (NTP)

1.5.1. The impact of NICE

The evidence on short-term benefits and equivalent long-term outcomes when compared with open surgery, and described in chapters 1.3.1-3, has led the National Institute for Health and Clinical Excellence for England and Wales (NICE) to recommend laparoscopic colorectal surgery as an option for suitably selected patients. The NICE technology guidance appraisal issued in 2006 (TA105), states that *“laparoscopic surgery (including laparoscopically assisted surgery) is recommended as an alternative to open surgery for people with colorectal cancer”* (NICE 2008). It further recommends that it should be offered to all those suitable for both an open and a laparoscopic operation and it must be performed by a surgeon who “has been trained in laparoscopic surgery for colorectal cancer and performs the operation often enough to keep his or her skills up to date”. A technical guidance appraisal by NICE normally has to be implemented by hospital trusts within three months of being issued.

1.5.2. The uptake of LCS in England

In 2006, when the NICE guidance was issued, it was estimated that only 5% of English colorectal surgeons were trained in laparoscopic colorectal surgery. Whole regions in England did not have the numbers of surgeons required to offer LCS to every eligible patient. Three main reasons for the slow uptake of LCS in England can be identified. Firstly in the 1990s, during the early phase of LCS, there were concerns about increased rates of distant metastases, i.e. port-site metastases (Wexner 1995). It took almost ten years to overcome these concerns by more advanced techniques such as the use of wound protectors and improved dissection techniques. Secondly, due to these concerns, LCS was regulated by NICE and only allowed to be carried out within randomised controlled trial protocols until 2003 (Guillou 2005). This may

have discouraged a number of surgeons from starting LCS. Thirdly, LCS has a notoriously long learning curve (Tekkis 2005c). Although some of these factors also had to be faced by surgeons in other parts of the world, it was a particularly challenging situation in the UK. The regulatory impact of NICE is significant and may slow down the implementation of innovative technologies. In addition, the uptake of a novel technique can be more difficult in a consultant-led clinical environment. Consultants often work as independent team leaders within a clinical unit and therefore do not necessarily receive training support from colleagues. For most consultants it is not feasible to enter a fellowship programme, leaving their present assignment for several weeks or even months. Hence, short-term practical courses followed by an autodidactic process were usually the only pragmatic means of taking up LCS. For many consultants, particularly those in the latter half of their consultant careers with well established and safe open practices, LCS was not an attractive option when faced with the prospect of up to several years before competency could be achieved.

Because geographical inequality of service provision is against the basic principles of the NHS, the guidance could not be implemented within the designated time frame of three months, so it was waived until 2009 and subsequently extended to September 2010.

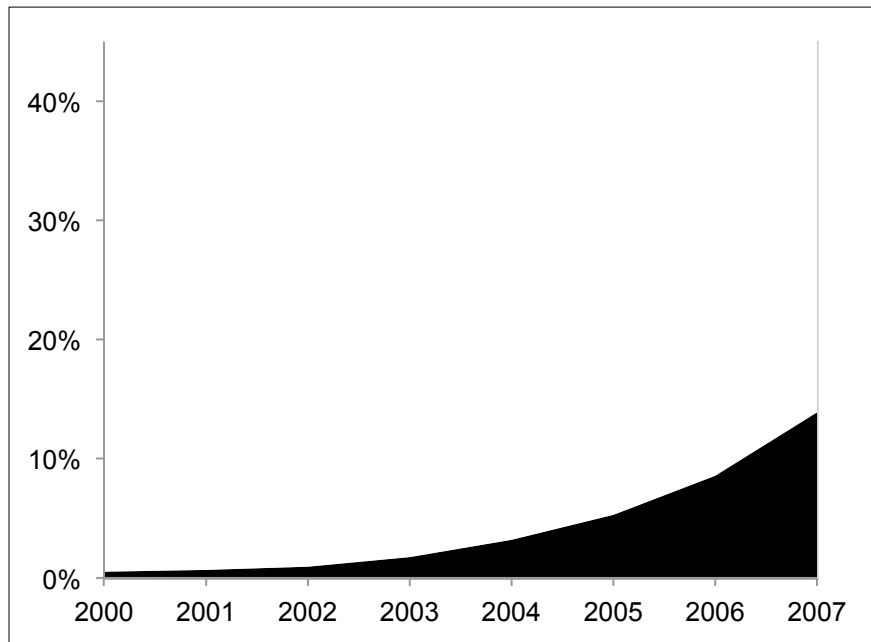


Figure 5. Uptake of LCS in England from 2000-2007 (Source: Hospital Episode Statistic (HES)).

1.5.3. Aim of the National Training Programme

In order to address the deficits in the uptake and provision of LCS, the Cancer Action Team at the Department of Health for England instigated the foundation of a National Training Programme for laparoscopic colorectal surgery (NTP or “Lapco”) in 2007. The aim of the programme was to provide appropriate training to a large number of colorectal consultants in order to fulfil the requirements of NICE TA105. After a formal application process, a national clinical coordinator with expertise in LCS and an educational centre (Imperial College London) with experience in surgical educational research were appointed to direct a structured process to implement the programme. A steering group formed by experts in LCS and an educational group with mixed experience in laparoscopic surgery and education were formed to create a structure and training plan. There was general agreement that it would be no longer acceptable for surgeons to gain experience in LCS at the expense of patient safety. Hence, it

was agreed that a period of supervised training with stringent assessment and audit must be a substantial part of the programme.

1.5.4. Administration and training structure

1.5.4.1. Organisational structure and administration

Initially, ten regional training centres including eighteen consultant trainers were appointed to provide laparoscopic colorectal surgery training, with a national coordinator and education centre. Over the course of the following three years an additional training centre was established and more trainers were appointed, now currently numbering eleven training centres and more than fifty trainers. The steering group, headed by the clinical and educational leads, consists of clinical experts and an educationalist acting as a decision making body for the mode of delivery of training and evaluation procedures. Each training centre is led by a training centre representative, usually the most senior laparoscopic surgeon of the unit (**Figure 6**). After successful application the delegates (“trainee” consultants) are allocated to a training centre and a trainer. Most delegates are trained by a single trainer, although some may have several trainers for supervised hands-on training.

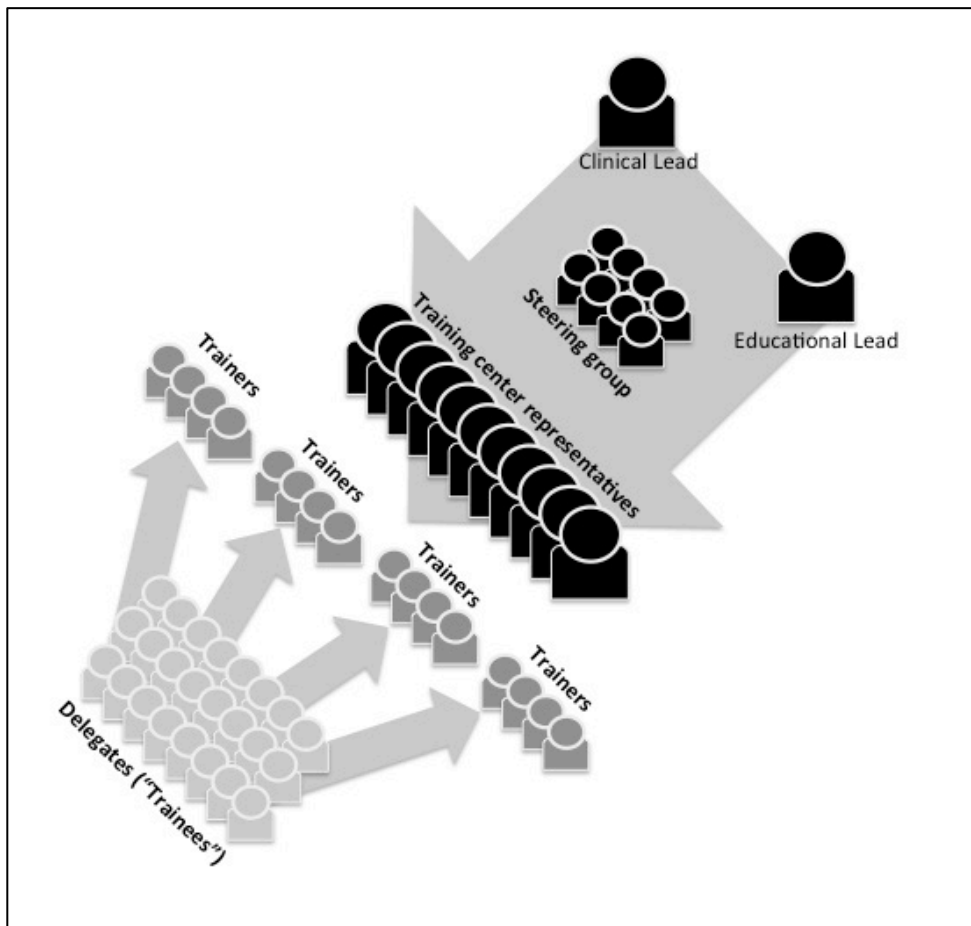


Figure 6. Organisational structure of the NTP. Delegates are allocated to one of the trainers who belong to one of the 11 training centres, represented by a senior laparoscopic surgeon. The clinical and educational chairmen lead the programme through a steering group.

1.5.4.2. Training structure

The main focus of the NTP is hands-on training in the operating theatre where the consultant in-training is directly supervised and mentored by an expert laparoscopic colorectal surgeon. The NTP also includes cadaveric, enhanced recovery and theatre team training courses (Wyles 2010a). The training centres were selected by open competition, based on specific criteria encompassing the expertise of the surgeons and the centres' infrastructure. The coordination process of such a programme is vital to ensure efficient delivery of clinical training and provision of competent laparoscopic colorectal surgeons. A web-based educational platform for administration, including application submission, training case assessments and logbooks, has been essential for the success of the programme (Lapco 2009).

The NTP training is delivered in different stages, and includes selection of candidates (eligibility), a 1-2 day course using cadaveric or animal models, supervised operating both at the training centre ("inreach" training) and at the hospital of the delegate ("outreach" training), a sign off procedure and finally independent performance with at least one year of clinical audit. Ideally these stages should represent different stages of proficiency according to the Dreyfus model (see also **Chapter 1.4.3.2.** and **Figure 7**) (Eraut 1994).

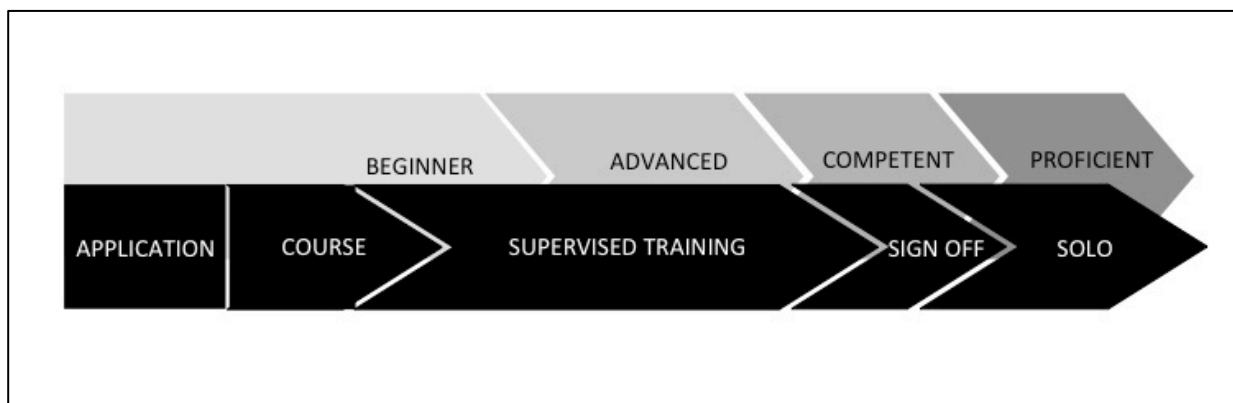


Figure 7. The training pathway for NTP delegates.

The stages of the programme include (**Figure 7**):

- I. **Application/ selection:** to fit the eligibility criteria, candidates must be consultant surgeons with experience in open colorectal surgery and basic laparoscopic procedures. Official support from NHS trusts is essential to ensure that their surgeons will have the opportunity to practice laparoscopic colorectal surgery once training is completed. After successful application a training plan is agreed by the delegate that includes the provision of assessment throughout the training and the sign-off at the end of supervised training.

- II. **Courses:** the NTP offers a selection of courses using both cadaveric models [Newcastle, Nottingham, Bristol (all UK)] and animal models [Norderstedt/ Hamburg (Germany) and Elaincourt/ Paris (France)]. In addition, some centres (Basingstoke, Bradford) offer immersion courses using supervised operating, video demonstrations and lectures in small groups. Each delegate needs to have attended at least one course before embarking on supervised training. In addition the NTP organises 2-3 master classes per year in different locations. These courses are aimed at a wider audience and focus on a specific subject, often accompanied by live operating by an expert laparoscopic surgeon. Master classes are not mandatory but are usually very well attended by delegates who have already started supervised training.

- III. **Supervised training:** this is the core stage of the training plan. Each delegate is allocated to a personal trainer. Although the details of training are left to the individual training centres, training usually starts as “inreach”. In an initial meeting,

the trainer and delegate would discuss the training plan, set certain ground rules and establish administrative and clinical responsibilities. Some trainers invite their delegates to observe a number of cases; others would start straight away with hands-on training. The delegates perform the operation under instruction and direct supervision of the trainer. “Inreach” training can be followed by “outreach” training, with supervision in their own theatre environment. Some training centres also offer support to the local theatre team by sending a dedicated senior scrub nurse to the “outreach” hospital. There is no set number of cases to be performed but there is a loose recommendation by the NTP that some 20 cases should be performed before embarking on the next stage. However, whenever the trainer feels the delegate is ready for the sign-off procedure, trainer and delegate request the invitation for sign-off from the NTP head office.

- IV. **Sign-off:** Before starting independent (solo) performance, delegates are required to be signed-off by the NTP. The steering group have determined the outlines of the sign-off procedure (see **chapter 6**). The procedure requires the submission of two video-recorded, unedited and independently performed cases by the delegate (right- and left-sided resection) for assessment. The trainer may be present in theatre but not scrubbed and at this stage of training substantial instruction by the trainer should not be necessary.
- V. **Independent practice (solo):** After completion of the sign-off assessment the delegates are provided with a letter from the head office, confirming successful completion of supervised training and independent practice is recommended. Delegates are required to audit their data for at least one year using a structured audit form. Data is available to the NTP for the evaluation of the programme. The

delegates are encouraged to maintain a good rapport with their trainers and discuss any problems arising, if required.

1.5.5. Conclusion

The setup of a National Training Programme in England in 2007 provided a unique opportunity to explore and scientifically assess the mechanisms and outcomes of a guided and structured induction process of a technically complex surgical procedure at a national level. Despite a small number of studies investigating the effects of unstructured training on clinical outcomes, little is known about the impact of a structured training programme on the development of proficiency in laparoscopic colorectal surgery.

1.6. Aims of the thesis

Based on the research problems outlined in the previous chapters, the aim of this thesis was to investigate the proficiency gain curve and competency achievement within a laparoscopic colorectal training programme for specialist colorectal surgeons.

The objectives of this thesis were to contribute evidence and close gaps for the following interrelated topics:

Objective 1: to identify shortfalls in the evaluation of the proficiency gain curve in laparoscopic colorectal surgery

Objective 2: to develop methods to describe the proficiency gain curve in laparoscopic colorectal surgery

Objective 3: to analyse proficiency and competency gain for specialist colorectal surgeons within the National Training Programme

1.7. Structure of the thesis

The thesis is divided into the following chapters.

I. Chapter 1 – Introduction

II. **Chapter 2 – Learning curve and case selection in LCS:** this chapter is a report of a systematic review of the literature summarising the current evidence in the literature about the proficiency gain curve (“Learning curve”) in LCS. It demonstrates that clinical outcome parameters are the only metrics that have been used to date. Data sets of several major studies are then synthesized to construct an average learning curve. This data is also used to suggest a case selection framework to prevent negative clinical outcomes in the early learning curve. (*Objective 1*)

III. **Chapter 3 – Impact of supervised training on clinical outcomes:** a systematic review and meta-analysis of the evidence in the literature aimed to measure the impact of direct supervision in the operating theatre on clinical outcomes. It demonstrates that surgeons with direct supervision achieve significantly better clinical outcomes than autodidact surgeons during the early “learning curve”. These results imply that morbidity and mortality data do not adequately describe the educational impact of a training programme. (*Objective 1*)

IV. **Chapter 4 – Development, validation and implementation of a monitoring tool for the proficiency gain process in LCS:** as shown in the two previous chapters, clinical outcome parameters are inappropriate and do not fit the purpose of monitoring the progress of skills acquisition in a supervised training programme. In

this study the structured development of a novel assessment tool to define the proficiency gain curve is demonstrated. This Generic Assessment Scale (or “GAS form”) provides a structure to assess the trainee’s performance in a non-judgmental way. It has been validated for both trainer and trainee (self-assessment) and implemented into the NTP. Examples of “learning curves” are given. (*Objective 2*)

- V. Chapter 5 – Observational Clinical Human Reliability Approach (OCHRA) for the assessment of competency:** this study describes an experiment to test the validity of a human reliability assessment, using error modes to describe competency in LCS by means of video material of operations. The videos were blindly assessed using scientific video tagging software, to identify errors and near misses committed by the operating surgeons. Surgeons at different levels of experience were compared and cut-off levels for competency determined. (*Objective 2*)
- VI. Chapter 6 – Development of the Competency Assessment Tool (CAT) for LCS:** expert interviews and findings from the previous study were used to develop a practical assessment tool for the assessment of competency in LCS. This tool is implemented in the NTP for trainees at the transition of unsupervised to supervised training. (*Objective 2*)
- VII. Chapter 7 – Educational and clinical outcomes of the NTP:** in this preliminary analysis of the NTP, educational outcomes are used to describe the proficiency gain process within the programme. Clinical outcomes are used to demonstrate the safety and clinical standards throughout the learning phase. (*Objective 3*)

VIII. Chapter 8 – Discussion: each of the previous sections is discussed in detail within the chapter. In this chapter, the objectives and its interrelations are discussed and future research is recommended.

2. PROFICIENCY GAIN CURVE AND CASE SELECTION IN LAPAROSCOPIC COLORECTAL SURGERY

2.1.	Introduction	63
2.1.1.	Background	63
2.1.2.	Aim	64
2.2.	Methods.....	65
2.2.1.	Summary of methods.....	65
2.2.2.	Systematic review.....	65
2.2.2.1.	Search strategy	65
2.2.2.2.	Inclusion and exclusion criteria	65
2.2.2.3.	Data extraction	66
2.2.2.4.	Parameters.....	66
2.2.2.5.	Study quality assessment	67
2.2.3.	Proficiency gain curve analysis.....	69
2.2.3.1.	Statistical software	69
2.2.3.2.	Risk adjusted CUSUM.....	69
2.2.3.3.	Change point analysis	70
2.2.4.	Risk prediction analysis.....	70
2.2.5.	Expert survey and validation.....	71
2.3.	Results	72
2.3.1.	Systematic review.....	72
2.3.2.	Proficiency gain curve.....	75
2.3.3.	Risk prediction	78
2.3.4.	Expert opinion and recommendations.....	80
2.4.	Discussion	83
2.4.1.	Summary	83
2.4.2.	Comparison with results from previous studies	83
2.4.3.	Ethical implications	84
2.4.4.	Value of case selection	84
2.4.5.	Limitations.....	85
2.4.6.	Conclusion.....	86
2.5.	Disclosure.....	87
2.6.	Acknowledgments.....	88

2.1. Introduction

2.1.1. Background

In the previous chapter the evidence for clinical benefit and oncologic safety of LCS has been demonstrated. It was shown that despite these advantages the uptake in England and other countries was slow (COST 2004; Guillou 2005; Jayne 2007; Faiz 2009a; Hedrick 2009; Neudecker 2009; Schwab 2009). Three main reasons could be identified: (1) concerns about oncologic safety before evidence from large RCTs was available, (2) in England, LCS was approved only as part of the CLASICC trial until 2003 and (3) surgeons were concerned about the notoriously long proficiency curve (Martel 2006). Since the publication of a series of large RCTs, the first two factors have been addressed. The long proficiency curve however remains a hurdle for those wishing to begin LCS. In this chapter two areas on the proficiency gain curve in LCS are investigated. Firstly, the current evidence on the proficiency curve in LCS is summarised and parameters used to describe the curve are critically analysed. There is great disparity about the length of the suggested proficiency curves, ranging from 11-110 cases before proficiency has been achieved, and it was shown that inexperience is an independent factor for negative clinical outcomes ((Simons 1995; Dincler 2003; Tekkis 2005c). The increased rate of adverse clinical outcomes at an early stage on the proficiency gain curve raises ethical questions and highlights the need for mechanisms which reduce complications and conversions. The impact of training and supervision will be further investigated in the following chapters, but there is also evidence that careful patient selection may also reduce preventable adverse outcomes (Schlachta 2003; Miskovic 2010b). Currently, there is no consensus on how patient profiles should be selected for LCS over the proficiency gain curve. Proposed prediction models for conversion rates are based on single centre experience and are cumbersome for use in daily clinical practice (Schlachta 2000; Tekkis 2005b). It has been

shown that these prediction models are unreliable and not transferrable between different hospital environments (Cima 2011). Hence, in the second part of this chapter evidence from multiple centres and opinion from the wider surgical community is sought in order to recommend a case selection framework for guidance of surgeons through their individual proficiency gain process in LCS.

2.1.2. Aim

The aims of this chapter are to systematically search the literature in order to analyse how the proficiency gain curve was measured, how many cases it takes to reach consistent clinical results (length of proficiency gain curve) and to recommend a case selection framework in order to reduce complication and conversion rates at the early stage of independent practice.

2.2. Methods

2.2.1. Summary of methods

The systematic review was carried out along the guidelines of the PRISMA statement in order to analyse the current literature on learning curves in LCS comprehensively (Moher 2009), **Table 5**). Data sets were requested from the original authors and the overall proficiency gain curve analysed using risk-adjusted CUSUM curves. Multilevel regression models and qualitative expert group consensus were used to recommend guidelines for case selection during the learning curve.

2.2.2. Systematic review

2.2.2.1. Search strategy

A systematic search of the literature was performed using the databases MEDLINE (1988-2010, October week 4) and EMBASE (1988-2010). The search strategy included three different areas combined with “AND” [(“learning” and “curve”) or (“proficiency” and “gain” and “curve”) AND (“laparoscopy”) AND (“colorectal surgery” or “colon” or “colectomy” or “rectum” or “proctocolectomy”)]. Reference lists of identified papers were searched for additional papers.

2.2.2.2. Inclusion and exclusion criteria

Two authors assessed all studies from the primary search independently for their relevance by reading the paper titles and abstracts (DM and SMW). Reports in the English language

describing the learning curve of laparoscopic or laparoscopically assisted colorectal surgery were included. Case reports, letters or comments, editorials, bulletins, reviews, descriptions of techniques and studies on hand-assisted, open or robotic surgery were excluded from this review. For all studies meeting potential inclusion criteria, full text was retrieved and analysed for eligibility. A Cohen's $k = 0.93$ indicated an excellent interrater agreement (Kim 2007).

2.2.2.3. Data extraction

Information on statistical methods, outcome measures and lengths of the proficiency gain curve were extracted from the studies. All authors were contacted by email or telephone with a request to provide anonymous raw data of the published case series. The request was repeated until a positive or negative answer was received. For four studies no answer was obtained after repeated contact attempts, of which three were from more than ten years ago, and one was a small single centre study from 2006 (Bennett 1997; Kockerling 1998; Marusch 2001; Avital 2006). In studies where the same data or subsets of data were used, only one main dataset was used for analysis (Reissman 1996; Park 2007; Park 2009b; Scheer 2009). All included studies were non-randomized case series.

2.2.2.4. Parameters

For databases we requested minimum information on successive case number or operating date, surgeon identifier if multiple surgeons were involved, resection type, BMI (body mass index) and ASA classification (American Society of Anesthesiologists). Outcome parameters were defined as conversion (any unplanned conversion to open surgery) (Shawki 2009), postoperative complications (in-hospital or 30-day complication rate), operating time (incision

to closure) and mortality (in-hospital 30-day mortality). Further information was sought if available (lymph node count, recurrence rate, blood loss).

2.2.2.5. Study quality assessment

In order to measure the studies' quality, the framework on the methodological requirements for the description of the learning curve proposed by Ramsay et al. was used (Ramsay 2001). According to this, each study was screened for (i) prospective data collection; (ii) detailed description of the experience of the surgeons (operators); (iii) data source described; (iv) proportion by each operator (if more than one); (v) statistician or expert in data analysis involved and (vi) best statistical method for type of data (hierarchy of methods). Each criterion provided a sixth ($\frac{1}{6} = 0.1666\dots$) for a maximum overall score of one.

Topic	Checklist item	Reported
Rationale	Describe the rationale for the review in the context of what is already known	2.1.1.
Objectives	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	2.1.2.
Eligibility criteria	Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale	2.2.1.2.
Information sources	Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	2.2.2.1., 2.2.2.3.
Search	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	2.2.2.1.
Study selection	State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	2.2.2.2.
Data collection process	Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	2.2.2.3.
Data items	List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made	2.2.2.4.
Risk of bias in individual studies	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	n/a (no interventional RCTs)
Summary measures	State the principal summary measures (such as risk ratio, difference in means).	n/a (raw data was available)
Synthesis of results	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I^2 statistic) for each meta-analysis	2.2.3.
Risk of bias across studies	Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)	2.2.2.5.
Additional analyses	Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	2.2.5.
Study selection	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram	2.3.1.
Study characteristics	For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations	Table 6
Risk of bias within studies	Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).	n/a (no interventional RCTs)
Results of individual studies	For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot	Table 7
Synthesis of results	Present results of each meta-analysis done, including confidence intervals and measures of consistency	2.3.2-3., Figure 9, Tables 8-10
Risk of bias across studies	Present results of any assessment of risk of bias across studies (see item 15)	Table 6
Additional analysis	Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)	2.3.4., Figure 10, Tables 11-12
Summary of evidence	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as health care providers, users, and policy makers)	2.4.1.-4
Limitations	Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)	2.4.5.
Conclusions	Provide a general interpretation of the results in the context of other evidence, and implications for future research	2.4.6.

Table 5. Checklist of PRISMA guideline according to Moher 2009.

2.2.3. Proficiency gain curve analysis

2.2.3.1. Statistical software

Raw data analysis was performed using the open source environment R (R for Mac OS X Cocoa GUI, Version 2.11.1, R Foundation for Statistical Computing, 2010) and the Statistical Package for the Social Sciences software [PASW Statistics 18.0.2, IBM SPSS Chicago (IL), USA].

2.2.3.2. Risk adjusted CUSUM

Learning curves were analysed using risk-adjusted CUSUM curves. Binary outcomes (conversion, post-operative complications and mortality) were risk adjusted by fitting a two-level random effects logistic regression model that included surgeon and all patient-related independent variables [surgeon (level 1), operation type, diagnosis, age, ASA, BMI and gender (all level 2)]. Hospital was initially also included in a three-level random effect logistic regression model, however since the inclusion resulted in no difference in the fitness of the model, the two-level model was preferred. Continuous data such as blood loss, lymph node retrieval and hospital stay were log transformed. They were treated as continuous outcomes and were risk adjusted using a two-level random effects linear regression analysis. As multiple surgeons provided data, the mean values of the expected and observed outcomes for each case order were computed. CUSUM curves were plotted by comparing observed and expected outcome values, using the equation $S_i = S_{i-1} + (x_{i(\text{obs})} - x_{i(\text{exp})})$; whereas S =cumulative means, i =case order, x =outcome value.

2.2.3.3. Change point analysis

Raw CUSUM curves provide a graphical representation of data trends. They were originally designed to monitor expectedly stable processes with the aim of detecting a moment in time when the process is going out of control. This early warning system is well established in industry such as business procedures and factory production processes (Page 1954). Using CUSUM for proficiency gain curves in surgery is exactly the opposite: analysing a process that is initially out of control and determining the moment in time when it starts to produce stable outcomes. Graphically, any time point on the x-bar (time) where the curve changes direction on the y-axis is a potential change of trend, although the vast majority of these represent data fluctuation. Several methods have been described to analyse the relevance of a change, most of them indicating when a process is going out of control (Grigg 2003). Taylor recently described an alternative method (Taylor 2007). This method allows calculation of the probability if a change of the direction of the curve is coincidental by a bootstrapping approach. Theoretically any point on the curve could be investigated but for pragmatic reasons only the more extreme values are used. For this study, CUSUM curves were bootstrapped by randomly re-distributing the case order 10,000 times. By comparing the difference between the maximum and minimum S ($\Delta S = S_{\max} - S_{\min}$) of the original CUSUM with those obtained from the bootstrapping samples, a confidence level could be determined for the point of change. If, for example, 40% of all bootstrapped curves had a higher ΔS , the confidence level would be 60%. A confidence level of >95% was defined as moderate, and >98% as strong evidence that a change had occurred.

2.2.4. Risk prediction analysis

For each dependent variable, two different multilevel random effects logistic regression models were fitted, one including surgeon and hospital identities, and one including surgeon identity alone. The better model was selected after comparing them using the chi-squared test applied to

-2*log likelihood values (Gelman 2007). Although the total number of independent variables (patient and case-related factors, surgeons, hospitals), inclusive of multiple levels of categorical variables, is as many as forty four, this is justified by the high sample size of nearly 5000 cases (Peduzzi 1995). Missing data in patient-related independent variables was replaced by the corresponding averages. The selected model was used to predict conversion and complication rates, and the resultant regression coefficients were then used to compute risk-adjusted odds ratios of the significant independent variables.

2.2.5. Expert survey and validation

Most fields were unambiguous, binary (e.g. conversion) or continuous (e.g. age of patient) data and therefore consistent across the different databases. The field “diagnosis” was less consistent since it could be coded in different ways. In order to adjust coding across different databases the diagnostic categories had to be broadened and became very generic (e.g. “cancer” or “inflammatory bowel disease”). The risk prediction model showed that even these broad categories were independent factors for the prediction of negative outcomes but they were not practical for the purposes of guidelines. Hence, diagnoses were excluded from the model and consensus was sought through an expert group inquiry. Twenty-nine experts and opinion-leaders from eight different countries (Argentina, Canada, France, Germany, Japan, New Zealand, UK, USA) were consulted using an online survey (SurveyMonkey™, Palo Alto (CA), USA). Seventeen possible indications for laparoscopic colorectal resection, dependent on operative complexity, were rated on an interval scale from 0 (least complex) to 100 (most complex). The final recommendations were sent back to the same expert group for adjustment and approval.

2.3. Results

2.3.1. Systematic review

The primary search identified 401 papers. After removing duplicates and limiting the search to articles, two independent reviewers examined 203 titles and abstracts. Of those, 176 were excluded according to pre-set criteria (61 not on laparoscopic colorectal surgery, 54 reviews, 56 not on learning curves, 5 not in English). A further 3 studies were excluded after full texts were retrieved (same dataset such as an included study (Agachan 1996; Reissman 1996; Park 2007; Scheer 2009). One study was added through cross-referencing (Buchanan 2008). Twenty-four studies fulfilled the inclusion criteria. Of these, 10 anonymous raw data sets were retrieved, but the remaining 14 were unable to provide sufficient data, and in 8 studies the data bases were no longer retrievable (Senagore 1995; Simons 1995; Wishner 1995; Begos 1996; Bruch 1999; Lauter 2001; Chen 2007; Agha 2008). For 5 studies, there was no response after repetitive requests (Bennett 1997; Kockerling 1998; Marusch 2001; Avital 2006; Lordan 2008). For one study the request was declined (Agachan 1997). Three data sets could not be used for analysis: two contained only rectal resection data (Choi 2009; Ito 2009) and one did not provide sufficient information about the identity of the operating surgeons (Shah 2005)], leaving a total of seven for analysis (**Figure 8**). The average quality of published papers was low, with an average quality score of 0.31 (0-0.86) (**Table 6**). Six studies scored 0.5 or higher indicating that they met half or more of the required criteria (**chapter 2.2.2.5**); four of these datasets were retrieved for further analysis (Dincler 2003; Tekkis 2005c; Li 2009; Park 2009a). In the majority of studies inadequate methods (split group method) for analysis had been used. The most common outcome parameters were operating times (20 studies), conversion rates (15 studies), complication rates (11 studies), hospital stay (6 studies) and readmission rates (2

studies). Recovery of bowel function, case complexity, blood loss, lymph node retrieval and local recurrence rates were analysed by single studies (Senagore 1995; Park 2009a). The only non-clinical outcome measure used in one study was operating room and hospital stay costs (Senagore 1995).

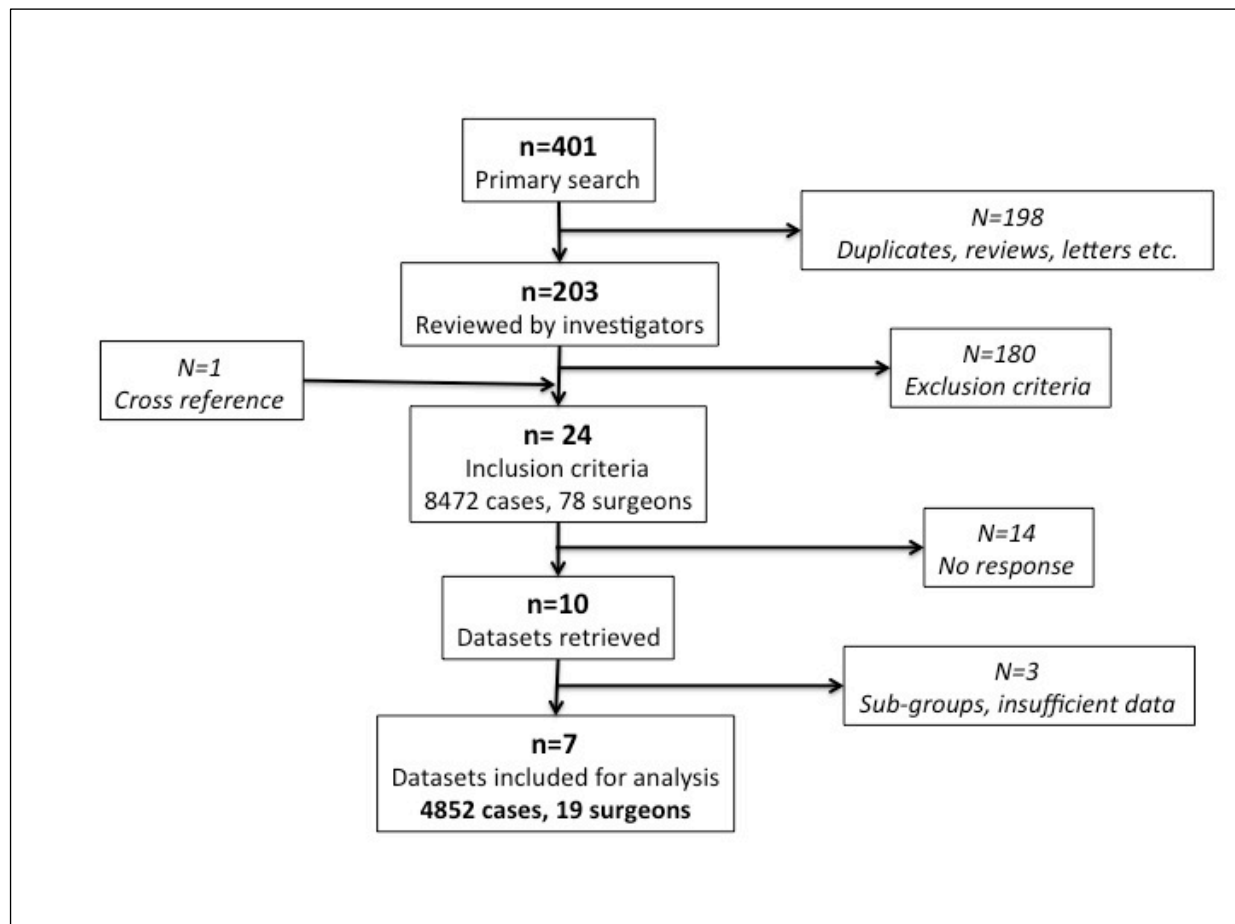


Figure 8. Flowchart for study selection.

Study	Country	Surgeons	Patients	Data organization	Statistical method*	Group size**	Quality score							
							Prospective	Experience of operators described	Data source described	Proportion by each operator**	Statistician expert involved	Best statistical method for type of data	Risk adjusted	Total score
(Simons 1995)	USA	4	144	Split group	descriptive	5		+		+				0.29
(Senagore 1995)	USA	1	60	Split group	comparative	20	+			N/A				0.17
(Wishner 1995)	USA	6	150	Split group	comparative	25	+	+		+			+	0.57
(Begos 1996)	USA	1	50	Split group	descriptive	5	+			N/A				0.17
(Agachan 1997)	USA	4	175	Split group	comparative	35	+			+				0.29
(Kockerling 1998)	Germany	18	500	Continuous	descriptive	N/A	+		+					0.29
(Bruch 1999)	Germany	1	300	Split group	comparative	100	+							0.14
(Lauter 2001)	USA	2	150	Split group	descriptive	50								0.00
(Schlachta 2001)	Canada	3	889*	Split group	comparative	30	+		+					0.29
(Dincler 2003)	Switzerland	2	1409*	Continuous	CUSUM	25	+	+	+	+	+	+		0.86
(Tekkis 2005c)	USA	4	889*	Continuous	CUSUM	N/A	+		+	+	+	+	+	0.86
(Shah 2005)	UK	3	80	Split group	descriptive	26			+	+				0.29
(Park 2007)	S.Korea	1	197	Continuous	descriptive	N/A	+			N/A				0.17
(Chen 2007)	USA	8	287	Split group	descriptive	36								0.00
(Agha 2008)	Germany	6	300	Split group	comparative	100		+		+				0.29
(Lordan 2008)	UK	1	152	Split group	comparative	30-65				N/A				0.00
(Kim 2007)	USA	1	100	Split group	comparative	N/A				N/A	+			0.17
(Fukunaga 2008)	Japan	1	410*	Continuous	descriptive	N/A	+			N/A				0.17
(Buchanan 2008)	UK	1	230*	Split group	comparative	104-149	+	+		N/A				0.29
(Choi 2009)	S. Korea	3	199	Continuous	CUSUM	N/A	+	+	+	+	+	+		0.86
(Park 2009a)	S. Korea	1	1014	Continuous	CUSUM	N/A	+		+	N/A		+		0.50
(Ito 2009)	Japan	1	200	Split group	comparative	20		+						0.14
(Scheer 2009)	Canada	4	487	Split group	comparative	5-30 ³	+							0.14
(Li 2009)	China	1	100	Split group	descriptive	5	+	+	+	N/A				0.50
TOTAL		78	8472				16/24	8/24	8/24	8/15	4/24	4/24	2/24	0.31

Table 6. Quality assessment of included studies. For underlined studies detailed data sets (raw data) were obtained. *The case numbers of these datasets were higher than published in the original papers.

2.3.2. Proficiency gain curve

Data was obtained for 4907 cases, performed by 26 surgeons in 7 institutions. Cases from 7 surgeons from one institution were excluded as they performed fewer than 20 procedures, leaving 4852 cases (19 surgeons) for inclusion in the analysis. The average experience of the remaining surgeons was 232 cases (range 21-1014). The overall mean outcomes were 8.75% conversions, 15.93% complications, 1.96% mortality, 18.65% recurrence, 180min operating time, 100ml blood loss, 14 harvested lymph nodes and a hospital stay of 7 days (**Table 7**). RA-CUSUM for conversion rate showed a significant learning curve after 152 cases [confidence level (CL) 99.96%], for complications after 143 cases (CL 98.5%), for operating time after 96 cases (CL 97.20%), for blood loss after 87 cases (CL 99.96%) and for hospital stay after 103 (CL 99.10%), and lymph node retrieval after 139 cases (CL 99.99%). Data on lymph node harvest (CL 47.55), length of stay and mortality (CL 32.68%) showed no significant learning curves at any point (**Figure 9, Table 8**). For recurrence, no significant change point could be found; however, data from only three surgeons were available and thus may not be representative.

Parameter	Overall outcome	<150 cases	>151 cases	n
Conversion	8.75%	13.22%	4.39%	4852
Complication	15.93%	21.40%	10.60%	4852
Operation time*	180 (140-235) min	190min	180min	4774
Blood loss*	100 (50-200) ml	100ml	98.5ml	1608
LN*	14 (9-22)	12	15	2517
LOS*	7 (5-11) days	7	7	3685
Mortality	1.96%	2.41%	1.35%	3002

Table 7. Summary of clinical outcomes.

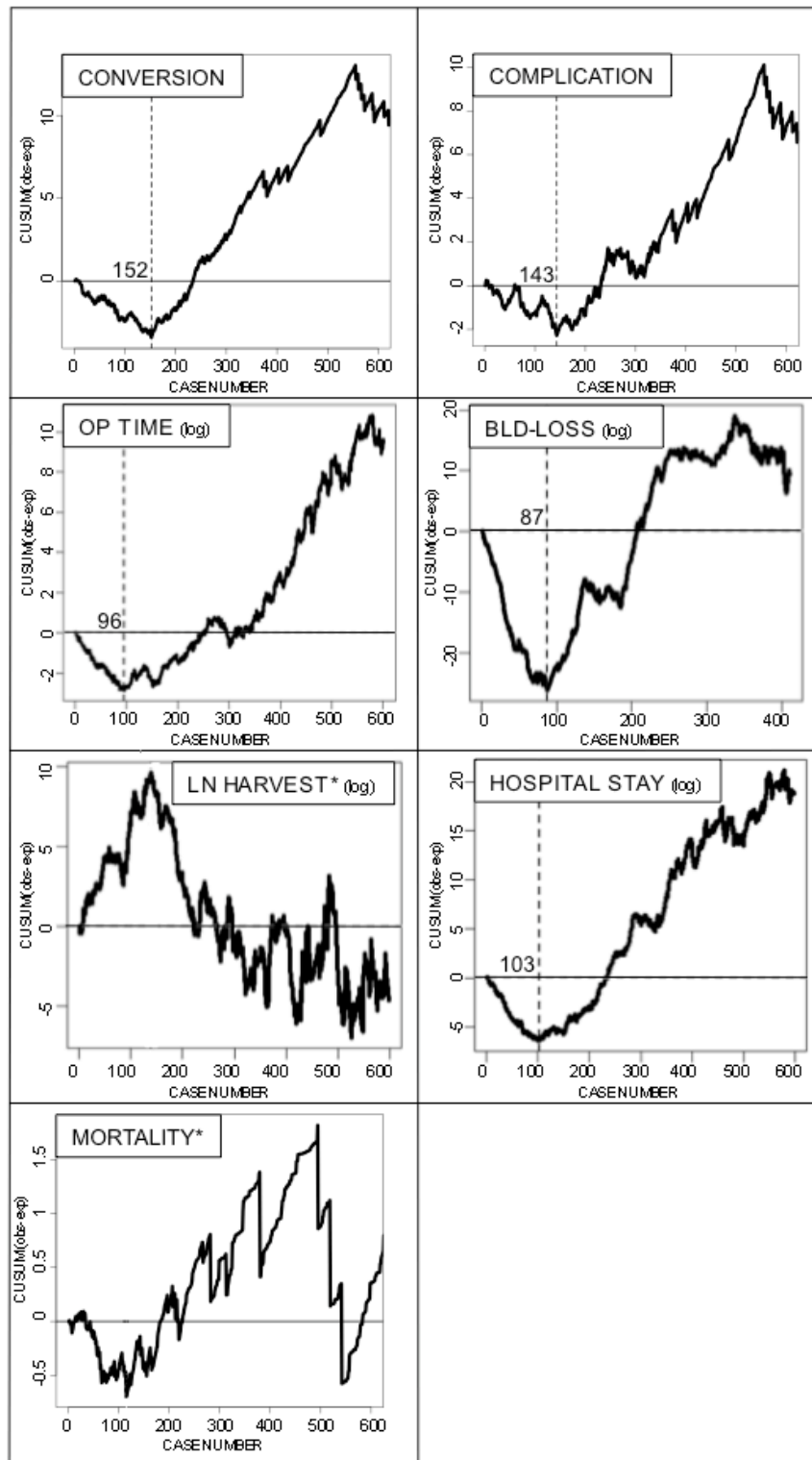


Figure 9. Proficiency gain curves (RA-CUSUM) for different parameters. Dashed lines and numbers indicate the length of the proficiency gain curve (change points). *Lymph node harvest and mortality did not reach sufficient confidence levels.

Parameter	Change point	Confidence
Conversion	152	99.96%
Complication	143	98.50%
Operation time	96	97.20%
Blood loss	87	99.96%
Lymph node retrieval	-	47.55%
Length of stay	103	99.10%
Mortality	-	32.68%

Table 8. Summary of confidence levels and change points for different parameters.

2.3.3. Risk prediction

Body mass index was a significant, independent variable predicting both conversion and complications. Procedures were categorised into colonic resections (right hemicolectomy, transverse colectomy, sigmoid resection) and those involving rectal resection (anterior resection, abdominoperineal resection, proctocolectomy), because it was assumed that there is a difference in complexity. Male gender was also a significant predictor, but mainly for operations involving pelvic (rectal) resection. This can be explained by the anatomical differences between the narrow male and the wide female pelvis. Hence, pelvic resections were further categorised into male and female.. Procedures involving pelvic dissection, especially in male patients, were the strongest predictors for both conversion and complications. In addition, the experience of the surgeon had an independently significant impact on outcome (**Tables 9 and 10**). Patient age and increasing American Society of Anesthesiologists (ASA) scores were also significant predictors. However, since these factors are unlikely to have added technical complexity to a case, they were excluded from the recommendation framework. Cancer, inflammatory bowel disease and emergency indications had a higher risk for conversion, but not complications (**Table 9 and 10**). Available data on previous abdominal surgery and recurrence of cancer were incomplete and poorly defined, and therefore not analysed.

	OR	95% CI	Estimate	Std Error	p-value
BMI					
>30 kg/m ² [intercept]	1		0		
27.5-30 kg/m ²	0.833	0.562-1.235	-0.182	0.197	0.353
<27.5 kg/m ²	0.516	0.368-0.724	-0.661	0.169	<0.001
Resection type					
Pelvic male [intercept]	1		0		
Pelvic female	0.542	0.340-0.865	-0.612	0.233	0.009
Non-pelvic	0.538	0.379-0.763	-0.620	0.175	<0.001
Diagnosis					
Benign tumour [intercept]	1		0		
Cancer	1.62	1.144-2.294	0.482	0.174	0.005
IBD	2.088	1.254-3.478	0.736	0.255	0.004
Diverticular disease	1.421	0.731-2.765	0.352	0.332	0.29
Acute*	2.045	1.193-3.506	0.716	0.27	0.008
Experience					
>150 cases [intercept]	1		0		
0-50 cases	3.495	2.312-5.150	1.251	0.194	<0.001
50-100 cases	2.581	1.798-3.704	0.948	0.181	<0.001
100-150 cases	2.248	1.540-3.280	0.810	0.189	<0.001

Table 9. Risk predictions for conversion.

	OR	95% CI	Estimate	Std Error	p-value
BMI					
>30 kg/m ² [intercept]	1		0		
27.5-30 kg/m ²	0.861	0.596-1.244	-0.150	0.184	0.416
<27.5 kg/m ²	0.709	0.530-0.949	-0.343	0.146	0.002
Resection type					
Pelvic male [intercept]	1.000		0.000		
Pelvic female	0.618	0.432-0.885	-0.481	0.179	0.007
Non-pelvic	0.436	0.330-0.577	-0.830	0.140	>0.001
Experience					
>150 cases [intercept]	1		0		
0-50 cases	1.489	1.132-1.959	0.398	0.137	0.004
50-100 cases	1.238	0.964-1.589	0.213	0.125	0.087
100-150 cases	1.310	1.016-1.689	0.270	0.127	0.033

Table 10. Risk prediction for complication (diagnosis were not significant)

2.3.4. Expert opinion and recommendations

Since descriptions of diagnosis in the database were very generic, expert group opinion regarding the complexity of more specific diagnostic conditions was obtained. The experts had a cumulative experience of 25,000 laparoscopic colorectal resections (range per surgeon 300-2200 cases). There was consensus that increasing T-stage and complicated inflammatory disease were likely to increase the complexity of the case (**Table 11**). Based on the evidence from the prediction models and the expert survey, a step-wise framework for case selection preference during the learning curve was established (**Table 12**). The learning curve was split into four levels I-IV. The according number of cases (0-50, 51-100, 101-150 and >150) reflects the learning curve for the average of all surgeon, based on the calculations above but may differ substantially for an individual and are therefore somewhat arbitrary. The framework was returned to the expert group in order to gauge their level of approval. A majority (76.2%) of the experts agreed with the concept (**Figure 10**). Seven surgeons commented that the guideline was, in parts, too restrictive, especially for BMI. However, another surgeon recommended that the upper BMI limit at stage I should be set at $23\text{m}^2/\text{kg}$. Two experts recommended using visceral obesity rather than BMI and another five surgeons suggested adding previous abdominal surgery to the guidelines. Since both factors are difficult to standardize, and only recommended by a minority of experts, they were not included in the current guidelines.

Complexit	Diagnosis	Median
Level I	Solitary benign polyp	0
	T1-2 cancer	10
Level II	Partial resection/stricturoplasty for Crohn's stricture	30
	Elective uncomplicated diverticular (no abscess on	40
	Cancer after stenting	50
	Perforation/bleeding due to penetrating trauma	50
	T3 cancer	50
	Partial resection for Crohn's fistula	55
	Tumour in transverse colon	55
Level III	Acute bleeding	64
	Polyposis (eg. FAP)	70
	Elective diverticular (with abscess on CT)	70
	Non-obstructive acute inflammation with perforation	73
	Restorative resection for UC	75
	Elective diverticular with colovesical fistula	75
Level IV	T4 cancer	80
	Acute obstruction	85
	Cancer with complicated fistula (eg. colovesical)	90

Table 11. Expert survey results. The interval scale ranged from 0 (least complexity) to 100 (highest complexity). The list was split into four categories (complexity level) according quartiles calculated from original data (30, 60, 80).

Complexity Level*:	I	II	III	IV
BMI [kg/m ²]	<27.5	<30	<30	>30
Resection	Colon	Female pelvic	Male pelvic	any
Diagnosis:				
- Cancer**	<T3 [†]	T3 [†]	T3	T4
- Inflammatory	none	uncomplicated ^{††}	complicated ^{††}	emergency

* Approximate case experience: I=1-50, II=51-100, III=100-150, IV=>150

**Preoperative staging (CT, MRI)

[†]Excluding transverse colon, proctocolectomy.

^{††} *Uncomplicated*: e.g. diverticular disease without abscess. *Complicated*: diverticular disease with abscess, intraabdominal fistula, restorative resection for UC.

Table 12. Recommendations for case selection during the learning curve.

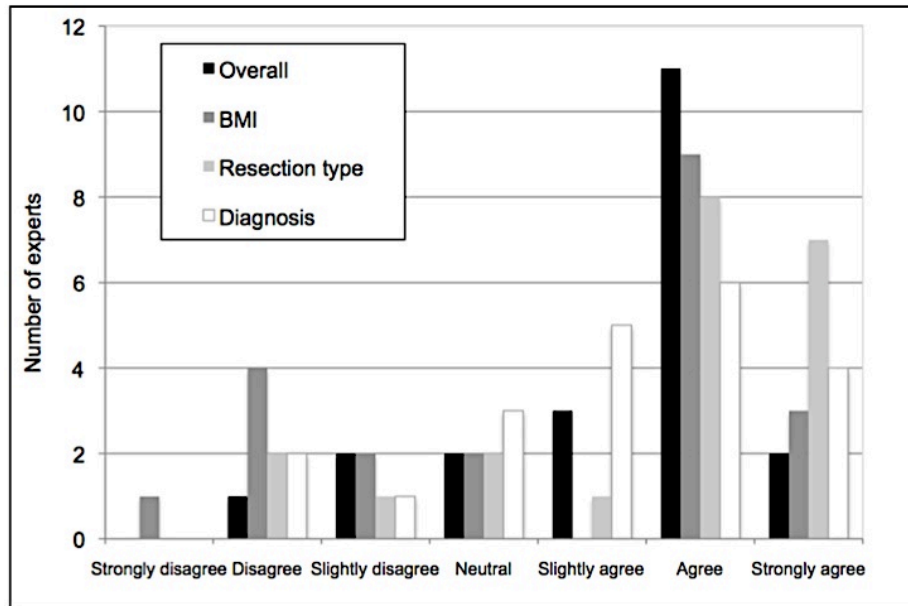


Figure 10. Agreement of experts responding to the survey regarding the guidelines and its different components. Most disagreement was that BMI limits were too restrictive and that previous abdominal surgery should be considered for inclusion in the guidelines.

2.4. Discussion

2.4.1. Summary

This chapter describes a systematic review of published literature on the learning curve in laparoscopic colorectal surgery. Synthesized evidence demonstrates with a high level of confidence using advanced RA-CUSUM methods that the learning curve takes up to 140-150 procedures. A regression analysis and data from an expert survey indicate that BMI, type of resection and diagnosis are crucial factors necessary to predict the risk of complications, conversions and, hence, case complexity. Based on this evidence and supported by expert opinion, a guideline for case selection during the early stage of independent practice was established.

2.4.2. Comparison with results from previous studies

The learning curve in LCS is longer than that suggested by earlier individual case series. The learning curve in the majority of previous studies had been analysed using inferior statistical methods, such as the split group analysis. This method, coupled with small sample size, is likely to result in imprecise learning curves (Ramsay 2001). For the current analysis the sample size was maximised by combining seven major data sets, resulting in a large multi-centre database, and advanced learning curve analysis methods were applied. The present results are reflected in single centre case series using a similar approach of data analysis. A single centre case series by two surgeons using CUSUM curves, showed learning curves of around 110 cases (Dincler 2003). In another study by four surgeons, using risk-adjusted CUSUM curves, right and left sided resections were analysed separately, resulting in learning curves of between 55

(right) and 70 (left) cases. A combined analysis of right- and left-sided colectomies revealed a learning curve of 120 procedures (Tekkis 2005c).

2.4.3. Ethical implications

The protracted length of the learning curve raises ethical questions. LCS currently is an established operating method and negative clinical outcomes resulting from the learning process of a surgeon are no longer acceptable. Although conversion rates, morbidity and mortality parameters are undeniably important quality measures in surgery, there are constraints when using them as an index for the learning curve. Using negative clinical outcome parameters to describe the learning process of a surgeon is akin to observing a training airplane pilot by counting the numbers of crashes and near misses. This example demonstrates that exceeding the “normal” rate of undesirable outcomes is ethically unacceptable at any point in the learning process. There is increasing evidence that supervised training in LCS is effective and outcomes of “training” procedures are comparable to those of surgeons who are already experts in the field (Dalton 2009; Miskovic 2010b). In addition, alternative methods of measuring proficiency gain in a supervised setting have been proposed (Miskovic 2010a). An autodidactic approach used to acquire the necessary skills for LCS should now be obsolete and unacceptable. Although cadaveric or animal tissue training courses and other forms of simulation may be an initial step towards proficiency (Wyles 2010b), operating on real patients in a structured supervised setting is unavoidable. This underlines the need for structured training programmes, fellowships and objective assessment of competency in LCS.

2.4.4. Value of case selection

Once surgeons start to perform independently, careful case selection is the key to avoiding unnecessary conversions and complications. The framework for case selection during the

learning curve as presented in this chapter is supported by the results from single-centre based prediction models for conversion in LCS (Schwandner 1999; Schlachta 2000; Tekkis 2005b; Kirchhoff 2008; Rotholtz 2008; Tan 2008; Miskovic 2010b). It represents a practical guideline for surgeons following completion of supervised training and supports a controlled increase in case complexity during the learning curve. Although the average numbers for overall case experience at each complexity level are rather arbitrary (because the learning rate may differ amongst individuals), future LCS surgeons need to be aware of the long learning process. Recent criticism about the validity of prediction models for conversion and how accurately statistical data analysis reflects reality, may well be justified (Cima 2011). Hence, the decision about the feasibility of any individual case depends on the clinical experience of the practising surgeon, taking into account additional factors such as previous surgery, visceral body fat, local tumour invasion, co-morbidities and patient's preferences..

2.4.5. Limitations

There are certain limitations to this study. Firstly, the data sets used for analysis are of inconsistent quality since they were neither collected prospectively, nor using an identical data collection framework. Nevertheless, the number of collated data exceeds any previous single- or multicentre study on proficiency gain in LCS. Secondly, not all experts agreed with the guidelines. To some experts the guidelines were too restrictive. This could have arisen from the proposed case experience numbers suggested for the different complexity levels, and which may have appeared rather high to many established LCS surgeons. It needs to be emphasized that these numbers are based on the average surgeon and that individuals may be quicker in developing their skills. The guideline should not be restrictive to surgeons who wish to progress and develop their skills but it should provide a framework to ensure maximum patient safety. In addition, sociological backgrounds seem to play an important role: a Japanese

surgeon tended to recommend lower BMI thresholds, whereas North American surgeons generally found them to be too restrictive. Several experts suggested adding previous abdominal surgery to the guidelines as a factor for increased case difficulty. However, the evidence in the literature is not uniform and not every abdominal operation adds a similar level of complexity. It was suggested in a recent study that only major prior gastrointestinal surgery increases complication rates (Fukunaga 2008).

2.4.6. Conclusion

The initial learning curve in LCS should be overcome within the safety zone of a structured, supervised training programme. Thereafter, surgeons can further develop their skills by a stepwise increase in case complexity and should continuously audit of their results to ensure best possible patient safety throughout this process. Future work needs to be carried out to investigate performance-enhancing interventions which might reduce the learning curve in LCS.

2.5. Disclosure

The contents of this chapter have been published and presented at the Annual Meeting of the American College of Surgeons (D Miskovic, M Ni, SM Wyles, P Tekkis, GB Hanna. Learning curve and case selection in laparoscopic colorectal surgery – an international multi-centre analysis of 4852 cases. 97th meeting of the American College of Surgeons (ACS), San Francisco, October 2011).

The original reference of the paper is:

D Miskovic, M Ni, SM Wyles, P Tekkis, GB Hanna. Learning curve and case selection in laparoscopic colorectal surgery – an international multi-centre analysis of 4852 cases. Dis Colon Rect (accepted, in print)

Many of the methods used in this chapter cannot be easily computed using commercially available statistical software and need to be programmed from scratch. Dr Melody Ni has performed the programming in the open source software R.

2.6. Acknowledgments

I would like to thank the following experts for their invaluable advice (in alphabetical order):

T. Cecil (Basingstoke, UK), C. Delaney (Cleveland, USA), A. Fingerhut (Poissy, France), J. Fleshman (St Louis, USA), N. Francis (Yeovil, UK), Y. Fukunaga (Tokyo, Japan), J. Griffith (Bradford, UK), A.M. Gudgeon (Frimley, UK), H. Hasegawa (Tokyo, Japan), A. Horgan (Newcastle, UK), M. Ito (Chiba, Japan), J. Jenkins (Harrow, UK), A. Kanazawa (Osaka, Japan), R. Kennedy (Harrow, UK), R. Longman (Bristol, UK), C. Maxwell-Armstrong (Nottingham, UK), R. Motson (Colchester, UK), C. Neal Ellis (Pittsburg, USA), K. Otsuka (Iwate, Japan), T. Rockall (Guildford, UK), C. Schlachta (London ON, Canada), W. Schwenk (Hamburg, Germany), A. Senagore (Los Angeles, USA), M. Soop (Auckland, New Zealand), N. Rotholtz (Buenos Aires, Argentina), E. Weiss (Miami, USA), S. Wexner (Miami, USA), S. Yamaguchi (Saitama, Japan), S. Yamamoto (Tokyo, Japan),

I would also like to thank those who provided their data: C. Schlachta, J. Mammazza and E.C. Poulin (Canada), J.C. Li (China), Y. Fukunaga, (Japan), I.J. Park (Korea), P. Buchmann and S. Dincler (Switzerland), R. Kennedy (UK), V.W. Fazio (USA).

3. SYSTEMATIC REVIEW AND META-ANALYSIS ON THE IMPACT OF SUPERVISION ON CLINICAL OUTCOMES IN LAPAROSCOPIC COLORECTAL SURGERY

3.1.	Introduction.....	89
3.1.1.	What is supervision?	90
3.1.2.	Aim	91
3.2.	Methods.....	92
3.2.1.	Systematic review.....	92
3.2.1.1.	Search strategy.....	92
3.2.1.2.	Inclusion criteria	93
3.2.1.3.	Exclusion criteria	94
3.2.1.4.	Study quality assessment	94
3.2.2.	Meta-analysis.....	96
3.2.2.1.	Data extraction.....	96
3.2.2.2.	Data analysis	96
3.3.	Results.....	97
3.3.1.	Results from the literature search.....	97
3.3.2.	Study description and publication bias.....	101
3.3.3.	Subgroup comparison and meta-analysis.....	102
3.4.	Discussion	105
3.5.	Summary	105
3.5.1.	Are morbidity and mortality parameters useful for performance assessment?	105
3.5.2.	The impact of clinical supervision	106
3.5.3.	Limitations.....	106
3.5.4.	Conclusion.....	107
3.6.	Disclosures.....	108
3.6.1.	Acknowledgments.....	109

3.1. Introduction

In the previous chapter the length of the proficiency gain curve was demonstrated and the need for mechanisms to reduce it was discussed. It has also been shown that mortality and morbidity data to monitor progress along the learning curve are ethically challenging. With increasing numbers of surgeons practising LCS, the training environment has changed over the past ten years. In the 1990s, LCS was mainly performed by pioneers with little support from the wider surgical community. In the early 2000s, the technique became increasingly popular and finally

accepted as a valid technique after the publication of large RCTs and publication of the NICE guidelines. In the wake of this development, there was a transition from autodidact skills acquisition towards supervised training. Dedicated fellowships and training within regular surgical curricula became available, although limited to only a few centres (Schlachta 2003; Hedrick 2009). In the previous chapter, all data originates from surgeons who taught themselves or who had limited supervised training before performing unsupervised LCS. Little is known about the impact of such training on clinical outcomes. Previous single-centre studies in other areas have shown that trainees achieve similar results to experts in open *supervised* colorectal surgery and carotid endarterectomy (Rijbroek 2003; Renwick 2005). These results support the hypothesis that supervised training has a direct impact on surgical outcomes.

3.1.1. What is supervision?

Supervision is a generic and not a specifically educational term. It can be defined as overseeing a process or an operation (from Medieval Latin “supervidere”, or Latin “super” and “videre”). In a clinical educational context, it is often referred to as clinical supervision, which can be defined as a range of educational interactions, such as mentoring or coaching, but also feedback and appraisal (McGill 2011). Teaching a complex procedure such as LCS to advanced trainees or specialists may require advanced teaching skills and a structured curriculum. Nevertheless, in reality, any form of educational encounter, such as simple instruction or individualised mentoring, is common. Numerous papers on teaching LCS skills (**Table 15**) refer to mentoring, although the term is often misused. Mentor originates from Greek mythology and is the name of a friend of Odysseus through which the goddess of wisdom Pallas Athena appeared to Telemachus, son of Odysseus (Homer 2009). Athena, in the figure of Mentor, advised Telemachus on how to find his father. Hence, a mentor is defined as an experienced, senior professional who may act as a teacher, but also as an advisor, a counsellor, or even a friend.

This definition generally implies more than a simple professional relationship limited to teaching in the operating theatre and may not be an appropriate term. Supervision or clinical supervision as a more generic term is more inclusive for different forms of educational interaction and will be used throughout this chapter.

3.1.2. Aim

The aim of this study was to systematically review the literature for outcome data of supervised surgeons and compare these with results of self-taught surgeons at the beginning of their proficiency gain curve.

3.2. Methods

3.2.1. Systematic review

The systematic review was performed in accordance with the PRISMA guidelines (Moher 2009, **Table 14**).

3.2.1.1. Search strategy

The results for this section are extracted from a systematic review on mentoring and simulation in laparoscopic colorectal surgery. Two investigators (DM, SMW) searched the databases MEDLINE (1950-2009, August Week 3), EMBASE (1980-2009 Week 34) and PsycINFO (1967-2009, August Week 3). Three different domains of MeSH-terms and keywords combined by “AND“, and in each domain the terms were combined by “OR” were used. The first domain contained terms on training, the second on surgical anatomy (colon and rectum) and the third on laparoscopy (**Table 13**). The selection was not restricted by language. Two investigators reviewed titles and abstracts resulting from the search, and those clearly unrelated to the topic and any duplicates were excluded. For those possibly fulfilling inclusion criteria, full text was retrieved and assessed for eligibility.

AREA 1 TRAINING ("learning" or "education" or "humans" or "preceptorship" or "attitude of health personnel" or "imitative behaviour" or "mental competency" or "motivation" or "achievement" or "cognition" or "visual perception" or "professional competence" or "clinical competence" or "mentors" or "teaching" or "educational technology" or "knowledge").mp. [mp=ti, ab, hw, tc, id, ot, nm, sh, tn, dm, mf] Results=12,134,308 AND AREA 2 SURGICAL ANATOMY ("colorectal surgery" or "colon" or "colectomy" or "rectum" or "proctocolectomy").mp. [mp=ti, ab, hw, tc, id, ot, nm, sh, tn, dm, mf] Results=305,167 AND AREA 3 LAPAROSCOPY ("laparoscopy" or "laparoscopic surgery" or "laparoscopic procedure" or "laparoscopic procedures").mp. [mp=ti, ab, hw, tc, id, ot, nm, sh, tn, dm, mf] Results=96,364
--

Table 13. Search strategy.

3.2.1.2. Inclusion criteria

For this chapter, studies describing outcomes of training cases in a supervised and non-supervised setting were used. Supervised, in this context, referred to whether an expert laparoscopic colorectal surgeon provided technical support, advice or guidance to the trainee during the operation, either by actually being present in the operating room, or remotely via video-links (“tele-mentoring”). Non-supervised was defined as surgeons performing their first series (early phase) of cases without a senior surgeon present during the procedure. The early phase was defined as the first 50 cases, representing the first level of the proficiency gain curve (see **Chapter 2**).

3.2.1.3. Exclusion criteria

Case reports, letters or comments, editorials, bulletins, reviews, descriptions of techniques and studies on hand-assisted, open or robotic surgery, simulation training, subgroup analyses and reports with no extractable data were excluded. If there was incomplete published data we asked the corresponding author for this by electronic mail. If investigators failed to respond to a repeat request or were unable to provide the desirable data, their study was excluded. The search was performed with a good interrater reliability (Cohen's $\kappa=0.77$).

3.2.1.4. Study quality assessment

It was not possible to apply a classical bias risk assessment method for the included articles as no randomized controlled trials were identified. All studies were case series, some could be considered to be interventional cohort studies, but usually without a proper control group. Due to this inconsistency, no appropriate quality assessment tool could be applied and the study quality remained descriptive.

Topic	Checklist item	Reported
Rationale	Describe the rationale for the review in the context of what is already known	3.1.1.
Objectives	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)	3.1.2.
Eligibility criteria	Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale	3.2.1.1.
Information sources	Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched	3.2.1.1.
Search	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated	Table 13
Study selection	State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	3.2.1.2-3.
Data collection process	Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	3.2.2.1.
Data items	List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made	3.2.2.1.
Risk of bias in individual studies	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis	n/a
Summary measures	State the principal summary measures (such as risk ratio, difference in means).	n/a
Synthesis of results	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I^2 statistic) for each meta-analysis	3.2.2.2.
Risk of bias across studies	Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)	3.3.2.
Additional analyses	Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified	n/a
Study selection	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram	Figure 11
Study characteristics	For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations	Table 15
Risk of bias within studies	Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12).	n/a
Results of individual studies	For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot	Table 15
Synthesis of results	Present results of each meta-analysis done, including confidence intervals and measures of consistency	2.3.2-3., Figure 9, Tables 8-10
Risk of bias across studies	Present results of any assessment of risk of bias across studies (see item 15)	3.2.2.
Additional analysis	Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression) (see item 16)	3.3.3.
Summary of evidence	Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as health care providers, users, and policy makers)	3.5.1-2.
Limitations	Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)	3.5.3.
Conclusions	Provide a general interpretation of the results in the context of other evidence, and implications for future research	3.5.4.

Table 14. PRISMA guidelines according to Moher 2009.

3.2.2. Meta-analysis

3.2.2.1. Data extraction

Two reviewers applied inclusion criteria independently and extracted data into a standardised electronic format (Excel 2008 for Mac, Microsoft Corporation, Redmond (WA), USA), guided by the Cochrane Reviewers' Handbook (Higgins). The outcome measures were determined by conventional clinical outcomes. End points for reports on clinical outcomes for supervised and non-supervised surgeons were morbidity (complications), mortality and conversion from laparoscopic to open surgery.

Two authors extracted conversion rates and morbidity data from reports on three subgroups *supervised*, *non-supervised* and *control group*. The control group entails results from surgeons with experience of at least 50 cases, representing those who have overcome the initial proficiency gain curve.

3.2.2.2. Data analysis

For the analysis, the software Comprehensive Meta-Analysis (Version 2.0, Biostat, Englewood, NJ, USA) was used. A meta-analysis for each subgroup was performed separately, using a fixed and random effects model for event rates. A subgroup analysis using the mixed model (method of moments) was carried out for levels of significance. A p-value of <0.05 was considered to be significant. Sensitivity analysis was performed in each subgroup by running the analysis several times and each time removing one study. No outliers could be identified in this analysis.

3.3. Results

3.3.1. Results from the literature search

After removing duplicates, 3420 citations from the literature search were retrieved (MEDLINE 3060 studies, EMBASE 360 studies, PsycINFO no studies). One hundred and fifteen reports potentially fulfilled our inclusion criteria and one study was considered through cross-referencing. Twenty two studies were included for analysis (**Figure 11**). Ninety-four reports were excluded. Sixty-four studies had non-usable or non-relevant data. In 28 studies there was insufficient data on initial experience (early phase). One study on laparoscopic rectal resections was excluded, as the surgeons operating on 36 of the first 50 first cases already had experience of more than 50 colon resections (Ito 2009). One further study could not be included as the investigators were unable to provide more detailed data on individual patients (Chikkappa 2009).

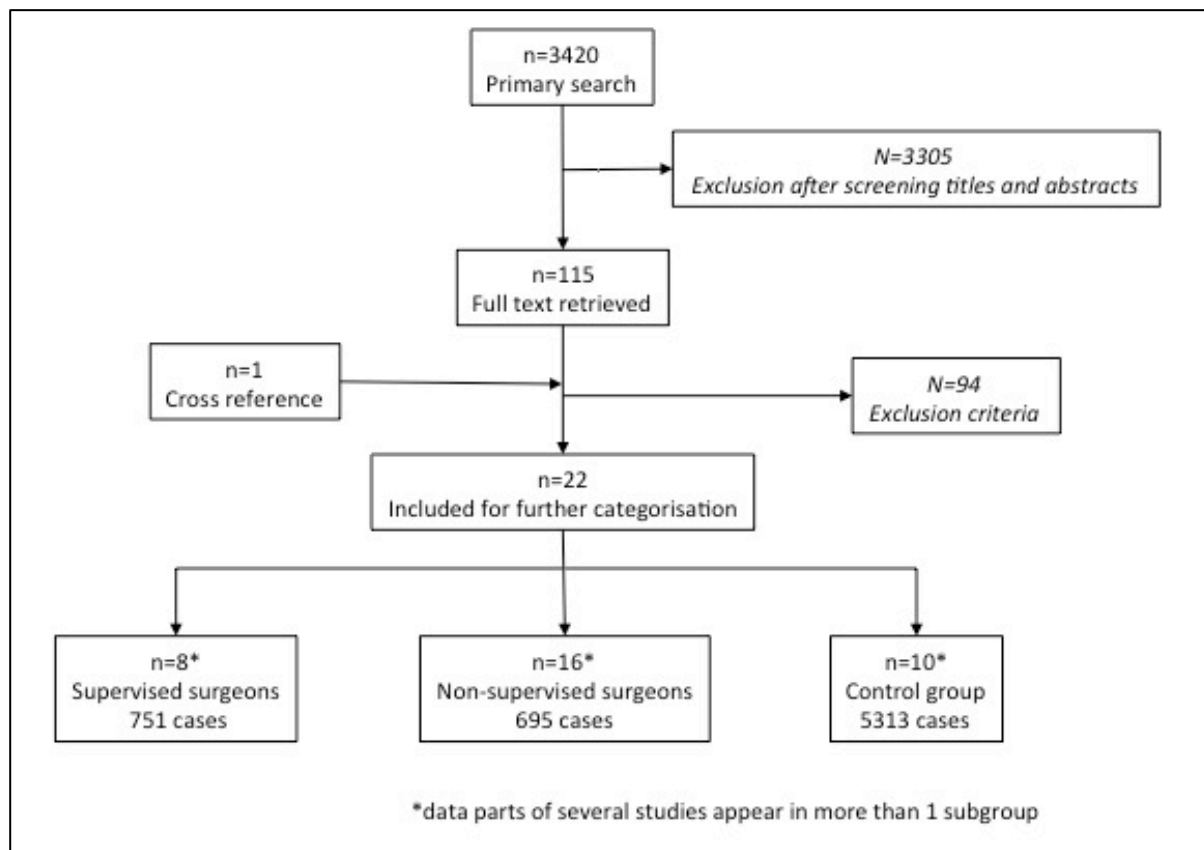


Figure 11. Flowchart of inclusion and exclusion of studies.

	Year	Study Type	Supervised	Non-supervised	Control	n (surgeons)	Study aim (as declared by authors)	Main outcome measures
Wishner	1995	Case series		X		6	Learning curve	Length of procedure, complication rate, conversion rate
Begos	1996	Case series		X		1	Assess the results of laparoscopic surgery for rectal carcinoma during the learning curve. Outcome compared to open cases	Operating time, length of stay, postoperative pain, lymphnode harvest
Reissman	1996	Case series		X		n/a	Assessing the learning curve, for first 100 cases	Intra and postoperative complications, conversion to open, length of ileus, hospitalisation
Agachan	1997	Case series		X		4	Assess the operative time for different colorectal procedures and see how it changes with experience	Operating time complication rate conversion rate
Pandya	1999	Case series		X	X	n/a	Assess the indications for conversion to laparotomy	Conversion rate
Schwandner	1999	Case series			X	3	Identify predictive factors for conversion	Morbidity,mortality, conversion
Hamel	2000	Case series			X	n/a	To identify if there is a difference in outcome for R/hemicolectomies with and without previous abdominal surgery	Morbidity and mortality, conversion rate
Schlachta 2000 (35)	2000	Case series			X	3	To identify preoperative factors that might predict operative time, conversion, intra and postop complications	Operative time, conversion rate, intra and postop complications
Lauter	2001	Case series		X	X	5	Initial experience of 150 cases of laparoscopic assisted colectomy	Operating time, conversion rate, length of stay, complications, port site metastases
Schlachta	2001	Case series		X	X	3	Define the learning curve for laparoscopic colorectal resections	Complications, conversion rate, operating time
LeMoine	2003	Case series		X	X	n/a	Analyse the causes and consequences of conversion in lap sigmoidectomy for diverticular disease	Conversion rate, morbidity, mortality
Lin	2003	Intervention study	X			6	Retrospective review on the impact of a teaching model in LCS in an urban community hospital. Model for teaching laparoscopic colectomy to surgical residents	trainee satisfaction scores, morbidity Data, Operating time
Mehall	2005	Intervention study	X			2	To compare the technical success and outcomes of laparoscopic colectomy performed by resident surgeons and attending surgeons	Operating time, complication rate, conversion rate
Tekkis	2005	Case Series		X	X	4	Identification of risk factors for conversion	Regression analysis of potential risk factors, RA-CUSUM for learning curve
Arteaga Gonzalez	2006	Case series		X		1	Assess the results of laparoscopic surgery for rectal carcinoma during the learning curve	Clinical, anatomopathological, economic variables
Reichenbach	2006	Case series		X		3	Assess whether there is a difference between outcomes of cases performed at a university hospital rather than a district general hospital	Patient demographics, indication for surgery, op time, morbidity and mortality, conversion rate, LN count, path
Sebajang	2006	Intervention study	X			4	Assess whether telementoring and telerobotic assistance improves the range and quality of lap colorectal surgery performed by community surgeons	Conversion rate, complication rate, length of hospital stay, 30d mortality
Birch	2007	Intervention study	X	X		7	Assess the impact of mentoring on a single institution's advanced MIS practice	Adoption rate, conversion rate, operative time, intraoperative complications, postoperative complications, mortality and length of hospital stay
Daetwiler	2007	Intervention study	X			13	Assess if it is safe to introduce easy cases early into residency	Conversion, complication, hospital stay, operative time, conversion rate, intraoperative blood loss
Kim	2007	Case series		X		1	Assessing the learning curve	Patient demographics, indication for surgery, operation time, morbidity and mortality, conversion rate
Agha	2008	Case series		X	X	6	Analyse the conversion rate and its influence on early postoperative morbidity/ short-term outcome	Conversion rate
Champagne	2008	Case series	X			1	To determine what approach (SLC vs. HAC) is most effective for training LCS? Aim to see if resident does better with HAC vs straight laparoscopy	Completion of operation by trainee
Kirchhoff	2008	Case series			X	13	Identification of predictive factors for intra and postoperative complications	operation time, morbidity and mortality, conversion rate
Pigazzi	2008	Intervention study	X			10	Assess impact of a fulltime lap trained preceptor on department	Conversion rate, complication, length of stay, operative times, blood loss
Rotholtz	2008	Case series		X	X	n/a	Identification of predictive factors for conversion and hence determine pattern in conversion rate	Univariate analysis for identification, multivariate for confirmation
Tan	2008	Case series			X	n/a	Determine risk factors for conversion in laparoscopic cases	Conversion rate
Park	2009	Case series		X	X	1	Identification of risk factors for conversion	Morbidity, mortality conversion
Schlachta	2009	Intervention study	X			2	Description of experience with telementoring	Conversion, operating time, hospital stay, complication, mortality

[Previous page]

Table 15. Summary of included studies

3.3.2. Study description and publication bias

In 6 studies the impact of a mentor's physical presence in the operating room on the trainees' performance was assessed, taking mortality and morbidity data as end points from a total of 713 cases (Lin 2003; Mehall 2005; Birch 2007; Daetwiler 2007; Champagne 2008; Pigazzi 2008). In a further 2 studies representing 38 cases the effect of a mentor present in the operating room only at the beginning of the training phase, and then with subsequent training via a remote video-link device were evaluated (tele-mentoring) (Sebajang 2006; Schlachta 2009). No publication bias was detected for these studies as no significant asymmetry was shown on the funnel plot [Figure 12, (Egger 1997)].

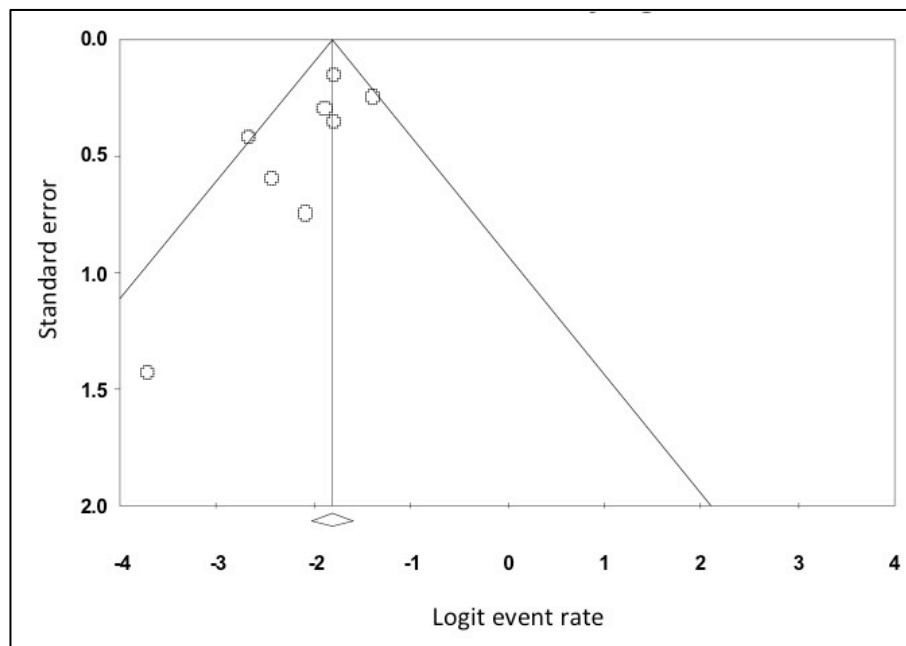


Figure 12. Funnel plot for studies of the supervised subgroup

3.3.3. Subgroup comparison and meta-analysis

A comparison of the pooled weighted outcomes of all supervised cases (n=751) compared with those of the control group (n=5,313) showed no significant difference in conversion, anastomotic leak or mortality rate (**Table 16**). Apart from conversions, data for other parameters was insufficient in the non-supervised group. The comparison of conversion rates revealed a significant difference in favour of the mentored group (13% for mentored, 20% for non-mentored, mixed model $p=0.033$, fixed model $p<0.001$). Furthermore, the conversion rates were significantly different for non-mentored and expert surgeons (20% vs. 11%, both models $p<0.001$), but no clear difference was found between mentored trainees and expert surgeons (13% vs. 11%, mixed model $p=0.284$, fixed model $p=0.013$) (**Table 17, Figures 13 and 14**).

Parameter	Supervised* (n=751)	Control* (n=5,313)	p-value**
Conversions	0.13 (0.10-0.17)	0.10 (0.77-0.13)	0.285
Complications	0.20 (0.16-0.24)	0.25 (0.19-0.31)	0.493
Anastomotic leak	0.03 (0.01-0.06)	0.04 (0.02-0.08)	0.368
Mortality	0.02 (0.01-0.04)	0.02 (0.01-0.02)	0.568
* pooled mean rates, random effects model (95% confidence intervals) ** mixed effect model (method of moments)			

Table 16. Comparison of different parameters between the supervised and the control subgroups

Statistical model	Supervised (n=751)	Control (n=5,313)	Non-supervised (n=665)
Fixed effects	0.14 (0.12-0.17)	0.12 (0.11-0.13)	0.22 (0.19-0.25)
Random effects	0.13 (0.10-0.17)	0.12 (0.10-0.13)	0.21 (0.15-0.26)

$p=0.284$ $p<0.001$
 $p=0.033$

Table 17. Comparison of conversion rates between subgroups and control group.

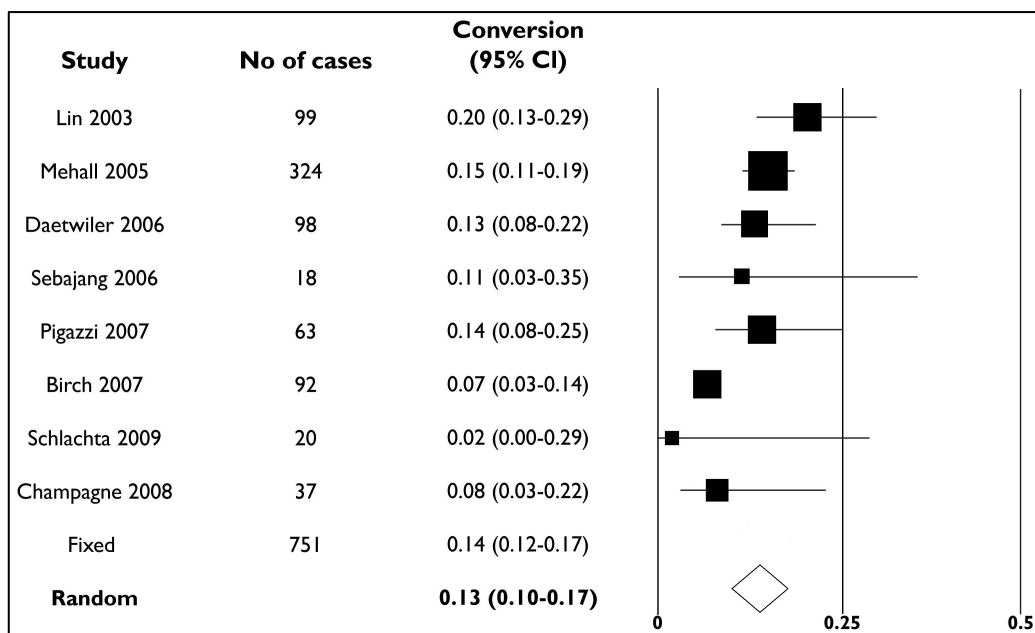


Figure 13. Forrest plot for weighted conversion rates within the supervised subgroup

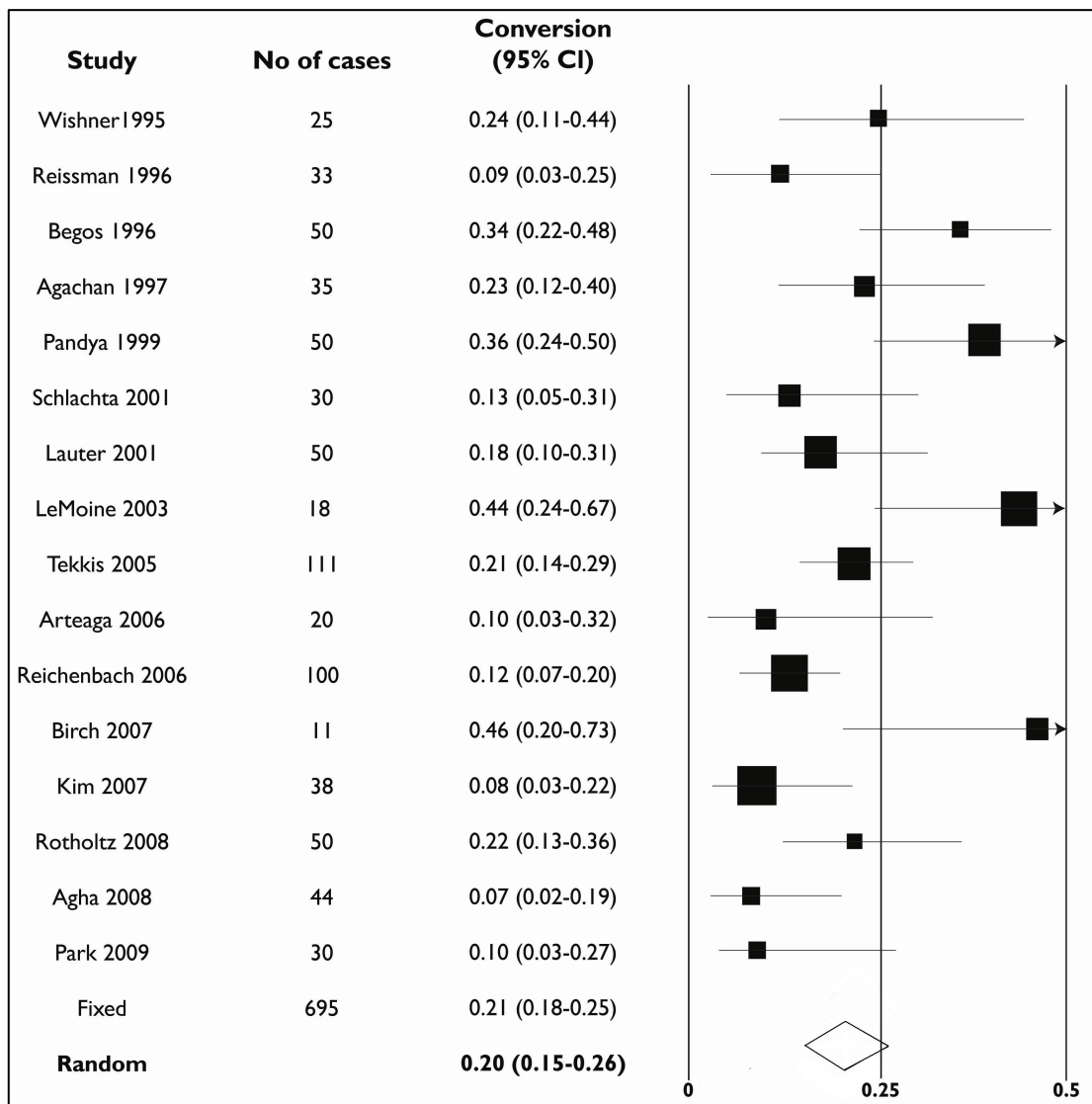


Figure 14. Forrest plot for weighted conversion rates within the non- supervised subgroup

3.4. Discussion

3.5. Summary

This review shows that supervised trainees achieve good clinical results in LCS. Analysed data of 6,064 patients reflects that trainees with an appropriate level of supervision generate the same complication, conversion and mortality rates as expert surgeons. Similar conclusions have been drawn from a large single centre study for carotid endarterectomy (Rijbroek 2003).

3.5.1. Are morbidity and mortality parameters useful for performance assessment?

Using morbidity and mortality data as a global indicator of performance has several limitations. Some of them have already been explained in the previous chapter. The analysis of this chapter highlights further limitations. Firstly, audit results can only be assessed when both large numbers and long-term follow-up data are available. This takes time and does not avoid unfavourable outcomes accumulating whilst the audit is in progress. Secondly, surgical competence is multi-dimensional and includes operative skills, cognitive factors, personality traits and decision-making. Thirdly, morbidity and mortality data cannot provide a prescriptive method for specifying how the performance of a procedure can be improved. Despite these shortfalls, an audit process remains essential as it promotes monitoring to ensure the maintenance of safe clinical practice throughout training. In many studies conversion rate is used as a performance parameter, though some authors state that this should not be considered to be a complication in itself since the decision to convert may also reflect the experienced surgeon's sound judgement (Dincler 2003; Tekkis 2005a; Tekkis 2005c; Faiz 2009b; Schlachta 2009).

3.5.2. The impact of clinical supervision

It would be expected that surgeons have similar complication and conversion rates at the same stage of training. Interestingly, this review demonstrated a significant difference in conversion rates between mentored and non-mentored trainees and hence this is likely to represent a difference in technical ability to achieve completion of a case laparoscopically. Although trainees at this level are often already proficient in basic technical skills, an experienced trainer may further aid intra-operative decision-making and comprehension of the anatomy, and guide the trainee towards error reduction mechanisms and thus potentially minimize the rate of unnecessary conversions. This cognitive apprenticeship model is a combined process of teaching craft and supporting strategic thinking by providing continuous feedback to the trainee during or after each training session (Collins 1987). The effects on clinical outcome observed in this meta-analysis can be explained by the presence of an expert who actively supports the trainee with verbal or physical intervention when appropriate. However, it might also represent a faster learning process in which itself leads to improved clinical outcome. Vygotsky described the *Zone of Proximal Development* as the range of tasks someone is able to learn. A capable teacher can significantly enhance the trainee's ability to move up along this scale, and boost performance (Vygotsky 1978; Zorga 2002).

3.5.3. Limitations

This review also has several limitations. Firstly, a direct comparison between the training groups and the experts as a control subgroup may be problematic as there was no case control. Expert surgeons are more likely to operate on more complex cases, which may explain why no significant difference between supervised trainees and experienced surgeons (control group) was found. Nevertheless, a significant difference was found when the mentored and non-mentored subgroups were compared. Secondly, there are difficulties in synthesizing data from studies of different design and methodological strength.

3.5.4. Conclusion

In conclusion, interpretation of clinical outcome data as a measure for technical performance is ambiguous. Nevertheless, it is crucial to collect this data prospectively as it ensures and documents the clinical effectiveness and safety of an educational programme. There is a distinct need for alternative assessment methods that allow any individual proficiency gain curve to be depicted.

3.6. Disclosures

The contents of this chapter were published as part of an article in the journal *Annals of Surgery*. The paper was written in collaboration with another researcher (two first authors). The analysis of this part of the study was mainly performed by DM, with statistical support by Dr Melody Ni and Dr Elena Kulinskaya (Imperial College London).

The original reference of the paper is:

Miskovic D and Wyles SM, Ni M, Darzi AW, Hanna GB. Simulation and Mentoring in Laparoscopic Colorectal Surgery as an illustration of training in technically advanced procedures. A systematic review. *Ann Surg* 2010 252(6):943-951. PMID: 21107103

Parts of this paper were also presented at the meeting of the European Association of Endoscopic Surgeons in Geneva in 2010. The reference is:

D Miskovic, SM Wyles, A Darzi, GB Hanna. Fifteen years of learning curves in laparoscopic colorectal surgery – what have we learnt? EAES (European Association of Endoscopic Surgeons) June 2010, Geneva/ Switzerland.

3.6.1. Acknowledgments

I would like to thank Daniel Birch (Edmonton, Canada), Selim Dincler and Peter Buchmann (Zurich, Switzerland), Omar Faiz (London, UK), In Ja Park (Daegu, South Korea), Nicolas Rotholtz (Buenos Aires, Argentina), and Christopher Schlachta (Toronto, Canada) for providing us with non-published data of their series. I am grateful to Matthias Egger, Head of the Department of Epidemiology and Public Health, University of Bern, for his helpful comments.

4. DEVELOPMENT, VALIDATION AND IMPLEMENTATION OF A MONITORING TOOL FOR TRAINING IN LAPAROSCOPIC COLORECTAL SURGERY

4.1.	Introduction	110
4.1.1.	Background	110
4.1.2.	Formative and summative assessment	111
4.1.3.	Is there a perfect assessment tool?	112
4.1.3.1.	Classical test theory	112
4.1.3.2.	Advanced concepts: generalizability theory	113
4.1.3.3.	Feasibility and implementation.....	113
4.1.4.	Aim	114
4.2.	Methods.....	115
4.2.1.	Scale development.....	115
4.2.2.	Classical test theory	115
4.2.2.1.	Validity	115
4.2.2.2.	Reliability.....	116
4.2.3.	Feasibility and implementation	116
4.2.4.	Statistical analysis	116
4.3.	Results.....	118
4.3.1.	Scale development.....	118
4.3.2.	Construct validity and reliability	121
4.3.3.	Feasibility and implementation	123
4.4.	Discussion	125
4.4.1.	Strengths of the tool in the context of formative assessment	125
4.4.2.	The use of the tool for proficiency gain evaluation.....	126
4.4.3.	Keys for successful implementation.....	126
4.4.4.	Limitations of the study.....	127
4.4.5.	Conclusion.....	127
4.5.	Disclosures.....	128
4.6.	Acknowledgments.....	129

4.1. Introduction

4.1.1. Background

The National Training Programme in England provides structured laparoscopic training to established consultant colorectal surgeons, with the aim of training a large number consultants to a level of competence in laparoscopic colorectal surgery (LCS) within 5 years (Lapco 2009).

Monitoring the training progression of trainees is an essential component of such an educational programme and requires regular structured assessment (van der Vleuten 2005). This approach not only provides information on the effectiveness of the programme, but also permits the detection of abnormally prolonged proficiency gain curves of individual trainees who may require additional support. Previous reports on attempts to implement assessment tools into LCS training programs lacked validity and educational impact (Sidhu 2006; Wohaibi 2007). Although observational assessment tools have been shown to be reliable and valid for basic surgical skills, their value is restricted to advanced laparoscopic surgery because of a ceiling effect (Martin 1997; Munz 2004; Vassiliou 2005). In the previous chapters it was demonstrated that clinical outcome parameters have been used to analyse the proficiency gain curve retrospectively. There are several problems with this approach, which include ethical issues, inaccuracy for individuals, the impact of case selection and the influence of an experienced surgeon present in theatre. Furthermore, clinical outcomes data do not provide a descriptive analysis of the operative areas that require improvement (see **chapters 2 and 3** for details).

4.1.2. Formative and summative assessment

Assessment of medical professionals may have different aims (Epstein 2007). Firstly, it can provide feedback to the learner and promote quality and efficiency of training by targeting teaching and learning in areas that require further improvement (Black 2009). Secondly, assessment can be used to identify incompetency and therefore is a tool to protect the public from bad clinical practice. Thirdly, assessment can be used as a selection tool to select potential candidates for a training curriculum. The terms formative and summative assessment were introduced by Scriven in 1967 and since then have become well established in the educational literature with a variety of definitions available, depending on the learning context (Scriven 1967). In a medical or surgical environment, formative assessment can be seen as formal or

informal feedback, guiding future learning and inspiring learners to achieve higher goals (Ben-David 2000). Summative assessment is a formal assessment aiming to define competence (see **Chapter 5**). This type of assessment does not often provide detailed feedback on which areas of training need to be improved (Schuwirth 2004).

4.1.3. Is there a perfect assessment tool?

The requirements for a perfect assessment tool are somewhat self-limiting. One might argue that a perfect assessment method is “a standardised evaluation tool that accurately mirrors reality”. However, this statement is almost a contradiction in itself, since reality hardly ever appears in a standardised form. The ability to react appropriately to the unexpected i.e. may also be part of competence and very difficult to assess with a standardised tool. Hence, it is probably more appropriate to scale down expectations to achieve a more fair assessment tool and define some basic and practical, rather than ideal, principles. It is remarkable that test theory almost exclusively originates from psychometrics and little is known whether or not these theories apply to performance assessment in a surgical context.

4.1.3.1. Classical test theory

In classical test theory assessment tools can be evaluated in several dimensions, describing different aspects of validity and reliability. The hypothesis is that any measurement entails a degree of error: $Reality = Measurement + Error$. The key concept of classical test theory is to measure different types of errors separately. A summary of different error dimensions and definitions can be found in **Table 18** (Ahmed 2011c). In addition, validity can also be defined as *convergent* validity i.e. the degree to which the tool represents other tests that theoretically measure the same construct. On the other hand, *discriminant* validity describes the degree to which a method differs from another tool that measures another dimension of

performance (Streiner 2008). These terms may overlap with other forms of validity, such as predictive or concurrent validity.

4.1.3.2. Advanced concepts: generalizability theory

The problem with classical test theory is that only one dimension at a time can be measured. Although this can be informative about the areas where a test may be limited, it fails to inform a great deal about the overall usefulness of a test. What does it precisely mean if a test turns out to be excellent in all dimensions apart from inter-test reliability? Can we use such a test? These are difficult questions with no definitive answer. The generalizability theory was introduced by Cronbach (Cronbach 1963) and addresses some of the shortfalls of classical test theory. It describes a test in multiple dimensions simultaneously and may be a more accurate description. Further details and an example can be found in **Chapter 6**.

4.1.3.3. Feasibility and implementation

Requirements for a good assessment tool need to be taken further, beyond validity and reliability. Feasibility is a prerequisite for successful implementation of any assessment method. Will those who expect to use it either as assessors or trainees accept the test? Scientific evidence for the validity and reliability of a test may not be sufficient to convince a group of adult learners to use a test, but other factors such as assessment time, tangibility and the availability of assessor training may be hidden factors to allow implementation of a tool successfully (Ahmed 2009; Ahmed 2011c). A good test, especially as a formative assessment, may need to provide useful feedback (educational impact) and it should be cost-effective (van der Vleuten 2005). Little is known about the impact of availability and incentive needed to

employ such a test, but these may be very important issues, especially in an adult learning environment.

Dimension	Definition
Face validity	Extent to which the examination resembles the situation in the real world
Content validity	Extent to which the intended content domain is being measured by the assessment exercise.
Construct validity	Extent to which a test is able to differentiate between a good and bad performer
Concurrent validity	Extent to which the results of the test correlate with gold standard tests known to measure the same domain
Predictive validity	Extent to which this assessment will predict future performance
Convergent validity	Extent to which a test correlates with another test, measuring the same area of performance
Discriminant validity	Extent to which a test discriminates from another test which measures another area of performance
Inter-rater Reliability	Extent of agreement between two or more assessors/ observers
Interitem Reliability	Extent to which different components of a test correlate (internal consistency)
Intertest Reliability	Ability of a test to generate similar results when applied at two different time points
Acceptability	Extent to which an assessment procedure is accepted by the subjects involved in the assessment.
Educational Impact	Extent to which test results and feedback contribute to improve the learning strategy on behalf of the trainer and the trainee
Cost Effectiveness	Technical and non-technical requirements for the implementation of a test into clinical practice

Table 18. Classical requirements of a good assessment tool (from REF Ahmed 2010)

4.1.4. Aim

The aims of this chapter were to develop, validate and implement a method for monitoring training progression in laparoscopic colorectal surgery in order to provide a platform for formative assessment to trainers and learners.

4.2. Methods

4.2.1. Scale development

In order to provide a scoring system that is non-judgmental, a Juster scale was introduced, which describes the amount of support needed to perform a task. A proposal of the scale was sent to members of the educational committee of the National Training Programme, represented by 12 expert laparoscopic surgeons, and an educationalist, all from ten different centers in the UK. The draft was discussed and amended via two telephone conferences, and all members of the committee approved a final version.

4.2.2. Classical test theory

4.2.2.1. Validity

Face validity was not relevant, since it is a workplace assessment and not a constructed task performance analysis. For *content validity*, a selective literature search on operating techniques in LCS was performed using internet sources, educational videos and books (Champault 2007; Evans 2009; 2010). Based on these sources a generic task analysis for laparoscopic colorectal resection was created. *Construct validity* was calculated by comparing four different experience groups. Trainees who carried out more than 15 consecutive cases were extracted, and their individual data were split into four experience groups (cases 1-5, 6-10, 11-15 and >15). Due to lack of a gold standard method, it was not possible to establish *convergent or discriminant validity*.

4.2.2.2. Reliability

Trainees and trainers were required to complete the same scoring sheet independently, which facilitated the assessment of *inter-rater reliability*. *Inter-item reliability* was calculated by comparing different areas of the assessment form. Analysis of *test-retest* and *intra-rater reliability* was not possible due to the setting of the study.

4.2.3. Feasibility and implementation

Ethical approval was obtained from the National Ethics Committee (REC 04/Q040356). The validation period was defined as the first year after implementation. During the first stage of the pilot phase (October 2008-May 2009) the form was available only in paper format. A web-based electronic form was created for the second phase (June 2009-November 2009). In order to evaluate feasibility, each assessment form also contained a short survey on its perceived usefulness, and the time spent to complete it.

4.2.4. Statistical analysis

The Statistical Package for the Social Sciences software (Version 17.0.0, SPSS Chicago, IL, USA) and Excel (Microsoft® Excel® for Mac 2008, Microsoft Corporation, Redmond, WA) was used. For construct validity, a comparative test for non-parametric data (Wilcoxon rank test) was used. Overall effects were tested using one-way Analysis of Variance (ANOVA). Inter-rater reliability between the two was measured using the intraclass correlation coefficient (ICC) and demonstrated on a Bland-Altman plot (Bland 1986). Inter-item reliability was calculated using Cronbach's α (Bland 1997). CUSUM (cumulative sum) charts were plotted using the equation $S_i = S_{i-1} + (x_i - x_R)$; $S_0 = 0$; $x_R = 5$; S_i is the cumulative sum, x_i the average score per procedure number and x_R the target score (Grigg 2003). For feasibility, the mean time

(standard deviations), to complete a form was calculated. Analysis of Likert scales was performed using median and interquartile range (IQR).

4.3. Results

4.3.1. Scale development

Right hemicolectomy, sigmoid resection, anterior resection, low anterior resection, total and subtotal colectomy and laparoscopically assisted abdomino-perineal resection were all classified as laparoscopic colorectal resections. Four generic task zones and 12 generic task steps were identified (**Table 19**). For several operations and techniques, some of the steps were not applicable (e.g. step 9 for right hemicolectomy or step 5 for medial to lateral dissection without laparoscopic dissection of vascular pedicle). Some non-laparoscopic steps (e.g. patient positioning) were included in the list as they may substantially facilitate the procedure. Others (e.g. wound closure) were not included, as they were not deemed to be related to the quality of advanced laparoscopic surgical skills.

The rationale for the scoring system was to avoid judgmental expressions such as “good” or “poor”. Instead the degree of support required to complete the task was used as a more objective and reliable parameter. The amount of physical and verbal support required from the trainer was divided into six categories using a “Juster” scale (**Table 20**) (Hoek 1993). For all other situations (e.g. another trainee performing the step, or the step not necessary for the operation, or the trainer performing the step when restricted by time constraint, etc.), a “Non-applicable” option was added for each step. A full description of the assessment tool can be found in **Figure 15**.

A. Exposure	
1.	Theatre setup (position of surgeons, scrub nurse, drapes etc.)
2.	Patient positioning
3.	Laparoscopic access (open, Veress needle or other techniques, and insertion of ports)
4.	Exposure of operating field (moving of omentum, small bowel etc.)
B. Dissection of vascular pedicle	
5.	Dissection of vascular pedicle (incision of peritoneum, creation of window below and above, and dissection with stapler, clips, ultrasound dissection tool or other techniques)
6.	Retrocolic dissection of mesentry (right side towards hepatic flexure, left side towards splenic flexure)
7.	Identification of landmark (right side: duodenum, left side: left ureter)
C. Mobilization	
8.	Dissection of flexure (right side: hepatic, left side: splenic)
9.	Mesorectal dissection (including total mesorectal excision (TME), only for rectal resection)
10.	Dissection of bowel (transection, using stapler other similar device)
D. Anastomosis	
11.	Extraction of specimen (creation of incision, bringing out specimen, completion of resection)
12.	Anastomosis (intra- or extra-corporeal)

Table 19. Generic task analysis for laparoscopic colectomy

0 – Not applicable
1 - Not performed, step had to be done by trainer
2 - Partly performed, step had to be partly done by trainer
3 - Performed, with substantial verbal support
4 - Performed with minor verbal support
5 - Competent performance, safe (without guidance)
6 - Proficient performance, couldn't be better

Table 20. Generic Juster Scale for a non-judgmental assessment

A. PRECEPTOR						
A1 Name of Preceptor						
A2 Name of Trainee						
A3 Patient ID						
B4 Operating date						
B. ASSESSMENT				1 Not performed, step had to be done by trainer 2 Partly performed, step had to be partly done by trainer 3 Performed, with substantial verbal support 4 Performed with minor verbal support 5 Competent performance, safe (without guidance) 6 Proficient performance, couldn't be better		
EXPOSURE						
B1 Correct theater setup	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B2 Appropriate patient positioning	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B3 Safe access technique	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B4 Exposure of operating field	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
VASCULAR						
B5 Safe dissection of vascular pedicle	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B6 Dissection of mesentry (retrocolic)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B7 Safe identification of ureter or duodenum	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
MOBILISATION						
B8 Dissection of hepatic or splenic flexure	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B9 Mesorectal dissection (where applicable)	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B10 Safe dissection of bowel	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
ANASTOMOSIS						
B11 Safe extraction of specimen	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
B12 Anastomosis	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
OVERALL PERFORMANCE						
B13 Overall Performance	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐
C. COMMENT						
D. ASSESSING THE ASSESSMENT						
C1 How difficult was this operation	very easy 1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐ very difficult
C2 How useful was this form?	not useful 1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐ very useful
C3 How long did it take you to fill in this form?	minutes					

120

4.3.2. Construct validity and reliability

Within the pilot period 30 trainers and 52 trainees submitted 333 and 277 forms respectively, from a total of ten national training centres. Twelve trainees underwent more than 15 training sessions each during the pilot phase. Construct validity was analysed on 193 cases from 12 trainees who carried out more than 15 cases. One-way analysis of variance of four experience groups (groups 1-4) confirmed the difference between the experience levels [$F(3,40)=6.128$, $p<0.001$] (**Figure 16**). There were significant differences between groups 1 and 2 (4 vs. 4.6, $p=0.030$) and between 3 and 4 (4.7 vs. 5, $p=0.008$), but not between groups 2 and 3 ($p=0.850$). Inter-rater reliability analysis was conducted on 94 cases, for which both the trainer and the paired trainee returned a completed form. In the remaining cases, there was failure to either complete the form, or provide the patient identification number. These issues were resolved at a later stage after improving participant instructions and by introducing on-line submission using an electronic form. Comparison of complete datasets showed a high level of inter-rater reliability ($ICC=0.867$, $F(94, 94)=13.989$, $p<0.001$, **Figure 17**). On the analysis of the complete data set (610 forms), there was excellent inter-item reliability with $\alpha=0.937$ for the trainers' assessments and $\alpha=0.920$ for the trainees' assessments. The results are summarised in **Table 21**.

Parameter	Test	Value
Construct validity	ANOVA	<0.001
Reliability:		
- inter-rater	ICC	0.867
- inter-item (trainer)	Cronbach's α	0.937
- inter-item (trainee)	Cronbach's α	0.920
Feasibility	Score	5/6

Table 21. Summary of results.

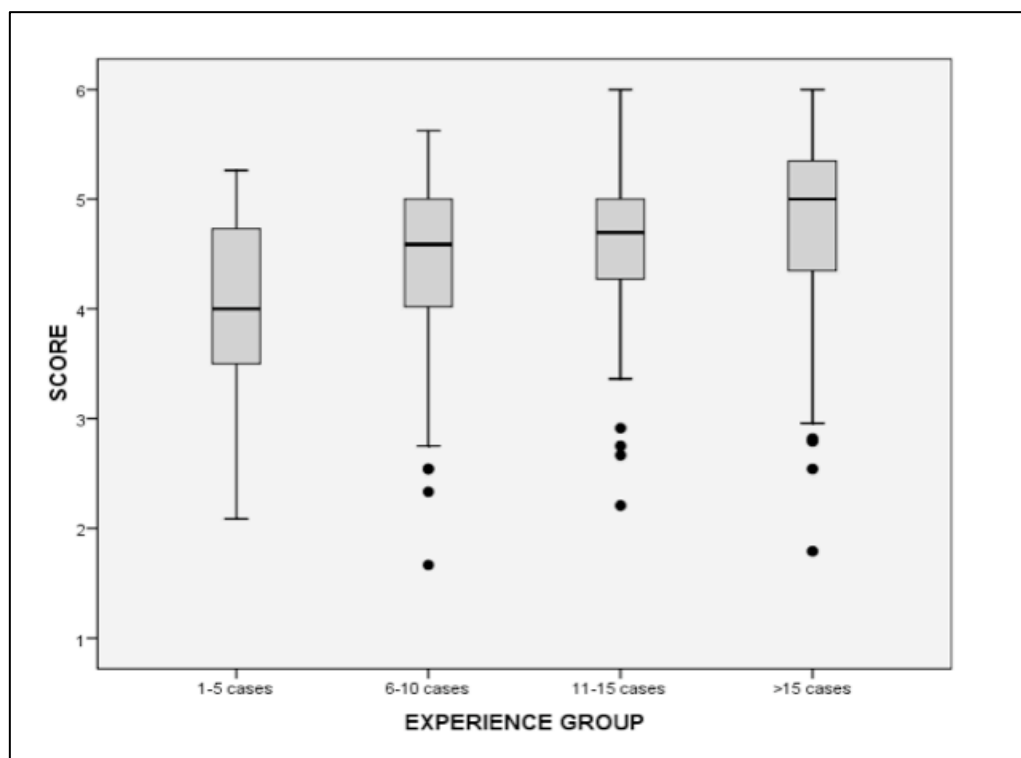


Figure 16. Construct validity for different experience groups.

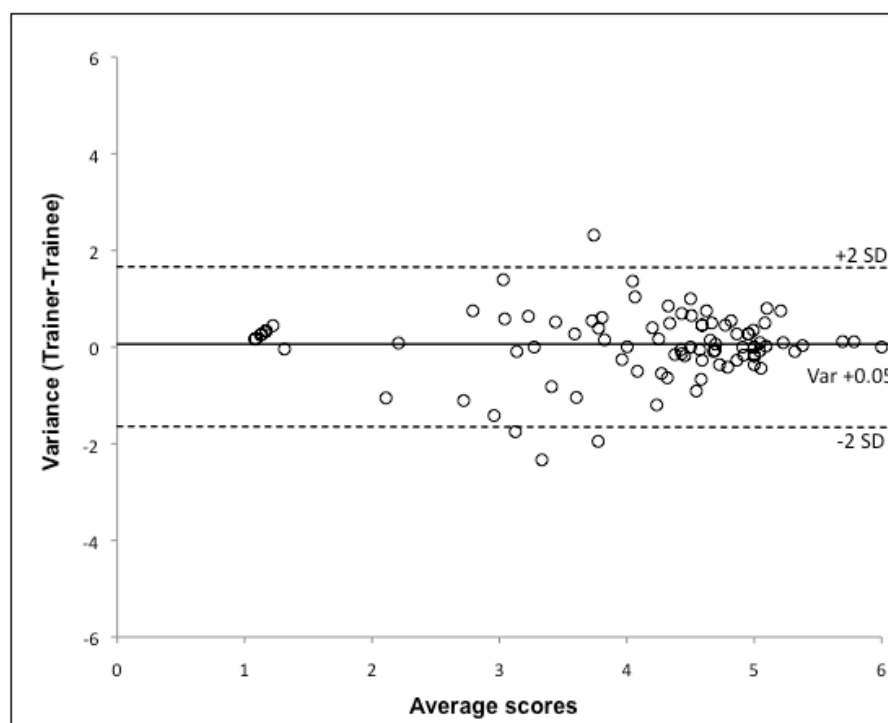


Figure 17. Bland-Altman plot for inter-rater reliability.

4.3.3. Feasibility and implementation

The median assessment time was 3.3 minutes [inter-quartile range (IQR)=1-5 minutes]. The median score for the usefulness of the form was 5 (IQR=4-5, 1-not useful at all, 6- very useful). For the construction of proficiency gain curves the trainers' scores of all 333 procedures were used. The target score was set at 5 [“competent performance, safe (without guidance)”]. The average score indicates that the target score was consistently achieved after the 21st procedure. However, this value is different for different components of the operation. For a detailed analysis the tasks steps were re-grouped and CUSUM charts of six different skill areas were computed. The areas “Set-up” and “Anastomosis” reached target scores from a very early stage. The curve for “Exposure” shows an inflection point around the 17th procedure, for “Vascular pedicle”, this occurred around the 22nd procedure. For “Mobilization” and “Hepatic/splenic flexure” no inflection point was observed until procedure 25 (**Figure 18**). Furthermore, proficiency gain curves can differ substantially between individual trainees, as shown in **Figure 19**.

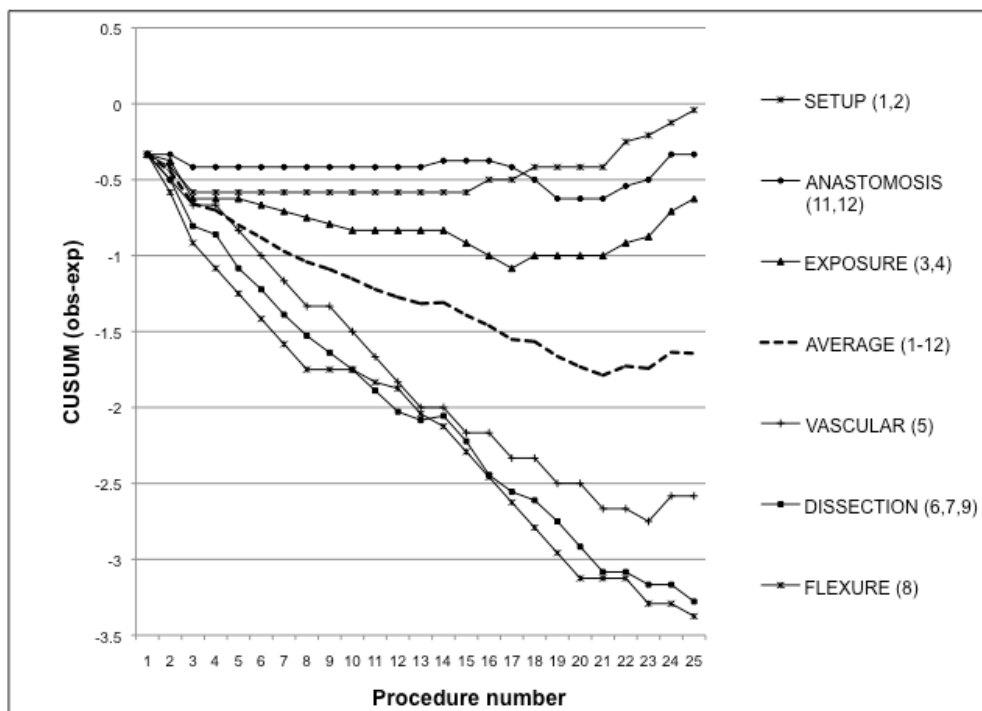


Figure 18. CUSUM proficiency gain curves for different areas of the procedure.

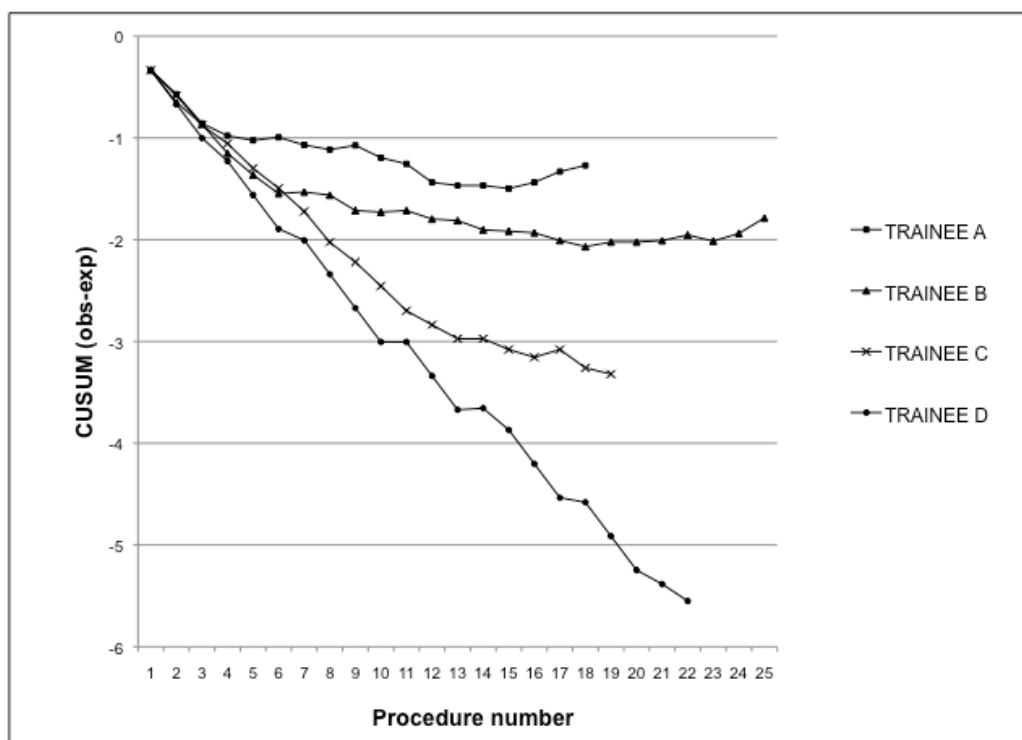


Figure 19. Examples of proficiency gain curves of four individual surgeons.

4.4. Discussion

In this chapter reliability, validity and feasibility of an observational assessment tool to monitor the proficiency gain process of laparoscopic colorectal surgery was demonstrated. There is no comparable data in a similar setting of a national educational program for advanced laparoscopic surgery.

4.4.1. Strengths of the tool in the context of formative assessment

The main strength of this tool is the combination of a task-specific checklist with a generic scale which is both highly practical and reliable. The use of generic task steps for laparoscopic colorectal resection allows the identification of areas within the operation that may be more difficult to master, and can thus help the trainers to focus their teaching on these areas. The descriptive scale (“Juster” scale) details the amount of support needed to complete a task which is a more objective parameter than judgmental scales (Hoek 1993; Streiner 2008). This may have added to the high level of inter-rater reliability. Other tools, such as the Objective Structured Assessment of Technical Skill (OSATS), also use descriptive scales, but each category has its specific descriptors, which makes the assessment more difficult and creates a learning curve for the assessor (Martin 1997; Sarker 2011). This requires assessor training, which is logistically difficult to coordinate for a national training programme with a high number of trainers in different locations. One previous study used a computer-based database to enter performance-related parameters of trainees in a single-centre study, showing construct validity when comparing trainees at different levels (Wohaibi 2007). However, the conclusions and the educational value remain unclear. Other generic assessment tools for laparoscopic surgery have shown construct validity, but none of them was used for the continuous description of the learning progress (Martin 1997; Vassiliou 2005). The monitoring tool presented in the current study has been widely accepted within the National Training Programme in England, as shown by the reported high trainer satisfaction levels. Another

indicator for its popularity is the fact that it is being used increasingly outside the National Training Program by general surgical trainees (registrars) and their trainers in the UK.

4.4.2. The use of the tool for proficiency gain evaluation

CUSUM charts were used to identify the point at which the trainees reach a satisfactory score. Although the proficiency gain curve for the overall score indicates in this preliminary analysis that the target score is generally achieved after the 21st procedure, this value is different for individual trainees and for various components of the operation. Certain steps of the procedure (vascular pedicle, mobilization, hepatic and splenic flexure) may take longer to learn than the overall score suggests. The analysis performed in **Chapter 3** suggests that the learning curve for self-taught surgeons is much longer, with up to 150 cases per surgeon. These proficiency gain curves, however, are hardly comparable to the present data since the current trainees are established consultants who have been trained actively by expert laparoscopic surgeons, hence the learning curve may be shortened substantially (Schlachta 2003; Birch 2007; Choi 2009; Li 2009). It can also be assumed that there was a certain degree of case selection in favour of more easy cases, suitable for beginners. The data also demonstrated variations between the different candidates in achieving their path to competency. This highlights the need for an individualized competency-based assessment rather than a set number of cases in order to be signed off for independent practice.

4.4.3. Keys for successful implementation

At the beginning of the implementation of the monitoring tool, the submission rate was lower than expected. Two events account for a significant improvement in the submission rate. The first was the introduction of a web-based assessment form by the coordinating centre, which facilitated the process substantially. The second was that the governing body of the National

Training Program used the number of completed forms as the evidence of training activity and consequently for funding allocation.

4.4.4. Limitations of the study

There are some limitations to this study. Content validity was not evaluated in a systematic way. Although multiple sources (literature, expert opinion, internet search) were used to establish content validity, certain procedural skills may have been missed. Also, we did not apply the generalizability theory to further explore reliability, as there were too few facets to implement into the equation (Crossley 2002; Streiner 2008). Furthermore, although trainees and trainers were instructed to fill in the form independently, there was no control and there is a potential risk for false positive interrater reliability.

4.4.5. Conclusion

The generic assessment scale (GAS) as presented in this chapter meets several requirements of a useful assessment tool and has been successfully implemented in a National Training Programme. Although the application of this tool enables the description of proficiency gain curves and may be helpful as a formative assessment tool, it does not necessarily provide reliable information on the competency level at the end of training. Further assessment methods need to be developed in terms of summative assessment.

4.5. Disclosures

Contents of this chapter have been published in the journal *Surgical Endoscopy*. The full reference of the paper is:

Miskovic D, Wyles SM, Coleman MG, Hanna GB. Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program. *Surg Endosc* (2011) 25:1136–1142. PMID: 20835723

The paper was also presented at the meeting of the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) in Washington DC in 2010. The reference is:

D Miskovic, SM Wyles, MG Coleman, GB Hanna. A National Training Program for laparoscopic colorectal surgery in the UK. Society of American Gastrointestinal and Endoscopic Surgeons (SAGES), Washington DC/ USA, April 2010

4.6. Acknowledgments

I would like to thank all members of the educational committee and the steering group representing the centres participating in the National Training Programme for their boundless effort to support the educational activities: Austin Acheson and Charles Maxwell-Armstrong (Nottingham), Tom Cecil (Basingstoke), Chris Cunningham (Oxford), Vivek Datta and Sav Papagrigoriadis (London), Nader Francis (Yeovil), John Griffith (Bradford), James Gunn (Hull), Alan Horgan (Newcastle-upon-Tyne), Robin Kennedy (Harrow, London), Roger Motson (Colchester), Amjad Parvaiz (Portsmouth) and Timothy Rockall (Guildford).

5. AN OBSERVATIONAL CLINICAL HUMAN RELIABILITY APPROACH (OCHRA) FOR COMPETENCY ASSESSMENT IN LAPAROSCOPIC COLORECTAL SURGERY

5.1.	Introduction.....	130
5.1.1.	Background	130
5.1.2.	Human reliability methods in surgery.....	131
5.1.3.	Aim	132
5.2.	Methods.....	133
5.2.1.	Surgeons and case selection	133
5.2.2.	Observational Clinical Human Reliability Analysis (OCHRA).....	133
5.2.3.		137
5.2.4.	Dissection-Exposure ratio	137
5.2.5.	Expert review	138
5.2.6.	Data extraction and statistical analysis.....	138
5.3.	Results.....	139
5.3.1.	Case complexity and operating time	139
5.3.2.	Construct validity	139
5.3.3.	D/E ratio	142
5.3.4.	Concurrent validity.....	142
5.4.	Discussion	144
5.4.1.	Summary	144
5.4.2.	Interpretation of results	144
5.4.3.	Limitations.....	146
5.4.4.	Conclusion.....	146
5.5.	Disclosures.....	147

5.1. Introduction

5.1.1. Background

In the previous chapter the development and validation of a method to monitor training progression and a platform to provide the trainee with formative feedback was demonstrated. Also discussed was the fact that this formative assessment does not necessarily provide a reliable guide to a trainee's capacity to perform the procedure in a competent way. It merely

suggests that the surgeon is capable of performing the operation without substantial input from the trainer. A crucial component of the National training programme is competency assessment at the end of supervised training before the surgeon embarks on independent clinical practice. Competency assessment was initially based on independent expert evaluation of operating videos, but novel methods are now required to further rationalize the process. Assessing competence in laparoscopic surgery has been attempted mainly for basic laparoscopic skills, using standardized simulated tasks or instruments to measure the path length of hand movements (Okraienec ; Hanna 1998; Moorthy 2003). In addition, observational assessment tools aimed at the junior level have been designed to assess global generic skills required in basic laparoscopic surgery (Martin 1997; Vassiliou 2005). Numerous studies have examined the construct validity of these methods by comparing junior and senior surgeons (Francis 2002; Vassiliou 2005; Datta 2006; Aggarwal 2008). However, their validity in assessing competent performance in advanced laparoscopic colorectal surgery among a group of specialist surgeons is doubtful, due to their inability to differentiate between performers at similar experience levels and a lack of procedure specific features (Munz 2004).

5.1.2. Human reliability methods in surgery

Human reliability methods use the evaluation of human factors in a work process (Kirwan 1992b; Kirwan 1992a). Its origins lie in the analysis of major industrial accidents, such as the Three Mile Island nuclear accident in the USA in 1979 (Hollnagel 1998). Subsequently, methods have been developed to identify errors and near misses in work processes in order to make the process safer and to prevent negative outcomes. This concept lends itself to surgical operations and has been introduced by Cuschieri and associates (Cuschieri 2006). The observational clinical human reliability analysis (OCHRA) has been developed to detect errors and near misses during surgical performance by means of operating video analysis. Significant

variations of error counts comparing surgeons of different training level were found (Joice 1998; Tang 2004b; Tang 2005). The method has been used for a variety of surgical procedures such as laparoscopic cholecystectomy, open rectal resection, laparoscopic pyloromyotomy, dacryocystorhinostomy and phacoemulsification (Tang 2004a; Gauba 2008). The concept implies that the fewer errors there are in a procedure, the less likely a negative outcome: even errors without an obvious direct adverse consequence may lead indirectly to an adverse outcome that is apparently disconnected to the initial error. An example of such an indirect consequence is a ureteric injury in a left sided colectomy. It is unusual that a surgeon simply transects an obscure structure, not knowing anything about the ureter. Hence, it is not a “transection error”. The true error usually started much earlier in the operation, i.e. entering the wrong tissue planes, complicated by unnecessary bleeding, eventually leading to a misperception of the anatomy and the transection of the ureter. This concept of identifying any deviation from the standard procedure (i.e. best surgical practice) can also be used as a measure for competence. In the example above, counting how many times the surgeon is entering the wrong tissue planes represent the risk of this surgeon cutting through the ureter, although this complication might not happen in the observed case. A low error count suggests safe and accurate operating, which is consistent with the definition of competent performance (van der Vleuten 2005).

5.1.3. Aim

The aim of this study was to investigate the potential of OCHRA in terms of construct and concurrent validity for competency assessment in laparoscopic colorectal surgery.

5.2. Methods

Operative video-recordings of laparoscopic colorectal resections were assessed using OCHRA. Expert assessment provided the gold standard for concurrent validity whereas the difference between passed and failed specialists was used for construct validity.

5.2.1. Surgeons and case selection

Trainees of the National training programme (NTP), termed “delegates”, were asked towards the end of their supervised training to submit two video-recorded cases for competency assessment, in order to be “signed off” from the program. Case selection was left to the surgeons, but they were advised to choose uncomplicated patients with one left- and one right-sided resection. These videos were submitted through an independent and blinded review process by two randomly selected NTP trainers. In addition to the sign-off videos, 6 experts were also asked to submit their own videos for the purpose of this study. These expert videos were then subjected to the same expert review process. NTP trainers were identified through a peer-reviewed structured process based on experience (>200 cases), case load and unit performance. By contrast, at the time of their sign-off, delegates had completed fewer than 50 resections.

5.2.2. Observational Clinical Human Reliability Analysis (OCHRA)

For OCHRA, a task analysis was performed on operative errors. The breakdown of tasks was created in accordance with the educational monitoring tool as described in **Chapter 4**, and which an expert group had approved previously. The aim was to create a task list that was generic and flexible enough to be applicable to different procedures including the approaches to colorectal resection, but specific enough to separate the tasks from one another. Some task areas, however, were not applicable for certain procedures [e.g. total mesorectal excision

(TME) in a right hemicolectomy]. For each task, the aim and a specific start point was defined (**Table 23**). Errors were defined and categorized, initially in accordance with previous descriptions of OCHRA (Joice 1998; Tang 2004b). For practical reasons and to facilitate scoring, two main error categories were created: *instrument usage errors* (“instrument”) and *tissue handling errors* (“tissue”). Instrument errors were defined as inaccurate use of a laparoscopic tool; tissue errors were defined as inappropriate interactions with the tissue. Each error was further described in detail, using “External Error Modes” (Joice 1998) (**Table 22**). If an error led to a direct consequence (e.g. bleeding), this was also recorded (Kirwan 1992a). Errors were recorded using scientific rating software (Observer XT, Version 8, Noldus Information Technology Inc., Leesburg (VA), USA). This software interface allows tagging point events and time periods on a video file using a customized rating scheme (**Figure 20**) For the purpose of this study, tasks and subtasks were treated as time periods and errors as point events. A researcher with a surgical background (DM) was trained by an expert in ergonomics and human reliability analysis techniques (GH), then gained experience in using the technique before rating all videos using OCHRA.

I.	Step not done
II.	Step partially completed
III.	Step repeated
IV.	Second step done in addition
V.	Second step done instead of first step
VI.	Step done out of sequence
VII.	Step done with too much force, speed, depth, distance, time, or rotation
VIII.	Step done with too little force, speed, depth, distance, time, or rotation
IX.	Step done in wrong orientation, direction, or point in space
X.	Step done on/with the wrong object (or plane)

Table 1. External Error Modes (EEM) according to Joice 1998.

	Task area	Task	Start point	Aim
1	Exposure	Port insertion	Bulging out and perforation of peritoneum	Ports inserted safely in adequate position
		Expose main vascular pedicle	Grasp small bowel	Small bowel removed from operating field, pedicle fully exposed
2	Dissection of vascular pedicle	Dissection around vascular pedicle	Incision of peritoneum and creation of lower window	Main vessel(s) sufficiently skeletonised at adequate level, landmarks (ureter, duodenum) defined
		Transsection of vessel	Clip-applier, vascular stapler or ultrasonic dissection device appears on screen	Vessel(s) secured and transected
3	Mobilization of colon	Medial mobilization	Mobilization medially to colon after vascular dissection	Mesentery released from Toldt fascia on medial side
		Lateral mobilization	Incision of peritoneal adhesions lateral to colon	Mesentery released from Toldt fascia on lateral side
4	Mobilization of flexure	Lateral mobilization	Incision of hepato-/splenic ligaments	Complete release of lateral attachments of colonic flexure
		Medial mobilization	Incision of gastrocolic ligament/ release of omentum	Complete release of medial attachments of colonic flexure
5	TME	Posterior/lateral mobilization	Entry into posterior TME plane	Rectum released from posterior and lateral, nerves protected, TME plane respected
		Anterior mobilization	Incision of anterior peritoneal pouch	Rectum released from anterior, structures (vagina, seminal vesicles) protected, TME plane respected
6	Resection	Dissection of colonic/rectal or ileal mesentery	Transsection of the mesentery towards colonic/ileal transsection zone	Mesentery safely transected, vessels secured, bowel cleared of mesentery, ready to transect
		Transsection of bowel	GI stapler appears on screen	Ileum, colon and/or rectum is safely transected
7	Anastomosis	Perforation of intestinal stump	Bulging out of stump by approaching circular stapler from inside	Tip of circular stapler safely and fully perforated through intestinal stump
		Approximation of bowel ends	Grasping of anvil (for stapled anastomosis)/ needle holder appears on screen (for hand sewn anastomosis)	Bowel ends completely approximated and secured, tested for leakage

Table 23. Generic task analysis for laparoscopic colectomies

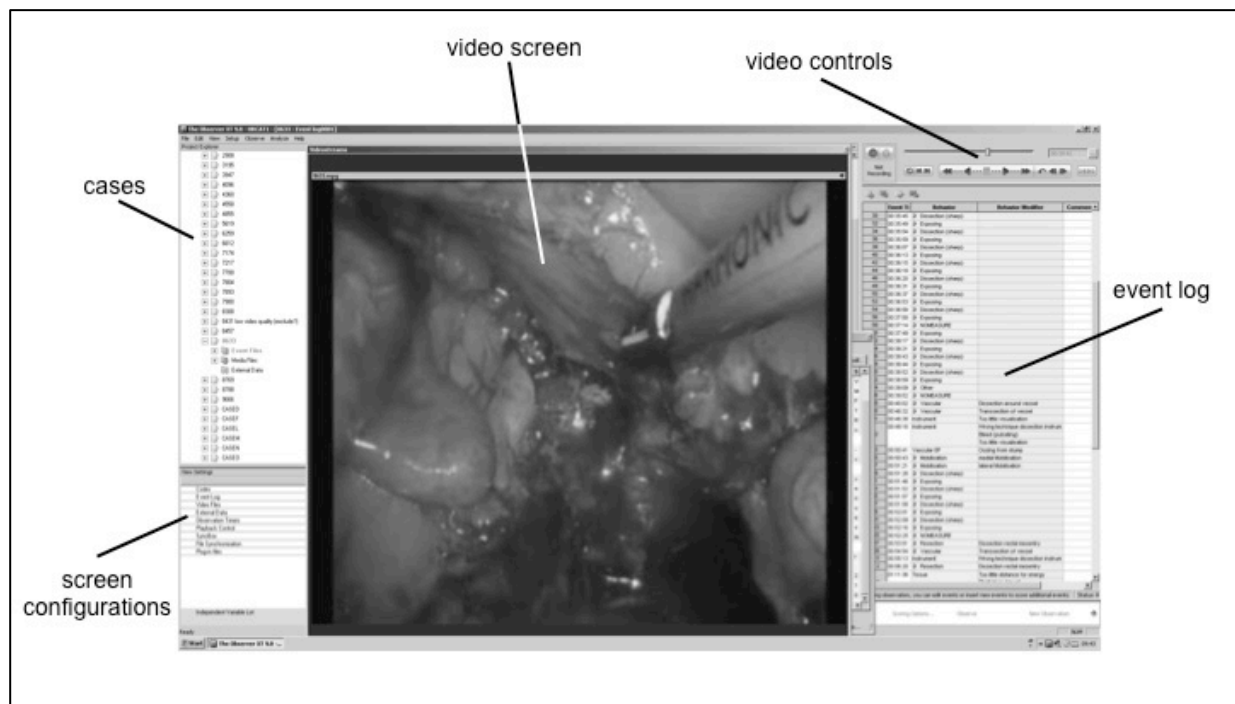


Figure 20. Screenshot of the video rating software used for analysis (Observer XT, Noldus software)

5.2.3.

5.2.4. Dissection-Exposure ratio

In addition to error count, an additional assessment modality was introduced to assess the level of efficiency. For a dissection task (lateral mobilization of the colon) that was performed consistently throughout all cases, the time spent on dissecting (“D”) and exposing (“E”) tissue was measured. Dissection was defined as actions that led to the separation of tissue, such as cutting or blunt dissection, whereas exposure were actions with the purpose of tissue plane definition, such as changing the direction of traction or grabbing a different piece of tissue. All other actions (e.g. cleaning the camera), were recorded as “other”. We hypothesized that the ratio between dissection and exposure (“DE ratio”) increases with experience, as the surgeons become more efficient.

5.2.5. Expert review

The minimum requirement for “passing” the expert review process was to perform the procedure in a safe and correct way. The surgeons submitted case videos to the NTP educational centre at Imperial College. Patient and surgeon identities were removed and the videos were converted to a standard video format [(“Moving Pictures Experts Group” (MPEG))] that could be replayed easily on any personal computer system. Two copies of each case video were created. Two independent judges, randomly chosen from the NTP trainer pool for each case, reviewed the videos fully blinded and independently. The judges provided an overall “pass/fail” global score on the quality of performance (“fail - needs more supervised training”/“pass - safe to continue independent training”). They also rated case complexity on a global Likert scale from 1 to 6 (1-very easy, 6-very difficult). The outcomes of this expert review process served as the gold standard for validating the concurrent validity of OCHRA. The difference in performance between passed and failed specialist was used for construct validity

5.2.6. Data extraction and statistical analysis

Error count was extracted from Observer XT and processed by the Statistical Package for the Social Sciences software [Version 18.0.2, SPSS Chicago (IL), USA]. Data normality was tested using detrended QQ plots. Task-specific and type-specific error counts were compared using non-parametric tests (Mann Whitney U test). Inter-rater and test-retest reliability was calculated using the intra-class correlation coefficient (ICC), Cohen’s kappa and Cronbach’s α , where applicable. For the prediction model a logistic regression model was built. Due to the moderate sample size, no more than four predictors were included in the model at any one time (Peduzzi 1995). Prediction accuracy was estimated using receiver operating characteristic (ROC) curves and classification tables were used to select cut-offs.

5.3. Results

Twenty-one consultant surgeons submitted a total of 33 full-length operating videos (17 right and 16 left colectomies). Six surgeons were experts in laparoscopic colorectal surgery and the remaining 15 were delegates. For 11 out of 15 delegates two different videos were available, and the remaining 10 surgeons had one case video each. The assessors scored 18 cases as satisfactory. In a further 14 cases, further supervised training was recommended. In one case they disagreed (inter-rater agreement: Cohen's $\kappa=0.939$). All expert cases passed the peer assessment. The case without agreement was excluded, leaving 32 cases for analysis. Based on these outcomes three groups were created for further analysis: *expert cases* (n=10), delegates who passed (*pass cases*, n=8) and those who failed (*fail cases*, n=14).

5.3.1. Case complexity and operating time

There was no significant difference in case complexity between the three groups (average complexity level for *experts*=2.0, *pass*=2.7, *fail*=2.3; $p=0.397$). Mean operating times (laparoscopic phase only) were 69 minutes (range 32-102 minutes) for right and 85 minutes (27-133 minutes) for left colectomies ($p=0.097$); there was no significant difference either between experts and delegates (66 vs. 83 minutes, $p=0.122$) or between the pass and the fail group (79 vs. 85 minutes, $p=0.539$). Delegates tended to spend proportionally more time on operating field exposure and dissection of the vascular pedicle than the experts, but none of these differences was significant.

5.3.2. Construct validity

A total of 399 events (errors and near misses) were identified, comprising 256 tissue-handling and 143 instrument-handling errors: 165/399 had a direct consequence, most of which were bleeding related (**Table 24**). There was no significant difference for total error counts (median

9 vs. 10, $p=0.665$) or for error sub-types when right and left colectomy was compared. There was a high level of consistency for the number of errors between two cases when performed by the same surgeon (test-retest reliability, Cronbach's $\alpha=0.838$). The difference of total error counts was significant when comparing the expert with pass and fail groups (median counts for experts=4, pass=10, fail=17; $p<0.001$). A significant difference was reached between the pass and fail group (10 vs. 17; $p=0.028$) (**Figure 21**). Detailed analysis of individual error types between the pass and fail groups demonstrated a significant difference for tissue-handling errors (7 vs. 12; $p=0.005$), a trend for consequential errors (4 vs. 7; $p=0.059$), but not for instrument-handling errors (4 vs. 5; $p=0.320$) (**Table 25**).

	Instrument	Tissue
Total errors (n=399)	143	256
-Examples of common errors (n)	- Poor visualisation (61) - Overshooting (16) - Little distance to structure (15) - Too little use of energy (12) - Additional attempt (11)	- Too much blunt force (70) - Wrong plane (50) - Too much tension (48) - Wrong direction of dissection (38) - Wrong direction of traction (24)
Errors with consequence (n=165)	28	137
- Bleed (ooze)	20	120
- Bleed (pulsating)	-	3
- Blunt bowel injury	5	8
- Perforating bowel injury	-	3
- Contamination	-	1
- Injury to mesentery	-	2

Table 24. Overview of identified errors

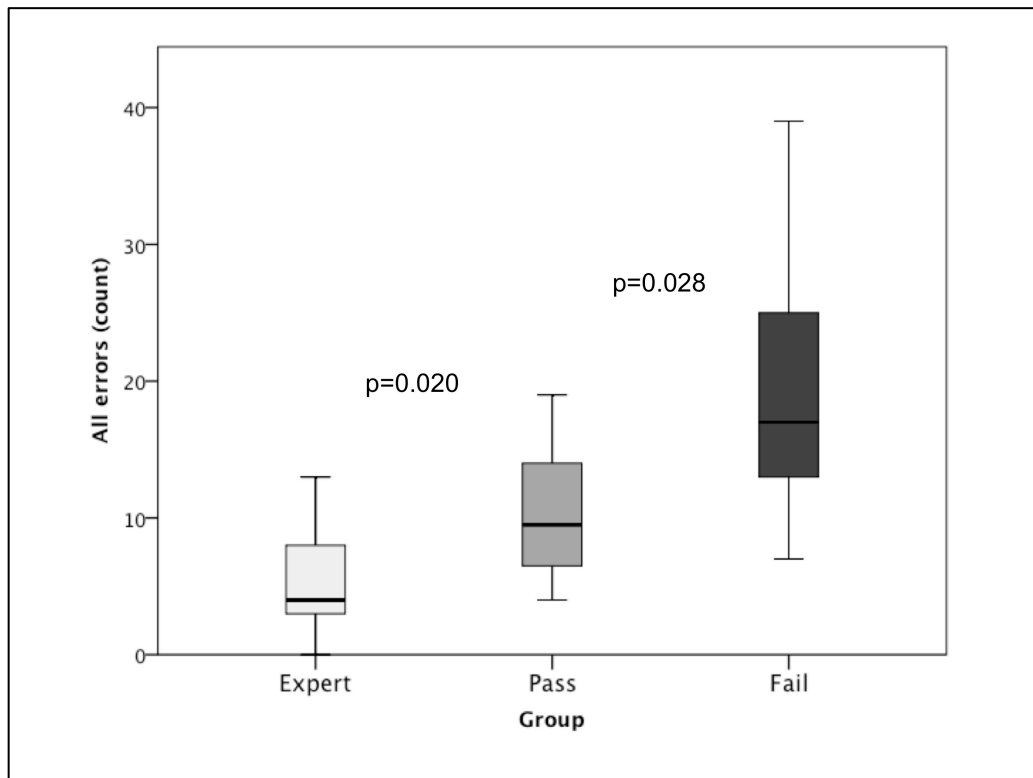


Figure 21. Construct validity for different groups for all errors

	Expert	p-value¹	Pass	Fail	p-value²
Cases n=32	n=10		n=8	n=14	
Total errors	4 (3-8)	>0.001	10 (6-14)	17 (12-25)	0.028
- Instrument	0 (0-2)	0.001	4 (1-7)	5 (3-10)	0.320
- Tissue	4 (2-6)	0.003	7 (3-9)	12 (9-15)	0.005
- Consequence	4 (2-5)	0.062	4 (2-5)	7 (4-10)	0.059
Task area⁴					
- Exposure	0 (0-1)	0.023	2 (0-4)	2 (0-6)	0.651
- Vascular p.	1 (0-2)	0.001	4 (2-5)	5 (2-8)	0.354
- Mobilization	2 (0-2)	0.034	1 (1-4)	5 (4-6)	0.029
- Flexure	2 (1-3)	0.374 ³	0 (0-1)	2 (0-4)	0.190
- Resection	0 (0-1)	0.017	2 (0-12)	6 (2-10)	0.621
Efficiency					
- DE ratio	1.0 (0.3)	0.001	0.6 (0.2)	0.6 (0.2)	0.717

Table 25. Summary of error counts for different error types and task areas and D/E ratios for different groups.

5.3.3. D/E ratio

In order to evaluate efficiency, the ratio between dissection time (D) and exposure time (E) in task 3, lateral mobilization of colon, was calculated. Experts used, on average, half the length of time for actual tissue dissection, resulting in a DE-ratio of 1, representing a highly efficient dissection technique. Delegates spent approximately 25% of the time on tissue dissection. This was significantly lower than experts (0.58 vs. 1.00; $p=0.001$). No difference was found between the pass and the fail group (0.59 vs. 0.56, $p=0.886$). (**Table 25, Figure 22**)

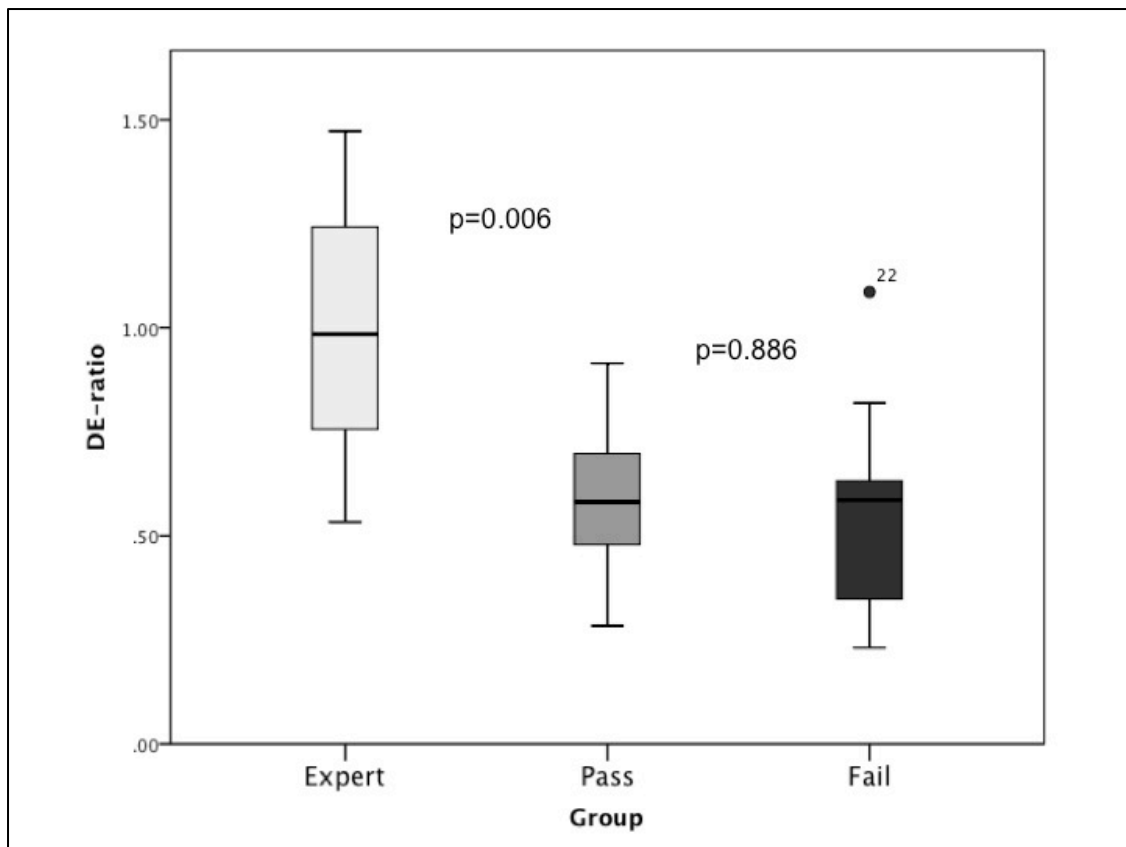


Figure 22. Construct validity using D/E ratio.

5.3.4. Concurrent validity

Based on the evidence above, instrument, tissue and consequential errors, as well as the DE ratio were included as independent variables in the logistic regression model. Dependent

variables were *competence* (all pass cases vs. all fail cases) and *expertise* (expert vs. delegate). When all four independent variables were used for *competence*, the area under the ROC curve (AUC) was 0.867, indicating a high level of predictive accuracy. Since tissue errors were the only significant predictor ($p=0.003$), the model accuracy did not change much when using it as a single independent variable (AUC=0.856) (**Figure 23**). This corresponds to an overall accuracy of 87.5% (sensitivity 100%, specificity 71.4%, **Table 26**). For *expertise*, tissue errors ($p=0.030$) and the DE ratio ($p=0.073$) were used as significant predictors and achieved an overall predictive accuracy of 84.4% (sensitivity 70%, specificity=90.9%), with an AUC of 0.900.

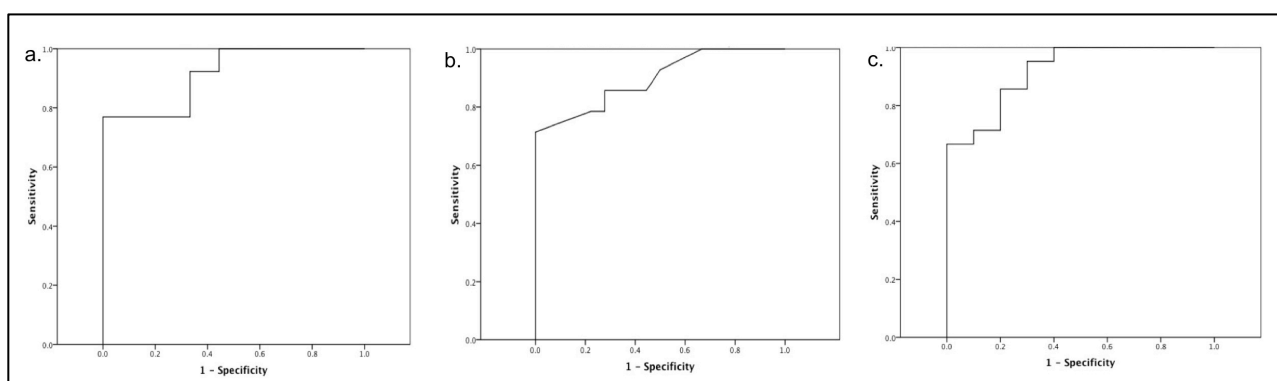


Figure 23. AUC curves for (a) overall errors, (b) tissue errors only and (c) tissue errors and D/E ratio.

		Predicted by OCHRA		
Observed		Expert	Delegate	
	Expert	7	3	Sensitivity 70%
	Delegate	2	20	Specificity 90.9%
		PPV 77.8%	NPV 87%	Accuracy 84.4%
		Predicted by OCHRA		
Observed		Pass	Fail	
	Pass	18	0	Sensitivity 100%
	Fail	4	10	Specificity 71.4%
		PPV 81.8%	NPV 100%	Accuracy 87.5%

Table 26. Prediction accuracy for tissue errors and D/E ratio.

5.4. Discussion

5.4.1. Summary

In this chapter a competency assessment method in advanced laparoscopic surgery at the specialist (consultant) level has been validated. The method provided significant discriminative power (construct validity) between competent and non-competent performance when using tissue-handling error counts. Furthermore, after achieving safe performance, as indicated by a low error count, experts differed from competent surgeons by a higher D/E-ratio, representing increased operative efficiency. It is worth noting that although operating time has often been used as a measure of increasing skill, it is clearly unsuitable for competency assessment (Scheer 2009). Although the sample size for this methodological study is moderate, the predictive potential (concurrent validity) of the method has been demonstrated using a logistic regression model.

5.4.2. Interpretation of results

Construct validity was demonstrated for tissue-handling errors but not for misuse of instruments. It is possible that this observation is related to the seniority of the subjects. Data from the present study suggest that only 19% of instrument usage errors had a consequence. Overall, only 36% of all errors were instrument usage related. In a study assessing 60 junior surgical trainees performing simulated laparoscopic cholecystectomies, 34% of instrument-related errors were consequential and 43% of all errors described wrong use of instruments (Tang 2005). By contrast, using OCHRA for laparoscopic pyloromyotomy performed by a group of Dutch specialist paediatric surgeons, none of the consequential errors was related to misuse of instruments (Tang 2004a). Although these studies describe different procedures of varying complexity, they demonstrate that with increasing surgical experience the number of

observed consequential errors due to instrument misuse decreases. The surgeons in this study were all experienced in basic laparoscopic procedures (cholecystectomy and appendicectomy) and they had already mastered the correct use of laparoscopic instruments.

The OCHRA approach used in this study to assess surgical competence is different from previous methods. It is more intuitive since experts, who were asked to assess a case, may subconsciously or consciously record errors and mishaps, and make their final judgment upon a summary of these events. Although previous assessment tools attempted to mirror the judges' thought processes by using numeric or descriptive scales, there is no evidence that these scales are sufficiently discriminative for advanced specialists (van Hove ; Ahmed 2011c). There is undoubtedly a ceiling (or floor) effect and the risk for observer bias, but it also remains unclear whether specialist operative competence is reliably measurable using a global assessment tool or a checklist. In addition, construct validity was usually established by comparing junior and senior surgeons; however, such a simple validation process is unlikely to be sufficient to justify high-stake judgments in certification procedures at the specialist level.

On the basic skills level there is increasing evidence for standardized tools to enable reliable assessment of technical skills (Fraser 2003; Vassiliou 2006). These methods measure generic laparoscopic skills, and therefore represent increasing levels of dexterity rather than competence of procedural performance. Since it has been shown in several studies that the majority of medical students can achieve competency in a simulated environment with little training without ever having operated on a single real patient, these methods do not qualify for representing the level of competency of very senior trainees and specialists (Gallagher 2004; Aggarwal 2006). Therefore it is essential to assess competence of surgeons at this level by evaluating real operations in the real environment. The identification of errors as proposed in the present study is a direct reflection of operative quality; and using a human reliability-type approach is a detailed description of an expert assessor's judgment process. This method is objective and directly linked to the observed procedure and not constrained by a scale. In

addition, since an adequately trained assessor can apply this method, time- and resource-consuming expert surgeon ratings may be avoided. Since technical errors are associated directly with potential surgeon-related morbidity, this method may be useful as part of certification and recertification processes at specialist level (Ahmed 2011b).

5.4.3. Limitations

Aside from the small sample size, the main limitation of this study is the lack of a gold standard method to describe competence. Expert performance and assessments were used as a benchmark, although these parameters may not be a precise representation of competent performance. It is currently impossible to overcome this limitation, yet OCHRA and related methods may create a detailed and robust description of the procedure and thus provide the backbone for a gold standard method in the future.

5.4.4. Conclusion

OUCHRA is a valid method for assessing surgical performance at specialist level. The method now needs to be integrated into a competency assessment system within a clinical training program. In addition, it can be potentially used for recertification and revalidation.

5.5. Disclosures

The contents of this chapter were published in the journal Surgical Endoscopy. The full reference of the paper, which is currently only available online (Epub ahead of print) is:

Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. Surg Endosc. 2011 Nov 1. [Epub ahead of print] PubMed PMID: 22042584.

6. DEVELOPMENT OF A COMPETENCY ASSESSMENT TOOL (CAT) FOR THE TRANSITION FROM SUPERVISED TO UNSUPERVISED PERFORMANCE

6.1.	Introduction.....	149
6.1.1.	Background	149
6.1.2.	Process.....	149
6.1.3.	Competency assessment in laparoscopic colorectal surgery	149
6.1.4.	Aim.....	150
6.2.	Methods.....	151
6.2.1.	Development of competency assessment tool (content validity)	151
6.2.1.1.	Interview studies	151
6.2.1.2.	Delphi method.....	151
6.2.2.	Validity and reliability studies	152
6.2.2.1.	Subjects and materials	152
6.2.2.2.	Assessors.....	152
6.2.2.3.	Concurrent validity	152
6.2.2.4.	Construct validity.....	153
6.2.2.5.	Reliability.....	153
6.2.3.	Implementation.....	155
6.2.4.	Software and analysis	156
6.3.	Results.....	157
6.3.1.	Content validity	157
6.3.1.1.	Outcomes from the interview study.....	157
6.3.1.2.	Outcomes of the Delphi method	160
6.3.1.3.	Scale development	160
6.3.2.	Validity and reliability studies	163
6.3.2.1.	Concurrent validity	163
6.3.2.2.	Construct validity.....	163
6.3.2.3.	Reliability.....	166
6.3.3.	Cut-off scores and implementation	167
6.4.	Discussion	168
6.4.1.	Summary	168
6.4.2.	Appraisal	168
6.4.3.	Limitations.....	169
6.4.4.	Clinical implications and conclusions.....	170
6.5.	Disclosures	171
6.6.	Acknowledgments.....	172

6.1. Introduction

6.1.1. Background

The method described in the previous chapter demonstrated features of a valid and reliable tool to describe competence, based on video recordings of a laparoscopic colectomy. However, there are certain drawbacks to the method. It is time consuming and requires the presence of a trained video analyst. An alternative method based on expert opinion may be more practical in the setting of a National programme with a high number of candidates.

6.1.2. Process

After a formal application procedure and attending a hands-on course in laparoscopic colorectal surgery, delegates are allocated to one of eleven training centres. They perform a number of supervised procedures at the training centre (in-reach training) and/or at their own hospitals (outreach training) in the presence of expert laparoscopic colorectal surgeons (trainers), as a continuous formative assessment process until both trainer and delegate are confident to continue in an unsupervised setting. The processes of the National training programme have been described in detail in **Chapter 1**.

6.1.3. Competency assessment in laparoscopic colorectal surgery

When the NTP was designed in 2008, the faculty, represented by all training centres, the national coordination office and the national educational centre, chose to implement an objective summative assessment process at the end of the supervised training period in order to ensure that the delegates adopt a safe surgical technique before embarking on independent practice. In 2008, no reliable validated tools were available to assess competency in laparoscopic colorectal surgery at a specialist level. Available generic tools for skills

assessment in laparoscopic surgery have been typically aimed at junior trainees and are unsuitable for high-stake competency assessment in advanced procedures (Martin 1997; Vassiliou 2005). There was a need to develop a transparent, objective and valid process for the reliable assessment of competency in laparoscopic colorectal surgery at specialist level.

6.1.4. Aim

The aim of this study was to develop, validate and implement a competency assessment tool (CAT) for technical surgical performance in laparoscopic colorectal surgery in the context of a summative assessment process for the NTP.

6.2. Methods

6.2.1. Development of competency assessment tool (content validity)

6.2.1.1. Interview studies

Detailed, semi-structured interviews were conducted with seven surgical trainers with extensive teaching experience in laparoscopic colorectal surgery (Britten 1995). The interviews focussed on skill domains that are important for a fair competency assessment and on the structure of a useful assessment sheet. All interviews were audio-recorded, transcribed and two investigators (DM and SMW) extracted data. Data were categorised and data items listed in random order.

6.2.1.2. Delphi method

A reiterative expert consensus process was applied (Delphi method) (Dalkey). Extracted interview data was sent to all members of the educational committee of the NTP, comprising 15 expert laparoscopic colorectal surgeons and one educationalist. They were asked to rate each item with respect to its relevance for summative assessment of competency. The rating scale ranged from (1) “should not be on a marking sheet” to (5) “should definitely be part of a marking sheet” with (3) being “undecided”. This process was repeated in a second Delphi round with the average results and the standard deviations of the previous round being disclosed to participants (Dalkey ; Romm 1979; Fink 1984; Cuschieri 2001). All items with a mean score of >3 or highest rank were sought to be integrated in the competency assessment tool (CAT). The marking sheet, detailed instructions and a catalogue of definitions of terms,

was given to the expert group for final evaluation and suggestions requested for amendments (third Delphi round).

6.2.2. Validity and reliability studies

6.2.2.1. Subjects and materials

Six NTP trainers (*expert group*) with an average experience of 730 cases (range 300-1500) were asked to submit DVD-recorded cases to serve as a control group; all other submitting surgeons were specialist colorectal surgeons (consultants) at various stages in their NTP training (*delegate group*).

6.2.2.2. Assessors

Assessors were randomly chosen from a pool of NTP trainers. The assessors were blinded for the performing surgeons and their trainers, the hospital and patient identifiers. Assessors received basic information on outcome parameters and written information on how to use the CAT including a detailed catalogue of definition of terms for each item on the assessment sheet (see **Appendix i**).

6.2.2.3. Concurrent validity

In addition to the CAT scoring, the assessors were asked to give an overall outcome statement (“Needs further training” = **Fail** or “Safe performance demonstrated” = **Pass**). If they felt that the provided video material was not sufficient to make a final statement they could choose the option “Not enough evidence for judgment”. An additional section for free text was also provided and assessors were encouraged to add additional comments if required.

The concurrent validity of CAT was examined against a detailed error analysis using observational clinical human reliability analysis (OCHRA) as described in the previous chapter. Each case was analysed using OCHRA by a trained researcher in human reliability analysis methodology (DM) who was blinded to the surgeon, hospital, patient and outcomes of the procedure. Video-recordings were rated by counting predefined errors using scientific rating software (Observer XT, Version 8, Noldus Information Technology Inc., Leesburg (VA), USA) (Joice 1998; Tang 2004b; Ahmed 2011a). Total numbers of errors committed per case are reported.

6.2.2.4. Construct validity

For construct validity CAT scores of delegates and experts were compared. In addition, delegates who passed the assessment according to the overall assessment were compared with those who failed (“pass” vs. “fail”).

6.2.2.5. Reliability

The CAT scoring sheet consists of 16 items and two randomly chosen assessors rated surgeons performing two cases (right hemicolectomy and left hemicolectomy). The generalizability theory analysis was applied for the overall reliability of the method. The analysis was performed with a fully crossed design, using the facets surgeon (S), case (C), items (I) and assessors (A) (Streiner 2008):

$$(S \times C \times I \times A)$$

Based on this design, the generalizability coefficient was computed, using the following equation:

$$\textit{Generalizability coefficient} = \sigma^2_S / (\sigma^2_S + \sigma^2_C + \sigma^2_I + \sigma^2_A)$$

Whereas σ^2_f stands for variance for facet f.

For criterion-referenced reliability a cut-off score was introduced into the generalizability equation. The cut-off scores were determined using specificity and sensitivity tables from receiver operating characteristics (ROC) curves. The equation to calculate the Generalizability coefficient was :

$$(\sigma_s^2 + (\mu - x)^2) / (\sigma_s^2 + (\mu - x)^2 + \sigma^2)$$

whereas μ : grand mean, x : cut-off score, σ_a^2 : sum of variances apart from σ_s^2 .

6.2.3. Implementation

Based on clinical experience there was a consensus by the educational programme committee that every delegate should submit two cases at the end of their training for the assessment by two independent assessors. These cases were performed independently, without trainer input. For practical reasons, only the intra-abdominal laparoscopic phase was used for assessment. The video-recordings were sent to the National Educational Centre at Imperial College, de-identified and forwarded to two randomly chosen NTP trainers at two different sites from a pool of twenty-seven NTP trainers (**Figure 24**).

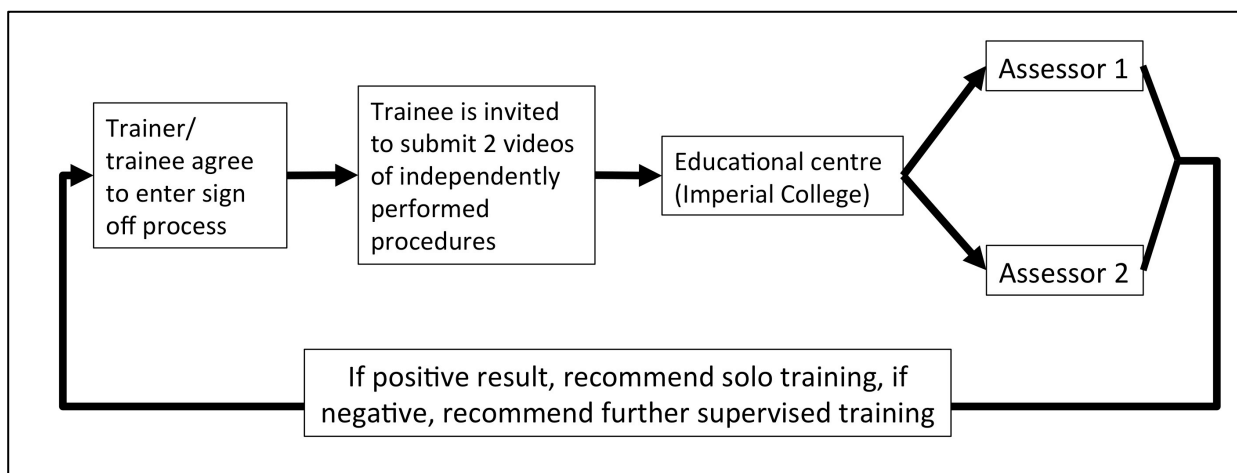


Figure 24. Process of competency assessment at the end of supervised training as suggested by the educational committee

6.2.4. Software and analysis

Generalizability theory analysis was performed using the G String III software, based on urGENOVA (version 5.5.2) (Bloch 2010). All other analyses were performed using the Statistical Package for the Social Sciences software [Version 18.0.2, SPSS Chicago (IL), USA]. Generalizability analysis was performed as described in **chapter 6.2.2.5**. For construct and concurrent validity the Wilcoxon signed-rank test and Spearman's rho were used. Test accuracy was analysed using ROC curves and is reported as sensitivity, specificity and negative (NPV) and positive predictive value (PPV).

6.3. Results

6.3.1. Content validity

6.3.1.1. Outcomes from the interview study

After reviewing the transcript of the interviews two main categories constantly reoccurring in the interviews could be identified:

- I. *Competency skills: what are the required skills on behalf of the surgeon to be considered as competent?*

Data extracted from the interview transcripts showed that the items “Tissue handling”, “Economy and flow of movements”, “Correct sequence of steps” and “Operating time” were mentioned by three or more experts (see **Table 27**). Careful tissue handling was described as “(being able of) *handling delicate structures...carefully with minimal trauma*” (Expert A) or “*...moving bowel around too aggressively*” (Expert B). One expert said that “*...prolonged operations*” (Expert C) should be avoided suggesting operating time as an assessment criterion. However, others addressed the difficulty of using time for assessment as it may be affected significantly by case complexity (Expert B). Examples for “Economy and flow of movements” are “*...not being too hesitant*” (Expert A), “*...avoiding unnecessary movements*” (Expert D) and “*...speed, economy and ease of movements*” (Expert E).

Three experts mentioned the importance of a sequential approach. However, Expert B noted that a step out of sequence not always equates with bad performance:

“This is very difficult to say, you know, if you’d look at fifty (...) surgeons, they wouldn’t all agree (...) it depends where you were taught. I think it is very difficult to criticize how someone does it logically, because someone could turn around and say: ‘Well actually, it may be done out of sequence but that was a piece of bowel that looked like it was suitable to be dealt with at that time!’” (Expert B)

II. Assessment structure: what is the ideal structure of the assessment form and what kind of scoring system should be used?

The interviews revealed proposals for two different assessment form structures: five experts proposed either a task-specific structure by breaking down the procedures into generic task areas. Two experts suggested a global assessment tool assessing generic skills of the surgeon.

Four different marking systems were proposed:

- *Likert scale*: e.g. “1-very poor” to “7-excellent”
- *Juster scale*: e.g. “1- Consistently operating in the wrong tissue plane, major bleeding” or “3- Good planes most of the time, control of bleeding”
- *Binary generic scale*: e.g. “Competent” and “Not competent”
- *Level of supervision*: e.g. “performed with verbal assistance” or “performed without assistance”

Rank	Item	n	Round 1	SD 1	Round 2	SD 2	Δ SD
1	Competent use of non-dominant hand (traction, exposure, positioning)	1	4.7	0.6	5.0	0.0	-0.6
2	Competent use of dominant hand (cutting, coagulation, blunt dissection)	1	4.6	0.7	4.7	0.5	-0.2
3	Correct demonstration of anatomy (Ureter, duodenum, vessels etc)	1	4.6	0.5	4.7	0.5	0.0
4	Degree and direction of traction	1	4.4	0.8	4.6	0.5	-0.3
5	Safe use of stapling/ clipping instruments	1	4.5	0.5	4.6	0.5	0.0
6	Stay in correct tissue plane	2	4.5	0.5	4.6	0.5	0.0
7	Control of bleeding, uncontrolled dissection of blood vessels	2	4.6	0.5	4.6	0.5	0.0
8	Safe use of grasping instruments of bowel and other anatomical structures	1	4.4	0.7	4.4	0.5	-0.2
9	Safe use of dissecting instruments with energy source (ultrasound, diathermy)	1	4.5	0.5	4.4	0.5	0.0
10	Basic Laparoscopic Skills (Depth Perception, Respect For Tissue, Time And Motion)	1	4.2	1.3	4.1	0.8	-0.5
11	Overall impression	1	4.7	1.3	4.1	1.5	0.2
12	Quality of Exposure	1	4.2	0.6	4.0	0.6	0.0
13	Unnecessary (near) avulsion of tissue, rough tissue handling	4	3.6	0.5	3.9	0.8	0.3
14	Collateral damage or potential collateral damage	2	3.8	1.4	3.7	1.0	-0.4
15	Manipulating instruments in the field of view	2	3.2	1.2	3.4	0.8	-0.4
16	Speed and economy of movements, number of unnecessary movements	3	3.7	0.6	3.3	1.0	0.4
17	Safe port insertion	1	3.9	1.1	3.3	0.8	-0.3
18	Use of the appropriate instruments	1	3.3	0.9	3.3	0.8	-0.1
19	Order of procedural steps, stepwise approach, finishing one step before starting the next	3	3.2	0.7	3.1	0.6	-0.1
20	Manoeuvring and positioning of camera in the abdomen	2	3.3	0.8	2.9	1.0	0.2
21	Decision for conversion to open	1	3.1	1.4	2.9	0.8	-0.6
22	Laparoscopic camera view (picture quality, camera angle, display window)	1	3.1	1.0	2.7	1.0	0.0
23	Route of access and port positioning	1	2.9	1.0	2.6	0.5	-0.5
24	Time for task step	1	2.8	1.1	2.6	0.8	-0.3
25	Operating time (case-specific), prolonged operation	3	2.7	1.0	2.3	0.8	-0.2
26	Operating strategy (laparoscopically vs. open, medial to lateral, etc.)	1	2.4	1.1	2.2	1.3	0.2

Table 27. Items mentioned for competency skills. Ranks according to final Delphi round (round 2). Number of experts giving the answer (n), average scores after each round and standard deviations (SD) are tabulated. Δ SD demonstrates increasing consistence after round 2.

6.3.1.2. Outcomes of the Delphi method

The surgical experts participating in the Delphi method had a median experience in laparoscopic colorectal surgery of 8.5 years (range 4-20) and have performed an average of 595 cases (range 300-1500). After the second Delphi round standard deviations (SD) decreased suggesting consolidation of agreement within the group (average $SD_{\text{round 1}}=0.86$, $SD_{\text{round 2}}=0.76$). Interestingly, after applying the Delphi method, the ranking is different to what the number of responses from the interview study suggests: although certain items were mentioned by several experts, they ranked relatively low after Delphi. Generic laparoscopic skills and outcome-related items (e.g. correct tissue plane) ranked significantly higher than operating time-related or sequence-related items (see **Table 27**). There was 100% agreement that the marking sheet should be structured in a task-specific way rather than a pure generic skills checklist. The group clearly preferred a descriptive scale (Juster), rather than Likert or other types of scale [mean rank=1.1, SD=0.48].

6.3.1.3. Scale development

Taking these results into account, a marking sheet was designed. In order to allow for a concurrent task-specific assessment and evaluation of generic skills a hybrid structure was created. The columns of the marking sheet represent four task components (i) access and exposure of the operating field, (ii) identification and dissection of the vascular pedicle, (iii) mobilisation of the colon and rectum and (iv) resection/ anastomosis of the bowel. There are four domains for each column to rate, representing generic skills, the process of performance and outcome-related factors: (i) use of instruments, (ii) tissue handling, (iii) errors, and (iv) the quality of the end product (**Table 28, Figure 25**).

		TASK STEPS			
		Exposure	Vascular Pedicle	Mobilisation	Resection/ Anastomosis
GENERIC SKILLS	Instruments	Safe port insertion	Safe use of vascular stapler device/ clip applicator/ energy-based dissection tool	Safe use of graspers/ Dissection tools (non-energy and energy-based)	Safe use of intestinal stapler device
	Skills	Safe and effective moving of bowel	Safe dissection technique	Identifying correct tissue planes	Safe preparation of bowel for resection
	Errors	No bleeding/ avulsion/ perforation	No bleeding/ avulsion/ perforation	No bleeding/ avulsion/ perforation	No bleeding/ avulsion/ perforation
	Outcome	Complete exposure of operating field	Safely dissected (sealed) level of dissection	Adequate mobilisation, anatomical landmarks demonstrated	Adequate resection, safe anastomosis

Table 28. Structure of the CAT form

The scoring system was descriptive, containing four scores, representing different skill levels according to the Dreyfus model (REF):

1. Novice
2. Advanced beginner
3. Competent
4. Proficient

Expert performance was not considered for this rating scale. Expert performance may deviate from basic rules since experts have developed their skills to a level that allows them to neglect certain guidelines. A typical example is the application of clips onto a vessel. The rule is that the tips of the clip should be visualised before application in order to prevent an injury to a structure that may lie behind the vessel. A very experienced surgeon however may apply a clip intuitively in the correct position, without visualising the tips of the clip in every situation. This example demonstrates that the rating of expert performance is extremely difficult and beyond the scope of this study.

TASK	EXPOSURE	PEDICLE CONTROL	MOBILISATION	RESECTION/ ANASTOMOSIS
INSTRUMENT USE	Ports are safely inserted in an ergonomically ideal position, operating field is exposed	Vascular pedicle is identified and vessels dissected out at an appropriate level	Mobilisation following anatomical planes, landmarks adequately defined and protected	Adequate and appropriate resection and safe performance of anastomosis
	Port insertion/ placement:	Use of haemostatic tool (clip applier/ diathermy/ stapler):	Use of graspers/ dissection tools:	Use of intestinal stapler (straight, circular):
	1. Dangerous technique (not visualised), hazardous or wrong position	1. Dangerous (uncontrolled moves/ views, wrong instrument)	1. Dangerous (traumatic grasping, hazardous use of dissection tool)	1. Dangerous (uncontrolled movements/ views, wrong cartridge)
	2. Incompetent (several attempts) or ergonomically poor position	2. Incompetent (awkward, repeated attempts, impaired views)	2. Incompetent (awkward and inefficient grasping, use of dissection tool)	2. Incompetent (awkward, repeated attempts, impaired views)
TISSUE HANDLING	Safe insertion and ergonomically good position	3. Safe (instrument accurately placed and engaged)	3. Safe (safe grasping, safe use of dissection tool)	3. Safe (instrument visualised, good position and safely engaged)
	4. Masterful insertion, ideal positioning	4. Efficient (minimal attempts, ideal position and safety)	4. Masterful (safe, ideal and efficient use of graspers, dissection tools)	4. Efficient (minimal attempts, ideal position and safety)
	5. Not applicable	5. Not applicable	5. Not applicable	5. Not applicable
	Manipulating small bowel:	Dissection around pedicle and division of blood vessels:	Anatomical dissection technique:	Preparation of bowel for resection, division of mesentery
TISSE HANDLING	1. Traumatic grasping, uncontrolled and forceful movements, insufficient view	1. Dangerous (views, uncontrolled movements, energy device)	1. Wrong tissue plane, unable to correct quickly	1. Dangerous (forceful grasping, dangerous use of energy device)
	2. Laborious and repeated attempts, ineffective patient positioning	2. Ineffective (several, hesitant cuts, jerky movements)	2. Repeatedly wrong tissue plane, able to correct	2. Ineffective dissection technique, safe tissue handling
	3. Safe and effective handling of bowel, effective patient positioning	3. Safe dissection technique, slightly laborious	3. Rarely wrong plane, able to correct quickly	3. Safe dissection technique, slightly laborious
	4. Safe and expeditious handling of bowel, effective patient positioning	4. Safe and expeditious dissection technique	4. Constantly stays in correct tissue plane	4. Safe and expeditious dissection technique
ERRORS	5. Not applicable	5. Not applicable	5. Not applicable	5. Not applicable
	This task was performed with:	This task was performed with:	This task was performed with:	This task was performed with:
	1. Bleeding, regular tissue avulsion, potential perforation of bowel	1. Uncontrolled bleeding, tissue avulsion, potential perforation	1. Uncontrolled bleeding, potential perforation, gross injuries to planes	1. Uncontrolled bleeding, tissue avulsion, potential perforation
	2. Minor bleeding/ occasional tissue avulsion, no perforation	2. Minor, controlled bleeding, occasional t. avulsion, no perforation	2. Minor, controlled bleeding, no perforation, repeated injuries to planes	2. Minor, controlled bleeding, occasional tissue avulsion, no perforation
END-PRODUCT	3. Minimal bleeding, no tissue avulsion, no perforation	3. Minimal bleeding, no avulsion, no perforation	3. Minimal bleeding, no perforation, planes mainly preserved	3. Minimal bleeding, no avulsion, no perforation
	4. No bleeding, no tissue avulsion, no perforation	4. No bleeding, no avulsion, no perforation	4. No bleeding, no perforation, planes completely preserved	4. No bleeding, no avulsion, no perforation
	5. Not applicable	5. Not applicable	5. Not applicable	5. Not applicable
	Was the operating field sufficiently exposed?	How was the management of the vascular pedicle?	Was the large bowel safely mobilised and were the landmarks identified?	How is the quality of the stump/ anastomosis?
QUALITY	1. Small bowel not removed, vascular pedicle not exposed	1. Vascular pedicle not secured or grossly at wrong level	1. Inadequate mobilisation, landmarks not identified (ureter, duodenum)	1. Likely to compromise integrity of anastomosis
	2. Small bowel loops partially obscure view for dissection, pedicle exposed	2. Secured, but divided to distal/ proximally	2. Inadequate mobilisation, landmarks identified, but incorrect correct plane	2. Too short, mesentery inadequately dissected, safe anastomosis
	3. Safe and adequate exposure, occasional fall back of small bowel	3. Vessels safely secured, correct level of division	3. Adequate mobilisation, landmarks correctly identified	3. Adequate resection level, safe anastomosis
	4. Expeditious and safe exposure, no fall back of small bowel	4. Vessels efficiently secured, ideal level of division	4. Ideal mobilisation, landmarks clearly demonstrated	4. Ideal stump management, ideal anastomosis
QUALITY	5. Not applicable	5. Not applicable	5. Not applicable	5. Not applicable

Figure 25. Finalised version of the CAT form.

6.3.2. Validity and reliability studies

6.3.2.1. Concurrent validity

There is a moderate inverse correlation between CAT scores and the number of errors committed as determined by the OCHRA method (Spearman's $\rho = -0.520$, $p < 0.001$ **Figure 28**).

6.3.2.2. Construct validity

The average CAT scores for experts were significantly better than delegates with mean (SD) of 3.19 (0.32) vs. 2.60 (0.50) respectively ($p = 0.004$). Moreover, the apprentices who passed the assessment had a significantly higher score than those who failed; mean scores (SD) were 2.95 (0.17) vs. 2.28 (0.49) respectively, $p < 0.001$. Although the tool can distinguish between passing and failing apprentices, the difference between apprentices who passed and the experts was not significant ($p = 0.075$) (**Figure 26**). Construct validity can be found for most scored items between passing and failing apprentices; however, the tool was insufficiently sensitive to differentiate between experts and apprentices who passed (**Table 29**).

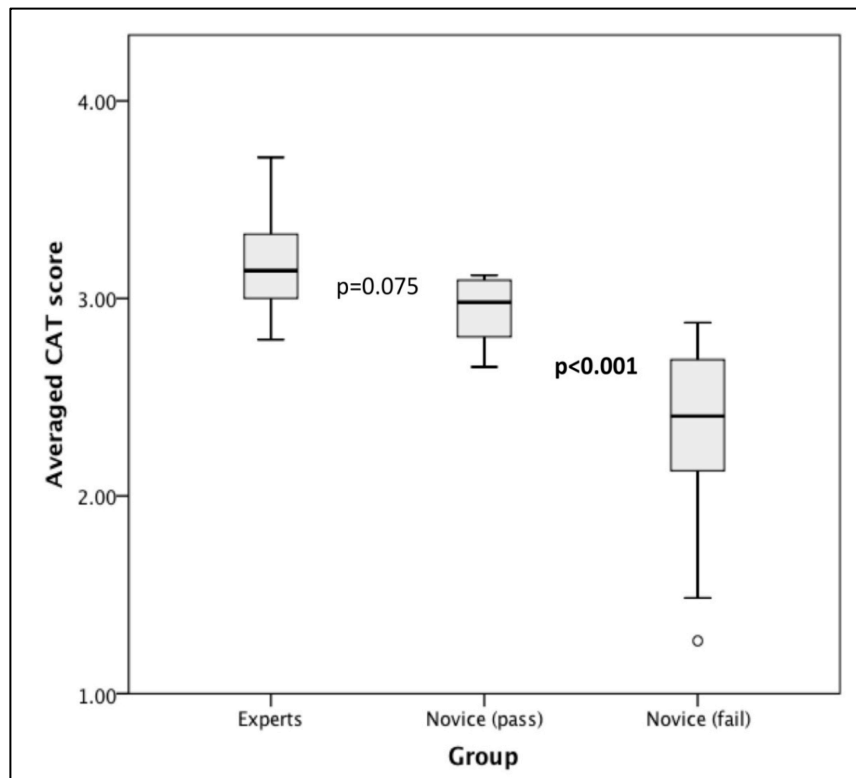


Figure 26. There was a significant difference of averaged CAT scores between the expert and novice group ($p=0.004$). The main effects could be found between apprentices who passed and those who failed, whereas the difference between experts and apprentices who passed was only moderate.

<i>Domains</i>	Exposure		Pedicle control		Mobilisation		Resection/ Anastomosis	
	<i>Mean (SD)</i>	<i>p</i>	<i>Mean (SD)</i>	<i>p</i>	<i>Mean (SD)</i>	<i>p</i>	<i>Mean (SD)</i>	<i>p</i>
Instrument use	Expert	3.30 (0.44)	Expert	3.50 (0.50)	Expert	3.27 (0.25)	Expert	3.50 (0.50)
	Pass	2.92 (0.28)	Pass	3.14 (0.22)	Pass	3.00 (0.29)	Pass	3.04 (0.14)
	Fail	2.64 (0.56)	Fail	2.95 (0.49)	Fail	2.81 (0.24)	Fail	3.04 (0.49)
Tissue handling	Expert	2.67 (0.87)	Expert	3.33 (0.31)	Expert	3.20 (0.27)	Expert	3.30 (0.27)
	Pass	2.88 (0.22)	Pass	2.71 (0.48)	Pass	2.94 (0.45)	Pass	2.96 (0.25)
	Fail	2.26 (0.72)	Fail	2.43 (0.54)	Fail	2.42 (0.49)	Fail	2.69 (0.68)
Errors	Expert	3.17 (0.55)	Expert	3.37 (0.42)	Expert	3.13 (0.36)	Expert	3.40 (0.55)
	Pass	3.18 (0.47)	Pass	2.97 (0.68)	Pass	2.92 (0.61)	Pass	3.14 (0.45)
	Fail	2.62 (0.82)	Fail	2.95 (0.76)	Fail	2.64 (0.69)	Fail	2.74 (0.60)
End-product quality	Expert	3.23 (0.60)	Expert	3.27 (0.43)	Expert	3.27 (0.25)	Expert	3.34 (0.42)
	Pass	2.97 (0.30)	Pass	2.85 (0.52)	Pass	2.92 (0.61)	Pass	3.00 (0.20)
	Fail	2.73 (0.25)	Fail	2.40 (0.45)	Fail	2.48 (0.22)	Fail	2.69 (0.41)

Table 29. Summary of mean scores (standard deviation) for each item of the scoring sheet.

6.3.2.3. Reliability

A total of 108 assessments in 54 cases (29 right and 25 left colectomies), performed by 31 different surgeons (6 experts, 25 apprentices) were used for analysis. Raw scores of the 16 item marking sheets were generalized over assessor (A), case (C) and item (I). The overall reliability for two assessors rating two cases per surgeon is represented by a generalizability coefficient of $G(ACI)=0.784$. Criterion-referenced generalizability, using a cut-off score of 2.55 (see below), increases reliability slightly, with a coefficient of $G(ACI)=0.803$. D-studies were performed to examine the effect of increasing numbers of assessors and cases (**Figure 27**).

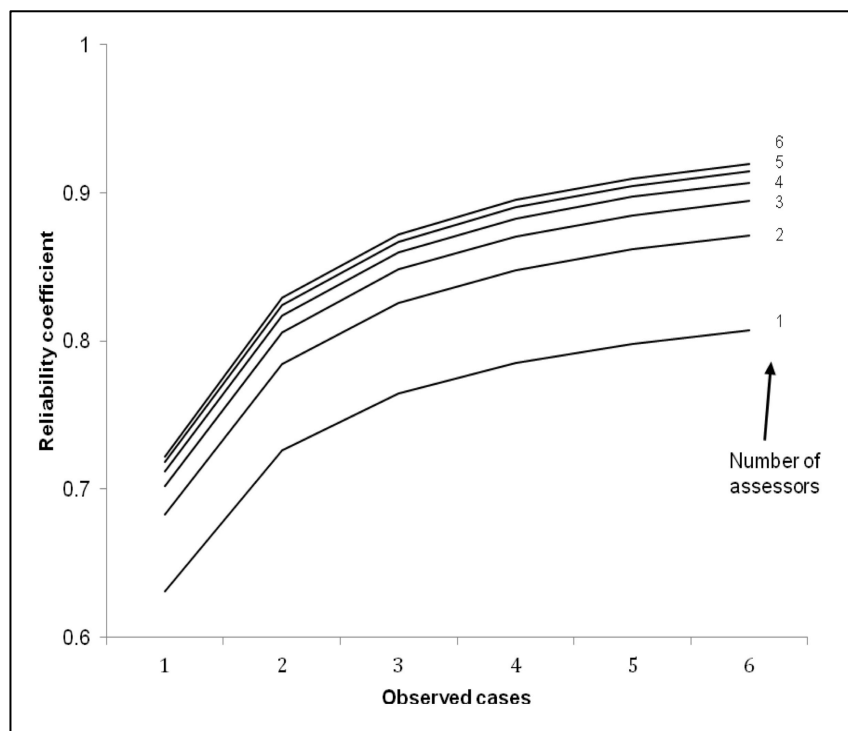


Figure 27. D study for increasing numbers of assessors and observed cases. The critical G coefficient of 0.8 is reached with three assessors observing two cases, or two assessors observing three cases.

6.3.3. Cut-off scores and implementation

Analysis of the ROC curve revealed that a cut-off score with maximum sensitivity and specificity is at an average score of 2.7 (out of 4) (sensitivity 94.4%, specificity 84.6%, PPV 93.8%, NPV 10.9%). A combination of both methods (CAT and OCHRA) can improve the sensitivity and specificity of the test. This can be achieved by choosing a lower cut-off at 2.55 for CAT to maximise sensitivity (100%). In contrast, for OCHRA a maximum of 16.75 errors on average are allowed in order to maximise specificity of the test. By combining the two methods with the above-mentioned benchmarks, an overall sensitivity reaches 100%, specificity 83.3%, a positive predictive value of 93.8% and a negative predictive value of 100% (Figure 28).

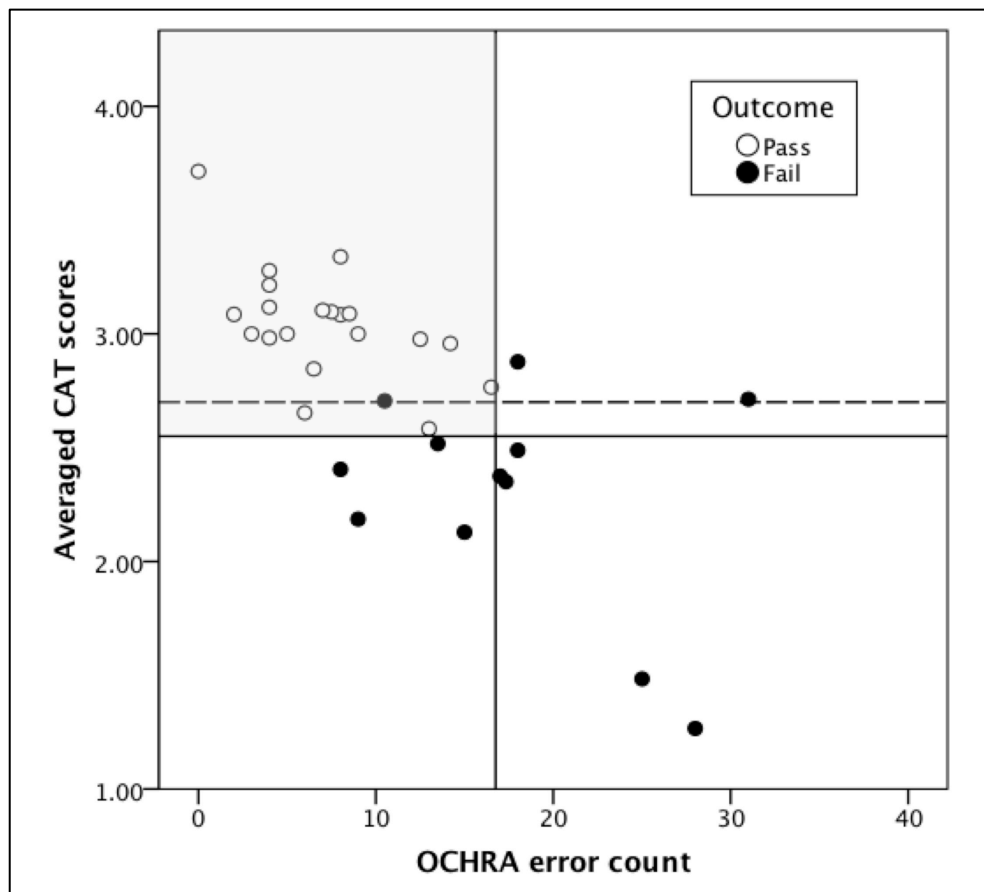


Figure 28. The scattered plot of individual cases suggests decreasing CAT scores with increasing number of observed errors using the OCHRA method (CAT scores are averaged for two assessors). The solid lines indicate the cut-offs for CAT at 2.5 and for OCHRA at 16.75. The dotted line represents the ideal cut-off for CAT. The grey rectangle represents the “passing” area.

6.4. Discussion

6.4.1. Summary

This is the first study on competency assessment of specialists that has been fully implemented in a national training programme and it has been widely accepted by assessors and apprentices. CAT allows the national training programme to identify consultant surgeons who can safely perform laparoscopic colorectal surgery before embarking on independent practice. The content validity of CAT was ensured by a structured process to elicit expert opinion on all relevant skills for competence assessment at specialist level. Construct validity was demonstrated within a group of consultant surgeons of different competency levels while the concurrent validity was shown against detailed error analysis using OCHRA. The difference between competent consultants and experts is less distinct because the target of the assessment is competency level which apprentices have also achieved. The correlation between CAT and OCHRA was only moderate because the two methods are based on different principles and hence, a perfect correlation cannot be expected. The combination of both methods increased the accuracy of assessment.

6.4.2. Appraisal

The authors carried out two recent systematic reviews. The first review on mentoring, simulation and assessment in laparoscopic colorectal surgery highlighted the need for further research into technical skills assessment.(Miskovic 2010b) The second on observational assessment tools for technical skills concluded that most studies did not fulfill the requirements for valid and reliable assessment tools and a step-wise approach for developing new assessment tools was recommended.(Ahmed 2011c) Another systematic review by van Hove et al evaluating the evidence of assessment tools with regards to their feasibility in clinical practice

concluded that only few can be used for examining or credentialing.(Van Hove 2008) When the NTP was introduced in 2008, no procedure specific assessment tool for laparoscopic colorectal surgery was available. Recently, two studies on the development of procedure specific assessment tools have reported using Objective Structured Assessment of Surgical Skills (OSATS).(Palter 2011; Sarker 2011) However, neither was specifically designed for summative competency assessment at specialist level. The CAT is a summative assessment method for technical competency in LCS at specialist level that has been developed and validated in a systematic and scientific process. Assessment tools in surgery have been previously validated by demonstrating construct validity between surgeon groups with different levels of experience, typically experts and trainees or junior and senior trainees. (Ahmed 2011c; Sarker) This intergroup difference is not practical because summative assessment methods aim to discriminate between different levels of performance within the same experience group. Also, as the CAT evaluates both the process of performance (instrument use, tissue handling and committed errors) and the quality of the end product of component tasks of laparoscopic colorectal procedures, it can provide detailed feedback on areas of concern that need to be addressed.

6.4.3. Limitations

The study has some limitations. Firstly, the CAT assesses only technical skills but clinical outcomes depend also on non-technical skills of the surgeon and the holistic care process within a surgical unit; both were beyond the scope of the study. Secondly, the apprentices submitted self-selected cases for the sign-off process and case complexity may be a cause for bias. This approach was adopted because the aim of the NTP was to train them to a stage where they are capable of performing a straight-forward case in a safe and competent way and hence appropriate case selection by the apprentice was considered to be a positive attribute of an insightful surgeon.

6.4.4. Clinical implications and conclusions

This study has several implications. The concept of a competency-based training curriculum with an exit competency-based assessment has been proven to work at a national level. Also, the assessment of technical competency of consultant surgeons before embarking on independent practice of new procedures has been shown to be practical and acceptable to both assessors and apprentices. The CAT is a workplace assessment method that also has the potential to be used for quality control within certification and revalidation programmes of specialist surgeons. Although morbidity and mortality data will remain the ultimate method to judge surgeons in clinical practice, there are circumstances that necessitate employing other approaches. For instance, the CAT could be used to evaluate performance without waiting for poor outcomes to accumulate while the audit is in progress. This study shows that the use of competency assessment tools to judge specialist technical performance is achievable on a national scale. Future research should focus on investigating the correlation of such tools with clinical outcomes. The modular structure of CAT allows modification of the tool for any other surgical or interventional procedure.

6.5. Disclosures

Full ethical approval was obtained for the use of de-identified patient video material and patients were consented accordingly by the operating surgeon (NREC 04/Q0403/54).

6.6. Acknowledgments

I would like to thank all trainers, training centre administrators and apprentices of the National Training Programme (Lapco) for their invaluable contributions to the results of this study. For the training centres: Tom Cecil (Basingstoke), John Griffith and Matt Clarke (Bradford), James Gunn and John Hartley (Hull), Savvas Papagrigoriadis and Vivek Datta (King's and St Thomas'), Hugh Gallagher and Mark Katory (Newcastle-upon-Tyne), Selva Sakar and David Watson (North West), Charles Maxwell-Armstrong and Austin Acheson (Nottingham), Ian Lindsey (Oxford), Jim Khan (Portsmouth), Rob Longman, Tony Dixon, Steve Mansfield, Nick Kenefick (South West), Iain Jenkins, Roger Motson, Tan Arulampalam, Ralph Austin, Zulfiqar Khan (St Marks/Colchester/Guildford).

7. IMPACT OF THE NATIONAL TRAINING PROGRAMME ON PROFICIENCY GAIN AND COMPETENCY IN LAPAROSCOPIC COLORECTAL SURGERY

7.1.	Introduction and aims.....	173
7.2.	Methods.....	175
7.2.1.	Database	175
7.2.2.	Proficiency gain curve analysis.....	176
7.2.2.1.	Clinical data (RA CUSUM).....	176
7.2.2.2.	Educational data (GAS forms).....	176
7.2.3.	Statistical analysis	177
7.3.	Results.....	178
7.3.1.	Clinical outcomes	178
7.3.2.	Proficiency gain curves	181
7.3.2.1.	Proficiency gain curve using clinical data	181
7.3.2.2.	Proficiency gain curve using GAS scores.....	185
7.3.2.3.	Learning rate	187
7.3.3.	Competency assessment	189
7.3.3.1.	Data summary	189
7.3.3.2.	Determinants of successful competency assessment.....	190
7.4.	Summary of results	193

7.1. Introduction and aims

By the end of 2011, 109 delegates performed more than 1300 training cases within the National Training programme, supervised by 49 different NTP trainers. In due course, the website has also been opened to surgical registrars who now can voluntarily submit GAS forms describing the proficiency gain in LCS during their training. Twenty-four trainees have taken this opportunity so far and have been used as a comparative subgroup for this study. In the previous chapters, methods have been developed and validated to describe the proficiency gain and to assess competency in laparoscopic colorectal surgery. It has been shown that the proficiency gain process can take up to 150 procedures in order to reach stable clinical outcomes when analysing data from unsupervised surgeons and preliminary evidence for the benefit of

supervised training has been established. However, up to now, no large single database has been available to investigate these effects further in more detail. Current data from the National Training Programme offers this opportunity for the first time. The aim of this chapter is to analyse the impact on clinical and educational outcomes of the programme, almost three years after its implementation. The objectives of this chapter are:

- (1) To describe the clinical outcomes
- (2) To analyse the effect of the programme on the clinical proficiency gain curve
- (3) To analyse the educational proficiency gain curve (level of independence)
- (4) To analyse the factors predicting successful completion of the programme

7.2. Methods

7.2.1. Database

Data included in this study originated from the beginning of the programme (January 2008) up to end of November 2011. The website to enter data was opened for registrars at a later stage and the first registrar entry was recorded in February 2010. All clinical data was self-reported by NTP delegates or registrars. In addition, a self-assessment using the GAS form (see **chapter 4**) was filled in for each case. Concurrently, the according supervisor (trainer) filled in another GAS form independently. A description of the instructions given to the participants can be found in **Appendix 2**. Clinical data required from the surgeons included patient demographics (age, gender, weight, height, ASA grade), case specifics (resection type, urgency, diagnosis) and outcome data (conversion, complications, length of hospital stay (LOS), reoperation, readmissions, mortality).

The screenshot displays the NTP website interface. On the left, a 'Specialist Registrar Account' login form is visible, with fields for First Name (Danilo), Last Name (Miskovic), Hospital Name (Queen Alexandra Portsmouth NHS), Hospital Address (Queen Alexandra Hospital, Cosham, Portsmouth PO6 3LY), Trust (Portsmouth Hospitals NHS Trust), Telephone (07858154988), and Email (d.miskovic@imperial.ac.uk). A black arrow points from the 'Add an Assessment' link in the top navigation bar to the 'A. SURGEON' section of the assessment form on the right. The assessment form is titled 'FILL IN SECTIONS A-D IMMEDIATELY AFTER THE OPERATION!' and contains several sections: A. SURGEON (A1 Operating Hospital, A2 Trainer), B. PATIENT (B1 Patient identification number, B2 Initials, B3 DOB, B4 Gender, B5 Operating date, B6 Height, B7 Weight, B8 ASA Grade), C. OPERATION (C1 Urgency, C2 Diagnosis, C3 Prior abdominal surgery, C4 Resection, C5 Anastomosis, C6 Intraoperative findings, C7 Conversion to open, C8 Reason for conversion), and D. OUTCOME (D1 Equipment problems, D2 Bleeding, D3 Exposure/Anatomy, D4 Bowel perforation).

Figure 29. Each surgeon has a personalised account to enter the data for each training case.

7.2.2. Proficiency gain curve analysis

7.2.2.1. Clinical data (RA CUSUM)

Proficiency gain curves for clinical data were constructed using complications and conversions as primary outcome parameters. Both are binary categorical data and hence a CUSUM curve was used for analysis. In order to adjust for individual case- and patient-related risks, a risk-adjustment (RA) model in accordance with the findings from **chapter 2** was created, including the covariates patient age, patient gender, ASA grade, diagnosis, BMI, resection type and emergency cases. In addition data were adjusted for hospital, surgeon and supervisor. Data was not adjusted for experience of the surgeon since this was the effect to be investigated by the CUSUM curve. The accuracy of the risk adjustment model was tested using ROC curves. An area under the curve of >0.7 was considered to be acceptable. For the CUSUM curve, the sum of all events was compared with the expected sum of events, using the CUSUM equation $S_i = S_{i-1} + (\Sigma_i - \Sigma_R)$; $S_0 = 0$; S_i is the cumulative sum, Σ_i the sum of events at procedure number i , and Σ_R the sum of expected events at procedure number i . The upper and lower borders of the CUSUM curve were chosen as 4 standard deviations (Ramsay 2001).

7.2.2.2. Educational data (GAS forms)

For scores from GAS forms, data were analysed with a simple CUSUM curve. A detailed description of the methods can be found in **chapter 4.2.4**. For the graphical description of the learning rate, a curve fit model was used (Feldman 2009). Raw data were plotted on a XY scattered graph and a curve was fitted using Excel software. A power law $[f(x) = ax^k]$ was assumed since learning processes are mostly follow this pattern (Ramsay 2001). Learning rates were described by the rate defining variable k .

7.2.3. Statistical analysis

For statistical analysis, the SPSS software was used (Statistical Package for the Social Sciences software [Version 18.0.2, SPSS Chicago (IL), USA]. For risk-adjustments and prediction models univariate models were applied to identify potential covariates. Variables with a $p < 0.1$ were included in a multivariate logistic regression model. If necessary, multiple levels of hierarchy were applied. For comparison of continuous data, the Mann Whitney U test was used, and for categorical data the chi-squared test used.

7.3. Results

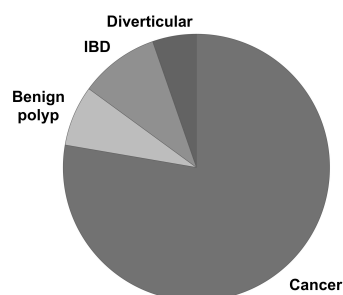
7.3.1. Clinical outcomes

By November 2011 a total of 1650 case records were collected (1309 NTP delegates, 341 registrars). Patient demographics are summarised in **Table 30**. There are no significant differences for NTP delegates and registrars.

	NTP	Registrar	p-value
Cases	1309	341	
Elective	1079	329	
Emergency	27	10	
Unknown	203	2	
M:F ratio	1.1 (573:523)	1.14(180:158)	0.603
Age median (IQR)	68 (60-76)	67 (55-74)	0.080
BMI median (IQR)	26.3 (23.9-29.3)	26.4 (23.8-29.4)	0.619
ASA median (range)	2 (1-4)	2 (1-4)	0.238

Table 30. Summary of patient demographics

Indication for surgery and types of procedures are summarised in **Table 31** and **32**. The proportion of cancer patients was slightly lower for registrars (80% vs. 71%). On the other hand, the proportion for inflammatory bowel disease cases was higher for registrars (7% vs. 17%).



	NTP	Registrar	p-value
Cancer	883 (80%)	239 (71%)	<0.001
Benign polyp	78 (7%)	28 (8%)	0.263
IBD	82 (7%)	57 (17%)	<0.001
Diverticular	62 (6%)	15 (4%)	0.242
Unknown	204	2	

Table 31. Proportions of indications for surgery (pie chart represents results from both NTP and registrars)

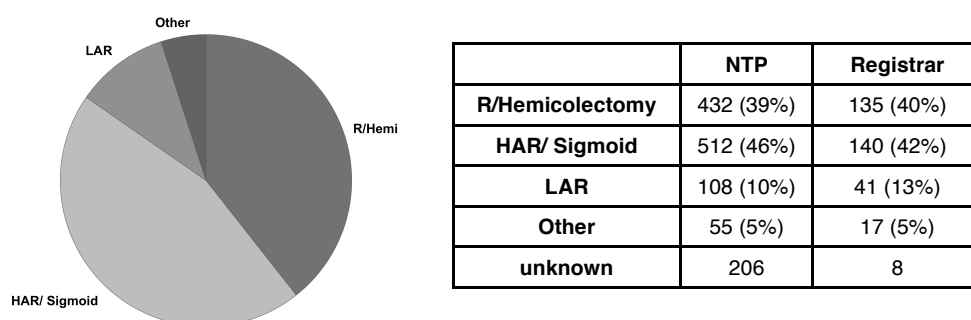


Table 32. Types of procedure performed (pie chart represents results from both NTP and registrars)

All data was self-reported and at the beginning of the NTP data collection was not mandatory. Hence, a significant proportion of data were incomplete. For further analysis, all cases with missing data on patient age, gender and/or diagnosis were excluded. A total of 413 (314 NTP, 99 registrar) fulfilled these exclusion criteria, leaving 1237 cases for analysis. Clinical outcomes are summarised in **Table 33**. The overall complication rate was 5.3% and the complication rate was 15.5%. There were no differences between NTP delegates and registrars apart from hospital stay which was slightly shorter for registrar cases (geometric mean 4.56 vs. 5.2 days, $p < 0.001$, **Figure 30**).

	NTP	Registrar	p-value
n=1237	995	242	
Conversions	51 (5.3%)	15 (6.2%)	0.506
Equipment problem	1	0	
Bleeding	3	1	
Exposure	49 (96%)	14	
Complications			
Total events	155 (15.6%)	37 (15.3%)	0.912
Ileus	41	7	0.375
Obstruction	8	2	0.972
Leak	24	3	0.263
Bleeding	12	5	0.303
Sepsis	15	5	0.537
SB injury	5	0	0.269
Medical	31	6	0.630
Other	20	9	0.400
Reoperation	3.62%	3.72%	0.940
Readmission	2.11%	3.31%	0.271
Mortality	0.60%	0.41%	0.724

Table 33. Summary of clinical outcomes

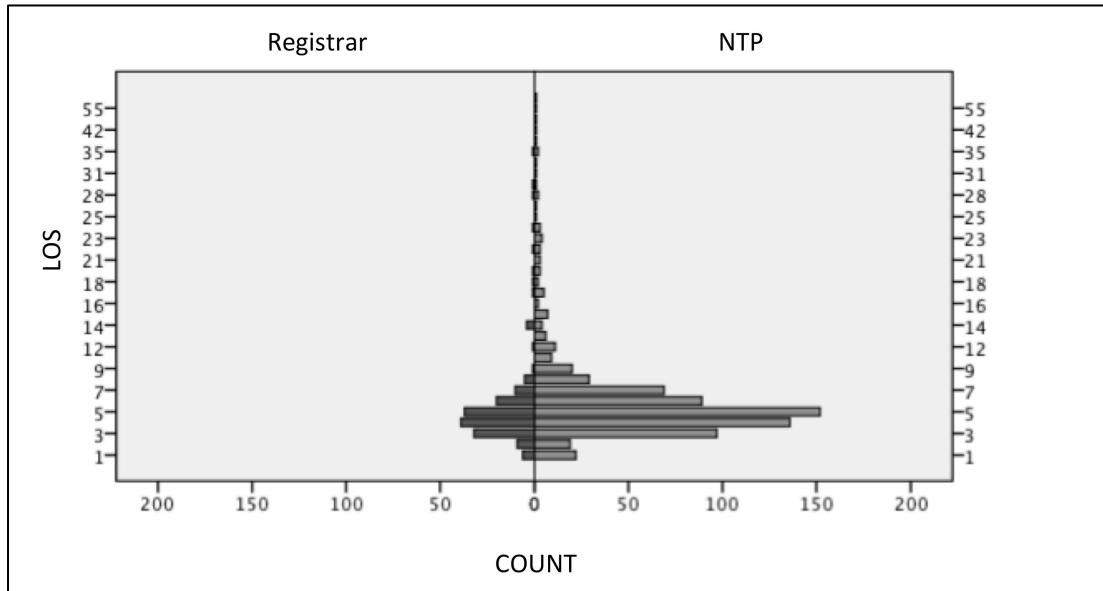


Figure 30. Graphical illustration (histogram) of hospital stay patients operated by NTP delegates and registrars.

7.3.2. Proficiency gain curves

7.3.2.1. Proficiency gain curve using clinical data

Proficiency gain curves were analysed for conversion and complication rates. The hierarchical logistic regression model for conversion and complication as dependent variables included hospital (block 1), surgeon and trainer (block 2) and patient age, patient gender, ASA grade, diagnosis, BMI, resection type and emergency cases (all block 3) as independent variables. These variables included, the prediction model achieved maximal accuracy with an area under the ROC curve of AUC=0.843 for conversion and AUC=0.721 for complications (see **Figure 31**)

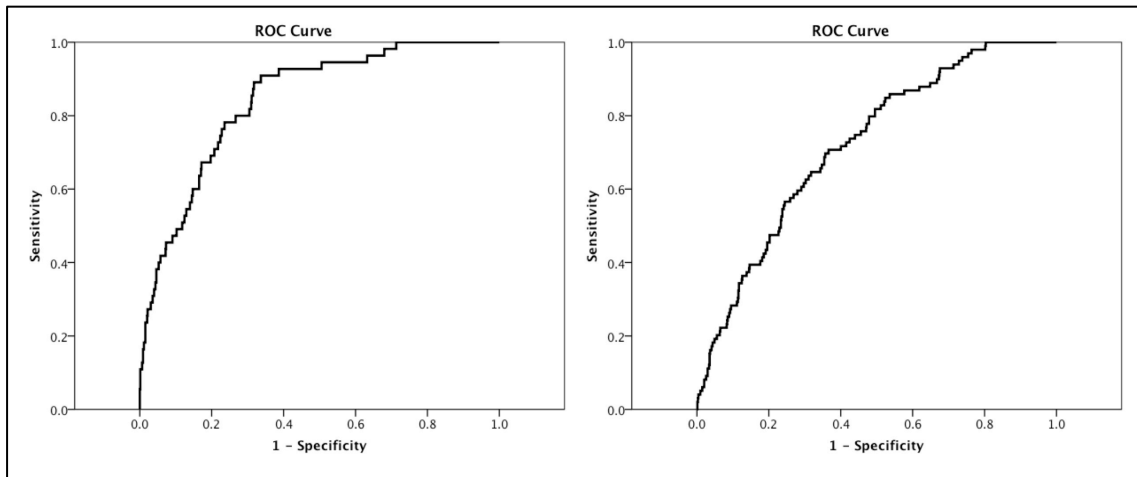


Figure 31. ROC curves for the prediction model using logistic regression for conversion (left) and complication (right).

The prediction models were used for the construction of risk-adjusted CUSUM curves to analyse the clinical proficiency gain curve. The number of cases performed by a single surgeon was 35. **Figure 32** illustrates the number of available assessments at different experience levels. As the programme is still continuing with a number of surgeons still in training and

since the length of training is different for different individuals, naturally more cases at the beginning of the learning curve were available at the time of analysis.

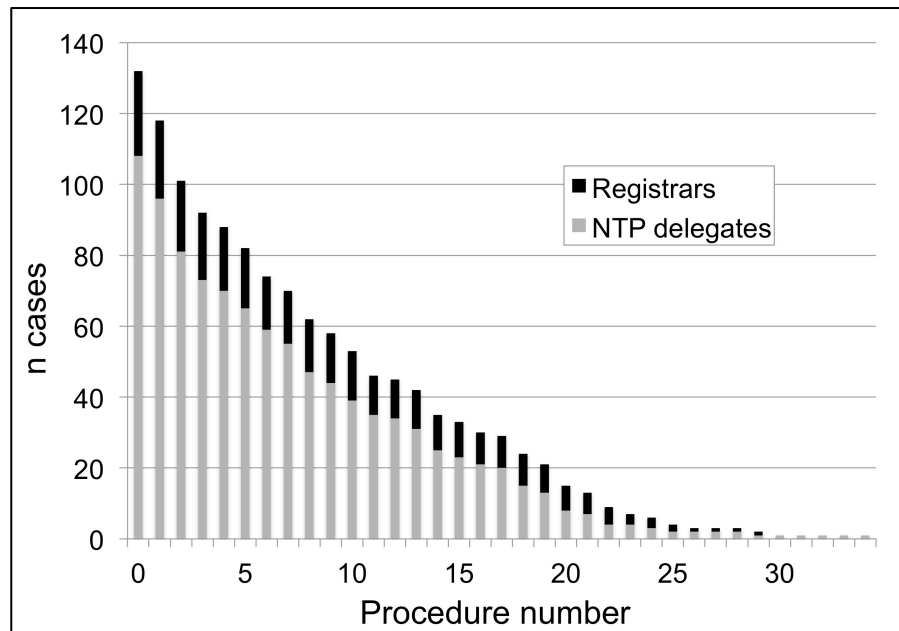


Figure 32. Number of surgeons submitting data at different stages during the proficiency gain curve.

The CUSUM curve for conversions did not show a change with increasing experience (**Figure 33**). The curve remains well within the upper and lower border and does not show an inflection point.

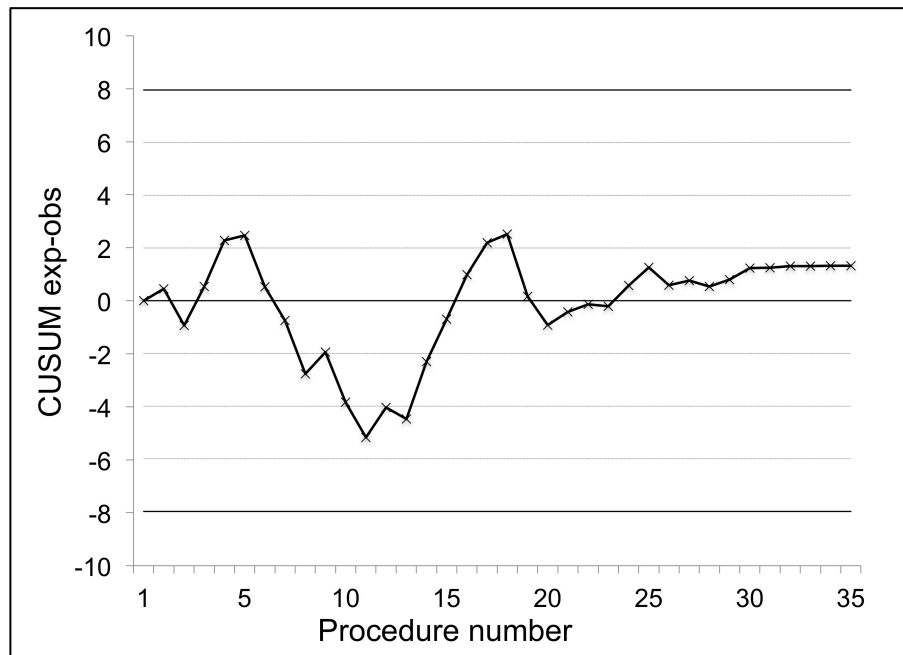


Figure 33. Proficiency gain curve for conversions. The curve remains within the boundaries of 4 standard deviations, indicating that the process within limits.

Equally, the proficiency curve for complications does not show an inflection point and stays within the safety limit of 4 standard deviations (**Figure 34**).

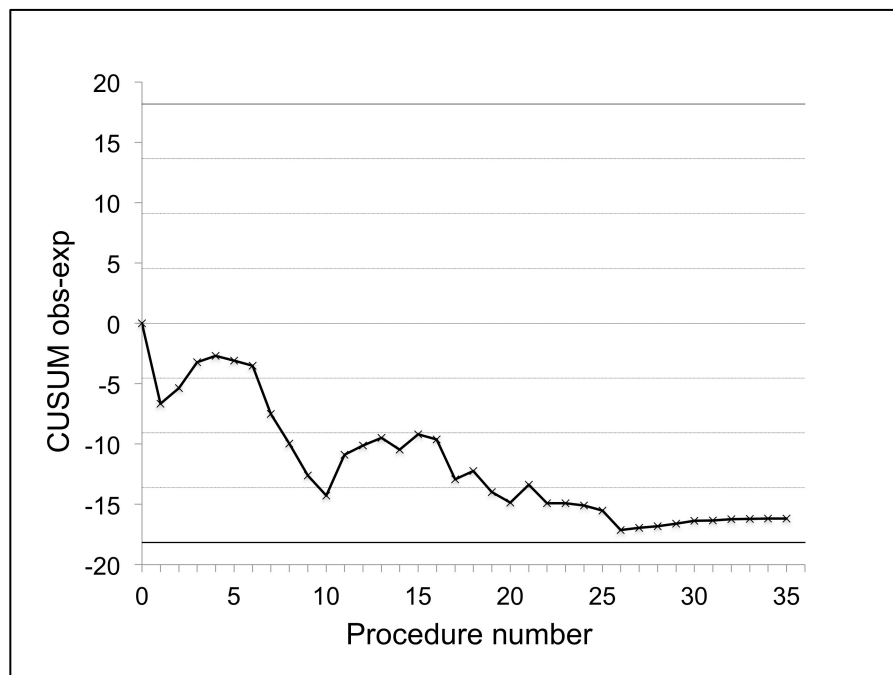


Figure 34. Proficiency gain curve for complications. The curve remains within the safety limits of 4 standard deviations.

Figure 35 summarises the clinical proficiency gain curves separately for registrars and NTP delegates. There is no obvious difference between the two subgroups.

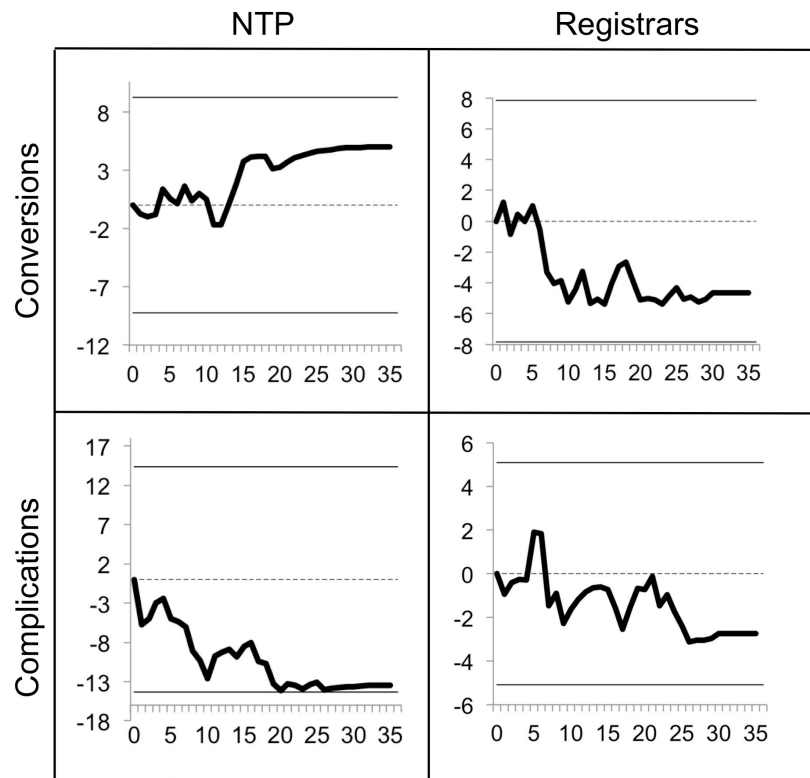


Figure 35. Subgroup analysis for NTP delegates and registrars for conversions and complications. None of the curves is crossing the boundary of 4 standard deviations (x: procedure number, y: CUSUM scores)

7.3.2.2. Proficiency gain curve using GAS scores

GAS scores were analysed separately for various subtasks and as the overall score (see **chapter 4** for details). Two different sets of CUSUM curves were constructed for consultants (**Figure 36**) and for registrars (**Figure 37**).

NTP delegates

The overall score showed an inflection point at 23 procedures, suggesting that a stable score of 5 or higher is reached on average after this number of training cases. For certain tasks, such as for theatre setup/ patient positioning this point is reached earlier (10 training cases) and for others later. Scores for the mobilisation of the flexure and TME do not show a definite inflection point.

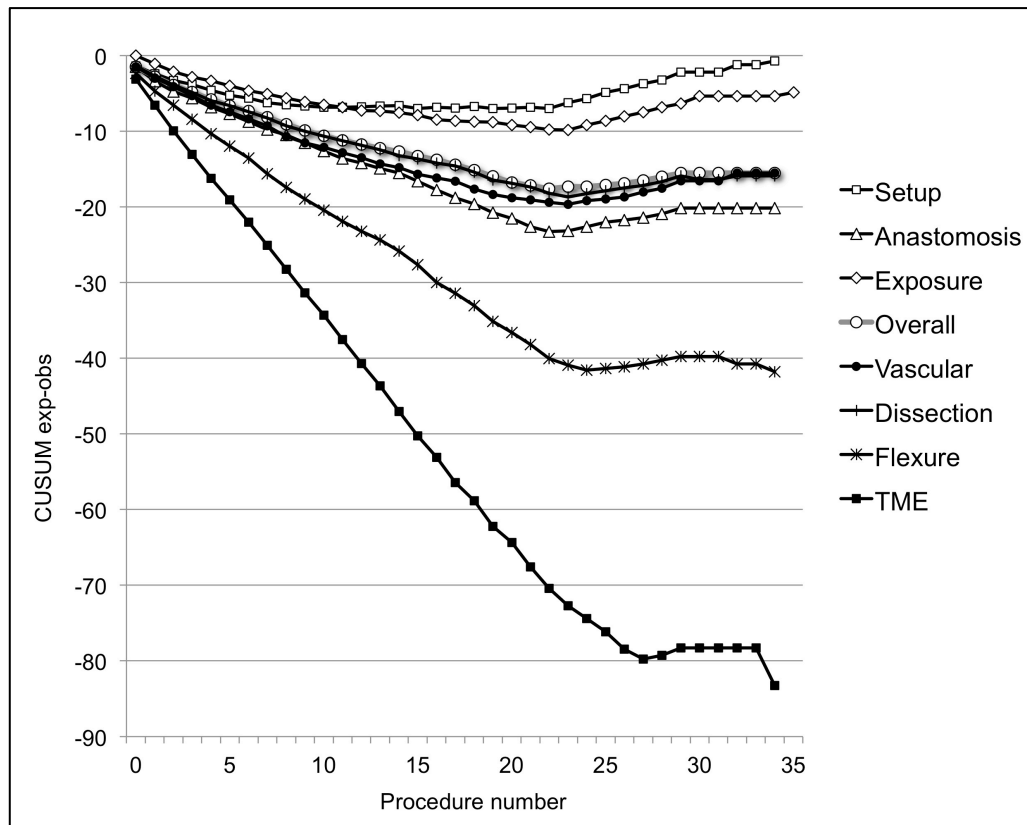


Figure 36. Summary of GAS scores for NTP delegates.

Registrars

For registrars data were only analysed up to the 25th procedure due to the limitations of available data beyond this point. No clear inflection point could be identified within this area.

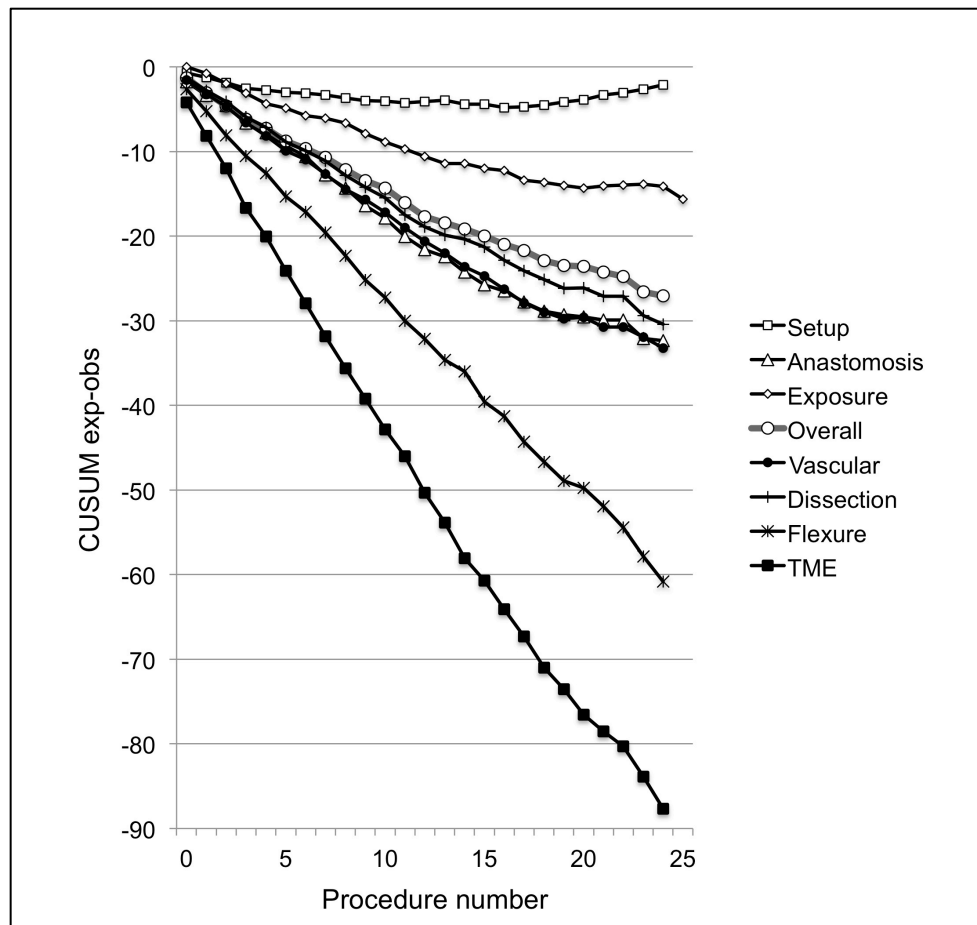


Figure 37. Summary of GAS scores for registrars

7.3.2.3. Learning rate

Due to the limitations of available data towards beyond 25 cases, the proficiency gain curves were analysed using a curve fit method. This allows projection of the expected proficiency increase based on the assumption that learning can be approximated using the power law.

Figure 38 shows a summary of the proficiency gain curves for NTP delegates (consultants) and registrars. **Table 34** is a summary of the learning rates, indicating that consultants seem to learn faster for a majority of the tasks.

Task	NTP	Registrar	Δk
Theatre setup	0.081	0.0643	0.0167
Patient positioning	0.0853	0.0586	0.0267
Access	0.0524	0.0629	-0.0105
Exposure	0.0796	0.0481	0.0315
Vascular pedicle	0.0908	0.0765	0.0143
Identification ureter	0.0769	0.0779	-0.001
Dissection mesentery	0.0789	0.068	0.0109
Flexure mobilisation	0.088	0.0665	0.0215
TME	0.0759	0.0616	0.0143
Bowel dissection	0.0758	0.0368	0.039
Specimen evacuation	0.0671	0.0438	0.0233
Anastomosis	0.0346	0.0336	0.001
Overall	0.0904	0.0599	0.0305

Table 34. Learning rates according to the power law curve ($y=ab^k$). The numbers represent the variable k (learning rate).

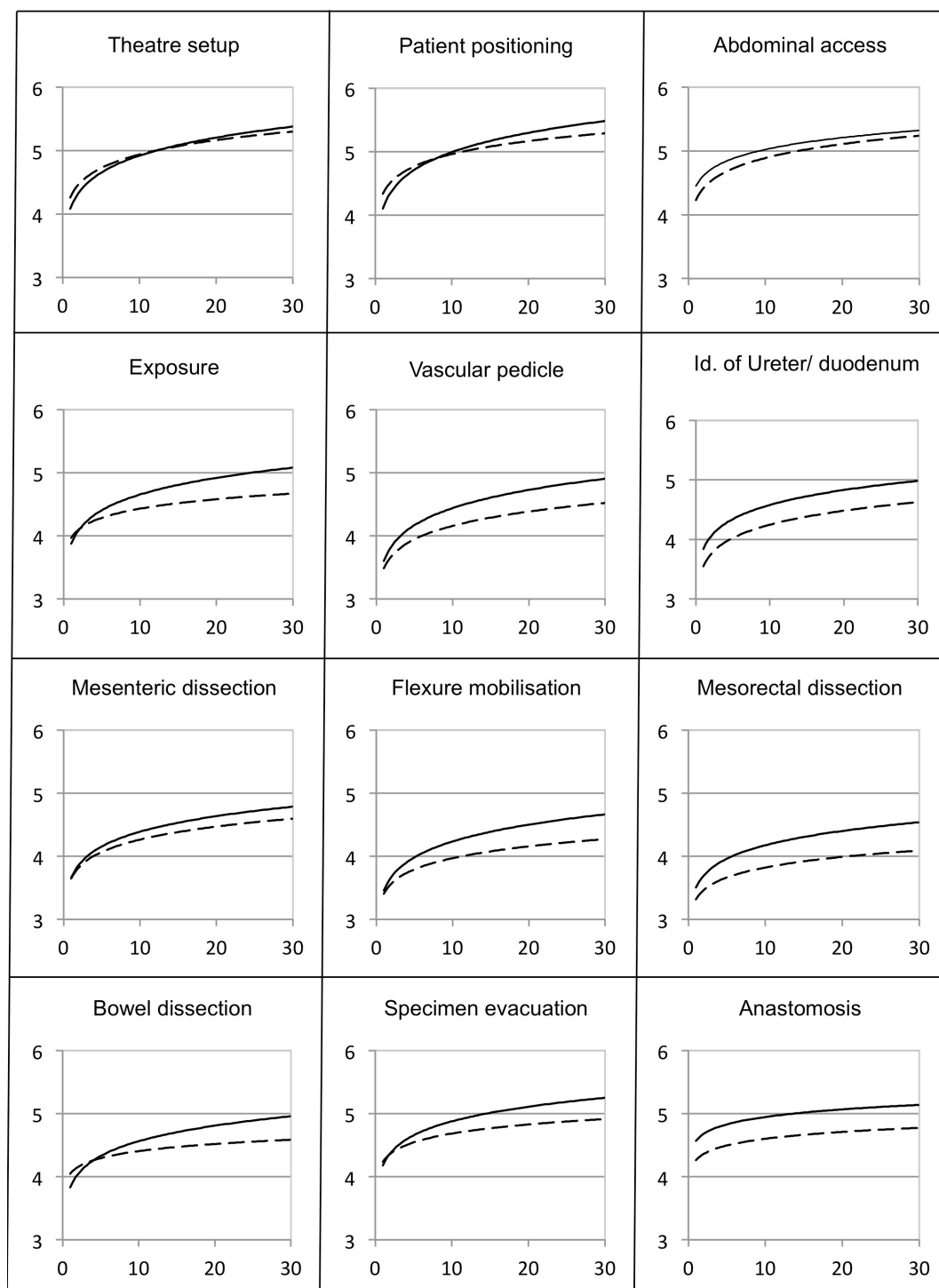


Figure 38. Power curve fit model for learning rates for NTP delegates (continuous lines) and registrars (dashed lines) for different subsections of the GAS form (x: procedure number, y: GAS score).

7.3.3. Competency assessment

7.3.3.1. Data summary

By November 2011 a total of thirty-one delegates from the NTP submitted videos of two independently performed cases which were sent for assessment (see **chapter 6** for the description of the process). Twenty-two passed the assessment, and nine failed. Re-submissions of delegates who failed at the first attempt were not included in this analysis.

There were submissions from all training centres. A summary can be found in **Table 35**.

	Pass	Fail
Basingstoke	0	2
Bradford	2	0
Colchester	2	1
Hull	0	2
Kings	1	0
Newcastle	1	1
Nottingham	4	1
NW	1	0
Oxford	3	1
Portsmouth	4	0
St Mark's	1	0
SW	3	1
TOTAL	22	9

Table 35. Numbers of trainees from different training centres (NW – North West, SW- South West training region).

7.3.3.2. Determinants of successful competency assessment

Delegates who passed performed on average more training cases than those who failed.

Nevertheless there was no difference between the two groups in terms of the overall experience (including unsupervised cases outside the NTP) (**Table 36**).

	Pass	Fail	p-value**
n delegates	22 (71%)	9 (29%)	
n training cases*	21 (11-26)	7 (3-12)	0.003
n cases at start*	15 (0-32)	0 (0-44)	0.638
n cases at end*	41(29-65)	33 (13-50)	0.317
*Median (interquartile range) **Mann-Whitney U test			

Table 36. Summary of competency assessment outcomes. The pass rate was 71%. There was a significant difference for numbers of training cases performed within the programme (supervised training). There was no difference for the total numbers of cases trainees have performed before entering or at the end of the programme.

There was no difference between the pass and fail group with respect to the average GAS scores. There was a trend that those in the pass group had a higher proportion of GAS scores >5 within the last three supervised cases before they submitted their sign-off cases. Delegates who failed seemed to have a greater discrepancy between their self-assigned scores and those given by their trainers. Whether this was a positive or negative discrepancy made no difference. As expected, the CAT scores were significantly different comparing those who passed and those who failed (**Table 37** and **Figure 39**).

	Pass	Fail	p-value
n delegates	22	9	
GAS score	4.49 (4.73-5.06)	5.00 (4.39-5.12)	0.610
Discrepancy (trainer vs. self)	0.21 (0.09-0.50)	0.44 (0.28-0.64)	0.053
>1 out of 3 last GAS scores >5*	10 (45%)*	1 (11%)*	0.098*
CAT score	3.03 (2.83-3.16)	2.42 (2.38-2.68)	<0.001
CAT R/Hemi	3.08 (2.80-3.18)	2.48 (2.73-2.81)	0.001
CAT L/hemi	3.07 (2.84-3.29)	2.47 (2.37-2.81)	<0.001
*Two or three out of the last three cases before sign-off scored 5 or higher. All values are median (interquartile range), except *total number (percentage). Mann Whitney U test, except *Chi-square test			

Table 37. Summary of GAS and CAT scores. CAT scores are significantly different between delegates who passed and those who failed the assessment. Overall GAS scores did not differ. However, delegates who failed tended to have a higher discrepancy between the scores given by their trainers and their self-assessment.

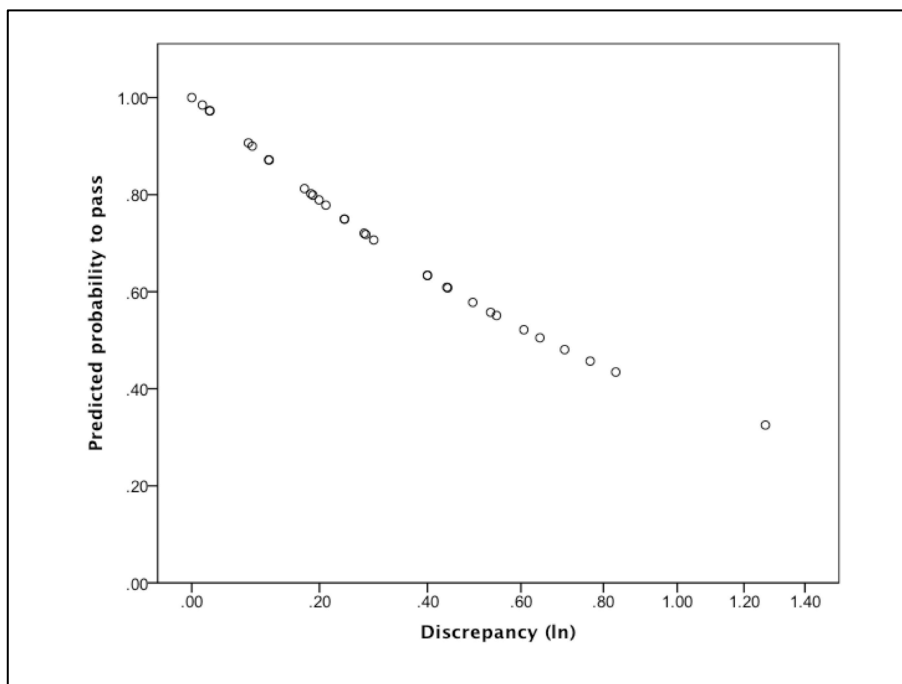


Figure 39. Predicted probability using the discrepancy as a sole factor for the prediction of passing the sign off. A cut off at a discrepancy of 0.3 reaches highest sensitivity (72.7%) and specificity (77.8%).

A univariate analysis was performed in order to identify parameters likely to predict successful completion of the sign off procedure. Variables with a $p < 0.1$ were included in a multivariate logistic regression model. These parameters were number of supervised training cases,

discrepancy between self- and trainer-assessment and at least two GAS scores >5 out of the final three cases before sign off (**Table 38**).

Parameter	p-value
Training centre	0.223
Trainer	0.181
<i>No. of training cases</i>	<u>0.026</u>
No. of cases at start of programme	0.550
No. of cases at end of programme	0.946
Average GAS score	0.979
<i>Discrepancy (ln)</i>	<u>0.073</u>
<i>>1 out of last 3 GAS scores >5</i>	<u>0.077</u>

Table 38. Univariate analysis of factors for prediction of successful sign off.

Using these three covariates, the prediction model for successful completion of the sign-off procedure proved to be moderately reliable with an AUC the ROC curve of 0.909 (**Table 39** and **Figure 40**).

	B	S.E.	Exp(B)	95% CI	p-value
Training cases	0.122	0.065	1.13	0.994-1.285	0.061
Discrepancy GAS scores (ln)	-1.219	0.791	0.295	0.063-1.391	0.123
>1 out of 3 last GAS scores >5	-1.467	1.28	0.231	0.019-2.835	0.252

Table 39. Multivariate analysis including covariates that had a $p < 0.1$ in the univariate analysis.

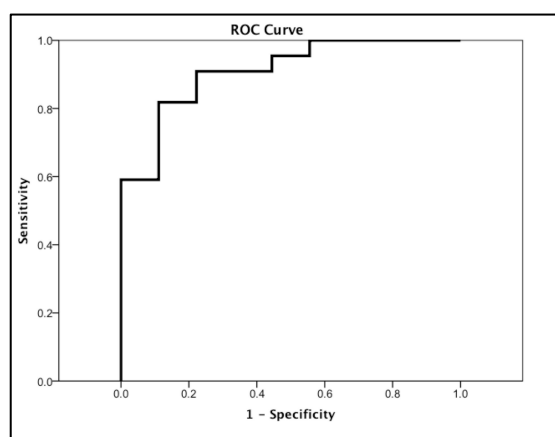


Figure 40. ROC curve for multivariate prediction model (Area under the curve (AUC)=0.909)

7.4. Summary of results

In this chapter a preliminary analysis of the current data set of the National training programme was performed. In addition, data submitted by surgical trainees (registrars) was used for comparison. The main outcomes, in accordance with the objectives described in the introduction, are as follows:

- (1) The clinical outcomes were not different between NTP trainees and registrars. Overall, these results are favourable and comparable with data series published by experienced surgeons.
- (2) There was no significant proficiency gain curve for clinical outcomes suggesting that the presence of an expert trainer may prevent negative clinical outcomes during the training period.
- (3) Analysis of GAS scores was used to measure the proficiency gain process. The GAS score represents the independency of a trainee, i.e. their ability to perform the case without substantial help of a trainer. Most tasks seem to be mastered by NTP delegates after 20-30 supervised cases. This figure may represent the average length of the proficiency gain curve in a supervised setting. It was shown that NTP delegates learn most tasks quicker than registrars, at least when analysing their self-assessment scores. A comparison of trainer assessments was not possible due to a lack of available data for registrars.
- (4) The number of supervised training cases was the single most important factor for a successful completion of the programme. Other factors were high GAS scores towards the end of training and a close match of the self-assessment score with the score given by the trainer.

8. DISCUSSION

8.1.	Summary of results	195
8.2.	Discussion of objectives.....	197
8.2.1.	Objective 1: <i>to identify shortfalls in the evaluation of the proficiency gain curve in laparoscopic colorectal surgery</i>	197
8.2.1.1.	Ethical concerns	197
8.2.1.2.	Reliability of clinical outcome data for the description of the proficiency gain curve	198
8.2.2.	Objective 2: <i>to develop methods to describe the proficiency gain curve in laparoscopic colorectal surgery</i>	199
8.2.3.	Objective 3: <i>to analyse proficiency and competency gain for specialists colorectal surgeons within the National Training Programme</i>	201
8.2.3.1.	Epidemiological implications	201
8.2.3.2.	Educational implications.....	202
8.3.	Future research	204

8.1. Summary of results

Firstly, the literature was systematically searched for evidence on the proficiency gain curves. Previous descriptions are predominantly based on clinical data published by surgeons starting LCS in an autodidactic setting. Although uniform evidence on structured teaching programmes was sparse, a meta-analysis of these data suggested that supervised training had a significantly positive impact on clinical outcomes during training. Secondly, novel methods of assessment were developed. Since clinical data alone seemed to be unsuitable to monitor proficiency gain during supervised training, a novel method was developed and validated. This method (the generic assessment scale – GAS) is based on the level of independence of a trainee when performing an operation under supervision for different areas of the task. It was shown that this method is reliably portraying the proficiency gain during training. In addition, novel methods were developed to measure competency. Competency was defined as safe, independent and correct performance of a straightforward case. This definition was based on the requirements as determined by the steering group of the National training programme. The methods included

the analysis of video-recorded operations on real patients using a human reliability approach and a structured assessment form that was implemented as a substantial component in the sign-off procedure of the NTP. Thirdly, a preliminary analysis of the outcomes of the NTP with more than one hundred trainees who performed around 1500 supervised procedures was performed. A third of the trainees have reached competency level and entered the sign-off procedure so far. Analysis of these data revealed that clinical outcomes were not affected by training level. The average number required to reach competency level for sign off (length of proficiency gain curve) was 20-30 training cases. The number of supervised training cases was the single most important factor for successful passing of the sign-off procedure. Increasing numbers of additional (individually performed) cases did not increase the chances of passing the test. The level of agreement between trainer and trainee on the quality of the performance was also identified as a positive predictor for successful completion of the competency assessment procedure.

8.2. Discussion of objectives

8.2.1. Objective 1: *to identify shortfalls in the evaluation of the proficiency gain curve in laparoscopic colorectal surgery*

In chapters 3 and 4 the current literature has been searched methodically to investigate the evidence on the proficiency gain curve in laparoscopic colorectal surgery. It has been shown that current concepts are based on retrospective analysis of prospectively collected clinical outcome parameters of patients. It took up to 150 cases per surgeon to achieve stable clinical outcomes and the potential benefits of careful case selection and supervised training have been demonstrated.

This evidence raises two relevant questions. Firstly, do ethical concerns prohibit the use of morbidity and mortality data for the description of individual proficiency gain curves?

Secondly, are clinical outcome parameters reliable for the description of the proficiency gain curve?

8.2.1.1. Ethical concerns

The National training programme in Laparoscopic colorectal surgery is the first of its kind (Coleman 2011). Unlike the development of a novel pharmaceutical innovation, many surgical procedures were not developed in a structured way, for example by means of artificial and animal models prior to their use on patients (Reitsma 2002). On the contrary, these techniques evolved through an organic process of “trial and error”. Many surgical innovations in cardiac surgery for example, and later used in civilian medicine, were developed under war conditions in so-called Mobile Army Surgical Hospitals (MASH) (King 2005). Many surgeons would argue that this process is necessary for surgical innovation.. However, is it ethically appropriate to use patient outcomes data to measure individual proficiency gain curves of surgeons learning

an established surgical technique? In chapter 4 this concept was compared with observing a trainee airplane pilot and counting the numbers of crashes and near misses arising as a consequence of training. This example demonstrates the absurdity of the concept. Evidence from literature analysis and clinical data analysis from the National Training Programme (chapter 8) demonstrates that training in an appropriate setting avoids increased rates of negative outcomes during proficiency gain; hence proficiency gain by means of clinical outcome data is obsolete and ethically unjustified.

8.2.1.2. Reliability of clinical outcome data for the description of the proficiency gain curve

Clinical outcome data are multifactorial and despite attempts to risk-adjust data, there may be a substantial amount of noise in the data that remains undetected by the observer. In laparoscopic colorectal surgery, Tekkis et al. were the first to use risk-adjusted data in the context of the proficiency gain curve, although this technique has not been used consistently since (Tekkis 2005c). The quality of risk adjustment is dependent on the quality and comprehensiveness of the available data. Traditionally, risk factors such as patient-related (e.g. age, gender, BMI, comorbidities) and case-specific factors (tumour stage, diagnosis, tumour site) are collected prospectively. However, other factors, such as anatomical variation, performance shaping factors, team factors and changes in theatre setup or postoperative care may also influence outcomes and are difficult to standardise. The use of clinical data in a supervised training setting has been shown to be useless. The influence of an experienced surgeon present during the operation (supervised training) has significant impact on clinical outcomes and makes a proficiency gain analysis using these data impossible. Nevertheless, measuring clinical outcomes in terms of data audit remains a crucial tool in the evaluation of the quality of a

clinical service. The use of clinical data for the description of individual proficiency gain curve of established surgical techniques however is unreliable.

8.2.2. Objective 2: *to develop methods to describe the proficiency gain curve in laparoscopic colorectal surgery*

Three novel methods of proficiency gain and competency assessment have been developed. For proficiency gain a simple assessment tool describing the level of supervision required to perform the task has been developed and validated (**chapter 4**). This Generic Assessment Scale (GAS) has proved to be practical and useful as a formative assessment tool *during* supervised training, but was not designed for competency assessment at *the end* of training; it indicates when a trainee reaches a level that is adequate for entering the sign-off procedure. Hence, other assessment methods for the evaluation of video-recorded laparoscopic colorectal resections were developed. Firstly, a method to objectively analyse the number of errors, as a surrogate for competency, was adapted for laparoscopic colorectal surgery. Secondly, a practical assessment tool was developed to measure competency for laparoscopic colorectal surgery. This Competency Assessment Tool (CAT) was also validated and is now fully implemented into the National Training Programme. Although the methodological details and limitations were analysed in the respective chapters, the context of these assessment methods within a structured training programme requires further discussion.

A variety of assessment tools are already available for direct observation of several surgical procedures. These include global assessment scales (e.g OSATS), task-specific checklists or mixed assessment methods, such as procedure based assessments (PBA) (van Hove ; Winckel 1994; Martin 1997; Ahmed 2011c). Most of these tools have been designed for the assessment of junior surgical trainees and are not suitable for the assessment of specialist colorectal surgeons. Global assessment tools are tailored to evaluate generic surgical skills, such as the economy of movement or the use of the assistant. Specialist surgeons have already developed

these basic surgical skills and stand alone generic skills are not sufficient to assess competency comprehensively. PBAs account for these shortfalls by mixing global and task-specific items. In addition perioperative skills, such as consenting the patient, establishing a postoperative care plan or sending a specimen for pathology, are assessed (Beard 2011). For specialist surgeons, perioperative topics may be less important for assessment since these skills are already established. Hence, novel assessment tools were specifically designed for this programme. They are an integral part of the training curriculum. GAS forms are used during supervised training and act as a formative assessment tool to monitor and encourage training of different task-steps, whereas CAT forms are used for a summative assessment before embarking on independent practice. These tools have shaped the programme considerably and also serve as a quality control for the success of the national initiative. Van der Vleuten defined the requirements for a good assessment tool (van der Vleuten 2005). The utility index of an assessment tool can be described as reliability, validity, educational impact, cost and acceptability (van der Vleuten 2010). Reliability and validity have been demonstrated for both tools and are described in detail in the appropriate chapters. The costs of the tools are not to be underestimated. The tools themselves are cost-neutral since no additional costs are involved apart from minimal additional assessors' time. However, the funding for the training centres was directly dependent on the numbers of GAS forms submitted and each sign-off assessment using the CAT form was also compensated. Although this funding was aimed primarily towards the cost of training and not the assessment, it was undoubtedly a strong incentive for the trainers to complete and submit the assessment forms. From the intercollegiate surgical training curriculum, trainees are responsible for completion of workplace-assessments (PBAs) . By the time of their annual review of competency progression (ARCP), they must produce a number of assessments and if they are unable they can be penalised. The NTP clearly demonstrates that acceptability and cost-effectiveness of workplace assessment can be reached through an incentive- rather than a penalty-based system.

8.2.3. Objective 3: *to analyse proficiency and competency gain for specialists colorectal surgeons within the National Training Programme*

The National Training Programme has epidemiological and educational implications. Clinical and educational data were analysed and presented in **chapter 7**.

8.2.3.1. Epidemiological implications

Colorectal cancer is the third most frequent cancer in the UK with an incidence of 46.1 per 100,000 and an annual prevalence of 32,644 cases (UK 2011). In recent years almost 20,000 colorectal resections have been performed annually (NICE 2007). The clinical outcomes of the cases within the National training programme are satisfactory. A conversion rate of 5.3%, complications of 15.6% and a median hospital stay of 5 days are acceptable clinical outcomes, even for an experienced laparoscopic colorectal surgeon. When comparing the outcomes of the NTP with the results presented in **chapter 2** of the first 150 cases (“learning curve”) of surgeons who trained unsupervised, the difference between a structured training programme and an autodidact approach becomes clearly evident. The conversion rate for autodidact surgeons was 13.2%, the complication rate 21.4% and median hospital stay 7 days. In a direct comparison, this corresponds to a number needed to treat of 13 to prevent one conversion and 17 to prevent one complication. The potential epidemiological impact of the NTP with regards to reduction of negative clinical outcomes, hospital bed days and costs are significant.

According to the Hospital Episode Statistics (HES) for the last completed year (May 2009-April 2010), 31% of all elective colorectal resections were performed laparoscopically. This is more than double compared to the year prior to the start of the NTP in 2007. With about 600 training cases per year, supervised training cases performed within the NTP already contribute to more than 5% of all colorectal resections in England.

8.2.3.2. Educational implications

The findings of this thesis raise several questions about educational issues.

Firstly, different learning rates for NTP delegates and surgical trainees (registrars) were demonstrated. Although clinical outcomes were comparable for both groups, NTP delegates seemed to learn faster than registrars. Interestingly, there was little or no difference in generic tasks, such as patient positioning, theatre setup or access. However, there was a clear difference in learning rates for procedure-specific tasks such as exposure, vascular pedicle dissection or flexure mobilisation. A possible reason for this difference may be previous experience in open colorectal surgery. In an experimental setting, this relationship could not be shown (Brown 2011). However, the subjects in this experiment were all very junior surgeons, performing a relatively easy task on a virtual reliability simulator (laparoscopic cholecystectomy). Advanced colorectal resections are more complex and previous open experience may increase the understanding of anatomical and pathological complexities which may be an advantage when learning laparoscopic colorectal surgery.

Secondly, the length of the proficiency gain curve was shown to be only 20-30 procedures. This is significantly shorter than the 90-150 procedures as demonstrated for autodidact surgeons in **chapter 4**. The two groups cannot be compared directly. Evidence from the literature is based on clinical outcomes data and partly originates from the 1990s. Technological advances in imaging technology and instrumentation may have accelerated proficiency gain. Nevertheless, data as presented in this thesis support the significant impact of supervised training on the shortening of the proficiency gain curve. As shown in **chapter 7**, the number of supervised training cases was the single most important factor in predicting successful performance in the sign-off process, whereas the overall number of procedures was no different when the pass and the fail group were compared. This evidence supports the significant beneficial influence of supervised training. There is also evidence from the educational literature to support this

concept. Vygotsky described the zone of proximal development in child education, describing the increase in learning when guided by an adult teacher (Vygotsky 1978). Similar observations were also made in adult learning (Zorga 2002).

Thirdly, three main factors have been identified to be predictive for successful completion of the programme and reaching competency status in laparoscopic colorectal surgery. As mentioned before, the strongest indicator was the number of training cases performed within the programme. In addition, higher GAS scores towards the end of the supervised training period were beneficial, supporting the validity of the method. Interestingly, an absolute difference between GAS scores submitted by the trainer and those made in a self-assessment by the trainee was also predictive for competency. The lower the difference, the more likely the candidate would pass the final assessment. It should be emphasised that both positive and negative differences predicted unfavourable outcomes. The interpretation of this finding remains speculative. A candidate who is constantly overestimating his or her own skills importantly may lack insight. On the other hand, a candidate who constantly underestimates him- or herself may have been lacking self-confidence. Both characteristics might be regarded as disadvantageous features for the successful completion of the competency assessment procedure. Non-surgical literature supports this hypothesis. Dunning and Kruger asked psychology students to rate themselves in an examination. They found that poor performers significantly overrated their performance, whereas good performers made an adequate self-assessment (Dunning 2003).

8.3. Future research

The National Training Programme is still running and more data from the programme and beyond can be expected over the next two years. Beyond completion of supervised training, the delegates will start their independent practice. This has several implications on future research. Firstly, the long-term training effects of the NTP need to be investigated. Are the excellent results achieved during supervised training transferrable to independent practice? Are delegates, who are considered to be competent, able to produce satisfactory clinical outcomes after active training and in independent practice? Secondly, the epidemiological impact of the programme needs further understanding; especially socio-economic factors such as cost effectiveness are of major importance in the current climate of limited resource in the public healthcare sector. Thirdly, evidence on factors predicting competence need further investigation. This has important implications on the selection process of surgeons in such educational programmes. In addition, the competency assessment process needs further development. It would be desirable to find methods for workplace-based assessments that are more objective and more automated than the currently available developed tools.

The National Training Programme in Laparoscopic Colorectal Surgery in England will provide many more opportunities to achieve these goals.

Appendix i. Instruction Manual for CAT

WHAT IS L-CAT?

L-CAT is a marking sheet for the assessment of technical surgical skills in Laparoscopic Colorectal Surgery. It is designed to assess the surgeon's performance by watching a live, life-streamed or a videorecorded operation. The aim is to determine the level of competency of a training surgeon in order to recommend further training with or without a supervisor (signing off procedure).

WHAT ARE THE POSSIBLE OUTCOMES?

The outcomes of an assessment are: 1. *Needs further improvement* or 2. *Safe performance*. In case 1 the level of competency has not been reached yet and further supervised training is recommended. In case 2 the surgeon's performance was solid and safe and further training without a supervisor (independent training) is recommended.

HOW TO RATE A TECHNICALLY DIFFICULT CASE?

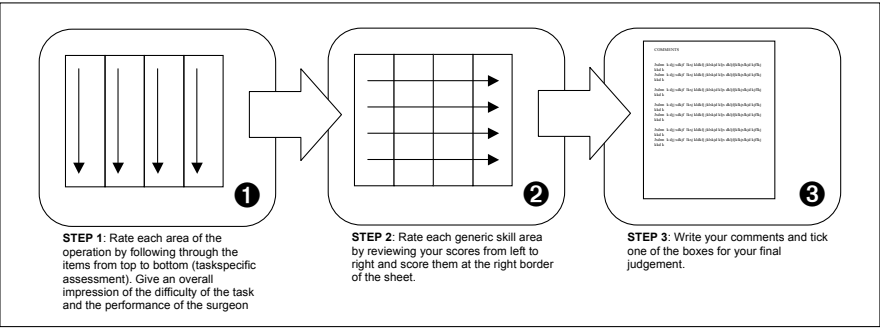
For each task area you also have to rate the difficulty of the case. Your overall impression of the surgeon's skills should always be in relation to the difficulty of the case; eg. the (competent) surgeon may have caused bleeding due to difficult anatomy in an obese surgeon vs. the (incompetent) surgeon caused bleeding in a slim patient with normal anatomy.

WHAT DOES L-CAT NOT RECORD?

L-CAT does not include non-technical skills assessment (teamwork, leadership, communication etc). It also does not record the influence of the assistant.

HOW TO USE L-CAT

L-CAT is designed to support the assessors by streamlining their thoughts and help document the assessment. Follow the three following steps:



How the scoring works:

- A. The scoring system can be thought of like traffic lights: incompetent ("red"), novice ("amber") competent ("green") and expert ("star") performance. You should rate each item independently, if you are not sure about the terminology read the Definitions of Terms.
- B. For your ratings of each horizontal and vertical column (task-specific and generic sections) you have to review your item ratings vertically and horizontally. To reach competency in a section a majority of the items should be "green" and there should be no "reds".
- C. For the overall rating (Step 3) you have to review your section ratings. To suggest "safe performance", a majority of the sections should be "competent". If one section is red it mandates "needs improvement".

"RED"	☐	Dangerous dissection technique
"AMBER"	☐	Ineffective dissection technique
"GREEN"	☐	Safe dissection technique
"STAR"	☒	Expert dissection technique
	☐	Not applicable

DEFINITIONS OF TERMS

TASK STEP 1	EXPOSURE
Ports are inserted in an ergonomic position, operating field is exposed	Task Step 1 begins with the first port insertion and ends when the exposure of the operating field is completed and dissection commences. The aim is to insert the ports in a correct, ergonomic position utilising a safe technique under direct vision. Using meaningful movements the small bowel should then be swiftly and effectively moved out of the operating field by using both atraumatic graspers (type Johann, Babcock etc) and also by changing the patient position.
Dangerous port insertion, wrong position	Ports are inserted without being visualised or ports are inserted tangentially. Tip of port forcefully touches intraabdominal organs with potential or resultant injury. Ports are inserted in an non-ergonomic position (wrong angles, wrong directions, evidence of torque).
Incompetent port insertion or poor port position	Ports are visualised on insertion however several attempts are made to achieve placement (abdominal wall is depressed >3 times before successful puncture, or more than one hole is made in the peritoneum for one port). Ports are inserted suboptimally for ideal ergonomic operating (angles, directions), however no evidence of torque.
Safe port insertion and good port position	Ports are inserted perpendicularly under direct vision without harming any intraabdominal organs. The ports are in a position that allows ergonomic operating (good angle, correct direction, no torque).
Expert port insertion and ideal position	Ports are swiftly inserted perpendicularly under direct vision in one attempt without harming any intraabdominal organs. The ports are in a position that allows ergonomic operating (ideal angle, correct direction, no torque). Adhesions are taken down, if necessary to enable perfect port placement.
Dangerous and ineffective moving of bowel	Bowel is grasped with wrong instrument (traumatic grasper, (hot) ultrasound grasper etc.). Bowel is repeatedly grasped forcefully with atraumatic grasper. Mesentery or bowel is grasped with such force that it causes bleeding or perforation. Bowel has to be moved several times and repeatedly falls back into the same position. Adhesions are dissected causing potential harm to adjacent organs. Lack of meaningful movements.
Ineffective moving of bowel	Safe handling of small/ large bowel at most times. Sometimes bowel is grasped too forcefully with atraumatic grasper. Bowel has to be moved several times and repeatedly falls back into the same position. This is observed throughout the task. Adhesions are dissected without causing potential harm to adjacent organs.
Safe and effective moving of bowel	Safe handling of small/ large bowel. There are episodes when bowel has to be moved several times although the majority of movements of instruments are effective, meaning that the bowel stays in a different position when moved. Adhesions are dissected without causing potential harm to adjacent organs. Evidence of attempts to change patient position.
Efficient and expeditious moving of bowel	Safe handling of small/ large bowel. Efficient moving of bowel meaning that the bowel stays where it is placed when moved into a different position. Adhesions are dissected without causing potential harm to adjacent organs. Effective use of tilt/head-down etc.
Causes bleeding/ avulsion/ potential perforation	Grasping of bowel and mesentery or dissection of adhesions causes bleeding OR on several occasions, avulsion of tissue by crushing or tearing due to a lack of respect for the tissue. The bowel is handled in a dangerous way that may cause immediate or delayed perforation
Causes minor bleeding/ avulsion, no perforation	Grasping of bowel and mesentery or dissection of adhesions causes minor bleeding OR occasional avulsion of tissue (by crushing or tearing). No danger of perforation.
Causes minimal bleeding/ no avulsion or perforation	Grasping of bowel and mesentery or dissection of adhesions causes minimal bleeding/ no avulsion of tissue (by crushing or tearing). No danger of perforation.
No bleeding/ avulsion/ potential perforation	Grasping of bowel and mesentery or dissection of adhesions causes no bleeding/ no avulsion of tissue (by crushing or tearing). Evidence of total respect for the tissue and understanding of the amount of force necessary to achieve goal. No danger of perforation.
Incomplete exposure of operating field	The surgeon starts with the dissecting of the next task (vascular pedicle or mobilisation) without appropriate exposure of the operating field (small bowel not sufficiently removed, vascular pedicle not retracted).
Ineffective but safe exposure of operating field	The exposure of the operating field is sufficient but non-ideal and laboriously undertaken. The small bowel loops are close to the dissecting area.
Safe and adequate exposure operating field	The operating field is safely exposed before dissection starts with evidence of a paucity of movements. The small bowel is removed although occasionally fall back into the field.
Expeditious and safe exposure of operating field	The operating field is perfectly exposed before dissection starts using efficient and meaningful movements and minimal delay. The small bowel is removed and doesn't fall back into the field.

page3/6

TASK STEP 2	VASCULAR PEDICLE
Vascular pedicle is identified and vessels dissected out at an appropriate level	Task Step 2 begins with the retraction of the vascular pedicle (ileocecal, middle colic, IMA etc.) and ends with the complete dissection of the vein and artery. If more than one main vessel (eg. ileocecal and middle colic) is dissected at different stages of the operation all vascular dissections are summarized in this section. If the quality of performance differs (eg. good for ileocecal, but bad for middle colic), the worst performance has to be rated. The aim is to safely dissect out the main vessels without causing immediate or potential delayed bleeding and without collateral damage to bowel and other adjacent organs.
Dangerous use of stapler/ clip-applier	Wrong stapler is used (not vascular). Stapler/clip-applier is fired without clearing the far side OR additional tissue (bowel, mesentery etc) is caught in the stapler. No staples applied. Clips not safe (not circumferential, only one clip (rather than two)).
Incompetent use of stapler/ clip-applier	Awkward and inefficient handling of the stapler gun. Stapler far side not visualised before firing but presumably clear. Clips have to be removed/ re-applied.
Safe use of vascular stapler/ clip-applier	Far side of stapler/clip-applier visualised and cleared before firing, some attempts of positioning the gun may have been performed before firing.
Expert use of stapler/ clip-applier	Far side of stapler/clip-applier visualised and cleared before firing with minimal positioning attempts.
Dangerous dissection technique	Mesentery around the vascular pedicle (window above and below the vessel) is dissected with the wrong instrument/technique (eg blunt rather than sharp, with or without diathermy etc). Heat is applied too close to bowel/ vessel or other adjacent organs.
Incompetent dissection technique	Mesentery around the vascular pedicle (window above and below the vessel) is dissected with the right instrument but awkward technique (uncontrolled or jerky movements, some dissection attempts are without any apparent effect).
Safe dissection technique	Mesentery around the vascular pedicle (windows above and below the vessel) is dissected with the right instrument and safe technique (effective tissue dissection), although may take marginally longer than an expert surgeon.
Expert dissection technique	Mesentery around the vascular pedicle (windows above and below the vessel) is dissected with a high grade of efficiency and safety.
Causes bleeding/ avulsion/ potential perforation	Dissecting the mesentery around the pedicle causes bleeding OR dissection of pedicle causes squirting bleeding OR on several occasions avulsion of tissue (by crushing or tearing) with inappropriate control of bleeding. The bowel is handled in a dangerous way that may cause immediate or delayed perforation.
Causes minor bleeding/ avulsion, no perforation	Dissecting the mesentery around the pedicle causes bleeding OR dissection of pedicle causes minor bleeding OR occasional avulsion of tissue (by crushing or tearing). Bleeding controlled effectively. No danger of bowel perforation.
Causes minimal bleeding/ no avulsion or perforation	Dissecting the mesentery around pedicle causes minimal bleeding OR dissection of pedicle causes minimal bleeding/ no avulsion of tissue (by crushing or tearing). Any bleeding controlled quickly, appropriately and effectively. No danger of bowel perforation.
No bleeding/ avulsion/ potential perforation	Dissecting the mesentery around pedicle causes no bleeding OR dissection of pedicle causes no bleeding/ no avulsion of tissue (by crushing or tearing). No danger of perforation.
Dangerous pedicle dissection or grossly at the wrong level	The pedicle is dissected but not secured (delayed bleeding possible: eg. only one clip proximally, clips applied in a wrong fashion, energy source not used effectively etc). The pedicle is dissected either too distal or proximal.
Safe, but ineffective pedicle dissection or non-ideal level	The pedicle is dissected and secured although not with ease. The pedicle is dissected too distal or proximal to the ideal point.
Safe pedicle dissection at the at the right level	The pedicle is dissected and secured safely. The pedicle is dissected within range of the ideal dissection point.
Bloodless and efficient pedicle dissection at right level	The pedicle is dissected and secured safely and efficiently at the ideal point.

page 4/6

TASK STEP 3	MOBILISATION
Large bowel mobilisation following anatomical planes/ landmarks adequately defined and protected	This task step starts with the separation of tissue planes after dissecting the vascular pedicle (or after „Exposure“ for procedures without vascular dissection). It ends when the segment of large bowel is fully mobilised and ready to be resected. The aim is to identify landmarks which have to be protected (ureter, duodenum) and to mobilise the colon along anatomical planes (eg. Toldt's fascia, Gerota fascia etc). If required the hepatic and/or splenic flexure have to be mobilised without injuring the spleen, liver, stomach, pancreas etc.
Dangerous use of graspers/ dissection tools	Traumatic graspers are used to grasp the bowel or other sensitive structures. Forceful blunt dissection causes tissue bleeding. Hot dissection tool touches sensitive structures causing potential harm.
Incompetent use of graspers/ dissection tools	Awkward and inefficient grasping and exposure for dissection. Bowel grasped forcefully several times with atraumatic grasper. Inefficient dissection (eg. movements too short, in the wrong direction, and without apparent effect). Inadequate dissection techniques (eg. blunt rather than sharp or vice versa). Hot dissection tool used in close proximity to sensitive structures causing potential harm.
Safe use of graspers/ dissection tools	Safe grasping of bowel. Adequate use of blunt and sharp dissection. Safe use of hot dissection tool.
Highly efficient use of graspers/ dissection tools	Ideal and efficient use of graspers and dissecting tools, paucity of movements and without causing potential harm to adjacent structures.
Wrong tissue plane, unable to correct quickly	Dissection in or along the wrong tissue plane most of the time. Does not seem to realise that is in the wrong tissue planes and insufficient attempt is made to refine the correct plane.
Repeatedly in wrong tissue plane, able to correct quickly	Repeated dissection in or along the wrong tissue plane but with prompt realisation and correction
Correct tissue plane, able to correct quickly if occasionally strays out of plane	Maintains correct tissue plane the majority of the time, and if plane is lost, corrects quickly
Constantly stays in correct tissue plane	Stays in correct tissue plane almost all the time. Corrects immediately if plane lost.
Causes bleeding/ avulsion/ potential perforation	Dissection along the tissue planes causes profuse squirting bleeding OR on several occasions avulsion of the tissue (by crushing or tearing). Inappropriate control of bleeding. The bowel is handled in a dangerous way that may cause immediate or delayed perforation.
Causes minor bleeding/ avulsion, no perforation	Dissection along the tissue planes causes bleeding OR dissection of pedicle causes minor bleeding OR occasional avulsion of tissue (by crushing or tearing). Bleeding controlled effectively. No danger of bowel perforation.
Causes minimal bleeding/ no avulsion or perforation	Dissection along the tissue planes causes minimal bleeding OR dissection of pedicle causes minimal bleeding/ no avulsion of tissue (by crushing or tearing). Any bleeding controlled quickly, appropriately and effectively. No danger of bowel perforation.
No bleeding/ avulsion/ potential perforation	Dissection along the planes causes no bleeding OR dissection of pedicle caused no bleeding/ no avulsion of tissue (by crushing or tearing). No danger of perforation.
Bowel inadequately mobilised or anatomical landmarks not defined	The mobilised bowel segment is too short. The mesentery is dissected clearly too peripherally or centrally. The landmarks are not demonstrated.
Bowel inadequately mobilised, anatomical landmarks defined	The mobilised bowel segment is too short. The landmarks are demonstrated although again the mesentery is inadequately dissected.
Bowel adequately mobilised, anatomical landmarks defined	The mobilised bowel is of sufficient length, anatomical landmarks demonstrated, appropriate dissection of mesentery.
Ideal mobilisation of bowel, anatomical landmarks defined	The mobilised bowel is of ideal length. Landmarks clearly demonstrated and perfect preparation for next step.

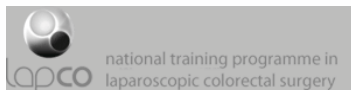
page 5/6

TASK STEP 4	RESECTION/ ANASTOMOSIS
Large bowel is adequately and appropriately resected and anastomosis is performed safely	Task Step 4 starts with the preparation of the bowel for dissection (eg. clearance of mesentery around terminal ileum or circumferential dissection of mesorectum for rectum). It ends with the complete dissection of bowel and the creation of the anastomosis (if applicable).
Dangerous use of intestinal stapler	Wrong stapler is used (not intestinal) . Stapler is fired without clearing the far side OR additional tissue (ovary, mesentery etc) is caught in the stapler. No staples applied. Stapler is fired in the wrong angle (not perpendicular to the bowel).
Incompetent use of intestinal stapler	Awkward and inefficient handling of the staple gun. Both stapler jaw tips not visualised before firing but presumably clear. Several attempts made to place the stapler gun.
Safe and competent use intestinal stapler	Tips of stapler visualised and cleared before firing, few attempts of positioning the gun may have been performed before firing, but bowel eventually in optimum position.
Expert use of intestinal stapler	Tips of stapler visualised and cleared before firing with slick positioning and optimal resection, and appropriate tissue compression prior to firing for stapler type.
Dangerous preparation of bowel for resection	Grasping the bowel/stump/vessels forcefully with grasper. Inappropriate use of diathermy close to bowel/stump/vessels.
Incompetent preparation of bowel for resection	Mesentery around bowel dissected ineffectively, tissue handled safely and respectfully.
Safe and competent preparation of bowel for dissection	Safe and effective but slightly laborious tissue dissection for preparation of bowel for dissection.
Efficient and optimal preparation of bowel	Efficient and effective tissue dissection for preparation of bowel for dissection.
Causes bleeding/ avulsion/ potential perforation	Dissection around the bowel causes squirting bleeding OR on several occasions avulsion of the tissue (by crushing or tearing). Inappropriate control of bleeding. The bowel is handled in a dangerous way that may cause immediate or delayed perforation.
Causes minor bleeding/ avulsion, no perforation	Dissection around the bowel causes minor bleeding OR dissection of pedicle caused minor bleeding OR occasional avulsion of tissue (by crushing or tearing). Any bleeding controlled effectively. No danger of bowel perforation.
Causes minimal bleeding/ no avulsion or perforation	Dissection around the bowel causes minimal bleeding OR dissection of pedicle caused minimal bleeding/ no avulsion of tissue (by crushing or tearing). Any bleeding controlled quickly, appropriately and effectively. No danger of bowel perforation.
No bleeding/ avulsion/ potential perforation	Dissection around bowel caused no bleeding OR dissection of pedicle caused no bleeding/ no avulsion of tissue (by crushing or tearing). No danger of perforation.
Resection level likely to compromise stump perfusion or anastomosis	The bowel is resected causing perforation, with faecal spillage into the abdominal cavity OR bowel is resected at a level with inadequate blood supply. This is not corrected. Anastomosis under tension.
Resection suboptimal, anastomosis safe	The bowel is resected with minimal faecal spillage which is promptly cleaned. Resection level is suboptimal but this is corrected. Anastomosis is safe.
Adequate resection, safe anastomosis	The bowel is resected cleanly and adequately with a good anastomosis and blood supply.
Ideal resection and anastomosis technique	The bowel is efficiently resected at an ideal point, with an adequate blood supply and perfect anastomosis technique.

page 6/6

Appendix ii. GAS form online instructions

[Home](#) | [Coordination Office](#)



Please see new links
to all section below

GAS Forms

Upon completion of each clinical procedure the trainer and trainee are required to complete and upload their own respective Global Assessment Score (GAS) forms through the Lapco web site.

Trainer Section (Form A)

Please see attached example. ([Download Form A](#))

Trainee Section (Form B)

Please see attached example ([Download Form B - Incorporating Mini STTAR](#))

Uploading a GAS Form

To upload a GAS form go to the Lapco web site (www.lapco.nhs.uk)

Go to the top right hand corner of the home page, click on the Lapco 'Log In' icon.

Enter your email address and password

Go to 'Add an Assessment' which will bring up the form to complete and click on submit when ready.

Please ensure the name of the trainer/trainee is correctly entered as this will automatically create a 'pending' record in their respective account for the correlating case.

Your account can be viewed at any time which keeps an on line record of all your assessments.

Sharing a Clinical Session/GAS Form

If a case is shared, both trainees can upload a GAS form as long as both trainees have to be present from start to end and be involved in sections B1-B3 and B10-12 (see GAS form). In addition, each trainee needs to perform a 'significant proportion' of the laparoscopic part ie at least two of the sections B4 to B9 (See GAS form). Those parts that are not performed by a trainee as primary surgeon will be scored as N/A.

For any further queries on GAS Forms please contact the Coordination Office (tania.dorey@nhs.net)

>> Training Video



Click on the links within each button below;

Lapco Section (Consultant Trainees only)
LOGIN to upload GAS forms
TRAINER Registration | TRAINEE Registration

Fellow Section
VIEW | LOGIN to upload GAS forms
TRAINER Registration | TRAINEE Registration

SPR Section
VIEW | LOGIN to upload GAS forms
TRAINER Registration | TRAINEE Registration

Train the Trainer
VIEW 60 sec video course overview
Details & Information

>> NTP Training Courses

- Lapco Train The Trainer 
- Clinical Immersion Courses
- Master Classes
- Theatre Practitioner Courses
- Other Courses

>> NTP Information

- About The Programme
- NICE Guidelines
- Peer Review
- In Reach & Out Reach Training
- Training Agreement
- GAS Forms
- Learning Curve
- Sign Off Process
- DVD Procedure
- Post Sign Off Requirements

9. REFERENCES

- ABMS. (2012). "American Board of Medical Specialties." from <http://www.abms.org/>.
- Abraham, N. S., J. M. Young and M. J. Solomon (2004). "Meta-analysis of short-term outcomes after laparoscopic resection for colorectal cancer." *Br J Surg* **91**(9): 1111-24.
- Agachan, F., J. S. Joo, M. Sher, E. G. Weiss, J. J. Noguerras and S. D. Wexner (1997). "Laparoscopic colorectal surgery. Do we get faster?" *Surg Endosc* **11**(4): 331-5.
- Agachan, F., J. S. Joo, E. G. Weiss and S. D. Wexner (1996). "Intraoperative laparoscopic complications. Are we getting better?" *Dis Colon Rectum* **39**(10 Suppl): S14-9.
- Aggarwal, R., T. Grantcharov, K. Moorthy, J. Hance and A. Darzi (2006). "A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill." *Am J Surg* **191**(1): 128-33.
- Aggarwal, R., T. Grantcharov, K. Moorthy, T. Milland and A. Darzi (2008). "Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room." *Ann Surg* **247**(2): 372-9.
- Agha, A., A. Furst, et al. (2008). "Conversion rate in 300 laparoscopic rectal resections and its influence on morbidity and oncological outcome." *Int J Colorectal Dis* **23**(4): 409-17.
- Ahmed, A. R., D. Miskovic, T. Vijayaseelan, W. O'Malley and G. B. Hanna (2011a). "Root cause analysis of internal hernia and Roux limb compression after laparoscopic Roux-en-Y gastric bypass using observational clinical human reliability assessment." *Surg Obes Relat Dis* (Epub ahead of print).

- Ahmed, A. R., D. Miskovic, T. Vijayaseelan, W. O'Malley and G. B. Hanna (2011b). "Root cause analysis of internal hernia and roux limb compression following laparoscopic Roux-Y Gastric Bypass using Observational Clinical Human Reliability Analysis (OCHRA)." Surg Obes Relat Dis (in print).
- Ahmed, K., H. Ashrafian and D. Miskovic (2009). "Lack of laparoscopic training at the junior level." Ann R Coll Surg Engl **91**(4): 352-3; author reply 353.
- Ahmed, K., D. Miskovic, K. Nagpal, T. Athanasiou and G. B. Hanna (2011c). "Observational assessment of technical skills – A systematic review." Am J Surg (in press).
- Archer, S., D. Brown, C. Smith, G. Branum and J. Hunter (2001). "Bile duct injury during laparoscopic cholecystectomy. Results of a national survey." Ann Surg **234**(4): 549-559.
- Avital, S., H. Hermon, R. Greenberg, E. Karin and Y. Skornick (2006). "Learning curve in laparoscopic colorectal surgery: our first 100 patients." Isr Med Assoc J **8**(10): 683-6.
- Beard, J. D., S. Choksy and S. Khan (2007). "Assessment of operative competence during carotid endarterectomy." Br J Surg **94**(6): 726-30.
- Beard, J. D., J. Marriott, H. Purdie and J. Crossley (2011). "Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology." Health Technol Assess **15**(1): i-xxi, 1-162.
- Begos, D. G., J. Arsenault and G. H. Ballantyne (1996). "Laparoscopic colon and rectal surgery at a VA hospital. Analysis of the first 50 cases." Surg Endosc **10**(11): 1050-6.
- Ben-David, M. F. (2000). "The role of assessment in expanding professional horizons." Med Teach **22**(5): 472-7.
- Bennett, C. L., S. J. Stryker, M. R. Ferreira, J. Adams and R. W. Beart, Jr. (1997). "The learning curve for laparoscopic colorectal surgery. Preliminary results from a

- prospective analysis of 1194 laparoscopic-assisted colectomies." Arch Surg **132**(1): 41-4; discussion 45.
- Birch, D. W., A. H. Asiri and C. J. de Gara (2007). "The impact of a formal mentoring program for minimally invasive surgery on surgeon practice and patient outcomes." Am J Surg **193**(5): 589-91; discussion 591-2.
- Black, P. and D. Wiliam (2009). "Developing the theory of formative assessment." Educational Assessment, Evaluation and Accountability **21**(1): 5-31.
- Bland, J. M. and D. G. Altman (1986). "Statistical methods for assessing agreement between two methods of clinical measurement." Lancet **1**(8476): 307-310.
- Bland, J. M. and D. G. Altman (1997). "Cronbach's alpha." BMJ **314**(7080): 572.
- Bloch, R. (2010). "G_String_III [computer program]. Version 5.5.2 for Windows. Accompanied by: 1 user manual (pdf). Available from http://fhspcrd.mcmaster.ca/g_string/download.html.
- Britten, N. (1995). "Qualitative interviews in medical research." BMJ **311**(6999): 251-3.
- Brown, D. C., D. Miskovic, B. Tang and G. B. Hanna (2011). "Impact of established skills in open surgery on the proficiency gain process for laparoscopic surgery." Surg Endosc **24**(6): 1420-6.
- Bruch, H. P., T. H. Schiedeck and O. Schwandner (1999). "Laparoscopic colorectal surgery: A five-year experience." Dig Surg **16**(1): 45-54.
- Buchanan, G. N., A. Malik, A. Parvaiz, J. P. Sheffield and R. H. Kennedy (2008). "Laparoscopic resection for colorectal cancer." Br J Surg **95**(7): 893-902.
- Buunen, M., H. J. Bonjer, et al. (2009a). "COLOR II. A randomized clinical trial comparing laparoscopic and open surgery for rectal cancer." Dan Med Bull **56**(2): 89-91.

- Buunen, M., R. Veldkamp, et al. (2009b). "Survival after laparoscopic surgery versus open surgery for colon cancer: long-term outcome of a randomised clinical trial." Lancet Oncol **10**(1): 44-52.
- Ceulemans, R., M. Henri, E. Dutson, J. Leroy and J. Marescaux (2003). "Open versus laparoscopy assisted colectomy." Lancet **361**(9351): 73-4; author reply 75-6.
- Champagne, B. J., E. C. Lee, B. Valerian, D. Armstrong, W. Ambroze and G. Orangio (2008). "A novel end point to assess a resident's ability to perform hand-assisted versus straight laparoscopy for left colectomy: is there really a difference?" J Am Coll Surg **207**(4): 554-9.
- Champault, G. and H. Schimmelpenning (2007). Laparoscopic Sigmoidectomy for Cancer, Springer.
- Chen, W., E. Sailhamer, D. L. Berger and D. W. Rattner (2007). "Operative time is a poor surrogate for the learning curve in laparoscopic colorectal surgery." Surg Endosc **21**(2): 238-43.
- Chikkappa, M. G., S. Jagger, J. P. Griffith, J. R. Ausobsky, M. A. Steward and J. B. Davies (2009). "In-house colorectal laparoscopic preceptorship: a model for changing a unit's practice safely and efficiently." Int J Colorectal Dis **24**(7): 771-6.
- Choi, D. H., W. K. Jeong, et al. (2009). "Learning curves for laparoscopic sigmoidectomy used to manage curable sigmoid colon cancer: single-institute, three-surgeon experience." Surg Endosc **23**(3): 622-8.
- Cima, R. R., I. Hassan, V. P. Poola, D. W. Larson, E. J. Dozois, D. R. Larson, M. M. O'Byrne and M. Huebner (2011). "Failure of institutionally derived predictive models of conversion in laparoscopic colorectal surgery to predict conversion outcomes in an independent data set of 998 laparoscopic colorectal procedures." Ann Surg **251**(4): 652-8.

Coleman, M. G., G. B. Hanna and R. Kennedy (2011). "The National Training Programme for Laparoscopic Colorectal Surgery in England: a new training paradigm." Colorectal Dis **13**(6): 614-6.

Collins, A., Brown, J. S., & Newman, S. E. (1987). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics (Technical Report No. 403)., Centre for the Study of Reading, University of Illinois. January, 1987.

COST (2004). "A comparison of laparoscopically assisted and open colectomy for colon cancer." N Engl J Med **350**(20): 2050-9.

Cronbach, L. J., R. Nageswari and G. C. Gleser (1963). "Theory of generalizability: A liberation of reliability theory. ." Br J Stat Psychol **16**: 137-163.

Crossley, J., H. Davies, G. Humphris and B. Jolly (2002). "Generalisability: a key to unlock professional assessment." Med Educ **36**(10): 972-8.

Curet, M. J. (2004). "Port site metastases." Am J Surg **187**(6): 705-12.

Cuschieri, A. (2006). "Nature of human error: implications for surgical practice." Ann Surg **244**(5): 642-8.

Cuschieri, A., N. Francis, J. Crosby and G. B. Hanna (2001). "What do master surgeons think of surgical competence and revalidation?" Am J Surg **182**(2): 110-6.

Daetwiler, S., U. Guller, O. Schob and M. Adamina (2007). "Early introduction of laparoscopic sigmoid colectomy during residency." Br J Surg **94**(5): 634-41.

Dalkey, N. C. (1968). "The Delphi Method: A Methodology Used for the Elicitation of Opinion of Experts." RM-5888-PR. Santa Monica: Rand Corp.

- Dalton, S., A. Ghosh, N. Zafar, K. Riyad and A. Dixon (2009). "Competency in laparoscopic colorectal surgery is achievable with appropriate training but takes time: a comparison of 300 elective resections with anastomosis." Colorectal Dis.
- Datta, V., S. Bann, R. Aggarwal, M. Mandalia, J. Hance and A. Darzi (2006). "Technical skills examination for general surgical trainees." Br J Surg **93**(9): 1139-46.
- Delaney, C. P., R. P. Kiran, A. J. Senagore, K. Brady and V. W. Fazio (2003). "Case-matched comparison of clinical and financial outcome after laparoscopic or open colorectal surgery." Ann Surg **238**(1): 67-72.
- Delaney PF, R. L., Staszewski JJ, Ritter FE (1998). "The strategy-specific nature of improvement: The power law applies by strategy within task." Psychological Sci (9): 1-7.
- Dincler, S., M. T. Koller, J. Steurer, L. M. Bachmann, D. Christen and P. Buchmann (2003). "Multidimensional analysis of learning curves in laparoscopic sigmoid resection: eight-year results." Dis Colon Rectum **46**(10): 1371-8; discussion 1378-9.
- Dunning, D., K. Johnson, J. Ehrlinger and J. Kruger (2003). "Why People Fail to Recognize Their Own Incompetence." Curr Dir Psychol Sci **12**(3): 83-87.
- Egger, M., G. Davey Smith, M. Schneider and C. Minder (1997). "Bias in meta-analysis detected by a simple, graphical test." BMJ **315**(7109): 629-34.
- Epstein, R. M. (2007). "Assessment in medical education." N Engl J Med **356**(4): 387-96.
- Eraut, M. (1994). Developing professional knowledge and competence, Falmer Press.
- Evans, J., I. Jenkins and R. H. Kennedy (2009). Laparoscopic Right Hemicolectomy. UK, St. Marks Multimedia: 20 mins.

- Evrard, S., S. Mathoulin-Pelissier and A. Kramar (2003). "Open versus laparoscopy-assisted colectomy." Lancet **361**(9351): 73; author reply 75-6.
- Faiz, O., J. Warusavitarne, A. Bottle, P. P. Tekkis, A. W. Darzi and R. H. Kennedy (2009a). "Laparoscopically assisted vs. open elective colonic and rectal resection: a comparison of outcomes in English National Health Service Trusts between 1996 and 2006." Dis Colon Rectum **52**(10): 1695-704.
- Faiz, O., J. Warusavitarne, A. Bottle, P. P. Tekkis, A. W. Darzi and R. H. Kennedy (2009b). "Laparoscopically-assisted vs open elective colonic and rectal resection: a comparison of outcomes in English National Health Service Trusts between 1996 and 2006." Dis Colon Rectum (in print).
- Feldman, L. S., J. Cao, A. Andalib, S. Fraser and G. M. Fried (2009). "A method to characterize the learning curve for performance of a fundamental laparoscopic simulator task: defining "learning plateau" and "learning rate"." Surgery **146**(2): 381-6.
- Fink, A., J. Kosecoff, M. Chassin and R. H. Brook (1984). "Consensus methods: characteristics and guidelines for use." Am J Public Health **74**(9): 979-83.
- Francis, N. K., G. B. Hanna and A. Cuschieri (2002). "The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study." Arch Surg **137**(7): 841-4.
- Fraser, S. A., D. R. Klassen, L. S. Feldman, G. A. Ghitulescu, D. Stanbridge and G. M. Fried (2003). "Evaluating laparoscopic skills: setting the pass/fail score for the MISTELS system." Surg Endosc **17**(6): 964-7.
- Fukunaga, Y., M. Higashino, S. Tanimura, M. Takemura and H. Osugi (2008). "Laparoscopic colorectal surgery for neoplasm. A large series by a single surgeon." Surg Endosc **22**(6): 1452-8.
- Gallagher, A. G., A. B. Lederman, K. McGlade, R. M. Satava and C. D. Smith (2004). "Discriminative validity of the Minimally Invasive Surgical Trainer in Virtual Reality

- (MIST-VR) using criteria levels based on expert performance." Surg Endosc **18**(4): 660-5.
- Gaub, V., P. Tsangaris, C. Tossounis, A. Mitra, C. McLean and G. M. Saleh (2008). "Human Reliability Analysis of Cataract Surgery." Arch Ophthalmol **126**(2): 173-7.
- Gelman, A. and J. Hill (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge, Cambridge University Press.
- GMC. (2012). "The General Medical Council . Good Medical Practice ", from <http://www.gmc-uk.org>.
- Grigg, O. A., V. T. Farewell and D. J. Spiegelhalter (2003). "Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts." Stat Methods Med Res **12**(2): 147-70.
- Guillou, P. J., P. Quirke, H. Thorpe, J. Walker, D. G. Jayne, A. M. Smith, R. M. Heath and J. M. Brown (2005). "Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial." Lancet **365**(9472): 1718-26.
- Hanna, G. B., T. Drew, P. Clinch, B. Hunter and A. Cuschieri (1998). "Computer-controlled endoscopic performance assessment system." Surg Endosc **12**(7): 997-1000.
- Heathcote, A., S. Brown and D. J. Mewhort (2000). "The power law repealed: the case for an exponential law of practice." Psychon Bull Rev **7**(2): 185-207.
- Hedrick, T., F. Turrentine, H. Sanfey, B. Schirmer and C. Friel (2009). "Implications of laparoscopy on surgery residency training." Am J Surg **197**(1): 73-5.
- Higgins, J. P. T. and S. Green. "Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.1 (updated September 2008).", from The Cochrane Collaboration, 2008. Available from <http://www.cochrane-handbook.org>, last accessed 08/03/09.

- Hoek, J. and P. Gendall (1993). "A new method of predicting voting behaviour." Journal of the Market Research Society **35**(4): 361-373.
- Hollnagel, E. (1998). Cognitive reliability and error analysis method. Oxford, Elsevier Science Ltd.
- Homer (2009). Ilias. Odyssee: Zwei Bände im Schuber: In Prosa übertragen von Karl Ferdinand Lempp., Insel Verlag.
- Ito, M., M. Sugito, A. Kobayashi, Y. Nishizawa, Y. Tsunoda and N. Saito (2009). "Influence of learning curve on short-term results after laparoscopic resection for rectal cancer." Surg Endosc **23**(2): 403-8.
- Jacobs, M., J. C. Verdeja and H. S. Goldstein (1991). "Minimally invasive colon resection (laparoscopic colectomy)." Surg Laparosc Endosc **1**(3): 144-50.
- Jayne, D. G., P. J. Guillou, H. Thorpe, P. Quirke, J. Copeland, A. M. Smith, R. M. Heath and J. M. Brown (2007). "Randomized trial of laparoscopic-assisted resection of colorectal carcinoma: 3-year results of the UK MRC CLASICC Trial Group." J Clin Oncol **25**(21): 3061-8.
- Jayne, D. G., H. C. Thorpe, J. Copeland, P. Quirke, J. M. Brown and P. J. Guillou (2010). "Five-year follow-up of the Medical Research Council CLASICC trial of laparoscopically assisted versus open surgery for colorectal cancer." Br J Surg **97**(11): 1638-45.
- Joice, P., G. B. Hanna and A. Cuschieri (1998). "Errors enacted during endoscopic surgery--a human reliability analysis." Appl Ergon **29**(6): 409-14.
- Kennedy, R. H. (2008). "EnROL trial." from <http://www.octo-oxford.org.uk/alltrials/trials/EnROL.html>>.

- Kim, J., E. Edwards, W. Bowne, A. Castro, V. Moon, P. Gadangi and G. Ferzli (2007). "Medial-to-lateral laparoscopic colon resection: a view beyond the learning curve." Surg Endosc **21**(9): 1503-7.
- King, B. and I. Jatoi (2005). "The mobile army surgical hospital (MASH): A military and surgical legacy." J Natl Med Assoc **97**(5): 648-56.
- Kirchhoff, P., S. Dincler and P. Buchmann (2008). "A multivariate analysis of potential risk factors for intra- and postoperative complications in 1316 elective laparoscopic colorectal procedures." Ann Surg **248**(2): 259-65.
- Kirwan, B. (1992a). "Human error identification in human reliability assessment. Part 1: Overview of approaches." Appl Ergon **23**(5): 299-318.
- Kirwan, B. (1992b). "Human error identification in human reliability assessment. Part 2: Detailed comparison of techniques." Appl Ergon **23**(6): 371-81.
- Kitano, S., M. Inomata, A. Sato, K. Yoshimura and Y. Moriya (2005). "Randomized controlled trial to evaluate laparoscopic surgery for colorectal cancer: Japan Clinical Oncology Group Study JCOG 0404." Jpn J Clin Oncol **35**(8): 475-7.
- Klarenbeek, B. R., A. A. Veenhof, et al. (2007). "The Sigma-trial protocol: a prospective double-blind multi-centre comparison of laparoscopic versus open elective sigmoid resection in patients with symptomatic diverticulitis." BMC Surg **7**: 16.
- Kockerling, F., C. Schneider, et al. (1998). "Early results of a prospective multicenter study on 500 consecutive cases of laparoscopic colorectal surgery. Laparoscopic Colorectal Surgery Study Group (LCSSG)." Surg Endosc **12**(1): 37-41.
- Lacy, A. M., J. C. Garcia-Valdecasas, S. Delgado, A. Castells, P. Taura, J. M. Pique and J. Visa (2002). "Laparoscopy-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial." Lancet **359**(9325): 2224-9.

- Lacy, A. M., J. C. Garcia-Valdecasas, et al. (1995). "Short-term outcome analysis of a randomized study comparing laparoscopic vs open colectomy for colon cancer." Surg Endosc **9**(10): 1101-5.
- Lapco. (2009). "National Training Programme in Laparoscopic Colorectal Surgery", from <http://www.lapco.nhs.uk>.
- Lauter, D. M. and E. J. Froines (2001). "Initial experience with 150 cases of laparoscopic assisted colectomy." Am J Surg **181**(5): 398-403.
- Legorreta, A. P., J. H. Silber, G. N. Costantino, R. W. Kobylnski and S. L. Zatz (1993). "Increased cholecystectomy rate after the introduction of laparoscopic cholecystectomy." JAMA **270**(12): 1429-32.
- Leung, K. L., S. P. Kwok, S. C. Lam, J. F. Lee, R. Y. Yiu, S. S. Ng, P. B. Lai and W. Y. Lau (2004). "Laparoscopic resection of rectosigmoid carcinoma: prospective randomised trial." Lancet **363**(9416): 1187-92.
- Li, J. C., S. S. Hon, S. S. Ng, J. F. Lee, R. Y. Yiu and K. L. Leung (2009). "The learning curve for laparoscopic colectomy: experience of a surgical fellow in an university colorectal unit." Surg Endosc **23**(7): 1603-8.
- Lin, E., S. Szomstein, T. Addasi, L. Galati-Burke, J. W. Turner and H. I. Tiszenkel (2003). "Model for teaching laparoscopic colectomy to surgical residents." Am J Surg **186**(1): 45-8.
- Lordan, J. T., H. S. Tilney, S. Shirol, I. Jourdan and A. M. Gudgeon (2008). "Does the laparoscopic colorectal surgery learning curve adversely affect the results of colorectal cancer resection? A 3-year prospective study in a district general hospital." Colorectal Dis **10**(4): 363-9.
- Martel, G. and R. P. Boushey (2006). "Laparoscopic colon surgery: past, present and future." Surg Clin North Am **86**(4): 867-97.

- Martin, J. A., G. Regehr, R. Reznick, H. MacRae, J. Murnaghan, C. Hutchison and M. Brown (1997). "Objective structured assessment of technical skill (OSATS) for surgical residents." Br J Surg **84**(2): 273-8.
- Marusch, F., I. Gastinger, C. Schneider, H. Scheidbach, J. Konradt, H. P. Bruch, L. Kohler, E. Barlehner and F. Kockerling (2001). "Experience as a factor influencing the indications for laparoscopic colorectal surgery and the results." Surg Endosc **15**(2): 116-20.
- McGill, D. A., C. P. van der Vleuten and M. J. Clarke (2011). "Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review." Adv Health Sci Educ Theory Pract **16**(3): 405-25.
- Mehall, J. R., S. Shroff, S. A. Fassler, S. G. Harper, J. H. Nejman and D. M. Zebley (2005). "Comparing results of residents and attending surgeons to determine whether laparoscopic colectomy is safe." Am J Surg **189**(6): 738-41.
- Merriam-Webster. (2011). "Online Dictionary <http://www.merriam-webster.com>." from <http://www.merriam-webster.com>, last time accessed November 2011.
- Messick, S. (1984). "The psychology of educational measurement." Educational Measurement **21**: 215-237.
- Milsom, J. W., B. Bohm, K. A. Hammerhofer, V. Fazio, E. Steiger and P. Elson (1998). "A prospective, randomized trial comparing laparoscopic versus conventional techniques in colorectal cancer surgery: a preliminary report." J Am Coll Surg **187**(1): 46-54; discussion 54-5.
- Miskovic, D., S. M. Wyles, M. G. Coleman and G. B. Hanna (2010a). "Development, validation and implementation of a monitoring tool for training in laparoscopic colorectal surgery in the English National Training Program." Surg Endosc.

- Miskovic, D., S. M. Wyles, M. Ni, A. W. Darzi and G. B. Hanna (2010b). "Systematic review on mentoring and simulation in laparoscopic colorectal surgery." Ann Surg **252**(6): 943-51.
- Moher, D., A. Liberati, J. Tetzlaff and D. G. Altman (2009). "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement." BMJ **339**: b2535.
- Moorthy, Munz, Dosis, Bello and Darzi (2003). "Motion analysis in the training and assessment of minimally invasive surgery." Minim Invasive Ther Allied Technol **12**(3): 137-42.
- Munz, Y., K. Moorthy, S. Bann, J. Shah, S. Ivanova and S. A. Darzi (2004). "Ceiling effect in technical skills of surgical residents." Am J Surg **188**(3): 294-300.
- Nelson, H., J. C. Weeks and H. S. Wieand (1995). "Proposed phase III trial comparing laparoscopic-assisted colectomy versus open colectomy for colon cancer." J Natl Cancer Inst Monogr(19): 51-6.
- Neudecker, J., F. Klein, R. Bittner, T. Carus, A. Stroux and W. Schwenk (2009). "Short-term outcomes from a prospective randomized trial comparing laparoscopic and open surgery for colorectal cancer." Br J Surg **96**(12): 1458-67.
- Newell, A. and P. Rosenbloom (1981). Mechanisms of skill acquisition and the power law of practice. Cognitive skill and their acquisition. J. Anderson. Hillsdale, Erlbaum: 1-55.
- NICE (2007). "NICE implementation uptake report: laparoscopic surgery for colorectal cancer."
<http://www.nice.org.uk/media/411/18/ImplUptakeReportColorectalResectionsLaparoscopic.pdf>, last time accessed January 2012.
- NICE. (2008). "UK Department of Health. National Training Programme (NTP) in laparoscopic colorectal cancer surgery: supporting NICE Technology Appraisal (2006) recommendations for laparoscopic resection for colorectal cancer." from

http://www.dh.gov.uk/en/Publicationsandstatistics/Lettersandcirculars/Dearcolleagueletters/DH_091954, last time accessed November 2011.

Okraïnec, A., N. J. Soper, L. L. Swanstrom and G. M. Fried "Trends and results of the first 5 years of Fundamentals of Laparoscopic Surgery (FLS) certification testing." Surg Endosc.

Oxford (2005). The New Oxford American Dictionary, 2nd edition. Oxford, Oxford University Press.

Page, E. S. (1954). "Continuous inspection schemes." Biometrika **41**: 100-114.

Palter, V. N., H. M. MacRae and T. P. Grantcharov (2011). "Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology." Am J Surg **201**(2): 251-9.

Park, I. J., G. S. Choi, K. H. Lim, B. M. Kang and S. H. Jun (2009a). "Multidimensional analysis of the learning curve for laparoscopic colorectal surgery: lessons from 1,000 cases of laparoscopic colorectal surgery." Surg Endosc **23**(4): 839-46.

Park, I. J., G. S. Choi, K. H. Lim, B. M. Kang and S. H. Jun (2009b). "Multidimensional analysis of the learning curve for laparoscopic resection in rectal cancer." J Gastrointest Surg **13**(2): 275-81.

Park, J. S., S. B. Kang, S. W. Kim and G. N. Cheon (2007). "Economics and the laparoscopic surgery learning curve: comparison with open surgery for rectosigmoid cancer." World J Surg **31**(9): 1827-34.

Peduzzi, P., J. Concato, A. R. Feinstein and T. R. Holford (1995). "Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates." J Clin Epidemiol **48**(12): 1503-10.

- Pigazzi, A., C. Anderson, P. Mojica-Manosa, D. Smith, K. Hernandez, I. B. Paz and J. D. Ellenhorn (2008). "Impact of a full-time preceptor on the institutional outcome of laparoscopic colectomy." Surg Endosc **22**(3): 635-9.
- Rae, D., R. Bowyer, R. Wharton and I. Paterson (1995). "Port site metastasis after laparoscopic colorectal surgery for cure of malignancy." Br J Surg **82**(8): 1142.
- Ramsay, C. R., A. M. Grant, S. A. Wallace, P. H. Garthwaite, A. F. Monk and I. T. Russell (2001). "Statistical assessment of the learning curves of health technologies." Health Technol Assess **5**(12): 1-79.
- Reissman, P., S. Cohen, E. G. Weiss and S. D. Wexner (1996). "Laparoscopic colorectal surgery: ascending the learning curve." World J Surg **20**(3): 277-81; discussion 282.
- Reitsma, A. and J. Moreno (2002). "Ethical regulation for innovative surgery: the last frontier?" J Am Coll Surg **194**(6): 792-801.
- Renwick, A. A., E. L. Bokey, P. H. Chapuis, P. Zelas, P. J. Stewart, M. J. Rickard and O. F. Dent (2005). "Effect of supervised surgical training on outcomes after resection of colorectal cancer." Br J Surg **92**(5): 631-6.
- Rijbroek, A., W. Wisselink and J. A. Rauwerda (2003). "The impact of training in unselected patients on mortality and morbidity in carotid endarterectomy in a vascular training center and the recommendations of the European Board of Surgery Qualification in Vascular Surgery." Eur J Vasc Endovasc Surg **26**(3): 256-61.
- Romm, F. J. and B. S. Hulka (1979). "Developing criteria for quality of assessment: effect of the Delphi technique." Health Serv Res **14**(4): 309-12.
- Rotholtz, N. A., M. Laporte, G. Zanoni, M. E. Bun, L. Aued, S. Lencinas, N. A. Mezzadri and L. Pereyra (2008). "Predictive factors for conversion in laparoscopic colorectal surgery." Tech Coloproctol **12**(1): 27-31.

Sarker, S. K., I. Kumar and C. Delaney (2011). "Assessing operative performance in advanced laparoscopic colorectal surgery." World J Surg **34**(7): 1594-603.

Scheer, A., G. Martel, H. Moloo, E. Sabri, E. C. Poulin, J. Mamazza and R. P. Boushey (2009). "Laparoscopic colon surgery: does operative time matter?" Dis Colon Rectum **52**(10): 1746-52.

Schlachta, C. M., S. A. Kent, K. L. Lefebvre, M. L. McCune and S. Jayaraman (2009). "A model for longitudinal mentoring and telementoring of laparoscopic colon surgery." Surg Endosc **23**(7): 1634-8.

Schlachta, C. M., J. Mamazza, R. Gregoire, S. E. Burpee, K. T. Pace and E. C. Poulin (2003). "Predicting conversion in laparoscopic colorectal surgery. Fellowship training may be an advantage." Surg Endosc **17**(8): 1288-91.

Schlachta, C. M., J. Mamazza, P. A. Seshadri, M. Cadeddu, R. Gregoire and E. C. Poulin (2001). "Defining a learning curve for laparoscopic colorectal resections." Dis Colon Rectum **44**(2): 217-22.

Schlachta, C. M., J. Mamazza, P. A. Seshadri, M. O. Cadeddu and E. C. Poulin (2000). "Predicting conversion to open surgery in laparoscopic colorectal resections. A simple clinical model." Surg Endosc **14**(12): 1114-7.

Schuwirth, L. and C. van der Vleuten (2004). "Merging views on assessment." Med Educ **38**(12): 1208-10.

Schwab, K. E., H. M. Dowson, J. Van Dellen, C. G. Marks and T. A. Rockall (2009). "The uptake of laparoscopic colorectal surgery in Great Britain and Ireland: a questionnaire survey of consultant members of the ACPGBI." Colorectal Dis **11**(3): 318-22.

Schwandner, O., T. H. Schiedeck and H. Bruch (1999). "The role of conversion in laparoscopic colorectal surgery: Do predictive factors exist?" Surg Endosc **13**(2): 151-6.

- Schwenk, W. (2004). "[The LAPDIV-CAMIC Study. Multicenter prospective randomized study of short-term and intermediate-term outcome of laparoscopic and conventional sigmoid resection in diverticular disease]." Chirurg **75**(7): 706-7.
- Schwenk, W., B. Bohm, O. Haase, T. Junghans and J. M. Muller (1998). "Laparoscopic versus conventional colorectal resection: a prospective randomised study of postoperative ileus and early postoperative feeding." Langenbecks Arch Surg **383**(1): 49-55.
- Schwenk, W., O. Haase, J. Neudecker and J. M. Muller (2005). "Short term benefits for laparoscopic colorectal resection." Cochrane Database Syst Rev(3): CD003145.
- Scriven, M. (1967). The methodology of evaluation. Curriculum evaluation. R. E. Stake. Chicago, Rand McNally.
- Sebajang, H., P. Trudeau, A. Dougall, S. Hegge, C. McKinley and M. Anvari (2006). "The role of telementoring and telerobotic assistance in the provision of laparoscopic colorectal surgery in rural areas." Surg Endosc **20**(9): 1389-93.
- Senagore, A. J., M. A. Luchtefeld and J. M. Mackeigan (1995). "What is the learning curve for laparoscopic colectomy?" Am Surg **61**(8): 681-5.
- Sevdalis, N., R. Davis, M. Koutantji, S. Undre, A. Darzi and C. A. Vincent (2008). "Reliability of a revised NOTECHS scale for use in surgical teams." Am J Surg **196**(2): 184-90.
- Shah, P. R., A. Joseph and P. N. Haray (2005). "Laparoscopic colorectal surgery: learning curve and training implications." Postgrad Med J **81**(958): 537-40.
- Shawki, S., B. Bashankaev, P. Denoya, C. Seo, E. G. Weiss and S. D. Wexner (2009). "What is the definition of "conversion" in laparoscopic colorectal surgery?" Surg Endosc **23**(10): 2321-6.
- Short, E. (1984). "Competence reexamined." Educational Theory **34**(3): 201-207.

- Sidhu, R. S., E. Vikis, R. Cheifetz and T. Phang (2006). "Self-assessment during a 2-day laparoscopic colectomy course: can surgeons judge how well they are learning new skills?" Am J Surg **191**(5): 677-81.
- Silecchia, G., N. Perrotta, et al. (2002). "Abdominal wall recurrences after colorectal resection for cancer: results of the Italian registry of laparoscopic colorectal surgery." Dis Colon Rectum **45**(9): 1172-7; discussion 1177.
- Simons, A. J., G. J. Anthone, A. E. Ortega, M. Franklin, J. Fleshman, W. P. Geis and R. W. Beart, Jr. (1995). "Laparoscopic-assisted colectomy learning curve." Dis Colon Rectum **38**(6): 600-3.
- Streiner, N. and G. Norman (2008). Health Measurement Scales. A practical guide to their development and use. Oxford, Oxford University Press.
- Tan, P. Y., J. H. Stephens, N. A. Rieger and P. J. Hewett (2008). "Laparoscopically assisted colectomy: a study of risk factors and predictors of open conversion." Surg Endosc **22**(7): 1708-14.
- Tang, B., G. B. Hanna, N. M. Bax and A. Cuschieri (2004a). "Analysis of technical surgical errors during initial experience of laparoscopic pyloromyotomy by a group of Dutch pediatric surgeons." Surg Endosc **18**(12): 1716-20.
- Tang, B., G. B. Hanna and A. Cuschieri (2005). "Analysis of errors enacted by surgical trainees during skills training courses." Surgery **138**(1): 14-20.
- Tang, B., G. B. Hanna, P. Joice and A. Cuschieri (2004b). "Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy." Arch Surg **139**(11): 1215-20.
- Taylor, W. (2007). "Change-Point Analysis: A Powerful New Tool For Detecting Changes. <http://www.variation.com/cpa/tech/changepoint.html>, last time accessed November 2011.

- Tekkis, P. P., V. W. Fazio, et al. (2005a). "Evaluation of the learning curve in ileal pouch-anal anastomosis surgery." Ann Surg **241**(2): 262-8.
- Tekkis, P. P., A. J. Senagore and C. P. Delaney (2005b). "Conversion rates in laparoscopic colorectal surgery: a predictive model with, 1253 patients." Surg Endosc **19**(1): 47-54.
- Tekkis, P. P., A. J. Senagore, C. P. Delaney and V. W. Fazio (2005c). "Evaluation of the learning curve in laparoscopic colorectal surgery: comparison of right-sided and left-sided resections." Ann Surg **242**(1): 83-91.
- Tinmouth, J. and G. Tomlinson (2004). "Laparoscopically assisted versus open colectomy for colon cancer." N Engl J Med **351**(9): 933-4; author reply 933-4.
- Tjandra, J. J. and M. K. Chan (2006). "Systematic review on the short-term outcome of laparoscopic resection for colon and rectosigmoid cancer." Colorectal Dis **8**(5): 375-88.
- Tseng, L. N., F. J. Berends, P. Wittich, N. D. Bouvy, R. L. Marquet, G. Kazemier and H. J. Bonjer (1998). "Port-site metastases. Impact of local tissue trauma and gas leakage." Surg Endosc **12**(12): 1377-80.
- Tyras, D. H., H. B. Barner, G. C. Kaiser, J. E. Codd, D. G. Pennington and V. L. Willman (1980). "Bypass grafts to the left anterior descending coronary artery: saphenous vein versus internal mammary artery." J Thorac Cardiovasc Surg **80**(3): 327-33.
- UK, Cancer Research (2011). "News and Resources." from <http://info.cancerresearchuk.org/cancerstats/>, last time accessed November 2011.
- van der Vleuten, C. P. and L. W. Schuwirth (2005). "Assessing professional competence: from methods to programmes." Med Educ **39**(3): 309-17.
- van der Vleuten, C. P., L. W. Schuwirth, F. Scheele, E. W. Driessen and B. Hodges (2010). "The assessment of professional competence: building blocks for theory development." Best Pract Res Clin Obstet Gynaecol **24**(6): 703-19.

- Van Hove, C., K. Hardiman, B. Diggs, C. Deveney and B. Sheppard (2008). "Demographic and socioeconomic trends in the use of laparoscopic appendectomy from 1997 to 2003." Am J Surg **195**(5): 580-3; discussion 583-4.
- van Hove, P. D., G. J. Tuijthof, E. G. Verdaasdonk, L. P. Stassen and J. Dankelman "Objective assessment of technical surgical skills." Br J Surg **97**(7): 972-87.
- Vassiliou, M. C., L. S. Feldman, C. G. Andrew, S. Bergman, K. Leffondre, D. Stanbridge and G. M. Fried (2005). "A global assessment tool for evaluation of intraoperative laparoscopic skills." Am J Surg **190**(1): 107-13.
- Vassiliou, M. C., G. A. Ghitulescu, L. S. Feldman, D. Stanbridge, K. Leffondre, H. H. Sigman and G. M. Fried (2006). "The MISTELS program to measure technical skill in laparoscopic surgery : evidence for reliability." Surg Endosc **20**(5): 744-7.
- Veldkamp, R., E. Kuhry, et al. (2005). "Laparoscopic surgery versus open surgery for colon cancer: short-term outcomes of a randomised trial." Lancet Oncol **6**(7): 477-84.
- Vygotsky, L. (1978). Mind in Society. Cambridge, Harvard University Press.
- Wexner, S. D. and S. M. Cohen (1995). "Port site metastases after laparoscopic colorectal surgery for cure of malignancy." Br J Surg **82**(3): 295-8.
- Winckel, C. P., R. K. Reznick, R. Cohen and B. Taylor (1994). "Reliability and construct validity of a structured technical skills assessment form." Am J Surg **167**(4): 423-7.
- Wishner, J. D., J. W. Baker, Jr., G. C. Hoffman, G. W. Hubbard, 2nd, R. J. Gould, S. D. Wohlgemuth, W. K. Ruffin and C. F. Melick (1995). "Laparoscopic-assisted colectomy. The learning curve." Surg Endosc **9**(11): 1179-83.
- Wohaibi, E. M., D. B. Earle, F. E. Ansanitis, R. B. Wait, G. Fernandez and N. E. Seymour (2007). "A new web-based operative skills assessment tool effectively tracks progression in surgical resident performance." J Surg Educ **64**(6): 333-41.

Wozniak, R. (1999). Classics in Psychology, 1855-1914: Historical Essays. Bristol, Thoemmes Press.

Wyles, S. M., D. Miskovic, et al. (2010a). "Analysis of laboratory based Laparoscopic Colorectal surgery workshops within the English National Training Programme." Surg Endosc.

Wyles, S. M., D. Miskovic, et al. (2010b). "Analysis of laboratory-based laparoscopic colorectal surgery workshops within the English National Training Programme." Surg Endosc.

Yule, S., R. Flin, N. Maran, D. Rowley, G. Youngson and S. Paterson-Brown (2008). "Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system." World J Surg **32**(4): 548-56.

Zorga, S. (2002). "Supervision: the process of life-long learning in social and educational professions." J Interprof Care **16**(3): 265-76.